JAIST Repository
https://dspace.jaist.ac.jp/

Title	2者対話の返答音声におけるタイミングと自然さの関係分析
Author(s)	吉川, 禎洋
Citation	
Issue Date	2024-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/19367
Rights	
Description	Supervisor: 岡田 将吾, 先端科学技術研究科, 修士(情報科学)

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Quality Assessment of Voice Response Timing in Conversational Corpus through Timing Replacement

2230033 Sadahiro Yoshikawa

The timing estimation models in spoken dialogue systems (SDSs) have typically been trained by human responses in order to achieve the appropriate response timing. However, human response timings are not always appropriate: in a previous experiment in which annotators listened to responses with the timings replaced by fixed values, some responses with the mode value sounded more realistic than actual human responses. Since this previous experiment was a small-scale preliminary one that only showed that some speakers tended to be significantly preferred, in the current study, we conducted an experiment on about 1,700 human responses, and scored whether they could be replaced with the mode value. The results showed that the annotators tended to feel that mode (or perhaps from 0 ms to 400 ms) responses are more appropriate than actual overlappings. We determined the responses that could and could not be replaced with the mode value by a chi-square test and then formulated a detection task to predict them from the scores. The evaluation results showed that our proposed simple model outperformed random selection with the AUC of 0.650. On the basis of these results, we discuss about the challenging for implementing SDSs, using the score to predict which responses or response timings are appropriate for the SDS users. Our findings may suggest a more efficient way to determine the appropriate response timing for SDSs compared to training models by corpus data.