

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: 鷗木 祐史, 先端科学技術研究科, 博士

氏名	刘雨凝		
学位の種類	博士 (情報科学)		
学位記番号	博情第 540 号		
学位授与年月日	令和 6 年 12 月 24 日		
論文題目	Assessing the importance of social factors in dialogue processing		
論文審査委員	鶴木 祐史	北陸先端科学技術大学院大学	教授
	岡田 将吾	同	教授
	長谷川 忍	同	教授
	党 建武	中国科学院深セン先進技術研究所	教授
	SAKTI Sakriani Watiasri	奈良先端科学技術大学院大学	教授
	水町 光徳	九州工業大学	教授

論文の内容の要旨

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.

Keywords: social factors, conversational entrainment, social distance, Communication Accommodation, Speech act, Interactive Alignment Model

論文審査の結果の要旨

本研究は、日常会話における「同調現象」(entrainment)を通じて、会話中の社会的要因を定量化し、それがどのように対話シーンの分類や予測に役立つかを研究するものである。音声の同調現象は、対話者間の社会的距離や親密度に応じて発現する傾向がある。例えば、論争や説得の場面では距離を取り、慰めや喜びを分かち合うの場面では距離が縮まるなど、会話のトピックや目的に影響を受ける。このような「同調」は、相互理解を促進し、対話の成功に寄与する重要な社会的メカニズムである。本研究では、この音声対話における同調現象を用いて、会話中の社会的要因、特に社会的距離を音声特徴量や意味特徴量から定量化し、異なる対話の場面に応じた音声対話のパターンや動的な社会関係を解明することにある。

これまででは、個々の特徴量を独立に取り扱って音声の同調現象を評価する手法が一般的であった。本研究では、インタラクティブアラインメントモデルに基づき、複数の音響特徴量や意味特徴量との関連性を考慮した線形モデルを用いることで、音声対話における社会的要因を正確に捉えることを目指した。具体的には、中国語のテレビドラマ「Golden Marriage」から抽出した四つのシーンの音声対話を基に構築したオリジナルコーパスを用いて、社会的要因が異なるシーンでのパターンを分類した。次に、既存の英語コーパス IEMOCAP を用いることで、提案法がオープンソースのコーパスや異なる言語でも有効であることを証明した。さらにこれらのデータセットを通じて、対話における社会的要因のパターンが異なる会話のシーンごとに異なることを示し、音声の同調現象を用いて計算された社会的要因が異なる会話シーンの分類や予測に有用であることを明らかにした。

本研究の結果は、音声の同調現象を用いて計算された社会的要因、特に社会的距離を定量化することで、人と人とのコミュニケーション状況を適切に分類あるいは予測することができ、結果とし

て、話者がその状況に応じた最適な対応を取るための支援が可能であることを示している。また、社会的要因に基づく音声の同調の定量化により、自然で文脈適応型の対話システムが実現可能となり、ユーザに親しみやすく、応答精度の高い音声認識・応答システムの構築が可能となる。本研究の成果は、音声情報処理の枠組みで、音声の同調現象をどのように計算するのか、音響特徴量と意味特徴量からどのように社会的距離を定量化するのか、を明らかにしたものであるため、将来、高度な対話型 AI システムの開発に貢献するものと期待される。

以上、本研究は対話の同調現象に基づき、社会的要因の理解とその応用可能性に関する新たな知見を提供した。よって博士（情報科学）の学位論文として十分価値あるものと認めた。