



Title	Enhancing Speech Comprehension of Simultaneously Speaking Agents in noisy environments
Author(s)	Pourfannan, Hamed
Citation	大阪大学, 2023, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/92907
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

Abstract of Thesis

Name (Hamed Pourfannan)	
Title	Enhancing Speech Comprehension of Simultaneously Speaking Agents in noisy environments (雑音環境における複数エージェントの同時発話音声に対する理解の向上)
This thesis aims to improve the comprehension of a robot's speech when there is more than one robot speaking simultaneously, also known as cross-talking, in noisy environments. Cross-talking is a common occurrence in noisy social spaces and is shown to negatively affect human cognition. This thesis tackles the problem at hand by investigating the three key factors affecting the comprehension of concurrent speech in noise, namely properties of the foreground voice, the concurrent speeches, and the masking sound. Manipulating the first factor aims to increase the resilience of the foreground voice against background noise. For this aim, the effectiveness of manipulating the gender and pitch of the speaking voice to improve concentration on the robot's speech in a noisy environment is investigated. The obtained results indicate that using a low-pitch female voice increases the memory of the subjects about the spoken material significantly in both genders. Investigation of the second factor aims to utilize properties of simultaneously speaking voices to enhance the subject's attention to the target information via the presence of background speech. For this aim, the effectiveness of the speech time expansion technique is evaluated. The obtained results suggest that expanding both the speaking voices in time eliminates noise's negative effect and significantly increases subjects' speech comprehension. Consequently, this technique is applied in the real-world simultaneously speaking bilingual human-robot interaction scenario. The results indicate that the proposed technique can significantly improve speech comprehension in an otherwise adverse listening condition. Studying the third factor is concerned with improving the efficiency of sound masking for human-robot interaction in noise. Thus, a filter is created to induce low-frequency dominance in speech-shaped masking noise, as well as increase the amplitude of its consonant-dense range. The obtained filter is shown to successfully reduce the annoyance in the subjects and increase the efficiency of the masking sound. Consequently, the proposed masking noise is implemented in a real-world human-robot interaction in a multi-talker noisy condition and is shown to increase acoustic satisfaction and perceived sound privacy for the users of the robot significantly. In conclusion, the results obtained in this thesis reveal that when two robots speak at the same time, applying time expansion to both robots results in a significant improvement in the comprehension of the robot's speech. Furthermore, it was shown that using the proposed masking sound could significantly reduce distraction during a human-robot interaction in a noisy multi-talker environment. The observed effects suggest that selective attention in humans can be manipulated using the properties of background noise that carries language-type information. Neuroscientific theories such as the optimal stimulation model suggest that a moderate level of noise can enhance attention by increasing arousal in the human brain. However, this thesis is the first to show this is also possible using background noise that carries lingual information. Hence, the current thesis opens a door for future neuroscientific and psychological endeavors to investigate the mechanisms of human perception in a noisy cross-talking condition. The obtained knowledge in this thesis contributes to the well-being of humans living in modern society where noise is common.	

論文審査の結果の要旨及び担当者

氏 名 （ HAMED POURFANNAN ）		
	(職)	氏 名
論文審査担当者	主 査	教 授 石黒 浩
	副 査	教 授 佐藤 宏介
	副 査	教 授 飯國 洋二
論文審査の結果の要旨 本論文では、ノイズの多い環境下で複数のロボットが同時に発話する状況（クロストークキング）において、ロボットの発話内容の理解を向上させることを目的とし、三つの研究を行なった。騒音下でのクロストークキングに対する発話理解に影響を与える3つの重要な要因、すなわち、メイン音声の特性、同時発話音声の特性、そして騒音マスキング音声の特性をそれぞれの研究で順に調査し、これらがどのように発話内容の理解の向上に貢献できるのかを示した。 本論文の成果は以下のとおりである。研究1では、騒音環境下でのロボットの発話に対する理解を高めるために、メインの発話音声の性別とピッチを操作することが有効であることを示し、結果としてピッチの低い女性の声を使うことで、メイン音声に対する被験者の記憶力といった、発話理解を評価する指標の点数が有意に向上することを示した。研究2では、メイン音声に別言語の背景音声を加え、それぞれの音声において発話の速度を遅くし、また文章間の無言の時間を長くすることで、メイン音声に対する内容の記憶といった発話理解が向上できることを示した。さらに、提案手法を実際のロボットとのインタラクション環境に実装し、実世界における対話環境でもその有用性を確認した。研究3では、背景騒音に対するマスキング効果がより高いマスキングノイズの作成を目指し、従来より発話型騒音に対するマスキング効果が高いことが知られる Speech-shaped ノイズのスペクトラムに、低周波数領域を強化し、中・高周波数領域を弱化するフィルターを用いることで、また背景の発話型騒音に存在する子音の雑音効果を減らすために2.5kHzから4kHzの領域の周波数を強化することで、背景の発話音声をより効果的にマスキングできることを示した。さらに、提案されたマスキングノイズを人間とロボットの対話環境に実装し、背景の発話ノイズを効果的にマスキングできることや、被験者の主観的な音声プライバシーや満足度を向上させることを確認した。 以上、本論文は、ノイズの多いクロストークキング環境において、メイン音声、同時発話音声、および騒音マスキング音声の特性がどのように発話音声の理解を向上させることができるかを明らかにした研究であり、複数の発話エージェントによるクロストークキングが増えていく現代社会において、応用性の高い論文と言える。また、本論文で示された背景ノイズを用いた発話理解の向上は、人間の集中度を背景ノイズにより向上できた結果であると考えられ、今後の工学のみならず、認知科学における人間のさらなる理解に貢献できる知見に繋がると期待できる。よって本論文は博士（工学）の学位論文として価値のあるものと認める。		