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

Title	音声波形の振幅包絡線が緊迫感の知覚に及ぼす影響
Author(s)	川村，美帆
Citation	
Issue Date	2019-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15891
Rights	
Description	Supervisor:鵜木　祐史，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

修 士 論 文

音声波形の振幅包絡線が緊迫感の知覚に及ぼす影響

1710059　　川村 美帆

主指導教官　鵜木 祐史
審査委員主査　鵜木 祐史
審査委員　赤木 正人
党 建武
吉高 淳夫

北陸先端科学技術大学院大学

先端科学技術研究科
（情報科学）

平成31年2月

Abstract

The speech is one of the main tools when person communicates. Various information are included in speech. Speech has not only linguistic information but also non-linguistic information and para-linguistic information. For example, non-linguistic information is speaker identification and vocal emotion. Para-linguistic information is emphasis and intention. These are significant on communication. There are fundamental frequency and formant frequency as main speech features. The main reason is that speech information can be thought to be derived from human vocal organs. The contributions of typical acoustic features communicated in speech, such as F0, spectral envelope, and power fluctuation, were investigated. There are temporal amplitude envelope and temporal fine structure focus on temporal frequency of speech. Shannon et al. showed that temporal amplitude envelope of speech contribute linguistic information perception by the experiments conducted with noise-vocoded speech. Noise-vocoded speech is speech have only information of temporal amplitude envelope. Noise-vocoded speech synthesized temporal fine structure was changed to white Gaussian noise. Zhu et al. showed around 4 Hz of modulation frequency plays important rolls for linguistic information perception of temporal amplitude envelope. In addition, they showed temporal amplitude envelope of speech plays important rolls for non-linguistic information perception. The cues of non-linguistic perception on temporal amplitude envelope of speech were clarified by the experiments of speaker identification and vocal emotion. Zhu et al. of studies on noise-vocoded speech showed that temporal modulation cues provided by the temporal amplitude envelope affect how vocal emotion and speaker individuality are perceived. However, it is still unclear whether the temporal modulation cues also play an important role in the perception of para-linguistic. Understanding the mechanism of how human can perceive para-linguistic information from speech should be very important for solving of the mechanism of speech perception. This study to investigate the contribution of temporal amplitude envelope of speech on para-linguistic information focus on perception of urgency. Natural disaster be frequent in Japan. The study of disaster prevention and the refuge instruction advances. In the cases, it is important to communicate sense of urgency with speech. Noise-vocoded speech is explain as hearing of cochlea implant. The hearing of cochlear implant will be improve if it is clarified the roll of temporal amplitude envelope of speech on urgency perception.

This study aims to investigate whether temporal amplitude envelope of speech affects urgency perception or not. Here, we experimentally investigated if the temporal amplitude envelope of speech affects the perception of para-linguistic information, particularly urgency. There were two questions. First, it clarify whether the perception of urgency was

included in temporal amplitude envelope. The order of urgency perception were compared between evacuation calls and noise-vocoded speech. The order will be the same if the perception of urgency is included in the temporal amplitude envelope of speech. Second, it clarify that modulation frequency cues on the urgency perception. The phycology scale of urgency perception were compared with the stimuli were limit modulation frequency. It was clarify important components of modulation frequency for the urgency perception.

Synthesizing method of noise-vocoded speech as stimuli is explained. The input signal was divided a band by auditory filter bank. Auditory filter bank is frequency resolving power like normal hearing. It was multiplied that temporal amplitude envelope and noise divided into each band. The provided amplitude modulation noise was made to add in whole area. In first experiments, it clarify whether the perception of urgency was included in the temporal amplitude envelope. Four evacuation calls and corresponded noise-vocoded speech as stimuli were used. The stimuli were presented in order of AB. The experiment participants were asked comparison with A B about how much urgency. The experiment participant was ten. Each of determine the number of times were 12 times. Determine the number of times on each conditions were 12 times. The psychological scale were calculated from evaluation values of urgency perception with paired comparison of Sheffe. As results, the order of psychological scale on urgency perception was the same. Thus, it clarified that urgency perception of speech is included in amplitude envelope. Psychological scale of urgency perception of each stimuli were compared. The psychological scale of urgency perception on noise-vocoded speech was large. Sharpness was calculated to examine the factor. The sharpness of the noise-vocoded speech of the same original stimuli were high overall. It suggested that the high sharpness is factor to have felt that urgency perception. 4Hz of at least modulation frequency is important to linguistic information perception of the amplitude envelope of speech.

Second experiments, it find that the cues of modulation frequency relation on urgency perception. The temporal resolution of noise-vocoded speech was controlled by varying the upper limit of modulation frequency for the stimuli. In the condition, cut-off frequencies of low-pass filter were 7 conditions (2, 4, 6, 8 12, 16 and 32 Hz) and high-pass filter were 7 conditions (4, 6, 8, 12, 16, 24 and 32 Hz). The question was the same in the first experiments. Determine the number of times on each conditions were 756 times. These were compared noise-vocoded speech in which the temporal amplitude envelopes were identical to those of the original speech, and noise-vocoded speech in which the temporal amplitude envelopes had undergone low-pass or high-pass filtering. Urgent scales were derived from a paired comparison of the results and used to investigate the relationship between the temporal modulation components and urgency perception. Results of the experiment in modulation frequency on low components. The psychological scale of urgency was decline from less than 6 Hz of cut-off frequency. Results of the experiment in

modulation frequency on high components. The psychological scale of urgency was decline from more than 8 Hz of cut-off frequency. Thus, it was clarified that the components from more than 6 Hz to less than 8 Hz of modulation frequency contribute perception of urgency. In experiment using LPF, around 8 Hz to16Hz is important to the non-language information of amplitude envelope.

Three things that this study clarified were indicated. The degree of urgency of the noise-vocoded speech stimuli was perceived as being similar to that of the original. Temporal modulation components of noise-vocoded speech upwards of 6 Hz were significant cues for urgency perception. Temporal modulation components of noise-vocoded speech downwards of 8 Hz were significant cues for urgency perception. The results suggest that temporal modulation cues in the temporal amplitude envelope play an important role in urgency perception.

目 次

第 1 章　序論 **1**
1.1　はじめに 1
1.2　研究背景 2
1.3　本研究の目的 3
1.4　本論文の構成 3

第 2 章　音声波形の振幅包絡線が音声知覚に与える影響 **6**
2.1　聴覚の時間情報処理 6
2.2　音声波形の振幅包絡線が言語情報知覚に与える影響 8
2.3　音声波形の振幅包絡線が非言語情報知覚に与える影響 8

第 3 章　本研究の着目点 **9**
3.1　雑音駆動音声 9
3.2　音声の緊迫感 13
3.3　研究計画 13

第 4 章　原音声と雑音駆動音声の比較による緊迫感知覚の検討 **15**
4.1　目的 15
4.2　緊迫感知覚の順位の比較による検討 15
4.2.1　実験方法 15
4.2.2　実験結果 16
4.3　緊迫感知覚の心理量の比較による検討 19
4.3.1　実験方法 19
4.3.2　実験結果 19
4.4　考察 22

第 5 章　変調周波数の制限による緊迫感知覚の手がかりの検討 **23**
5.1　目的 23
5.2　低域の変調周波数成分の検討 23
5.2.1　実験方法 23
5.2.2　実験結果 24
5.3　高域の変調周波数成分の検討 27

5.3.1 実験方法 27
5.3.2 実験結果 27
5.4 考察 31

第 6 章 総合考察 **32**

第 7 章 結論 **34**
7.1 明らかにしたこと 34
7.2 残された課題 34

謝辞 **36**

参考文献 **37**

研究業績 **41**

付録 **43**

付 録 1 雑音環境が音声波形の振幅包絡線の非言語情報知覚に及ぼす影響 **44**
A1.1 目的 44
A1.2 話者弁別実験 44
A1.2.1 実験方法 44
A1.2.2 実験結果 45
A1.3 感情認識実験 47
A1.3.1 実験方法 47
A1.3.2 実験結果 47
A1.4 考察 50

付 録 2 雑音残響環境が音声波形の振幅包絡線の非言語情報知覚に及ぼす影響 **51**
A2.1 目的 51
A2.2 話者弁別実験 51
A2.2.1 実験方法 51
A2.2.2 実験結果 52
A2.3 感情認識実験 54
A2.3.1 実験方法 54
A2.3.2 実験結果 54
A2.4 考察 57

図 目 次

1.1 本論文の構成 5

2.1 耳の構造 7

3.1 雑音駆動音声作成の概略図 10
3.2 ERB_N-number 尺度に従った 16 帯域の帯域通過フィルタの周波数応答 . . . 11
3.3 本研究の実験計画 14

4.1 原音声と雑音駆動音声の緊迫感についての心理尺度 17
4.2 直接比較実験における緊迫感についての心理尺度 20

5.1 カットオフ周波数の異なる低域通過フィルタをかけた
雑音駆動音声の緊迫感についての心理尺度 25
5.2 カットオフ周波数の異なる高域通過フィルタをかけた
雑音駆動音声の緊迫感についての心理尺度 29

A1.1 話者弁別率 46
A1.2 感情認識率 48
A1.3 感情ごとの感情認識率 49

A2.1 話者弁別率 53
A2.2 話者弁別実験の平均弁別率の分布図 53
A2.3 感情認識率 56
A2.4 感情認識実験の平均認識率の分布図 56

表 目 次

3.1 ERB_N-number 尺度に従った各帯域通過フィルタのカットオフ周波数 . . . 12

4.1 原音声と雑音駆動音声の緊迫感についての心理尺度 17
4.2 原音声と雑音駆動音声の実験における分散分析表 18
4.3 直接比較実験における緊迫感についての心理尺度 20
4.4 直接比較実験の分散分析表 21
4.5 同じ原音声の刺激におけるシャープネスの違い 22

5.1 カットオフ周波数の異なる低域通過フィルタをかけた
雑音駆動音声の緊迫感についての心理尺度 25
5.2 カットオフ周波数の異なる低域通過フィルタをかけた
雑音駆動音声の実験における分散分析表 26
5.3 カットオフ周波数の異なる高域通過フィルタをかけた
雑音駆動音声の緊迫感についての心理尺度 29
5.4 カットオフ周波数の異なる高域通過フィルタをかけた
雑音駆動音声の実験における分散分析表 30

第1章　序論

1.1　はじめに

ヒトがコミュニケーションを図る上で，音声は重要なコミュニケーションツールの一つである．発話者が音声で何かを伝えようとし，聴取者がそれを知覚することで音声によるコミュニケーションが成り立つ．音声の主な音響特徴として，基本周波数やフォルマント周波数がある．しかし，これらは性別，年齢，発話者の状態によって異なる [1]．つまり，音声は，メッセージとしての情報だけでなく，個人性や感情などの多様な情報を含んでいる．Fujisaki は，音声情報を言語情報，非言語情報，パラ言語情報に分類し，言語情報と非言語，パラ言語情報は言語メッセージを含んでいるかどうか，非言語情報とパラ言語情報は話し手の意図の有無によって分類した [2]．しかし，音声によって伝わる情報に関しては聴取者の心理状態や能力にも依存する．例えば，年齢による捉え方の違いや聴覚の能力による場合などがある．

日本では，高度感音性難聴と呼ばれる重度の聴覚障がい者は約 35 万人といわれている．難聴には，鼓膜や耳小骨に障がいのある伝音性難聴と蝸牛に障がいのある感音性難聴がある．人工内耳はこの感音性難聴者に適応され，日本の人工内耳装用者の人数は約 1 万人程度といわれている．人工内耳とは，音を電気信号に変え，蝸牛の中に埋め込んだ電極で直接聴神経を刺激することで，音や言葉を認識させる装置である [3]．人工内耳は，重度の聴覚障がい者にとって，音声言語の獲得や活用が可能となる有効な機器である．聴神経に直接電気信号を送る人工内耳装用者の聴こえは音声波形の振幅包絡線情報のみを保存した音声の聴こえと解釈できることが報告されている [4]．この音声波形の振幅包絡線情報のみを保存した合成音声として雑音駆動音声がある．雑音駆動音声を用いた研究によって，音声波形の振幅包絡線情報の重要性や人工内耳装用者の聴こえが明らかにされてきた [4]．特に，音声波形の振幅包絡線情報が言語・非言語情報知覚に与える影響は検討されてきた [4–6]．しかし，音声波形の振幅包絡線が音声知覚に与える影響を議論するためには，パラ言語情報知覚の知見が必要である．

本研究では，音声波形の振幅包絡線がパラ言語情報知覚に及ぼす影響を検討し，音声波形の振幅包絡線が音声知覚に及ぼす影響の知見を広げることを目的とする．昨今，日本では災害に対する関心が高まっている．多くの犠牲者を出した東日本大震災はまだ記憶に新しい．避難喚起のために，東日本大震災では多くの防災行政無線放送が行われた．茨城県大洗町では，4 mもの津波に襲われながらも死者は一人もいなかった．井上は,「緊急避難命令，大至急，高台へ避難せよ」という異例とも言える命令調の防災行政無線が少なか

らず影響した結果だと報告している [7]．宮城県南三陸町危機管理課職員の遠藤さん（当時 24 歳）は防災行政無線で繰り返し住民へ避難を呼びかけ，命を落とした．住民らは遠藤さんの切羽詰まった声に救われたと話す [8]．この町の避難呼びかけでは「緊張感」を持たせるためにあえて男性の声でも呼びかけていた [9]．このような教訓から，音声の緊迫感に着目した研究が進んでいる．そこで，本研究では，パラ言語情報の一つとして音声の緊迫感をとりあげ，音声波形の振幅包絡線が緊迫感知覚に及ぼす影響を明らかにする．音声波形の振幅包絡線情報が緊迫感知覚に与える影響が明らかになれば，人工内耳装用者を含むあらゆる人々に緊迫感を伝えることに貢献できる．

1.2 研究背景

音声波形の振幅包絡線は音声知覚において重要である．音声情報には言語，非言語，パラ言語情報が存在する．これらは，Fujisaki らによって話し手の立場から分類されている．言語情報は言語メッセージ，非言語情報は感情や個人性などの話し手が意図していない情報，パラ言語情報は強調などの話し手が意図している情報である [2]．Shannon らは音声波形の振幅包絡線情報のみを保存した雑音駆動音声を用いて，言語情報知覚に音声波形の振幅包絡線が及ぼす影響を明らかにした [4]．雑音駆動音声とは，複数の帯域に分割し，各帯域の振幅包絡線を抽出し，白色性ガウス雑音と乗じることで得られる．Shannon らによると音声波形を少なくとも 4 つの帯域に分割すれば，音声波形の振幅包絡線情報のみで言語情報知覚が可能であることを明らかにした [4]．Drullman らは，音声の了解度について，音声の振幅包絡線の変調周波数成分の 4 Hz から 16 Hz までが重要であることを報告した [10, 11]．Dau は．変調フィルタの存在を支持し，変調フィルタバンクの機能モデルを提案した [12]．Zhu らは，音声の振幅包絡線の変調周波数を制限することで，言語情報，非言語情報知覚における音声波形の振幅包絡線の手がかりを明らかにした [6]．しかし，音声波形の振幅包絡線がパラ言語情報知覚に及ぼす影響は明らかにされていない．

本研究では，パラ言語情報の一つである音声の緊迫感について着目する．東日本大震災以来，災害時の情報伝達については多くの議論がなされてきた．特に，音声分野では，防災無線放送の改善のための研究が求められている．佐藤らは，多言語対応の情報伝達について，多言語一斉放送の可能性を検討した [13]．また，高野・土田は避難行動を喚起するための防災放送の実現に向けて検討した [14, 15] 音声の緊迫感は，災害時などの情報の伝達が限られた環境下で危険を伝えるために重要であり，検討が必要である．音声の緊迫感について，言語学的な観点からは，避難伝達文では短文，命令調であれば緊迫感が高いことが報告されている [17]．また，心理学，音響特徴の観点からは，緊迫感及び避難誘導効果の高い音声は，話速が速く，基本周波数が高く，中高域の周波数スペクトルが大きいことが報告されている [16, 18, 19]．

1.3　本研究の目的

本研究では，音声波形の振幅包絡線がパラ言語情報知覚に及ぼす影響を検討し，音声波形の振幅包絡線が音声知覚に及ぼす影響の知見を広げることを目的とする．そこで，音声の緊迫感に着目し，音声波形の振幅包絡線が緊迫感知覚に及ぼす影響について検討する．まず，音声波形の振幅包絡線に緊迫感知覚に関わる情報が含まれていると仮説を立て，緊迫感を有する音声刺激と雑音駆動音声の緊迫感の程度を比較することで検証する．

次に，音声波形の振幅包絡線の変調周波数成分の手がかりを明らかにするために，雑音駆動音声の変調周波数を制限させた刺激を用いて緊迫感の程度を比較し，検討する．

最後に，音声波形の振幅包絡線が音声知覚に与える影響について，言語情報，非言語情報，パラ言語情報をそれぞれについて論じ，総合的に考察する．

1.4　本論文の構成

本論文は 7 章で構成される．図 1.1 に本論文の構成を示し，以下で詳しく述べる．

第 1 章

音声知覚に注目した背景，人工内耳との関連を述べ，パラ言語情報の一つとして着目する音声の緊迫感について述べることで本研究の目的を明らかにする．

第 2 章

聴覚の生理的なメカニズムとその情報処理，主に時間情報処理に関して述べた後，音声波形の振幅包絡線が音声知覚に与える影響を詳しく説明する．

第 3 章

本研究の着目点について述べる．音声波形の振幅包絡線が音声の緊迫感にどのような影響を及ぼしているのかを検討するための実験計画を述べる．

第 4 章

音声波形の振幅包絡線が緊迫感知覚にどのような影響を与えているか解明すべく，緊迫感を有する音声と雑音駆動音声を用いて，そもそも音声の振幅包絡線が緊迫感知覚に影響を与えているのかを明らかにする．

第 5 章

音声波形の振幅包絡線の何を手がかりとして，音声の緊迫感を感じていいるのかを解明すべく，雑音駆動音声の変調周波数を制限させることで，緊迫感知覚の手がかりとなる変調周波数成分を明らかにする．

第 6 章

第 4 章，第 5 章での結果から，音声波形の振幅包絡線が緊迫感知覚に及ぼす影響を論じる．さらに，音声波形の振幅包絡線が音声知覚に与える影響として，言語情報，非言語情報，パラ言語情報のそれぞれの傾向を総合的に論じる．

第 7 章

本研究で明らかにしたこと，残された課題について述べる．

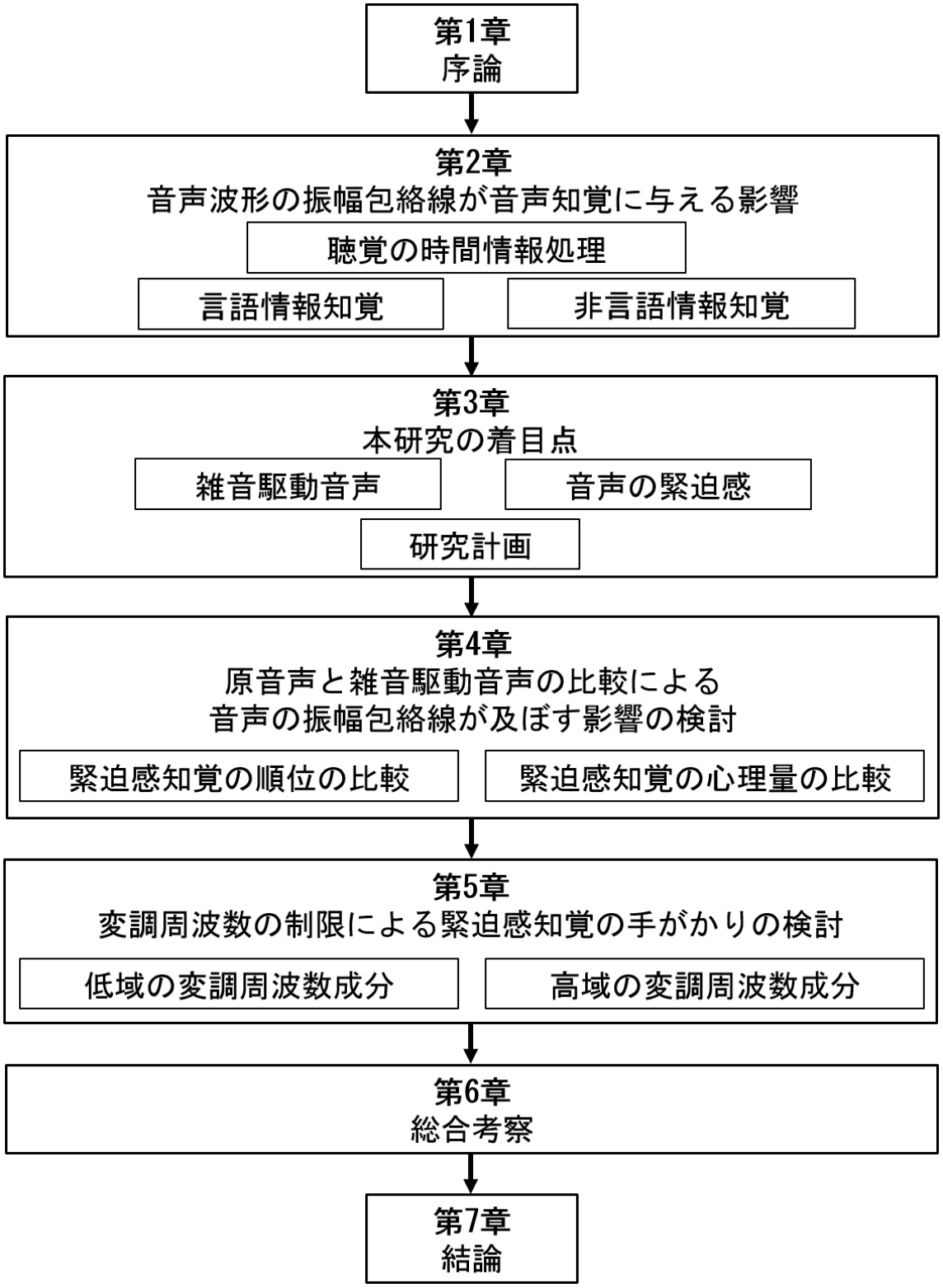

第1章
序論
第2章
音声波形の振幅包絡線が音声知覚に与える影響
聴覚の時間情報処理
言語情報知覚
非言語情報知覚
第3章
本研究の着目点
雑音駆動音声
音声の緊迫感
研究計画
第4章
原音声と雑音駆動音声の比較による
音声の振幅包絡線が及ぼす影響の検討
緊迫感知覚の順位の比較
緊迫感知覚の心理量の比較
第5章
変調周波数の制限による緊迫感知覚の手がかりの検討
低域の変調周波数成分
高域の変調周波数成分
第6章
総合考察
第7章
結論

図 1.1: 本論文の構成

第2章　音声波形の振幅包絡線が音声知覚に与える影響

2.1　聴覚の時間情報処理

音は圧力の時間的変化である．聴覚抹消系は，外耳，内耳および中耳で構成される．図2.1に，耳介から蝸牛までの概要図を示す．耳介で集約された音は，外耳道を通り，鼓膜で機械的振動に変換され，耳小骨で増幅される．増幅された音は，リンパ液の満たされた蝸牛に伝わり，周波数ごとに特定の場所を共振させる．そこで，電気信号が発生し，聴神経に伝わる．この周波数ごとの処理がフーリエ変換に置き換えられ，聴覚フィルタバンクとしてモデル化される．今回，聴覚フィルタバンクとして用いるのは，健聴者の聴覚フィルタの並びを模擬した $\mathrm{ERB_N}$-number 尺度であり，ノッチ雑音を用いた実験によって推定された [22]．さらに，聴覚の時間的な処理を理解するためには，音波形を振幅包絡線（変調波）と時間微細構造（搬送波）に分けて考える．音波形の振幅包絡線は時間的にゆるやかな変化であり，時間微細構造は，時間的に速い変化であり，時間微細構造に包絡線を重畳させたものといえる．振幅包絡線が語音の情報を表現するのに対し，時間微細構造はピッチ知覚や両時間差に基づく音源定位への貢献が優位である [20]．聴覚の末梢系では，音信号を聴覚フィルタバンクで帯域分割する周波数分解能が備わっている．また，内有毛細胞や神経発火のメカニズムによって半波整流と低域通過フィルタを通すのと同等の処理が行われている [12]．これが音波形の振幅包絡線にあたる処理である．また，聴覚の末梢系で振幅包絡線情報を処理する際に周波数選択性（変調フィルタバンク）があるとされ，Dau らは変調フィルタバンクの存在を示唆している [21]．変調フィルタバンクは脳内で変調スペクトル分析を行っているという仮説がある [12,21]．また，変調フィルタの存在を心理物理的に示唆した報告 [23,24] や変調周波数に選択性を持つニューロンの存在が示されている [25,26]．変調フィルタの比較的簡単な統計量で環境音が表現および再現できることも報告されている [27]．

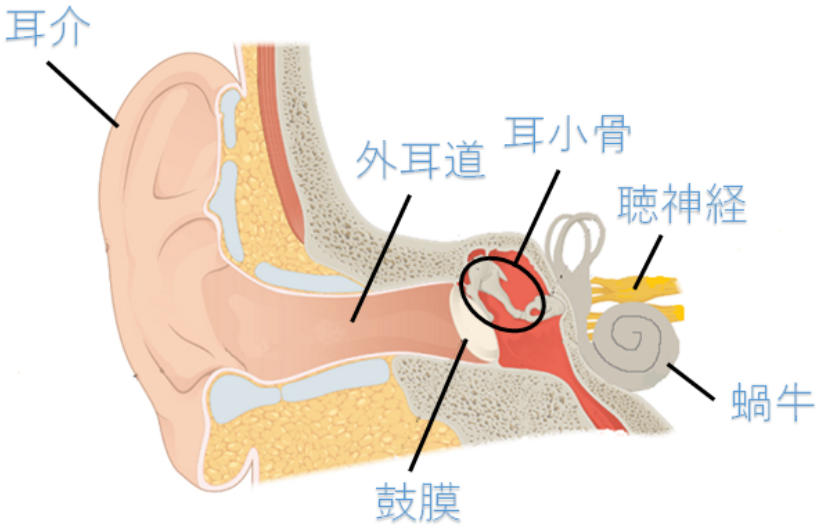


図 2.1: 耳の構造

2.2 音声波形の振幅包絡線が言語情報知覚に与える影響

雑音駆動音声は時間微細構造を雑音に置き換え，音声波形の振幅包絡線情報のみを保存した音声である．音声波形の時間微細構造には，音声の基本周波数やフォルマント周波数などの音声知覚に重要な成分が含まれている．しかし, 音声波形の振幅包絡線情報のみでも言語情報が知覚できる．Shannon らは帯域数を変化させた雑音駆動音声を用いて，音声波形の振幅包絡線における言語情報知覚への影響を検討した．その結果，雑音駆動音声を合成する際の帯域分割数が少なくとも 4 帯域あれば, 音声波形の振幅包絡線における言語情報を知覚できることを明らかにした [4]．また，Drullman は振幅包絡線の変調周波数を制限した刺激を用いて，音声了解度に重要な変調周波数成分を検討した．その結果，変調周波数の 4 Hz から 16 Hz までが重要であることを示唆した [10,11]．Zhu らは，音声波形の振幅包絡線の変調周波数の制限によって，言語情報に重要な変調周波数成分を調べた．その結果，5 Hz 未満の成分が重要であることを明らかにした．このことから，モーラの時間変動成分があれば，言語情報を知覚できるということが示唆された [6].

2.3 音声波形の振幅包絡線が非言語情報知覚に与える影響

音声波形の振幅包絡線における言語情報知覚については，多くの研究がなされてきた．しかし，音声情報には，言語情報だけでなく非言語情報やパラ言語情報が含まれている．Zhu らは，音声波形の振幅包絡線の変調周波数の制限によって，音声波形の振幅包絡線の非言語情報知覚に重要な変調周波数成分を検討した．その結果，変調周波数の 8 Hz から 16 Hz 付近が重要であることを明らかにした [6]．さらに，筆者らは，音声波形の振幅包絡線に含まれる個人性および感情の知覚における雑音や残響の影響を検討した．その結果, かなり劣悪な環境でない限り，音声波形の振幅包絡線に含まれる個人性や感情は知覚できることを明らかにした [28–30]．しかし，人工内耳装用者の聞こえにおける雑音や残響の影響はいまだに大きい [3].

第3章　本研究の着目点

3.1　雑音駆動音声

図3.1に雑音駆動音声作成の概略図を示す．まず，次式に示すように入力信号 $s(t)$ を様々な帯域に分割する．

$$s_k(t) = s(t) * h_{B,k}(t) \tag{3.1}$$

ここで，k は帯域通過フィルタのチャンネル番号を示し，$s_k(t)$ はチャンネル k における帯域信号である．また，$h_{B,k}(t)$ は帯域通過フィルタである．同時に，白色性ガウス雑音 $N(t)$ も同じ帯域通過フィルタで次式に示すように帯域分割する．

$$N_k(t) = N(t) * h_{B,k}(t) \tag{3.2}$$

$N_k(t)$ はチャンネル k における狭帯域雑音である．この処理はヒトの聴覚における周波数分解能を模擬した処理であり，以降，聴覚フィルタバンクとする．本研究では，聴覚フィルタとして，6次のIIR（Infinite Impulse Response）型のButterworth帯域通過フィルタを用いる．また，この帯域通過フィルタの帯域幅を $\mathrm{ERB_N}$(Equivalent Rectangular Bandwidth) となるように設計する．$\mathrm{ERB_N}$ は健聴者の聴覚フィルタの帯域幅を指し，聴覚フィルタの帯域幅を等価矩形帯域幅（Equivalent Rectangular Bandwidth）で近似したものである．帯域通過フィルタバンクにおけるフィルタの並びは，次式で定義される $\mathrm{ERB_N}$-number 尺度に従って，$\mathrm{ERB_N}$-number が3から35までの32個の帯域通過フィルタを配置する．図3.2に $\mathrm{ERB_N}$-number 尺度に従った16帯域の帯域通過フィルタの周波数応答を示す．

$$\mathrm{ERB_N} - \mathrm{number} = 21.4 \log_{10}\left(\frac{4.37f}{1000} + 1\right) \tag{3.3}$$

ここで，f は周波数 (Hz)，下付きの N は健聴者 (Normal Hearing) の特性であることを示す．本研究では，聴覚フィルタの帯域幅を $2\mathrm{ERB_N}$ とし，16帯域に分割する．表3.1に $\mathrm{ERB_N}$-number 尺度に従った各帯域通過フィルタのカットオフ周波数を示す．次に，Hilbert 変換 $H_k(t)$ と低域通過フィルタ $h_L(t)$ を用いて，音声信号の振幅包絡線 $e_k(t)$ を抽出する．低域通過フィルタとして，2次のIIR型Butterworth低域通過フィルタを用いる．

$$e_k(t) = |s_k(t) + jH_k(t)| * h_L(t) \tag{3.4}$$

低域通過フィルタのカットオフ周波数は実験に応じて異なるため，詳しい条件については各節で述べる．最後に，各帯域の振幅包絡線 $e_k(t)$ と狭帯域雑音 $N_k(t)$ を乗じ，各帯域の

振幅変調雑音を全帯域で足し合わせることで雑音駆動音声を作成した.

$$y(t) = \sum_{i=1}^{K} e_k(t) \times N_k(t) \tag{3.5}$$

$y(t)$ は作成された雑音駆動音声，K はチャンネル数を示す.

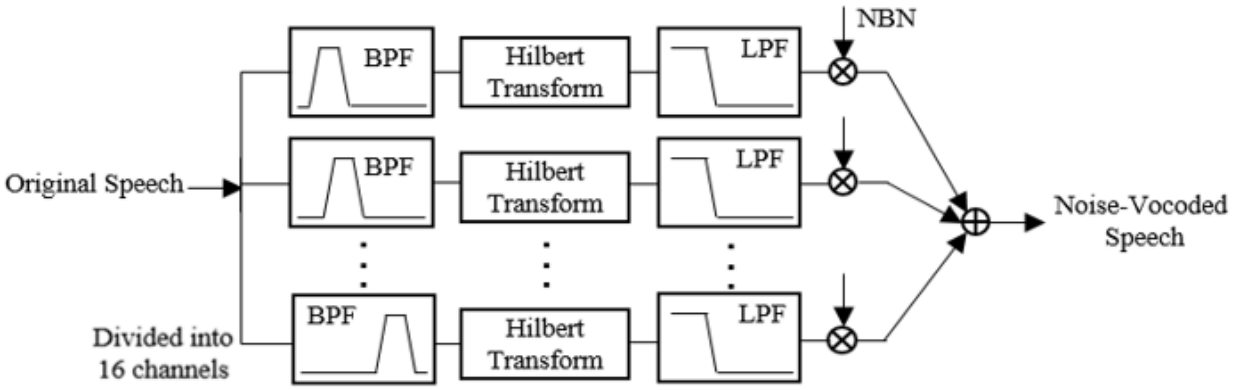


図 3.1: 雑音駆動音声作成の概略図

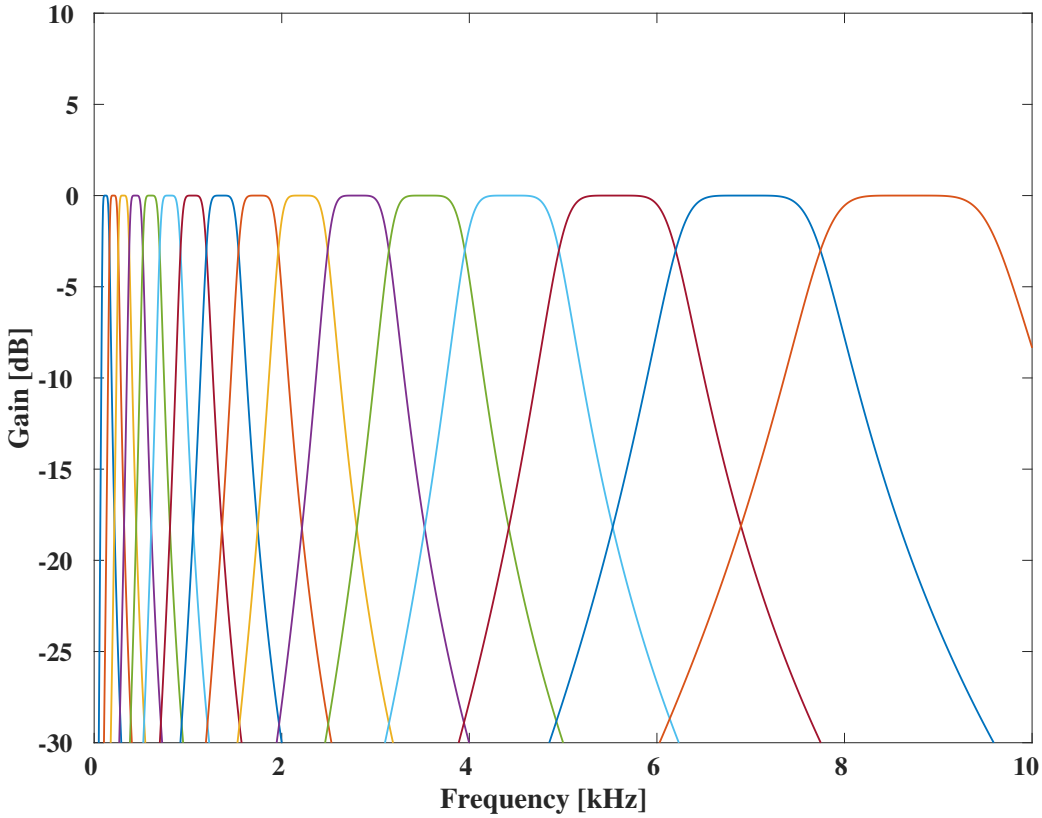


図 3.2: $\mathrm{ERB_N}$-number 尺度に従った 16 帯域の帯域通過フィルタの周波数応答

表 3.1: $\mathrm{ERB_N}$-number 尺度に従った各帯域通過フィルタのカットオフ周波数

チャンネル番号	$\mathrm{ERB_N}$-number	カットオフ周波数 [kHz]
1	3	0.0872
2	5	0.163
3	7	0.257
4	9	0.374
5	11	0.519
6	13	0.698
7	15	0.921
8	17	1.20
9	19	1.54
10	21	1.96
11	23	2.49
12	25	3.14
13	27	3.95
14	29	4.96
15	31	6.20
16	33	7.74
	35	9.66

3.2 音声の緊迫感

緊迫とは，きびしくさしせまる様子である [31]．それは，行動，表情，声など，様々な情報を駆使して伝えようとする．そもそも，音声の緊迫感はパラ言語情報に分類されるが，発話者が緊迫感を意図しているつもりであっても，受聴者に伝わるかどうかは定かではない．それも，受聴者の状態や心理的な状況が大きく関わっているからである．それでは，どのようにして音声で緊迫感を伝えれば良いのか．本研究では，緊迫感を有する音声刺激を用いて，音声波形の振幅包絡線にどのような影響を与えているのかを明らかにする．

本研究では，小林・赤木の研究で使用された男性アナウンサー１名による避難呼びかけ音声を刺激として用いた [18]．発話内容は,「今すぐ逃げてください」の１種類のみであった．本研究では，緊迫感の異なる４種類の避難呼びかけ音声を用いた．これら４種類の避難呼びかけ音声を A， B， C， D でラベル付けした．また，これらの避難呼びかけ音声に対応する雑音駆動音声を a， b， c， d でラベル付けした．

3.3 研究計画

本研究では，音声波形の振幅包絡線が音声知覚に及ぼす影響の知見を広げるために，パラ言語情報の一つである音声の緊迫感に着目し，音声波形の振幅包絡線がパラ言語情報知覚に及ぼす影響を検討する．どのようにこの研究を進めていくかを図 3.3 に本研究の実験計画として示す．

第４章では，原音声と雑音駆動音声の比較によって音声波形の振幅包絡線が緊迫感知覚に及ぼす影響を検討する．本実験の目的は，音声波形の振幅包絡線が緊迫感知覚に影響しているかを明らかにすることである．音声波形の振幅包絡線が緊迫感知覚に影響を及ぼしているならば，両刺激の緊迫感を感じる音声の順序が一致すると仮説をたてた．方法としては，原音声と雑音駆動音声を用いて，被験者に音声の緊迫感について評価させ，一対比較法によって音声の緊迫感についての心理尺度を算出し，順位と心理量を比較する．

第５章では，変調周波数を制限した雑音駆動音声を用いて，音声波形の振幅包絡線に含まれる緊迫感知覚の手がかりを検討する．本実験の目的は，音声波形の振幅包絡線に含まれる緊迫感知覚に重要な変調周波数成分を明らかにすることである．これまでの音声波形の振幅包絡線における言語・非言語情報知覚の結果から，緊迫感知覚においても特定の変調周波数成分が音声波形の振幅包絡線における緊迫感知覚に影響を及ぼしていると仮説をたてた．方法としては，カットオフ周波数を変化させた低域通過フィルタと高域通過フィルタをかけたそれぞれの雑音駆動音声を用いて，被験者に音声の緊迫感について評価させ，一対比較法によって音声の緊迫感についての心理尺度を算出し，両刺激で心理量を検討する．

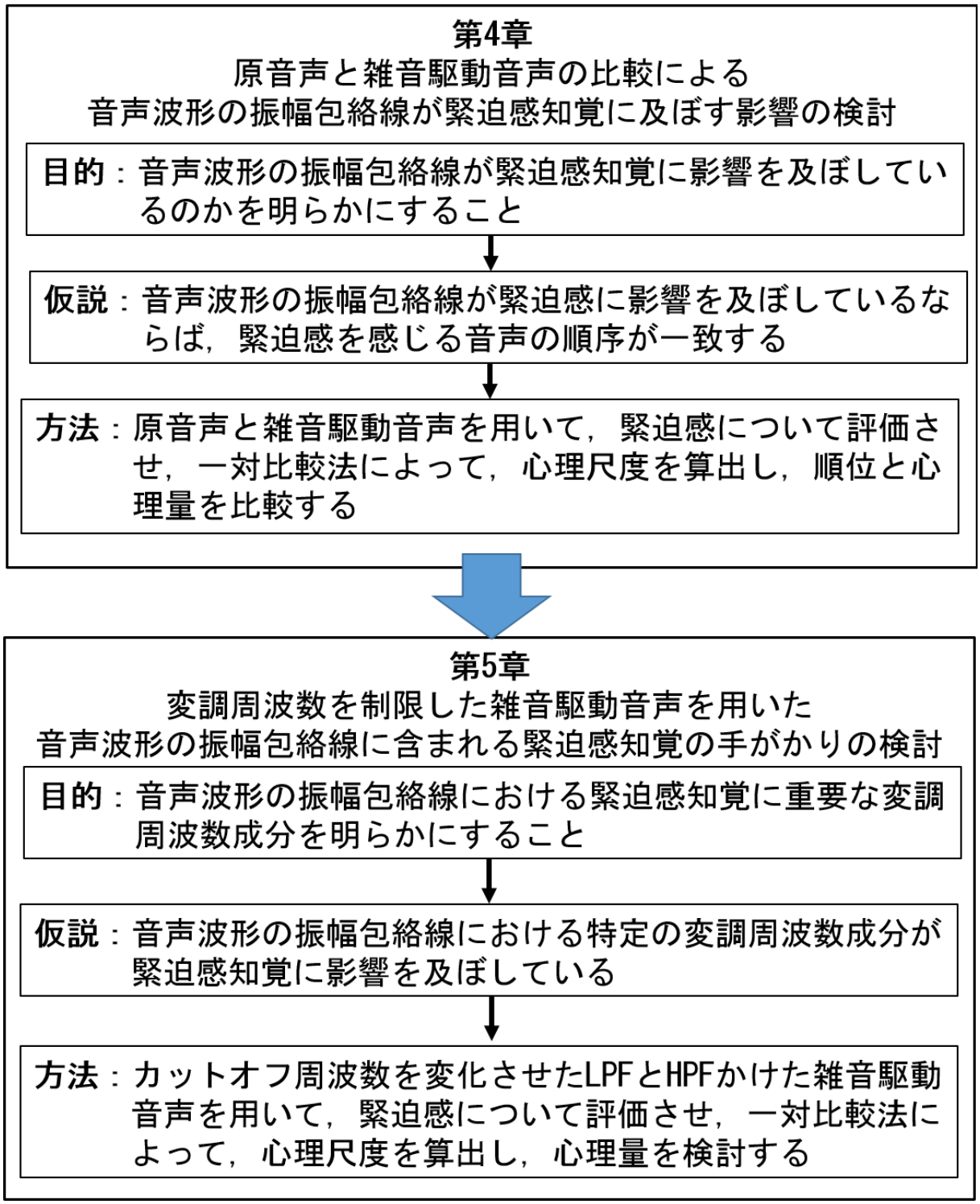


図 3.3: 本研究の実験計画

第4章　原音声と雑音駆動音声の比較による緊迫感知覚の検討

4.1　目的

本章は，音声波形の振幅包絡線が緊迫感知覚に寄与しているのかを明らかにすることを目的とする．4.2 では，振幅包絡線情報のみが保存された雑音駆動音声を用いて実験を行い検討することで，音声波形の振幅包絡線が音声の緊迫感に影響を及ぼしているのかを明らかにする．4.3 では，雑音駆動音声が緊迫感知覚に与える影響について検討する．

4.2　緊迫感知覚の順位の比較による検討

本実験では，音声波形の振幅包絡線が緊迫感の知覚に影響を及ぼしているのかを明らかにするために，振幅包絡線情報のみが保存された雑音駆動音声を用いて，音声の緊迫感について聴取実験を行い検討した．

4.2.1　実験方法

参加者

北陸先端科学技術大学院大学の 22 歳から 25 歳までの学生 10 名（男性 7 名，女性 3 名）が実験に参加した．被験者は，母国語が日本語であり，日常生活に支障のない程度の聴力を有していた．

刺激

第 3 節で記述した避難呼びかけ音声 4 種類とそれらをもとに作成した雑音駆動音声の計 8 刺激を実験に用いた．ここで，雑音駆動音声作成の際の低域通過フィルタのカットオフ周波数は 64 Hz であった．刺激の呈示には，PC（Windows 10， MATLAB），オーディオインターフェース（Fireface UCX），ヘッドフォン（SENNHEISER HDA 200）を用いた．

手続き

実験は防音室で行った．原音声（4 個）と雑音駆動音声（4 個）について，それぞれの刺激間で実験を行った．被験者には，A，B の順で音声刺激を呈示し，A に比べて B はどの程度緊迫してるか，あるいはしていないか（B の音声について答えてください）について質問した．その問いに対して，かなり緊迫している，やや緊迫している，同程度，やや緊迫していない，かなり緊迫しているの 5 段階で評価するように求めた．この緊迫感についての評価結果からシェッフェの一対比較法（浦の変法）[32–35] を用いて，刺激に対する音声の緊迫感についての心理尺度を算出した．順序効果を考慮して，それぞれの実験の総判断回数は 12 回であった．また，音声刺激の呈示順は被験者ごとにランダムとし，A と B の音声刺激の間隔は 0.5 秒であり，実験に要した時間は 10 分程度であった．

4.2.2 実験結果

原音声と雑音駆動音声のそれぞれの音声刺激を用いた実験における緊迫感についての評価結果から，シェッフェの一対比較法を用いて，音声の緊迫感についての心理尺度を求めた．その結果，原音声を用いた実験では，C， D， B， A の順に音声の緊迫感を高く感じることがわかった．分散分析を行った結果，刺激に対する主効果（$F(3, 77) = 34.42, p < 0.01$）が認められ，A と B の刺激間を除いて，有意差（$p < 0.01$）が認められた．雑音駆動音声を用いた実験では，c，d，b，a の順に音声の緊迫感を高く感じることがわかった．分散分析を行った結果，刺激に対する主効果（$F(3, 77) = 125.8, p < 0.01$）が認められ，すべての刺激間で有意差（$p < 0.01$）が認められた．心理量について着目すると，雑音駆動音声の緊迫感についての心理尺度は原音声の緊迫感についての心理尺度よりも 0 から離れていることがわかった．つまり，緊迫感の程度の差について，原音声よりも雑音駆動音声に緊迫感の程度の違いを感じていることがわかった．この要因については，次節で実験を行い検討する．

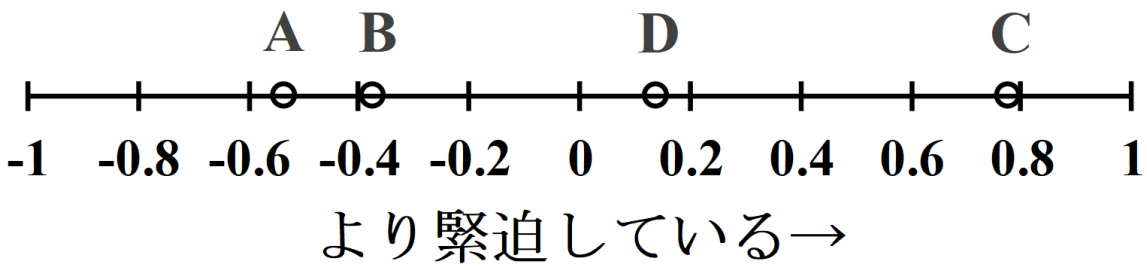
表 4.1: 原音声と雑音駆動音声の緊迫感についての心理尺度

(a) 原音声

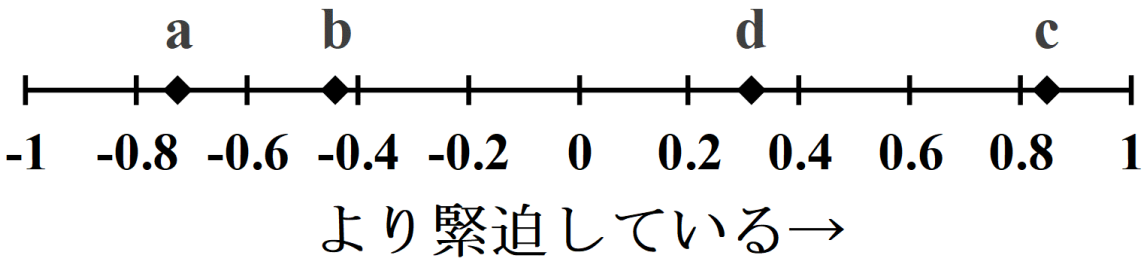
刺激ラベル	心理尺度
A	−0.5375
B	−0.3750
D	0.1375
C	0.7750

(b) 雑音駆動音声

刺激ラベル	心理尺度
a	−0.7250
b	−0.4375
d	0.3125
c	0.8500



(a) 原音声



(b) 雑音駆動音声

図 4.1: 原音声と雑音駆動音声の緊迫感についての心理尺度

表 4.2: 原音声と雑音駆動音声の実験における分散分析表

(a) 原音声の実験における分散分析表

要因	平方和	自由度	不偏分散	F 値	p 値
主効果	83.93	3	27.98	34.42	0.0000
主効果×個人	10.33	27	0.3824	0.4706	0.9849
組み合わせ効果	13.68	3	4.558	5.609	0.0016
順序効果	1.633	1	1.633	2.010	0.1603
順序×個人	5.867	9	0.6519	0.8021	0.6155
誤差	62.58	77	0.8127		
全体	178.0	120			
$F(3, 80; 0.01) = 4.036$					

(b) 雑音駆動音声の実験における分散分析表

要因	平方和	自由度	不偏分散	F 値	p 値
主効果	123.0	3	40.99	125.8	0.0000
主効果×個人	17.28	27	0.6398	1.963	0.0115
組み合わせ効果	2.325	3	0.7750	2.378	0.0763
順序効果	1.200	1	1.200	3.683	0.0587
順序×個人	3.133	9	0.3481	1.068	0.3956
誤差	25.09	77	0.3259		
全体	172.0	120			
$F(3, 80; 0.01) = 4.036,\ F(24, 80; 0.05) = 1.654$					

4.3 緊迫感知覚の心理量の比較による検討

4.2 の実験結果から，原音声と雑音駆動音声で緊迫感についての心理量が僅かに異なり，雑音駆動音声の緊迫感についての心理尺度が 0 から離れていることがわかった．本実験では、この要因について検討するために，原音声と雑音駆動音声計 8 個の刺激を直接比較する実験を行い検討した．

4.3.1 実験方法

参加者

北陸先端科学技術大学院大学の 22 歳から 25 歳までの学生 10 名（男性 8 名，女性 2 名）が実験に参加した．被験者は，母国語が日本語であり，日常生活に支障のない程度の聴力を有していた．

刺激

4.2 と同様の音声刺激 8 個を用いた．そのため，雑音駆動音声作成の際の低域通過フィルタのカットオフ周波数は，4.2 の実験と同様に 64 Hz であった．刺激の呈示も 4.2 と同様の機器を用いた，

手続き

実験は防音室で行った．避難呼びかけ音声の刺激（4 個）と雑音駆動音声の刺激（4 個）について，計 8 個の音声刺激について実験を行った．問いと評価項目は 4.2 の実験と同様であった．緊迫感についての評価結果からシェッフェの一対比較法（浦の変法）を用いて，刺激に対する音声の緊迫感についての心理尺度を算出した．順序効果を考慮して，実験の総判断回数は 56 回であった．また，音声刺激の呈示順は被験者ごとにランダムとし，A と B の音声刺激の間隔は 0.5 秒であり，実験に要した時間は 20 分程度であった．

4.3.2 実験結果

原音声と雑音駆動音声の計 8 個の音声刺激を用いた実験における緊迫感についての評価結果から，シェッフェの一対比較法を用いて，緊迫感についての心理尺度を求めた．その結果，音声刺激は，c， C， d， D， b， B， a， A の順に刺激の緊迫感を高く感じることがわかった．分散分析を行った結果，刺激に対する主効果（$F(7, 459) = 100.8, p < 0.01$）が認められた．有意差が認められなかった刺激間は，a と B， a と b， D と d， D と C， d と C， d と c， C と c であり，その他の刺激間には有意差（$p < 0.01$）が認められた．

表 4.3: 直接比較実験における緊迫感についての心理尺度

刺激ラベル	心理尺度
A	−0.7625
a	−0.4188
B	−0.3187
b	−0.2562
D	0.2188
d	0.3750
C	0.5125
c	0.3750

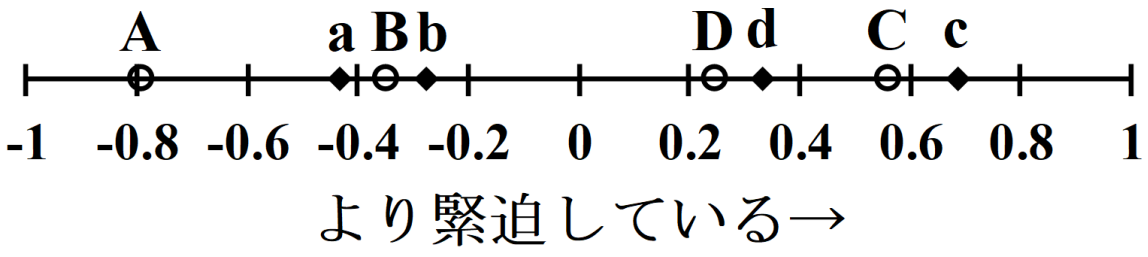


図 4.2: 直接比較実験における緊迫感についての心理尺度

表 4.4: 直接比較実験の分散分析表

要因	平方和	自由度	不偏分散	F 値	p 値
主効果	287.6	7	41.09	100.8	0.0000
主効果×個人	64.63	63	1.026	2.517	0.0000
組み合わせ効果	18.98	21	0.9036	2.217	0.0016
順序効果	8.257	1	8.257	20.26	0.0000
順序×個人	13.46	9	1.495	3.669	0.0002
誤差	187.1	459	0.4075		
全体	580.0	560			

$F(1, \infty; 0.01) = 3.841,\ F(7, \infty; 0.01) = 2.010,\ F(9, \infty; 0.01) = 1.880$
$F(20, \infty; 0.01) = 1.571,\ F(60, \infty; 0.01) = 1.318$

4.4 考察

原音声と雑音駆動音声の緊迫感の評価結果から心理尺度を求めた結果，原音声と雑音駆動音声で緊迫感を感じる刺激の順序が同じであることがわかった．このことは，音声波形の振幅包絡線に緊迫感の程度に関わる情報が含まれていることを示唆している．また，原音声と雑音駆動音声を直接比較させると，原音声に比べて雑音駆動音声の緊迫感を高く評価する傾向が見られた．雑音駆動音声を合成することで，音声の緊迫感を感じる成分が増加した，あるいは緊迫感を抑制している成分が減少したことが考えられる．濱田らは，同じ音声刺激を用いて，緊迫感および避難誘導効果の高い音声は，中高域の周波数スペクトルが大きいことを報告している [19]．このことから，雑音駆動音声に高い緊迫感を感じたのは，音の甲高さや鋭さに起因するものではないかと仮説をたて，次式に定義するシャープネスを算出した．

$$S = 0.11 \frac{\int_0^{24\mathrm{Bark}} N'(z) g'(z) dz}{\int_0^{24\mathrm{Bark}} N'(z) dz} \tag{4.1}$$

ここで，$N'(z)$ は臨界帯域ごとのラウドネス，$g'(z)$ はシャープネスの重み，z は臨界帯域番号である [36]．同じ原音声の刺激間のシャープネスの違いを表 4.5 に示す．その結果，同じ原音声の刺激間を比較したとき，原音声に比べて雑音駆動音声のシャープネスの値が高いことがわかった．同じ原音声の各刺激におけるシャープネスの差に着目すると，同程度であることがわかった．このことから，雑音駆動音声を合成することにより，音の甲高さや鋭さが増し，音声の緊迫感を高く感じていた可能性が考えられる．

表 4.5: 同じ原音声の刺激におけるシャープネスの違い

刺激ラベル	A(a)	B(b)	C(c)	D(d)
原音声	1.45	1.55	1.55	1.51
雑音駆動音声	1.73	1.84	1.82	1.77
差	0.28	0.29	0.27	0.26

第5章　変調周波数の制限による緊迫感知覚の手がかりの検討

5.1　目的

第4章で，音声波形の振幅包絡線が緊迫感の程度の知覚に寄与していることを明らかにした．本章では，音声波形の振幅包絡線の緊迫感知覚の手がかりを明らかにすることを目的とする．音声波形の振幅包絡線の変調周波成分について検討するために，カットオフ周波数を変化させた雑音駆動音声を用いて実験を行い検討した．5.2では，音声波形の振幅包絡線にカットオフ周波数の異なる低域通過フィルタをかけた雑音駆動音声を用いて実験を行い，低域の変調周波数成分の緊迫感知覚の手がかりを探索した．5.3でも同様に，カットオフ周波数の異なる高域通過フィルタをかけた雑音駆動音声を用いて実験を行い，高域の変調周波数成分の緊迫感知覚の手がかりを探索した．

5.2　低域の変調周波数成分の検討

本実験では，低域の変調周波数の成分における緊迫感知覚の手がかりを探索するために，音声波形の振幅包絡線にカットオフ周波数の異なる低域通過フィルタをかけた雑音駆動音声を用いて実験を行い検討した．

5.2.1　実験方法

参加者

北陸先端科学技術大学院大学の22歳から25歳までの学生10名（男性7名，女性3名）が実験に参加した．被験者は，母国語が日本語であり，日常生活に支障のない程度の聴力を有していた．

刺激

音声波形の振幅包絡線にカットオフ周波数の異なる低域通過フィルタをかけた雑音駆動音声を刺激とした．カットオフ周波数の条件は，2，4，6，8，12，16，32 Hzの7条件で

あった．刺激の呈示には，PC（Windows 10， MATLAB），オーディオインターフェース（Fireface UCX），ヘッドフォン（SENNHEISER HDA 200）を用いた．

手続き

実験は防音室で行った．被験者には，A，B の順で音声刺激を呈示し，A に比べて B はどの程度緊迫してるか，あるいはしていないか（B の音声について答えてください）と質問した．被験者には，かなり緊迫している，やや緊迫している，同程度，やや緊迫していない，かなり緊迫しているの 5 段階で評価するように求めた．音声の緊迫感についての評価結果からシェッフェの一対比較法（浦の変法）を用いて，刺激に対する音声の緊迫感についての心理尺度を算出した．順序効果を考慮して，実験の総判断回数は 756 回であった．また，音声刺激の呈示順は被験者ごとにランダムとし，A と B の音声刺激の間隔は 0.5 秒であった．756 回を 14 セクションに分け，1 セクション 54 回の判断とし，7 セクション後に 90 分以上の休憩をはさんだ．セクションごとの休憩（7 セクション後の休憩は除く）を踏まえると，実験に要した時間は 2 時間半程度であった．

5.2.2 実験結果

音声波形の振幅包絡線の低域の変調周波数成分について検討するために，カットオフ周波数の異なる低域通過フィルタをかけた雑音駆動音声の緊迫感の評価結果から一対比較法によって緊迫感についての心理尺度を算出した．各カットオフ周波数の条件における 4 種類の刺激の緊迫感についての心理尺度を図 5.1 に示す．縦軸は刺激の緊迫感についての心理尺度を示し，横軸はカットオフ周波数の条件を示す．各カットオフ周波数の条件において，刺激間の緊迫感について心理尺度の順序は変化していないことがわかった．分散分析を行った結果，刺激に対する主効果（$F(24, \infty) = 200.1$，$p < 0.01$）が認められた．カットオフ周波数の条件が 32 Hz から 4 Hz の条件においては同じカットオフ周波数の条件の刺激間すべてで有意差（$p < 0.01$）が認められた．カットオフ周波数の条件が 2 Hz の場合においては，b と d， d と c の刺激間では有意差は認められず，その他の刺激間で有意差（$p < 0.01$）が認められた．

表 5.1: カットオフ周波数の異なる低域通過フィルタをかけた
雑音駆動音声の緊迫感についての心理尺度

カットオフ周波数 [Hz]	刺激ラベル			
	a	b	c	d
2	−0.4888	−0.2254	0.0603	−0.1004
4	−0.4576	−0.2098	0.3504	0.1250
6	−0.5054	−0.1987	0.6161	0.1897
8	−0.5580	−0.2567	0.6429	0.1964
12	−0.5246	−0.3348	0.8036	0.2366
16	−0.5580	−0.2902	0.7321	0.2366
32	−0.4777	−0.2522	0.9062	0.3192

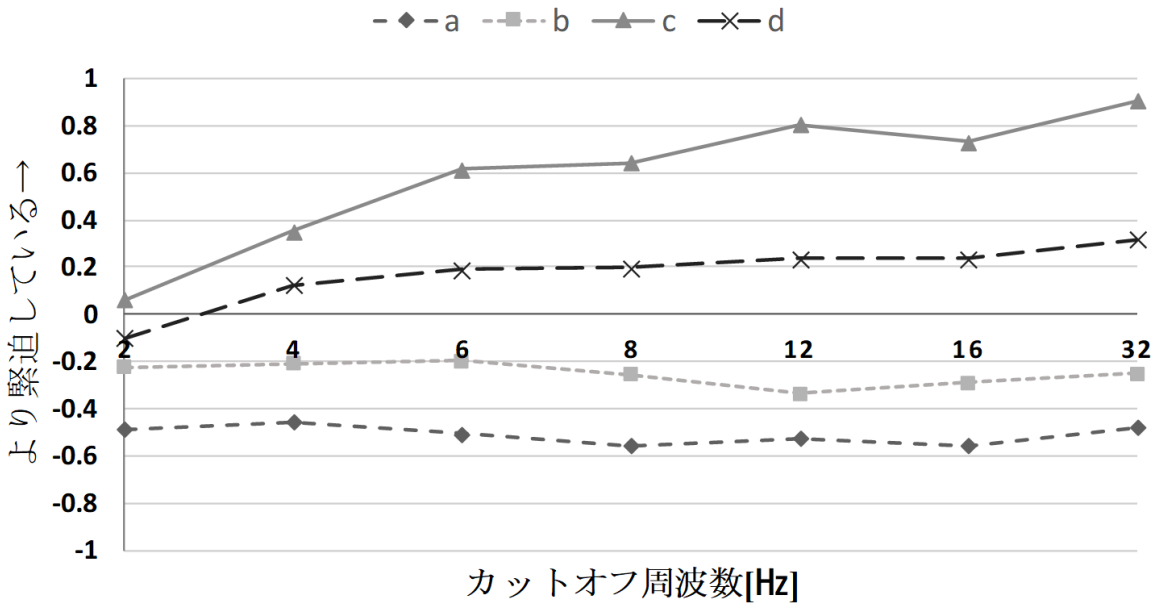


図 5.1: カットオフ周波数の異なる低域通過フィルタをかけた
雑音駆動音声の緊迫感についての心理尺度

表 5.2: カットオフ周波数の異なる低域通過フィルタをかけた雑音駆動音声の実験における分散分析表

要因	平方和	自由度	不偏分散	F 値	p 値
主効果	2480	27	91.87	200.1	0.0000
主効果×個人	696.3	189	3.684	8.022	0.0000
組み合わせ効果	205.7	351	0.5861	1.276	0.0005
順序効果	20.02	1	20.02	43.61	0.0000
順序×個人	30.34	7	4.335	9.440	0.0000
誤差	2513	5473	0.4592		
全体	5946	6048			

$F(1, \infty; 0.01) = 3.841,\ F(7, \infty; 0.01) = 2.010,$
$F(24, \infty; 0.01) = 1.517,\ F(\infty, \infty; 0.01) = 1.000$

5.3 高域の変調周波数成分の検討

本実験では，高域の変調周波数の成分における緊迫感知覚の手がかりを探索するために，音声波形の振幅包絡線にカットオフ周波数の異なる高域通過フィルタをかけた雑音駆動音声を用いて実験を行い検討した．

5.3.1 実験方法

参加者

北陸先端科学技術大学院大学の 22 歳から 25 歳までの学生 10 名（男性 8 名，女性 2 名）が実験に参加した．被験者は，母国語が日本語であり，日常生活に支障のない程度の聴力を有していた．

刺激

音声波形の振幅包絡線にカットオフ周波数の異なる高域通過フィルタをかけた雑音駆動音声を刺激とした．カットオフ周波数の条件は，4，6，8，12，16，24，32 Hz の 7 条件であった．雑音駆動音声に HPF をかける際，音声波形の直流成分を引いて，高域通過フィルタをかけ，振幅包絡線抽出の際に負の値があったものに関して負の値を 0 埋めし，その後，カットオフ周波数 64 Hz の低域通過フィルタをかけることで，高調波成分を除去した．刺激の呈示には，5.2 と同様の機器を用いた．

手続き

実験は防音室で行った．問いと評価項目は 5.2 と同様であった．緊迫感についての評価結果からシェッフェの一対比較法（浦の変法）を用いて，刺激に対する音声の緊迫感についての心理尺度を算出した．順序効果を考慮して，実験の総判断回数は 756 回であった．また，音声刺激の呈示順は被験者ごとにランダムとし，A と B の音声刺激の間隔は 0.5 秒であった．756 回を 14 セクションに分け，1 セクション 54 回の判断とし，7 セクション後に 90 分以上の休憩をはさんだ．セクションごとの休憩（7 セクション後の休憩は除く）を踏まえると，実験に要した時間は 2 時間半程度であった．

5.3.2 実験結果

音声波形の振幅包絡線の高域の変調周波数成分について検討するために，カットオフ周波数の異なる高域通過フィルタをかけた雑音駆動音声の緊迫感の評価結果から一対比較法によって緊迫感についての心理尺度を算出した．各カットオフ周波数の条件における 4

種類の刺激の緊迫感についての心理尺度を図 5.2 に示す．縦軸は音声刺激の緊迫感についての心理尺度を示し，横軸はカットオフ周波数の条件を示す．各カットオフ周波数の条件において，刺激間の緊迫感の心理尺度の順序は変化していないことがわかった．分散分析を行った結果，刺激に対する主効果（$F(24, \infty) = 1.517$，$p < .01$）が認められた．同じカットオフ周波数の条件で比較すると音声 c，d に有意差は認められないが，刺激 b との有意差はカットオフ周波数が 8 Hz まで認められた（$p < .01$）．刺激の種類に着目すると，音声 c，d において，カットオフ周波数の条件が 8 Hz と 12 Hz の間に有意差（$p < 0.01$）が認められた．また，音声 a はどのカットオフ周波数の条件においても他の音声と有意差（$p < 0.01$）が認められた．

表 5.3: カットオフ周波数の異なる高域通過フィルタをかけた
雑音駆動音声の緊迫感についての心理尺度

カットオフ周波数 [Hz]	刺激ラベル			
	a	b	c	d
4	−0.3393	−0.0268	0.5018	0.4893
6	−0.3232	0.0643	0.3786	0.3786
8	−0.3661	0.1429	0.3732	0.3018
12	−0.3786	−0.0054	0.1054	0.0804
16	−0.3946	−0.0357	0.0125	0.0804
24	−0.3161	−0.0018	−0.0929	0.0240
32	−0.4125	−0.0518	−0.1125	−0.0089

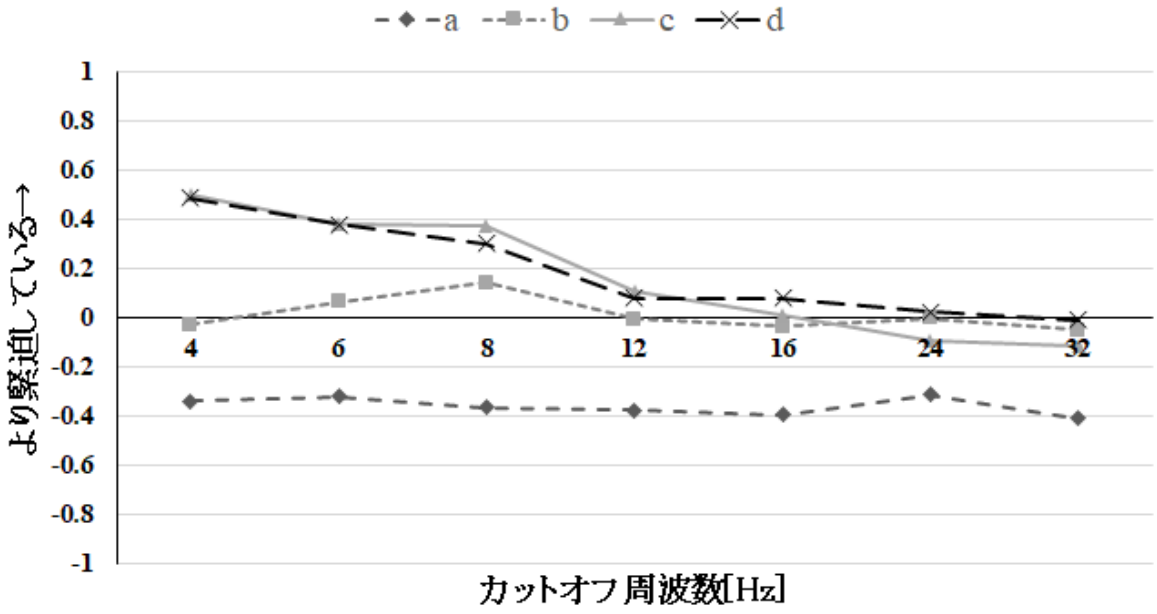


図 5.2: カットオフ周波数の異なる高域通過フィルタをかけた
雑音駆動音声の緊迫感についての心理尺度

表 5.4: カットオフ周波数の異なる高域通過フィルタをかけた雑音駆動音声の実験における分散分析表

要因	平方和	自由度	不偏分散	F 値	p 値
主効果	1120	27	41.48	83.61	0.0000
主効果×個人	855.0	243	3.519	7.092	0.0000
組み合わせ効果	176.8	351	0.5038	1.016	0.4119
順序効果	1.659	1	1.659	3.345	0.0675
順序×個人	154.9	9	17.21	36.68	0.0000
誤差	3438	6929	0.4861		
全体	5746	7560			

$F(1, \infty; 0.01) = 3.841,\ F(9, \infty; 0.01) = 1.880,$
$F(24, \infty; 0.01) = 1.517,\ F(\infty, \infty; 0.01) = 1.000$

5.4 考察

音声波形の振幅包絡線の変調周波数成分の手がかりを検討するために，カットオフ周波数の異なるフィルタをかけた雑音駆動音声を用いて実験を行った．低域の変調周波数成分では，どのカットオフ周波数の条件でも音声の緊迫感の程度の順位が変わらなかった．緊迫感の程度が最も高い音声 c について，カットオフ周波数の条件が 4 Hz から 6 Hz の間で有意差が認められた．このことから，音声波形の振幅包絡線の変調周波数成分が 6 Hz 以上あれば緊迫感知覚が可能であることがわかった．高域の変調周波数成分では，音声 c，d の緊迫感についての心理尺度が変わらなかった．緊迫感の高かった音声 c，d においてカットオフ周波数の条件が 8 Hz から 12 Hz の間に有意差が認められた．このことから，音声波形の振幅包絡線の変調周波数成分が 8 Hz 以下あれば緊迫感の知覚が可能であることがわかった．まとめると，音声波形の振幅包絡線の変調周波数が 6 Hz から 8 Hz の成分に緊迫感知覚の手がかりが存在していることが示唆された．

第6章　総合考察

音声波形の振幅包絡線は緊迫感知覚に影響を及ぼすのか

第4章では，音声波形の振幅包絡線が緊迫感知覚に影響を及ぼしているのかを明らかにした．原音声と雑音駆動音声の緊迫感についての評価結果から算出した心理尺度の結果から，音声の緊迫感の程度の順序が同じであった．このことから，音声の振幅包絡線に緊迫感の程度の知覚に関わる情報が含まれていることが示唆された．さらに，雑音駆動音声が緊迫感知覚に及ぼす影響を検討した結果，原音声に比べて雑音駆動音声の緊迫感の程度の差が大きかった．濱田らは，同様の音声刺激を用いて，緊迫感および避難誘導効果の高い音声は，中高域の周波数スペクトルが大きいことを報告している [19]．このことから，すべての刺激で音の甲高さや鋭さに関わるシャープネスを算出した．その結果，同じ原音声の刺激間で雑音駆動音声のシャープネスが高かった．同じ原音声の各刺激間のシャープネスの差は同程度であり，原音声よりも雑音駆動音声に緊迫感を感じたのは音の甲高さや鋭さに起因する可能性が考えられる．

音声波形の振幅包絡線に含まれる緊迫感知覚の手がかり

第5章では，音声波形の振幅包絡線に含まれる緊迫感知覚の手がかりを明らかにした．カットオフ周波数の異なるフィルタをかけた雑音駆動音声を用いて実験を行い検討した，その結果，低域の変調周波数成分では，カットオフ周波数の条件が 4 Hz と 6 Hz，高域の変調周波数成分では，カットオフ周波数の条件が 8 Hz と 12 Hz で緊迫感の知覚に差が見られた．このことから，変調周波数の下限は 6 Hz，上限は 8 Hz に緊迫感知覚の手がかりの存在が示唆された．両実験において刺激 a は常に緊迫感が低いと感じられた．音声の緊迫感について刺激の強さは重要な要因の一つである [16]．このことから，刺激 a の緊迫感が常に低い結果であることは，刺激 a のラウドネスが小さいことに起因している可能性がある．低域の変調周波数成分の結果と高域の変調周波数成分の結果を比べると，高域の変調周波数成分の実験において，緊迫感の高い音声 c，d の緊迫感の程度の違いが現れなかった．この要因として，変調周波数の 4 Hz 以下の成分が音声 c，d の緊迫感の程度の知覚に関連していたことが考えられる．

音声波形の振幅包絡線が音声知覚に与える影響

音声波形の振幅包絡線は音声知覚に重要な役割を果たしている．このことは，音声波形の変調周波数成分に含まれる手がかりを検討するために，カットオフ周波数の異なる低域通過フィルタをかけた雑音駆動音声を用いた実験によって明らかにされた．日本語音声の 4 モーラ単語を対象として，変調周波数成分の上限周波数と単語の認識率の関係を調査した．約 5 Hz 未満の変調成分を除去した際に平均正答モーラ数が減少する結果から，モーラの時間構造を再現する変調成分を保存さえしていれば，言語情報の取得が可能であることを明らかにした [6]．話者弁別と感情認識実験によって，信号の帯域分割数が 16 帯域の雑音駆動音声の非言語情報知覚では，個人性知覚が変調周波数の約 8 Hz，感情認識実験では，約 5 Hz が重要であることを明らかにした [40]．今回，音声波形の振幅包絡線におけるパラ言語情報知覚として音声の緊迫感に着目した実験を行った．本研究では，低域の周波数成分の検討だけでなく，高域の変調周波数成分も検討した．音声波形の振幅包絡線の変調周波数成分は 6 Hz 以上 8 Hz 以下が音声波形の振幅包絡線の緊迫感知覚に重要であることを示唆した．

第7章　結論

7.1　明らかにしたこと

本研究では，まず，緊迫感を有する音声と雑音駆動音声の緊迫感の程度を問う実験を行った．その結果，緊迫感を有する音声と雑音駆動音声の緊迫感の順位が一致した．このことから，音声波形の振幅包絡線は緊迫感の程度に関わる情報が含まれていることを明らかにした．

次に，変調周波数のカットオフ周波数を制限させた雑音駆動音声を用いて緊迫感の程度を問う実験を行った．その結果，低域の変調周波数成分では，6 Hz 以下で，高域の変調周波数成分では，8 Hz 以下で緊迫感の程度に差が認められた．音声波形の振幅包絡線の変調周波数の 6 H 以上 8 Hz 以下が緊迫感の知覚に寄与していることが示唆された．

最後に，音声波形の振幅包絡線が音声知覚に重要であること明らかにした．音声波形の振幅包絡線の言語情報知覚においては，約 5 Hz 以上，非言語情報知覚は言語情報より少し上の編徴周波数成分が重要である．緊迫感知覚は言語情報よりも上，非言語情報とおおよそ同じ変調周波数で知覚できることを明らかにした．

7.2　残された課題

変調スペクトル分析

Zhu らは変調スペクトルの分析と心理物理実験の結果から，振幅包絡線の非言語情報の役割を明らかにし，支持した [38,39]．本研究で，緊迫感知覚に重要な変調周波数成分を明らかにした．この結果と変調スペクトル分析の結果を照らし合わせさらに音声波形の振幅包絡線における緊迫感の知覚について議論を行う必要がある．

音声の振幅包絡線のパラ言語情報知覚における雑音残響の影響

筆者らは振幅包絡線における非言語情報の知覚について，雑音残響の影響を検討した．その結果，雑音や残響のある環境下でも基本的には，雑音駆動音声の比言語情報知覚は可能であることを報告した [28–30]．また，小林らは雑音や残響下における避難呼びかけ音声の緊迫感の知覚について検討した．その結果，雑音や残響があったとしても緊迫感はあ

まり低下しなかったことを報告した [37]．同様の手法で，雑音駆動音声の緊迫感について雑音や残響の影響を検討する必要がある．

時間微細構造

Drullman は音声波形の振幅包絡の言語情報知覚について時間微細構造と振幅包絡のどちらが優位かを検討した．その結果，言語情報知覚においては振幅包絡情報がより優位であることを示唆した [?]．この方法で，非言語情報知覚，パラ言語情報知覚について検討すれば，音声知覚における振幅包絡情報の知見がさらに深まる．

謝辞

本研究を進めるにあたり，厳しくも多大なる御指導，御助言ならびに多くの発表機会を賜りました指導教官の鵜木祐史教授に深く感謝いたします．また，熱心な御指導ならびに御助言を賜りました赤木正人教授に心より感謝致します．実験を進めるにあたり，御助言ならびに御討論を賜りました朱治博士，小林まおり博士，木谷俊介助教に大変感謝致します．また，実験に参加し，貴重な時間を割いてくださった被験者の皆様には心よりお礼申し上げます．苦楽や珈琲を共にした鵜木・赤木研究室の皆様と，楽しい時間を共有してくださった多くの先輩，同期ならびに同期のように闘ってくださった後輩に感謝いたします．最後に，本学での研究を支え，暖かく見守ってくれた家族に心から感謝いたします．

参考文献

[1] 森大毅, 粕谷英樹, 前川喜久雄, “音声は何を伝えているかー感情・パラ言語情報・個人性の音声科学（音響サイエンスシリーズ）,” コロナ社, 東京, 2014.

[2] H. Fujisaki, “Prosody, modeles, and spontaneous speech, in Computing Prosody,” Y. Sagiska, N. Campbell, and N. higuchi(Eds), Springer, pp. 27–42, 1996.

[3] 森尚彫, “日本における人工内耳の現状,” 保健医療学雑誌, vol. 6, No. 1, pp. 15–26, 2015.

[4] Robert V. Shannon, Fan-Gang Zeng, Vivek Kamath, John Wygonski, Michael Ekelid, “Speech Recognition with Primarity Temporal Cues,” Sciense, Vol. 270, Issue 5234, pp. 303–304, 1995.

[5] Rob Drullman, “Temporal envelope and fine structure cues for speech intelligibility,” The Journal of the Acoustical Sciety of America, 97, 585, 1995.

[6] Zhi Zhu, Yasutaka Nishino, Ryota Miyauchi , Masashi Unoki, “Study on linguistic information and speaker individuality contained in tenporal envelope of speech,” Acoustical Science and Technology, Vol. 37, No. 5, pp. 258–261, 2016.

[7] 井上祐之, “大洗町はなぜ「避難せよ」と呼びかけたのか,” 放送研究と調査, 9 月, pp. 32–53, 2011.

[8] NHK 総合テレビ, ニュースウォッチ 9（2011 年 3 月 23 日 21 時）.

[9] NHK 総合テレビ, ニュース（2012 年 3 月 9 日 19 時）.

[10] Rob. Drullman, J. M. Festen, and R. Plomp, “Effect of temporal envelope smearing on speech reception,” The Journal of the Acoustical Sciety of America, Vol. 95, No. 2, pp. 1053–1064, 1994.

[11] Rob. Drullman, J. M. Festen, and R. Plomp, “Effect of reducing slow temporal modulations on speech reception,” The Journal of the Acoustical Sciety of America, Vol. 95, No. 5, pp. 2670–2680, 1994.

[12] Torsten Dau, and Birger. Kollmeier, “Modeling auditory processing of amplitude modulation.II. Spectral and temporal integration,” The Journal of the Acoustical Sciety of America, Vol. 102, No. 5, pp. 2906–2919, 1997.

[13] 佐藤逸人, 森本政行, 栗栖清浩, “多言語一斉通知による緊急避難放送の可能性,” 日本音響学会講演論文集, 2–7–6, pp. 1609–1610, 2018.

[14] 土田義郎, 高野佐代子, “避難行動を喚起する防災放送の実現のための話者への教示に関する研究 その 1 主観的印象に与える教示の効果について,” 日本音響学会講演論文集, 2–7–1, pp. 409–412, 2017.

[15] 高野佐代子, 土田義郎, “避難行動を喚起する防災放送の実現のための話者への教示に関する研究 その 2 教示による音声の特徴の変化について,” 日本音響学会講演論文集, 2–7–2, pp. 413–414, 2017.

[16] Phil-Sik Jang, “Designing acoustic and non-acoustic parameters of synthesized speechwarnings to control perceived urgency,” International Journal of Industrial Ergonomics, Vol. 37, pp. 213–223, 2007.

[17] 小笠原奈保美, 大藤建太, “災害時避難伝達文の言語学的分析,” 日本音響学会講演論文集, 1–5–12, pp. 1531–1534, 2017.

[18] 小林まおり, 赤木正人, “避難呼びかけ音声の心理的評価,” 日本音響学会誌, Vol. 74, No. 12, pp. 633–640, 2018.

[19] 濱田康弘, 小林まおり, 赤木正人, “避難呼びかけ音声の音響的特徴の分析,” 日本音響学会講演論文集, 2–Q–38, pp. 373–374, 2018.

[20] Smith ZM, Delgutte B, Oxenham AJ, “Chimaeric sounds reveal dichotomies in auditory perception,” Nature416, pp. 87–90, 2002.

[21] Torsten Dau and Dirk. Puschel, “A quantitative model of the ”effective” signal processingin the auditory system. I. Model structure,” The Journal of the Acoustical Sciety of America, Vol. 99, No. 6, pp. 3615–3622, 1996.

[22] Brian C. J. Moore and Brian R. Glasberg, “Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,” The Journal of the Acoustical Sciety of America, Vol. 74, No. 750, pp. 750–753, 1983.

[23] Houtgast T, “Frequency selectivity in amplitude—modulation detection,” The Journal of the Acoustical Sciety of America, Vol. 85, pp. 1676–1680, 1989.

[24] Bacon SP and Grantham DW, “Modulation mask-ing : effects of modulation frequency, depth, andphase,” The Journal of the Acoustical Sciety of America, Vol. 85, pp. 2575–2580, 1989.

[25] Malone B and Schreiner CE, ”Time—varyingsounds : amplitude envelope modulations, in TheOxford handbooks of auditory science,” The auditory brain, A. Rees and A.R. Palmer, Editors. pp.125—148, Oxford University Press, New York, 2010.

[26] Langner G, Schreiner CE, “Periodicity codingin the inferior colliculus of the cat,” I. Neuronalmechanisms. J Neurophysiol60, pp. 1799–1822, 1988.

[27] McDermott JH, Schemitsch M, Simoncelli EP, “Simoncelli, Summary statistics in auditory perception,” Nat Neurosci16, pp. 493–498, 2013.

[28] 関谷伸一, 朱治, 鵜木祐史, “雑音駆動音声の言語・非言語情報知覚と室内音響特性による影響の検討,” 電子情報通信学会技術研究報告, Vol. 117, No. 255, pp. 41–46, 2017.

[29] 川村美帆, 朱治, 鵜木祐史, “雑音環境が雑音駆動音声の個人性・感情知覚に与える影響,” 日本音響学会聴覚研究会資料, Vol. 48, No. 2, pp. 175–180, 2018.

[30] 川村美帆, 朱治, 関谷伸一, 鵜木祐史, “雑音残響環境が雑音駆動音声の個人性及び感情情報の知覚に与える影響,” 電子情報通信学会技術研究報告, Vol.118, No. 190, pp. 83–88, 2018.

[31] 新村出, “広辞苑第四版,” 岩波書店, 東京, 1995.

[32] 高木英行, “使える！統計検定・機械学習–Ⅲ:主観評価実験のための有意差検定,” システム/制御/情報, Vol. 58, No. 12, pp. 258–261, 2014.

[33] 佐藤信, “統計的官能検査法,” 日科技連, 東京, 1985.

[34] 森敏昭, 吉田寿夫, “心理学のためのデータ解析テクニカルブック,” 北大路書房, 京都, 1990.

[35] 難波精一郎, 桑野園子, “音の評価のための心理学的測定法（音響テクノロジーシリーズ）,” コロナ社, 東京, 1998.

[36] 岩波眞一郎, “音色の感性学（音響テクノロジーシリーズ）,” コロナ社, 東京, 2010.

[37] 小林まおり, 赤木正人, “雑音・残響下での緊迫感がある音声の知覚,” 日本音響学会騒音・振動研究会資料, pp. 1–6, 2018.

[38] Zhi Zhu, Ryota Miyauchi, Yukiko Araki, Masashi Unoki, “Feasibility of vocal emotion conversion on modulation spectrogram for simulated cochlear implants,” 2017 25th European Signal Processing Conference (EUSPICO), pp. 1834–1838, 2017.

[39] Zhi Zhu, Ryota Miyauchi, Yukiko Araki, Masashi Unoki, "Modulation spectral features for predicting vocal emotion recognition by simulated cochlear implants," INTERSPEECH, pp. 262–266, 2016.

[40] Zhi Zhu, Ryota Miyauchi, Yukiko Araki, Masashi Unoki, "Contributions of temporal cue on the perception of speaker individuality and vocal emotion for noise-vocoded speech," Acoustical Science and Technology, Vol. 39, No. 3, pp. 234-242, 2018.

研究業績

本研究に関する研究業績

国際学会における発表

(口頭，査読有)

1. Unoki Masashi, Miho Kawamura, Kobaysahi Maori, Akagi Msato, How the temporal amplitude envelope of speech contributes to urgency perception, 23rd International Congress on Acoustics, Germanny, September, 2019.(abstract submitted)

国内学会における発表

(口頭，査読無)

1. 川村美帆, 小林まおり, 木谷俊介, 赤木正人, 鵜木祐史, 振幅包絡線に含まれる緊迫感の知覚, 日本音響学会聴覚研究会資料, 愛媛, 2019 年 2 月.
2. 川村美帆, 鵜木祐史, 雑音駆動音声の緊迫感知覚の検討, 日本音響学会 2019 年春季研究発表会, 東京, 2019 年 3 月.

その他の研究業績

国内学会における発表

(口頭，査読無)

1. 川村美帆, 朱治, 鵜木祐史, 雑音環境が雑音駆動音声の個人性・感情知覚に与える影響, 日本音響学会聴覚研究会資料, Vol. 48, No. 2, pp. 175–180, 沖縄, 2018 年 3 月.
2. 川村美帆, 朱治, 関谷伸一, 鵜木祐史, 雑音残響環境が雑音駆動音声の個人性及び感情情報の知覚に与える影響, 日本音響学会聴覚研究会資料, Vol. 48, No. 6, pp. 541–546, 宮城, 2018 年 8 月.

3. 朱治, 川村美帆, 関谷伸一, 鵜木祐史, 雑音残響環境における雑音駆動音声の個人性及び感情情報の知覚に関する検討, 日本音響学会 2018 年秋季研究発表会, 2-P-14, 大分, 2018 年 9 月.

4. 川村美帆, 朱治, 関谷伸一, 鵜木祐史, 雑音駆動音声の個人性・感情知覚における雑音残響環境の影響, 平成 30 年度電気関係学会北陸支部連合大会, G–15, 石川, 2018 年 9 月.

その他の業績

(受賞)

1. 川村美帆, 優秀論文発表賞, 電気関係学会北陸支部, 2018 年 9 月.

付録

音声波形の振幅包絡線が音声知覚に与える影響を議論するためには実環境の影響も考慮しなければならない．そのために行った二つの実験を以下にまとめる．一つ目は雑音環境が音声波形の振幅包絡線の非言語情報知覚に及ぼす影響，二つ目は，雑音残響環境が音声波形の振幅包絡線の非言語情報に及ぼす影響である．

付 録1　雑音環境が音声波形の振幅包絡線の非言語情報知覚に及ぼす影響

A1.1　目的

本研究は，雑音環境が雑音駆動音声の非言語情報知覚に与える影響を明らかにすることを目的とする．そのために，雑音を付加した音声から雑音駆動音声を合成し実験を行い検討した．個人性知覚の検討については，XAB 法を用いて話者の弁別を行う実験を行い，感情知覚の検討については，5 種類（平静，喜び，抑えた怒り，悲しみ，激しい怒り）の感情を判断させる実験を行った．

A1.2　話者弁別実験

本実験では，雑音環境が雑音駆動音声の個人性知覚に及ぼす影響を検討するために，話者を弁別させる実験を行い検討した．

A1.2.1　実験方法

参加者

北陸先端科学技術大学院大学の 22 歳から 25 歳までの学生 9 名（男性 5 名，女性 4 名）が実験に参加した．被験者は，母国語が日本語であり，日常生活に支障のない程度の聴力を有していた．

刺激

任意の雑音レベルを設定し，雑音レベルに応じた雑音を付加した音声から合成した雑音駆動音声を刺激とした．雑音レベルとして SNR を用い，条件は SNR が 20，15，10，5，0，−5 dB であった．刺激の呈示には，PC（Windows 10，MATLAB），オーディオインターフェース（Fireface UCX），ヘッドフォン（SENNHEISER HDA 200）を用いた．

手続き

実験は防音室で行った．被験者には，X，A，B の順で音声刺激を呈示した．X は基準となる刺激，A は X の話者と同じ話者で異なる文章の刺激，B は X の話者と異なる話者で異なる文章の刺激であった．これらの刺激を無音区間 0.5 s を挟んで呈示し，X の話者が A と B どちらの話者と同じであるかを強制判断させた．刺激の呈示回数は 1 回とし，繰り返しは許さなかった．順序効果を考慮して，実験の総判断回数は 280 回であった．また，音声刺激の呈示順は被験者ごとにランダムとし，280 回を 5 セクションに分け，1 セクション 56 回の判断とし，実験に要した時間は 1 時間程度であった．

A1.2.2　実験結果

図 A1.1 に話者弁別実験の結果を示す．横軸は SNR，縦軸は話者認識率であり，図中の丸印は認識率の平均値，縦棒は標準誤差を示す．SNR が低くなるにつれ，つまり雑音レベルが高くなるにつれ，話者弁別率の平均値が低くなる傾向が見られた．ここで，1 要因の分散分析を行ったところ，雑音条件に対する主効果（$p < 0.01$）が見られた．その下位検定とした多重比較の結果から，SNR が 5 dB の条件とその他の雑音条件との間に有意差（$p < 0.01$）が見られた．一方，他の雑音条件の間には有意差が見られなかった．

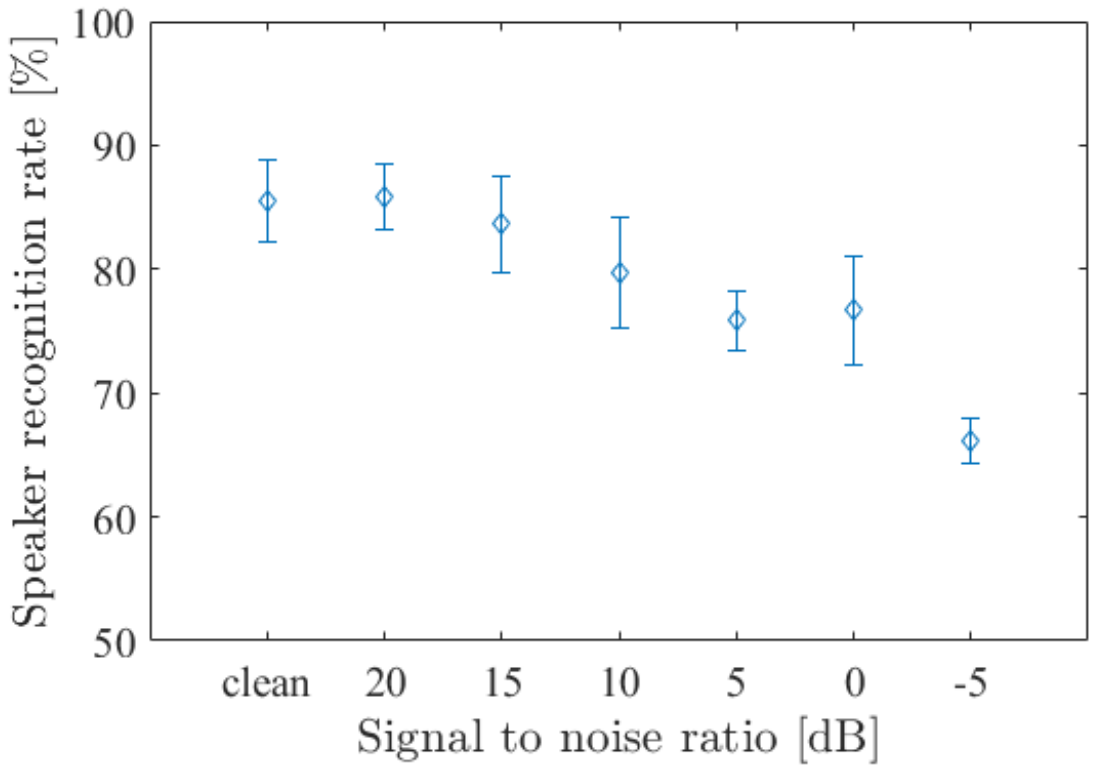

Speaker recognition rate [%]
100
90
80
70
60
50
clean
20
15
10
5
0
-5
Signal to noise ratio [dB]

図 A1.1: 話者弁別率

A1.3　感情認識実験

本実験では，雑音環境が雑音駆動音声の感情知覚に及ぼす影響を検討するために，5 種類（平静，喜び，抑えた怒り，悲しみ，激しい怒り）の感情を判断させる実験を行い検討した．

A1.3.1　実験方法

参加者

北陸先端科学技術大学院大学の 22 歳から 25 歳までの学生 11 名（男性 7 名，女性 4 名）が実験に参加した．被験者は，母国語が日本語であり，日常生活に支障のない程度の聴力を有していた．

刺激

任意の雑音レベルを設定し，雑音レベルに応じた雑音を付加させた音声から合成した雑音駆動音声を刺激とした．雑音レベルとして SNR を用いた．条件は SNR が 20，15，10，5，0，−5 dB であった．刺激の呈示には，PC（Windows 10，MATLAB），オーディオインターフェース（Fireface UCX），ヘッドフォン（SENNHEISER HDA 200）を用いた．

手続き

実験は防音室で行った．実験参加者に全ての刺激音をランダムに呈示し，5 つの感情から 1 つの感情を強制判断させた．刺激の呈示回数は 1 回で，繰り返しは許さなかった．刺激音は感情ごとに 5 種類の文章を用意し，前述の実験条件で作成した．5 文章，5 感情，実験条件 7 条件で総刺激数は 150 個であった．150 回を 3 セクションに分け，1 セクション 50 回の判断とし，実験に要した時間は 20 分程度であった．

A1.3.2　実験結果

図 A1.2 に感情認識実験の結果を示す．横軸は SNR，縦軸は感情認識率であり，図中の丸印は認識率の平均値，縦棒は標準誤差を示す．図 A1.2(a) に，5 種類すべての感情刺激を含んだ認識率の平均値と標準誤差を示す．話者認識実験と同様に，雑音レベルが高くなるとともに，感情認識率も低下していることがわかる．特に，SNR が 5 dB の時の感情認識率の平均値と 0 dB の時の感情認識率の平均値の間で最も差が大きかった．雑音条件と感情を要因とする 2 要因分散分析を行った結果，雑音条件に対する主効果（$p < 0.01$）が見られた．その下位検定とした多重比較の結果から，SNR が 5 dB と 0 dB の条件と 0 dB

より大きいSNRの条件の間に有意差（$p < 0.05$）が見られた．一方，他のSNRの条件の間には有意差が見られなかった．また，感情の種類による主効果（$p < 0.01$）と雑音条件と感情の相互作用も主効果（$p < 0.01$）が見られた．

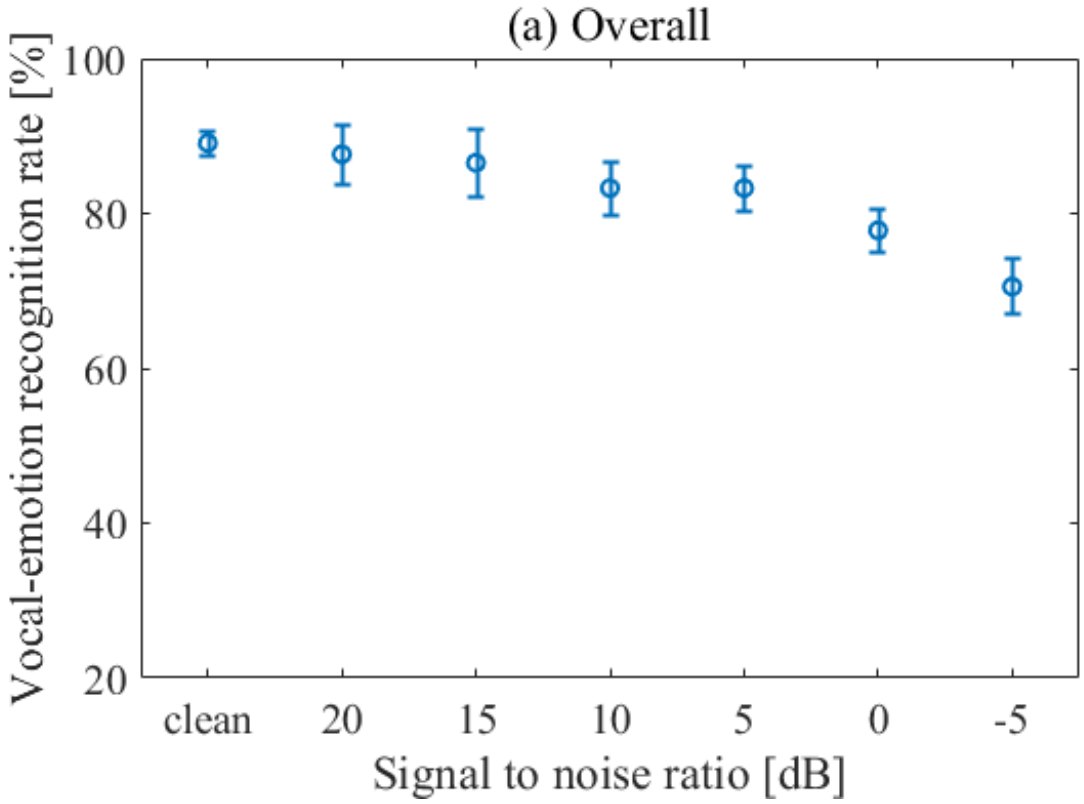


図 A1.2: 感情認識率

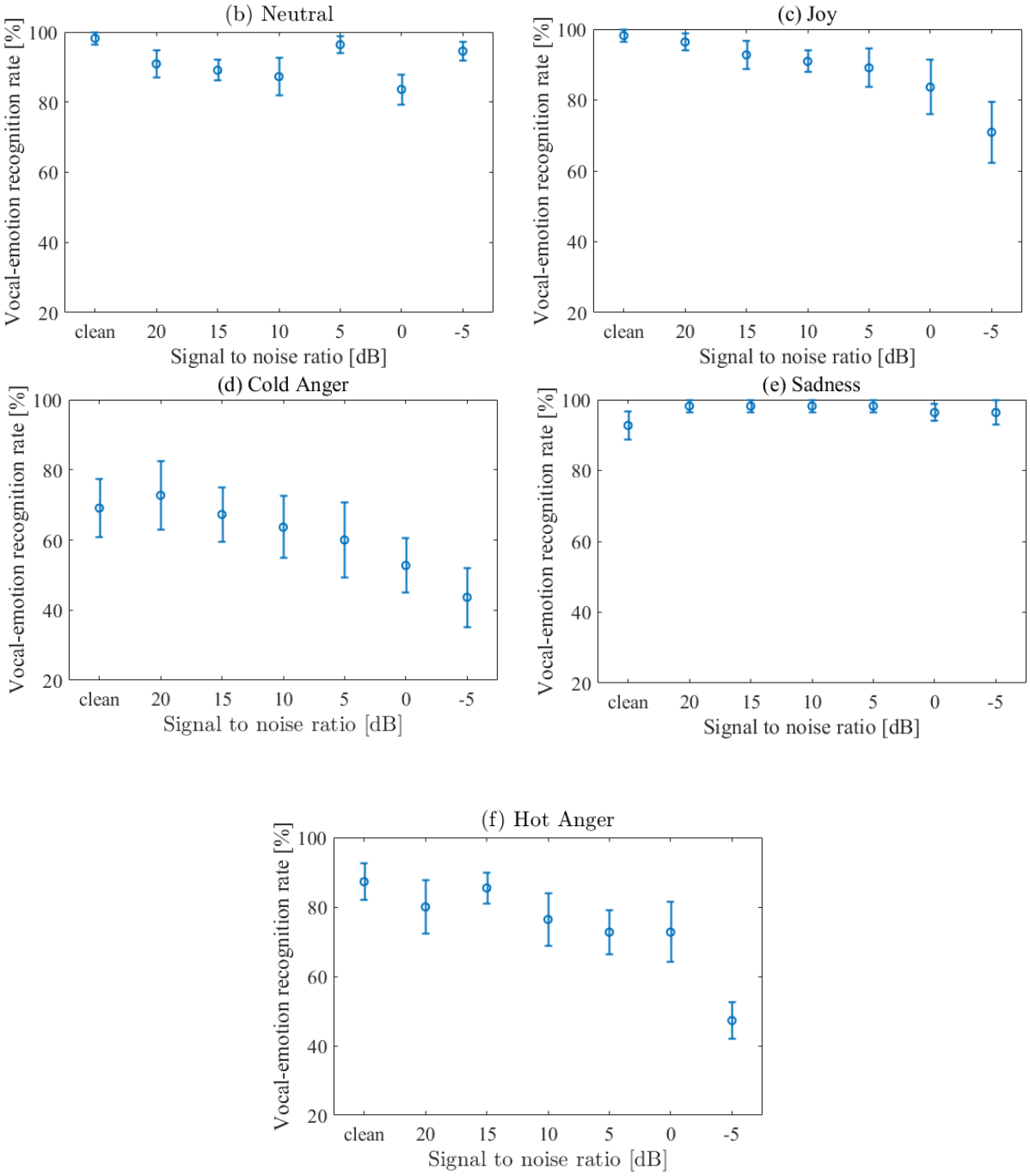


図 A1.3: 感情ごとの感情認識率

A1.4 考察

話者弁別実験の結果，雑音条件に対する主効果（$p < 0.01$）が認められた．下位検定として多重比較を行った結果，SNR が -5 dB より高い条件の時に限った場合，有意差は認められなかった．このことから，雑音環境下での雑音駆動音声の個人性知覚では，極端に劣悪な条件でなければ，雑音による影響を基本的に受けないとみなすことができる．感情認識実験の結果，雑音条件に対する主効果が認められた．下位検定として多重比較を行った結果，SNR が 0 dB に限った場合，有意差は認められなかった．このことから，雑音環境下での雑音駆動音声の感情知覚は，個人性知覚と同様に，極端に劣悪な条件でなければ，基本的に雑音の影響を受けないとみなすことができる．しかし，SNR が 0 dB 以下の条件とその他の SNR の条件の間では有意差（$p < 0.05$）が認められた．SNR が 0 dB 以下の条件では，雑音環境が感情知覚に影響を与えることがわかった．

付 録2　雑音残響環境が音声波形の振幅包絡線の非言語情報知覚に及ぼす影響

A2.1　目的

本研究は，雑音残響環境が雑音駆動音声に与える影響を明らかにすることを目的とする．そのために，雑音と残響を付加した音声から雑音駆動音声を合成し実験を行い検討した．個人性知覚の検討については，XAB 法を用いて話者の弁別を行う実験を行い，感情知覚の検討については，5 種類（平静，喜び，抑えた怒り，悲しみ，激しい怒り）の感情の認識実験を行った．

A2.2　話者弁別実験

本実験では，雑音残響環境が雑音駆動音声の個人性知覚に及ぼす影響を検討するために，話者を弁別させる実験を行い検討した．

A2.2.1　実験方法

参加者

北陸先端科学技術大学院大学の 22 歳から 25 歳までの学生 10 名（男性 8 名，女性 2 名）が実験に参加した．被験者は，母国語が日本語であり，日常生活に支障のない程度の聴力を有していた．

刺激

任意の雑音レベルを設定し，雑音レベルに応じた雑音と任意の残響条件に応じた残響を付加させた音声から合成した雑音駆動音声を刺激とした．雑音レベルとして SNR を用いた．条件は SNR が 20，15，10，5，0，−5 dB であった．残響条件は 0.5，1.0，2.0 s であった. 刺激の呈示には，PC（Windows 10，MATLAB），オーディオインターフェース

（Fireface UCX），ヘッドフォン（SENNHEISER HDA 200）を用いた．

手続き

実験は防音室で行った．被験者には，X，A，B の順で音声刺激を呈示した．X は基準となる刺激，A は X の話者と同じ話者で異なる文章の刺激，B は X の話者と異なる話者で同じ文章の刺激であった．これらの刺激を無音区間 0.5 s を挟んで呈示し，X の話者が A と B どちらの話者と同じであるかを強制判断させた．刺激の呈示回数は 1 回とし，繰り返しは許さなかった．順序効果を考慮して，実験の総判断回数は 600 回であった．また，音声刺激の呈示順は被験者ごとにランダムとし，600 回を 10 セクションに分け，1 セクション 60 回の判断とし，実験に要した時間は 2 時間程度であった．

A2.2.2　実験結果

図 A2.1 に話者弁別実験の平均弁別率と被験者による標準誤差を示す．横軸は SNR，縦軸は話者弁別率である．図中の実線，破線，鎖線はそれぞれ，残響時間が 0.5， 1.0， 2.0 s の話者弁別率である．また，図中の丸，四角，三角は弁別率の平均値，縦棒は標準誤差を示す．図 A2.2 には話者弁別実験の平均弁別率の分布図を示す．残響時間と SNR を要因として 2 要因の分散分析を行った結果，残残響時間と SNR に対する主効果がそれぞれ認められた．一方，残響時間 と SNR の相互作用は認められなかった．また，主効果が認められたものに対して，その下位検定として多重比較を行った．その結果，残響時間を要因とした場合では，残響時間が 0.5 s とその他の条件の間で有意差（$p < 0.05$）が認められた．SNR を要因とした場合では，SNR が 10 dB と 5 dB の条件の間に有意差（$p < 0.05$）が認められ，10 dB と 0 dB 以下の条件の間と 20 dB と 5 dB 以下の条件の間に有意差（$p < 0.01$）が認められた．

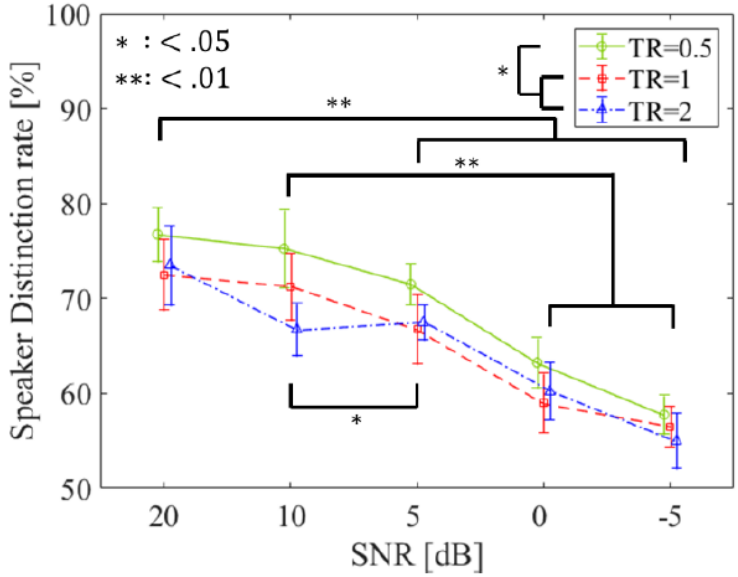


図 A2.1: 話者弁別率

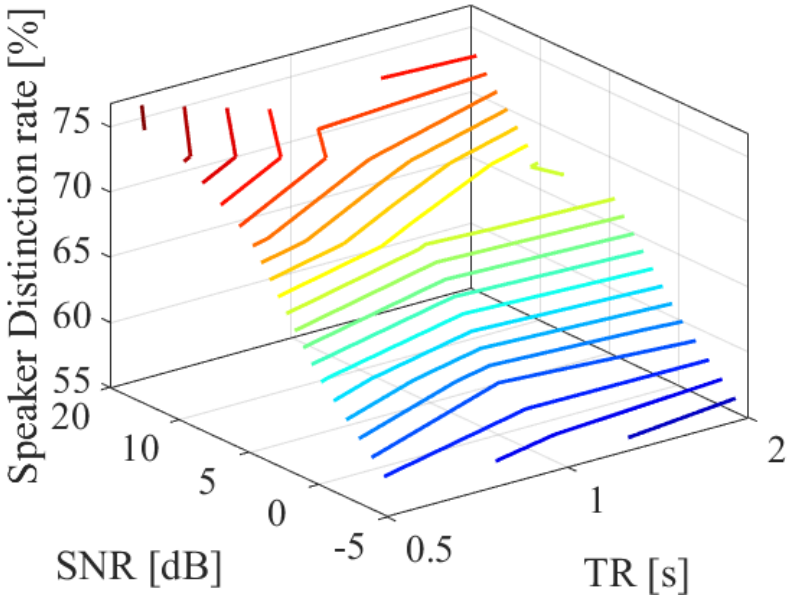


図 A2.2: 話者弁別実験の平均弁別率の分布図

A2.3　感情認識実験

本実験では，雑音残響環境が雑音駆動音声の感情知覚に及ぼす影響を検討するために，5 種類（平静，喜び，抑えた怒り，悲しみ，激しい怒り）の感情を判断させる実験を行い検討した．

A2.3.1　実験方法

参加者

北陸先端科学技術大学院大学の 22 歳から 25 歳までの学生 10 名（男性 8 名，女性 2 名）が実験に参加した．被験者は，母国語が日本語であり，日常生活に支障のない程度の聴力を有していた．

刺激

任意の雑音レベルを設定し，雑音レベルに応じた雑音を付加させた音声から合成した雑音駆動音声を刺激とした．雑音レベルとして SNR を用いた．条件は SNR が 20，15，10，5，0，−5 dB であった．刺激の呈示には，PC（Windows 10， MATLAB），オーディオインターフェース（Fireface UCX），ヘッドフォン（SENNHEISER HDA 200）を用いた．

手続き

実験は防音室で行った．実験参加者に刺激音声をランダムに呈示し，5 種類（平静，喜び，抑えた怒り，悲しみ，激しい怒り）の感情から 1 つの感情を強制判断させた．刺激の呈示回数は 1 回で，繰り返しは許さなかった．刺激音声は感情ごとに 5 種類の文章を用意し，前述の実験条件で作成した．5 文章，5 感情，実験条件 15 条件で総刺激数は 375 個であった．375 回を 5 セクションに分け，1 セクション 75 回の判断とし，実験に要した時間は 1 時間程度であった．

A2.3.2　実験結果

図 A2.3 に感情認識実験の平均認識率と実験参加者による標準誤差を示す．横軸は SNR，縦軸は感情認識率である．図中の実線，破線，鎖線はそれぞれ，残響時間が 0.5，1.0，2.0 s の感情認識率である．また，図中の丸，四角，三角は認識率の平均値，縦棒は標準誤差を示す．図 A2.4 には感情認識実験の平均認識率の分布図を示す．残響時間と SNR を要因として 2 要因の分散分析を行った結果，残響時間と SNR に対する主効果がそれぞれ認められた．一方，残響時間と SNR の相互作用は認められなかった．また，主効果が認めら

れたものに対して，その下位検定として多重比較を行った．その結果，残響時間を要因とした場合では，残響時間が 0.5 s と 2 s の条件の間で有意差（$p < 0.01$）が認められた．SNR を要因とした場合では，SNR が 10 dB，5 dB と 0 dB，−5 dB の条件の間と 20 dB と 10 dB 以下の条件の間に有意差（$p < 0.01$）が認められた．

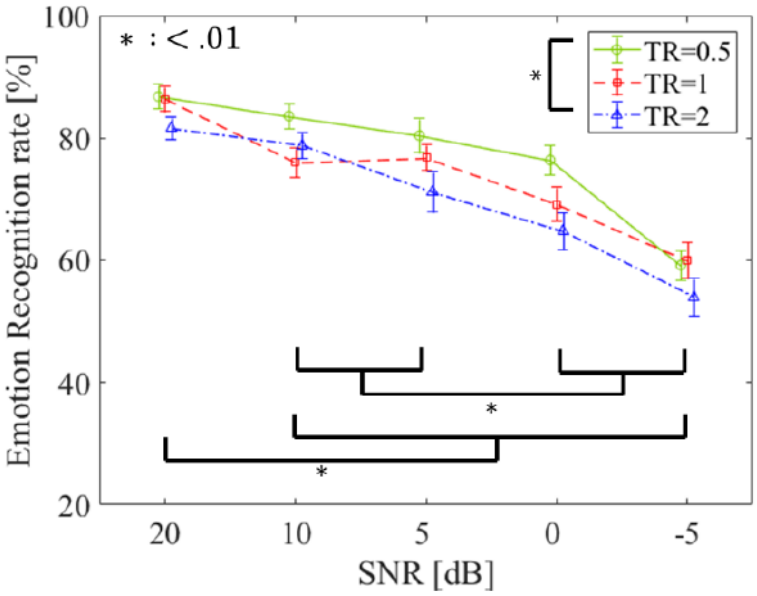


図 A2.3: 感情認識率

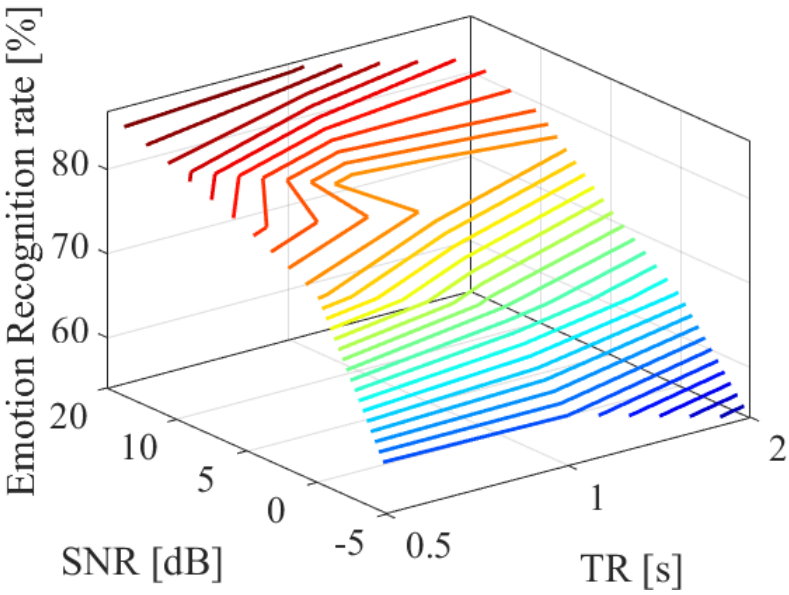


図 A2.4: 感情認識実験の平均認識率の分布図

A2.4　考察

平均話者弁別率と平均感情認識率の分布図（図 A2.2 と図 A2.3）より，認識率がほとんど低下していない部分が認められた．また，話者弁別率と感情認識率（図 A2.1 と図 A2.3）より，有意差が認められる条件間と分布図を見比べると，残響や雑音の影響をほとんど受けない条件範囲が存在する．これらの結果から，極めて劣悪な条件を除く雑音残響環境（1.0 s 未満，かつ SNR で 10 dB 以上）では，今回の実験の有意水準内で残響や雑音が雑音駆動音声の個人性・感情知覚にほとんど影響を与えず，残響時間と SNR がある閾値を越えると個人性・感情知覚に影響を与えることがわかった．日常的な雑音残響環境では，雑音駆動音声の個人性・感情知覚は残響や雑音の影響を受けにくいものと考えられる．