

Title	音声の時間振幅包絡線操作による緊迫感知覚の変化
Author(s)	房野, 早希
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/19813
Rights	
Description	Supervisor: 鷗木 祐史, 先端科学技術研究科, 修士 (情報科学)

修士論文

音声の時間振幅包絡線操作による緊迫感知覚の変化

房野 早希

主指導教員 鵜木 祐史

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)

令和7年3月

Abstract

Cochlear implant is one of the artificial organs used by people with hearing impairments. It converts input sound into electrical stimulation and stimulates the auditory nerve directly through electrodes. Insertion into the cochlea can significantly restore hearing. Treatment with cochlear implants is particularly effective for children with hearing loss. If patients restore hearing by early fitting of cochlear implant in infancy, they can communicate like that of normal hearing person. The number of cochlear implant users is expected to increase further, as the criteria for cochlear implantation in children are being revised to allow for earlier cochlear implantation. On the other hand, there are several issues in vocal communication using cochlear implant. Especially, Cochlear implant users find it difficult to get non-linguistic information. To realize rich communication for cochlear implant users, the transmission of non-linguistic information must be improved using speech synthesis processing for cochlear implants.

Noise-vocoded speech (NVS) is the simulator of cochlear implant speech that uses band-limited random noise (temporal fine structure) modulated with temporal amplitude envelope (TAE) information. Therefore, it can synthesize speech in which only the temporal information cues (information of TAE) are preserved, without any spectral information cues. If NVS can be used to reveal the acoustic features involved in the transmission of non-linguistic information, it can contribute to the improvement of speech synthesis for cochlear implants. In this study, the acoustic features involved in the urgency perception, one of the non-linguistic information are clarified.

It has been suggested by previous studies using NVS that TAE of speech includes not only linguistic information (intelligibility and comprehensibility), but also important cues related to non-linguistic information (emotion and individuality). Furthermore, it has revealed that the TAE of speech includes cues that are important for urgency perception. Furthermore, it has found that the temporal modulation frequencies for urgency perception important cues related to urgency perception were the temporal modulation frequency range from 4 Hz to 16 Hz. However, it has found that there was no difference in the urgency perception (i.e., it was difficult to distinguish the different of urgency perception) in the NVS with the time-reversed TAE included the modulation frequency band important for the urgency perception. This result has suggested that not only the long-time averaged modulation frequency components in the TAE of but also the temporal features of modulation frequency components are important cues for urgency perception. Therefore, it is necessary to examine the relationship between the instantaneous modulation components (IMCs) of TAE and the urgency perception to clarify the cause of the change in the degree of urgency of NVS to the time reversal processing.

On the other hand, it has investigated whether the IMCs of TAE contribute to emotion perception. It suggested that the cause of the reduction in the emotion recognition rate in NVS with time-reversed TAE. The reduction was reproduced by manipulating the IMCs using temporal stretching and compression on the TAE. If this result is not limited to emotion perception, but is also related to the urgency perception, it may be possible to clarify the cause of the reduction in urgency perception of the NVS with time-reversed TAE by manipulating IMCs of TAE.

This study aims to investigate the change in the urgency perception of the NVS by manipulating IMCs in TAE. The following four point of issues were investigated for this purpose. The first point is whether TAE contains cues for urgency perception. The second point is the modulation frequency components that are important for the urgency perception by using Modulation Filterbank (MFB) to manipulate the modulation frequency band of TAE. Before using MFB, it was investigated that whether MFB processing of TAE affects the urgency perception of NVS. Then, the modulation frequency band that is important for the urgency perception was investigated. The third point is the change urgency perception of the NVS through time-reversal processing of TAE. Finally, the fourth point is the change urgency perception of the NVS through manipulating of IMCs using temporal stretching and compression on TAE.

The above points were investigated l by conducting listening experiments with NVS. The modulation frequency domain and time domain of TAE were processed according to the conditions, and changes in the urgency perception of NVS with that TAE were investigated. The results of the experiment revealed the following:

1. TAE contain cues to the urgency perception in speech.
2. The modulation frequency components that are important cues for the urgency perception are from 4 Hz to 16 Hz.
3. Time-reversal processing of TAE reduce the urgency perception of NVS.
4. Manipulating IMCs using temporal stretching and compression on TAE can reproduce the change urgency perception of NVS caused by time-reversal processing on TAE, except for a few results.

These finding suggest that IMCs of TAE plays an important role in urgency perception in speech.

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 音声の時間振幅包絡線と音声知覚	5
2.1 時間振幅包絡線と音声知覚の関係	5
2.1.1 言語情報知覚	5
2.1.2 非言語情報知覚	6
2.2 本研究の新規性	6
第3章 緊迫感知覚実験の構成	8
3.1 実験について	8
3.1.1 実験の流れ	8
3.1.2 実験参加者, 実験機器, 実験手続き	9
3.2 音声刺激	11
3.3 雑音駆動音声の作成方法	11
3.4 時間伸長圧縮処理	12
第4章 時間反転処理による緊迫感の変化	38
4.1 仮説	38
4.2 実験結果	38
4.3 考察	38
第5章 時間伸長圧縮処理による緊迫感の変化	42
5.1 仮説	42
5.2 実験結果	42
5.3 考察	47
第6章 全体考察	48

第 7 章 結論	49
7.1 本研究で明らかになったこと	49
7.2 残された課題	49
付 録 A 雑音駆動音声の緊迫感知覚	50
A.1 仮説	50
A.2 実験結果	50
A.3 考察	54
付 録 B 緊迫感知覚に重要な変調周波数成分	55
B.1 変調フィルタバンクの影響	55
B.1.1 仮説	55
B.1.2 実験結果	55
B.1.3 考察	55
B.2 重要な変調周波数帯域について	59
B.2.1 仮説	59
B.2.2 実験結果	59
B.2.3 考察	59
謝辞	62
参考文献	62
研究業績	67

目 次

1.1	本論文の構成	4
3.1	実験の流れ	10
3.2	操作画面	16
3.3	NVS の作成方法の概略	17
3.4	TAE と IMCs (音声 A)	18
3.5	TAE と IMCs (音声 B)	19
3.6	TAE と IMCs (音声 C)	20
3.7	TAE と IMCs (音声 D)	21
3.8	Alpha (音声 A)	22
3.9	Alpha (音声 B)	23
3.10	Alpha (音声 C)	24
3.11	Alpha (音声 D)	25
3.12	時間圧縮処理した TAE と IMCs (音声 A)	26
3.13	時間伸長処理した TAE と IMCs (音声 A)	27
3.14	時間伸長圧縮処理した TAE と IMCs (音声 A)	28
3.15	時間圧縮処理した TAE と IMCs (音声 B)	29
3.16	時間伸長処理した TAE と IMCs (音声 B)	30
3.17	時間伸長圧縮処理した TAE と IMCs (音声 B)	31
3.18	時間圧縮処理した TAE と IMCs (音声 C)	32
3.19	時間伸長処理した TAE と IMCs (音声 C)	33
3.20	時間伸長圧縮処理した TAE と IMCs (音声 C)	34
3.21	時間圧縮処理した TAE と IMCs (音声 D)	35
3.22	時間伸長処理した TAE と IMCs (音声 D)	36
3.23	時間伸長圧縮処理した TAE と IMCs (音声 D)	37
4.1	実験 4 の結果	40
5.1	実験 5 の結果	43
5.2	実験 5 の結果 (時間反転処理と時間圧縮処理の比較)	44
5.3	実験 5 の結果 (時間反転処理と時間伸長処理の比較)	45
5.4	実験 5 の結果 (時間反転処理と時間伸長圧縮処理の比較)	46

A.1	実験 1-1（原音声）の結果	51
A.2	実験 1-2（雑音駆動音声）の結果	52
B.1	実験 2 の結果	57
B.2	実験 3 の結果	60

表 目 次

3.1	緊迫感知覚実験における評価カテゴリ	13
3.2	実験で利用した音声刺激の呼称と作成条件	13
3.3	各 MF の帯域幅	13
4.1	実験 4 の結果	40
4.2	実験 4 における分散分析表	41
5.1	実験 5 における分散分析表	43
5.2	実験 5 の結果 (時間反転処理と時間圧縮処理の比較)	44
5.3	実験 5 の結果 (時間反転処理と時間伸長処理の比較)	45
5.4	実験 5 の結果 (時間反転処理と時間伸長圧縮処理の比較)	46
A.1	実験 1-1 (原音声) の結果	51
A.2	実験 1-2 (雑音駆動音声) の結果	52
A.3	実験 1-1 (原音声) における分散分析表	53
A.4	実験 1-2 (雑音駆動音声) における分散分析表	53
B.1	実験 2 の結果	57
B.2	実験 2 における分散分析表	58
B.3	実験 3 の結果	60
B.4	実験 3 における分散分析表	61

第1章 序論

1.1 はじめに

音声は、人が意思伝達を行う上で重要なツールの1つである。人は音声を通して、その発話内容以外にも性別、年齢、感情、発話者の態度や意図など様々な情報を得ることができる。そして、これらの情報は環境や発話者の状態に合わせて変化する [1]。これらの情報を豊富に受け取ることで、齟齬のない円滑なコミュニケーションを行うことができる。

しかし、音声コミュニケーションはしばしば情報のやりとりに障害が生じる場合がある。その原因として、言語圏や文化圏の違い、相互の価値観のズレ、騒音環境下で音声の伝達がうまくいかないなどのほか、難聴によって音声の受け取りが困難になる場合が挙げられる。難聴は主に中耳の機能に障害が出る伝音性難聴と、内耳の機能に障害が出る感音性難聴に分類される。これらの難聴を抱える人々が、再び聴力を回復させる方法の1つに、人工内耳の利用が挙げられる。人工内耳は、感音声難聴を抱える人々が利用する人工臓器である。蝸牛に、蝸牛の聴覚フィルタ特性を模倣した電極を挿入する。入力された音を電気刺激に変換し、電極を通して直接聴神経を刺激する。これにより、脳に音や言葉を届けることが可能になる [2-4]。人工内耳は、特に小児の早期装用による効果が認められている。長南らは、未就学児から小学2年生までの人工内耳装用児の音節分解と音韻意識の発達について調査し、健聴者と同様の発達が見られることを明らかにした [5]。乳幼児期に装用手術を受け、聴力回復を図ることによって、健聴者と同様の言語発達及び言語獲得が期待できる [6]。人工内耳の適用基準が低年齢化していることもあり、今後小児の人工内耳装用者が増加すると考えられる。高齢化社会の中で難聴人口の増加が予想されることも踏まえると、人工内耳の需要は今後も増加すると見られる。

このように、聴覚障害に対し音の聞こえを回復させる効果を発揮する一方で、人工内耳装用者の音の聞こえには個人差が大きい [6, 7]。そのため、人工内耳装用者の聴取パフォーマンスは視覚からの情報に依存する部分が多いことが示されている [8, 9]。また、一部の感情の認識が難しい、音声から男女を識別することが困難といったように、音声の非言語情報及びパラ言語情報の知覚に課題を残している [10-15]。増える人工内耳装用者の、豊かな音声コミュニケーションを実現するためには、人工内耳用音声信号処理による非言語情報及びパラ言語情報の伝送の改善に向けて、それぞれの情報の知覚に係る音響特徴を明らかにし、効果的な伝

送が可能かどうかを検討する必要がある。

本研究では、非言語情報の1つである緊迫感に注目し、音声の緊迫感知覚に係る音響特徴について検討を行う。地震等の自然災害時において、迅速な避難行動を促すため、情報を適切に伝達する必要がある。避難呼びかけ音声の心理評価に関する検討にて、緊迫感のある音声はより避難行動を促すことが明らかとなっている [16]。これまで、緊迫感知覚に関わる音響特徴に関する検討は数多く行われてきた [17–19]。これらの検討に加えて、人工内耳用音声に含まれる手がかりの中で、緊迫感知覚に直接的に関わる音響特徴を明らかにすることで、人工内耳装用者を含む多くの人々に適切な緊迫感を伝送する音声の実現に貢献できる。

1.2 研究背景

これまで、人工内耳装用者の音声の聞こえを模擬したシミュレータを用いて、人工内耳用音声に関する検討を行う試みが行われてきた [20, 21]。この人工内耳用音声のシミュレータとして雑音駆動音声 (Noise Vcoded Speech: NVS) が知られている [22, 23]。NVS は、振幅変調モデルでみたときに、帯域制限されたランダム雑音を搬送波信号 (時間微細構造) として振幅変調信号 (時間振幅包絡線 (Temporal Amplitude Envelope : TAE)) を駆動して作成されたものである。そのため、NVS では、スペクトル情報の手がかりが無くなり、時間情報の手がかりだけが保存された状態の音声、つまり TAE 情報だけに手がかりが残った音声を合成できる。

先行研究では、NVS を利用して、音声の TAE に言語情報 [24, 25] だけでなく、非言語情報 [26] に関わる重要な手がかりが含まれていることを示した。更に、緊迫感に注目した研究では、TAE に緊迫感知覚に重要な手がかりが含まれることを示した [27]。また、緊迫感知覚に関わる重要な手がかりが 4 Hz から 16 Hz の変調周波数帯域にあることも明らかにした [28]。しかし、この重要な変調周波数帯域を含むが、時間方向を逆転させた TAE を持つ NVS の緊迫感知覚について検討したところ、元の NVS の緊迫感知覚とは様相が異なり、緊迫感の違いが分かりづらくなることがわかった [27]。この結果は、変調周波数帯域だけが緊迫感知覚に重要なのではなく、その時間的な変化も重要であることを示している。そのため、なぜ TAE の時間反転処理によって NVS の緊迫感知覚が変化したのか、その原因を解明する必要がある。

一方、Guo らは、NVS を用いた聴取実験にて、TAE の変調周波数帯域における瞬時変調成分 (Instantaneous Modulation Components: IMCs) が音声の感情知覚に寄与するかどうか検討した。その結果、TAE の時間伸長圧縮処理による IMCs の制御により、TAE の時間反転処理による感情認識率の低下を説明できることを示した [29]。もし、この結果が感情知覚だけに限定されるものではなく、緊迫感知覚にも関係するものであれば、IMCs を操作することで、上述の原因を解明することができるかもしれない。

1.3 本研究の目的

本研究では、TAE の IMCs の制御による NVS の緊迫感知覚の変化について検討する。そのため、次の 4 点について NVS を用いた聴取実験を通して検討する。

- (1) TAE が緊迫感知覚に関する手がかりを含むかどうか。
- (2) 緊迫感知覚に重要な TAE の変調周波数成分について。
- (3) TAE の時間反転処理による、NVS の緊迫感知覚の変化について。
- (4) 時間伸長圧縮処理により IMCs を制御することで、(3) の結果を説明できるか。

上記の 4 点を通して、音声の緊迫感知覚に TAE の IMCs が寄与するかどうか明らかにする。

1.4 本論文の構成

本論文は、7 章で構成される。

第 1 章は、本研究の研究意義、研究背景及び目的を述べる。

第 2 章では、TAE と音声知覚にまつわる知見と、本研究の新規性について述べる。

第 3 章では、本研究で実施した緊迫感知覚実験及び利用した音声刺激の詳細を述べる。

第 4 章では、実施した緊迫感知覚実験のうち、TAE に対する時間反転処理による NVS の緊迫感知覚の変化について検討した実験について、その詳細と実験結果、考察について述べる。

第 5 章では、TAE に対する時間伸長圧縮処理による NVS の緊迫感知覚の変化について検討した実験について、その詳細と実験結果、考察について述べる。

第 6 章では、第 4 章、第 5 章で述べた実験結果を中心に、TAE の IMCs の制御による音声の緊迫感知覚の変化について考察する。

第 7 章では、結論として本研究で明らかになったことと、また残された課題について述べる。

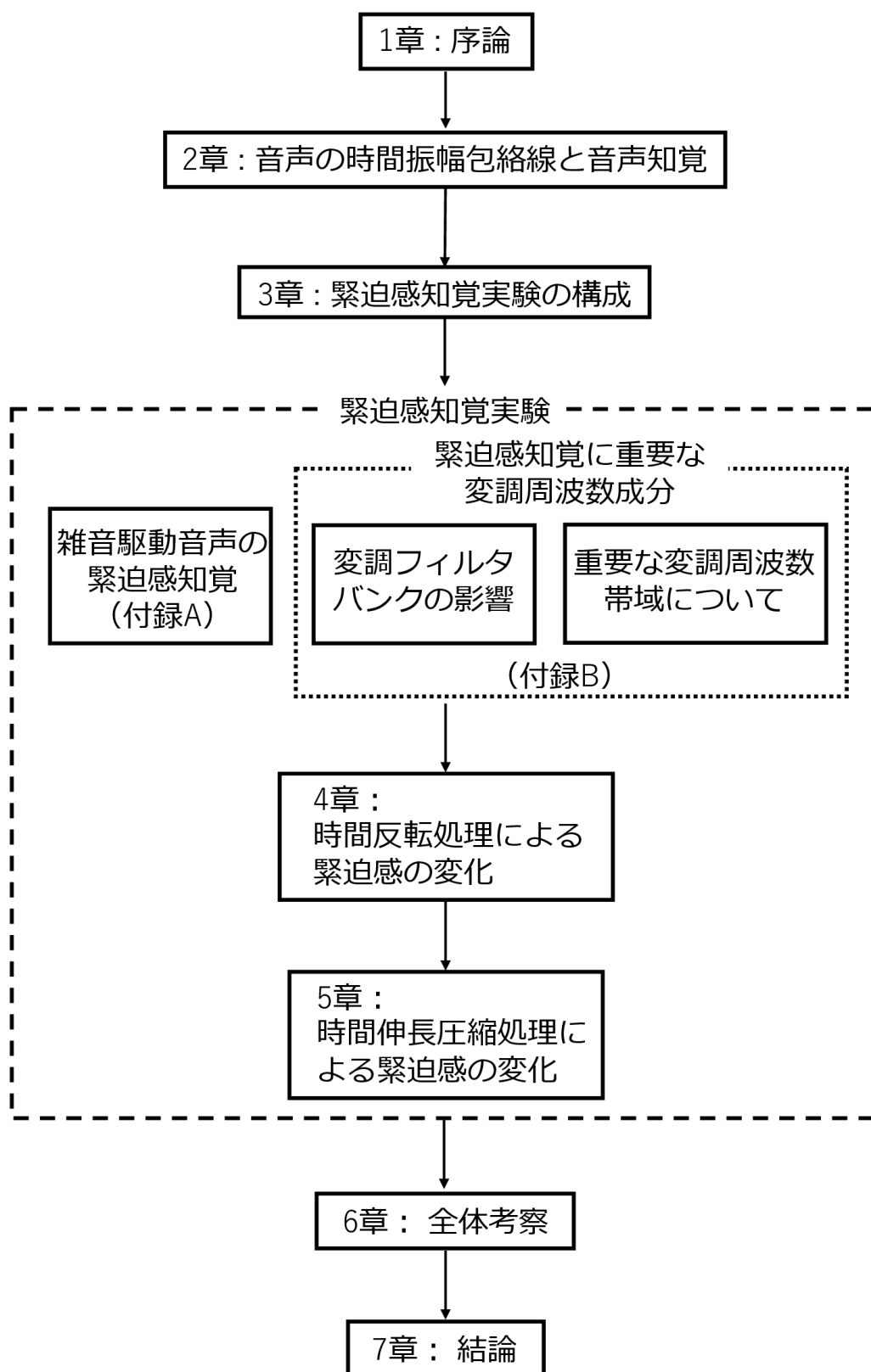


図 1.1: 本論文の構成

第2章 音声の時間振幅包絡線と音声知覚

2.1 時間振幅包絡線と音声知覚の関係

音波は、TAE（変調波）と時間微細構造（搬送波）に分けて考えることができる。音声において、TAE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音像定位への貢献が見られる [30]。先行研究より、TAEには言語情報のみならず、非言語情報についても豊富に含まれていることがわかった。これらの検討について、次より詳しく述べる。

2.1.1 言語情報知覚

これまで、NVSを用いた実験を通して、TAEに含まれる言語情報について検討されてきた。Shannonらは音声波形の包絡線について、NVSを用いた子音、母音及び文章の知覚実験を行った。NVS合成時に、3, 4帯域にわたる包絡線情報を合成することで、80%以上の認識率を得ることができた。このことから、TAEの情報を手がかりに音声の言語情報の知覚が可能であることを示した [24]。小畑らは、日本語の文章を用いたNVSの了解度試験を実施し、振幅包絡線が音声了解度（イントネーション知覚）に大きく関与していることを明らかにした [31]。また、Tachibanaらは、文章など高次の構造レベルをもつ言語情報ほど、音声認識においてよりTAEの貢献が優位になることを明らかにした。これは、TAEが韻律、リズムやアクセントなどの情報をふんだんに含むことを示唆した [32]。

Drullmanは、TAEが音声了解度において重要な情報であることに加えて、音声了解度に重要な変調周波数帯域が16 Hz以下の帯域であることを明らかにした [33]。Zhuらは、言語情報に重要な変調周波数成分について、5 Hz未満の成分が重要であることを示した。この結果より、モーラの音節構造を保存していれば言語情報を十分に知覚できることを示した [34]。なお、TAEから言語情報を知覚するメカニズムについて、生理学的な視点からも言及されている [22, 23, 35, 36]。

2.1.2 非言語情報知覚

音声に含まれる非言語情報について、TAE 情報を手がかりに知覚できることが示唆されている。

Qian らは、健聴者及び人工内耳装用者における、音声からの性別の識別について調査した。その結果、スペクトル的手がかりが乏しい場合、TAE が性別の識別の手がかりとして重要になることを明らかにした [37]。話者識別においても TAE は重要であることが示されている [38]。風間らは、音声の狭帯域包絡線から抽出した包絡線帯域間相関行列を用いることで、話者識別が可能なことを明らかにした [39]。Zhu らは、NVS を用いた話者識別に関する実験を行い、TAE 情報が個人性に関する重要な手がかりを含むことを明らかにした。また、個人性に関わる重要な変調周波数帯域は、8 Hz または 16 Hz 以下の帯域であることを示した [26,34]。Guo らは、感情知覚について更に、TAE の IMCs との関係に言及した [29]。TAE の瞬時変調周波数の時間変動成分に基づき TAE の時間伸長圧縮処理を行うことで、間接的に IMCs を制御した。処理した TAE をもつ NVS を用いて、感情認識に関する聴取実験を行った結果、感情認識率の低減が確認された。この結果は、TAE の IMCs が音声の感情知覚に寄与することを示唆した。

緊迫感知覚について、Unoki らは NVS を用いた聴取実験より、TAE に緊迫感知覚の手がかりが含まれていることを明らかにした。また、緊迫感知覚に重要な変調周波数帯域について、低域通過フィルタ及び高域通過フィルタを用いて TAE の変調周波数帯域を制御し、処理された TAE を持つ NVS の緊迫感知覚について調査した。その結果、6 Hz から 8 Hz の帯域が重要であることを明らかにした [27]。木谷らは、変調フィルタバンク (Modulation Filterbank: MFB) を用いた変調周波数帯域の制御を行い、緊迫感知覚に重要な変調周波数帯域について調査した。その結果、6 Hz から 8 Hz より広い 4 Hz から 16 Hz が重要な帯域であることを示した [28]。これらに加えて Unoki らは、時間反転した TAE を持つ NVS の緊迫感知覚について調査した [27]。低域通過フィルタ及び高域通過フィルタを用いて変調周波数帯域を制御した TAE に対して、時間反転処理を施した。処理された TAE を持つ NVS の緊迫感知覚について、音声刺激の条件に関わらず、知覚される緊迫感に差がないことがわかった。時間反転した TAE は、元の TAE と同じ振幅スペクトルを持つ。上述より明らかになった緊迫感知覚に重要な変調周波数帯域を含む TAE をもつ NVS だったにもかかわらず、元の NVS と違う様相が見られた。以上から、長時間平均的な変調周波数成分だけでなく、瞬時的に時間変化する変調周波数成分も緊迫感知覚において重要な手がかりであることが示唆された。

2.2 本研究の新規性

本研究の新規性は、TAE の IMCs に注目して、音声の緊迫感知覚の直接的な原因を解明することである。音声の緊迫感知覚について、音声提示の仕方による影

響や周波数成分との関係に言及した研究が数多く行われてきた。これらに加えて、TAEに含まれる緊迫感知覚の手がかりについて特定することができれば、緊迫感知覚に直接係る音響特徴について明らかにすることができる。また、感情知覚で明らかになったTAEのIMCsの寄与について、この寄与が感情知覚のみの限定的なものかどうかについて調査することで、複数の非言語情報に跨って共通する知覚のメカニズムを解明できるかもしれない。更に、緊迫感パラ言語情報の側面も持つため、非言語情報知覚での知見をパラ言語情報知覚にも拡張できるかもしれない。

第3章 緊迫感知覚実験の構成

3.1 実験について

本研究では、TAE の IMCs を制御することで、NVS の緊迫感知覚がどのように変化するかどうかを明らかにする。そのために、TAE に対する変調周波数帯域の制御及び時間反転処理、時間伸長圧縮処理による NVS の緊迫感知覚の変化を調査する。これらの調査は、次に説明する 5 つの聴取実験を通して行う。

3.1.1 実験の流れ

まず、TAE 情報のみが保存された音声の緊迫感知覚について検討する。実験 1 として、原音声と原音声を元にして作成した NVS について、それぞれの緊迫感知覚の様相について調査する。両者の様相を比較することで、TAE が緊迫感知覚に重要な手がかりを含むか検討する。

次に、緊迫感知覚において重要な変調周波数成分について検討する。TAE の変調周波数帯域の制御に MFB を用いるにあたって、まず、TAE に対する MFB 処理が NVS の緊迫感知覚に影響を与えるかどうかを検討する。そのために実験 2 として、MFB を利用して作成した NVS と、MFB を利用しないで作成した NVS の緊迫感知覚の様相の違いについて調査する。

実験 2 を踏まえて、MFB を利用して TAE の変調周波数帯域を制御し、緊迫感知覚において重要な手がかりとなる変調周波数成分について検討する。実験 3 として、木谷らの研究 [28] を踏まえて、0 Hz から 64 Hz までの変調周波数帯域を持つ NVS と、4 Hz から 16 Hz までの変調周波数帯域を持つ NVS の緊迫感知覚の様相の違いについて調査する。

ここまでの実験は、Unoki らの研究 [27] 及び木谷らの研究 [28] をなぞり、これらの結果を支持するかどうか確かめるための実験である。そこで、ここで実験 1 から実験 3 の結論を先に述べ、各実験の詳細については付録 A および付録 B に記載する。実験 1 より、TAE は緊迫感知覚に重要な手がかりを含むこと、実験 2 より、TAE に対する MFB 処理は NVS の緊迫感知覚に影響を与えないこと、実験 3 より緊迫感知覚に重要な手がかりとなる変調周波数成分は、4 Hz から 16 Hz までの変調周波数帯域であることがわかった。以上の結果は、Unoki らの研究及び木谷らの研究結果を支持することがわかった。以上を踏まえて、以下の章からは、実験 4 と実験 5 に関する議論を中心に行う。

緊迫感知覚に重要な変調周波数成分について、TAE に対する時間反転処理による NVS の緊迫感知覚の変化を検討する。Unoki らの検討 [27] を踏まえて、実験 4 として、元の NVS と時間反転した TAE を持つ NVS の緊迫感知覚の様相の違いについて調査する。両者の様相を直接比較し、TAE の時間反転処理によって、NVS の緊迫感が増加または低減するかについて明らかにする。

最後に、TAE の IMCs が NVS の緊迫感知覚に与える影響について検討する。実験 5 として、元の NVS、時間反転した TAE を持つ NVS と時間伸長圧縮した TAE を持つ NVS の緊迫感知覚の様相の違いについて調査する。時間伸長圧縮処理を利用して、時間反転した TAE の IMCs を模擬する。これにより、実験 4 の結果を説明できるかどうか、即ち、TAE の時間反転処理による NVS の緊迫感知覚の変化は、TAE の IMCs に依るものかどうか明らかにする。

以上全ての実験を踏まえて、TAE の IMCs と緊迫感知覚の関係について議論する。

3.1.2 実験参加者、実験機器、実験手続き

次より説明する実験参加者、実験機器及び実験手続きは、全ての実験に共通する。

実験参加者は、正常聴力を有する日本語母語話者 14 名（男性 7 名、女性 7 名）であった。実験は防音室の中で行われた。音声刺激の提示には PC（Windows10, matlab）、D/A コンバータ（RME, Fireface UCX）、ヘッドフォン（K240 STUDIO）を使用した。提示音圧レベルが 70 dB 前後になるように、ヘッドアンドトルソシミュレータ（B&K, 4128 型）と騒音計（B&K, 2231 型）を用いて校正した。

実験参加者は図 3.2 に示した操作画面を通して、音声を評価した。具体的には、実験参加者に音声を 2 個続けて提示し、1 番目に流れた音声と比べて 2 番目に流れた音声はどの程度緊迫しているかについて、表 3.1 の評価カテゴリに基づき 5 段階で評価させた。これらの評価結果から、シェッフェの一対比較法（浦の変法）を利用して各音声の緊迫感の程度を算出した [40]。

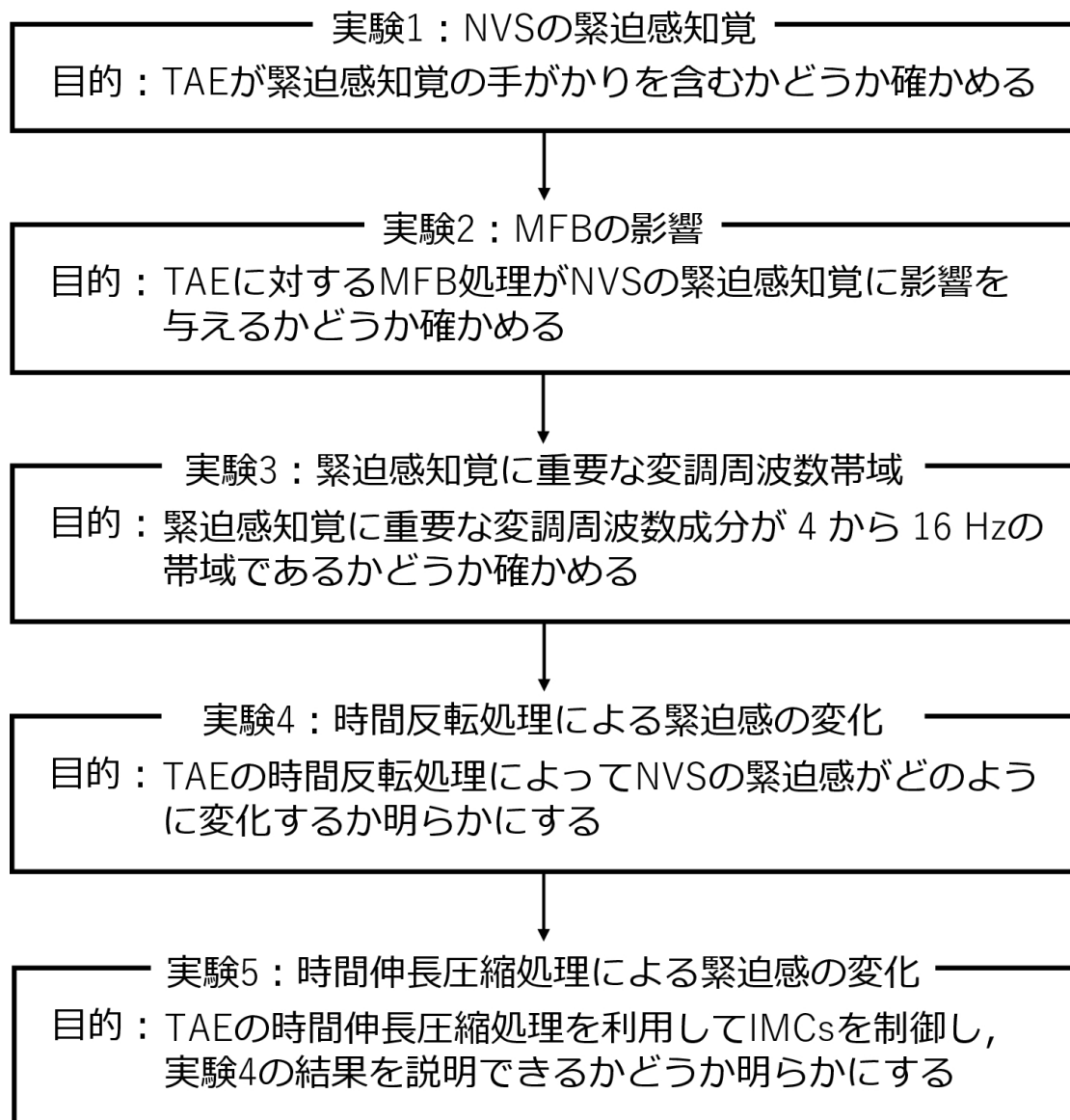


図 3.1: 実験の流れ

3.2 音声刺激

原音声として、小林・赤木らの研究 [16] で利用された音声刺激を利用する。この音声は、男性アナウンサー 1 名が発話した避難呼びかけ音声である。緊迫感の異なる 4 種の原音声 (A, B, C, D) に対して、各実験の目的に合わせて NVS を作成する。

表 3.2 に実験で利用する NVS の呼称と作成条件を示す。この表には、作成した NVS のラベルと、TAE の変調周波数領域の処理（変調周波数帯域の分割処理に用いた、MFB の変調フィルタ (Modulation Filter: MF) 番号）、TAE の時間領域の処理が順番に記されている。なお、“—” は、処理なしを示す。

3.3 雑音駆動音声の作成方法

図 3.3 に NVS の作成方法を示す。この作成方法は、主に 5 つの処理で構成される。

まず、処理 I では、ヒトの周波数分解機能を模擬した聴覚フィルタバンク (Auditory Filterbank: AFB) を利用して、入力信号 $s(n)$ を帯域分割する。ここで、聴覚フィルタ (Auditory Filter: AF) $\#k$ を利用して帯域分割された信号 $s_k(n)$ は、次式で表される。

$$s_k(n) = s(n) * h_k(n) \quad k = 1, 2, \dots, K \quad (3.1)$$

ただし、 $*$ は畳み込み演算子、 $h_k(n)$ は AF $\#k$ のインパルス応答を表し、 K は AFB の総チャンネル数を表す。ここでは、 $s(n)$ を 16 帯域 ($K = 16$) に分割する。AFB として IIR (Infinite Impulse Response) 型の 6 次 Butterworth 帯域通過フィルタ (Band-pass Filter: BPF) 群を利用する。ERB_N は健聴者の聴覚フィルタの帯域幅を表し、各フィルタの帯域幅は、等価矩形帯域幅 (Equivalent Rectangular Bandwidth: ERB) になるように近似した。AFB におけるフィルタの並びは ERB_N-number 尺度に沿うように調整した [41]。

次に、処理 II では、次式を利用して、帯域分割された信号 $s_k(n)$ の時間振幅包絡線 (TAE) $e_k(n)$ を抽出する。

$$e_k(n) = \text{LPF}[|s_k(n) + j\mathcal{H}[s_k(n)]|] \quad (3.2)$$

ただし、LPF は低域通過フィルタ (Low-pass filter: LPF) を、 $\mathcal{H}(\cdot)$ は Hilbert (HT) 変換を表す。ここでは、IIR 型の 2 次 Butterworth で構成された LPF を利用し、カットオフ周波数は、64 Hz とした。

処理 III では、MFB を利用して $e_k(n)$ を帯域分割する。ただし、フィルタリングによる直流成分への影響を考慮するために、事前に TAE の直流成分を除いた状態で、TAE を MFB に入力する。MFB の処理後に、MF $\#1$ の成分に改めて直流成分を足して、元の成分に戻す。MF $\#k$ における TAE $e_{k,m}(n)$ は以下の式で表される。

$$e_{k,m}(n) = e_k(n) * q_m(n) \quad m = 1, 2, \dots, M \quad (3.3)$$

ただし、 $q_m(n)$ は MF # k のインパルス応答を、 M は MFB の総チャンネル数を表す。ここでは、 $e_k(n)$ を 6 帯域 ($M = 6$) に分割する。表 3.3 に各 MF の帯域幅を示す。MFB は MF #1 に 0 Hz から 2 Hz の IIR 型の 5 次 Butterworth フィルタで構成された LPF を、MF # m ($m = 2, 3, \dots, M$) に 2 Hz から 64 Hz の間で (2^{m-1}) Hz の帯域幅をもつ IIR 型の 10 次 Butterworth フィルタで構成された BPF を利用する。その後、 $e_{k,m}(n)$ を足し合わせることで、変調周波数帯域を制御した TAE $\hat{e}_k(n)$ を作成する。 $\hat{e}_k(n)$ は次式で表される。

$$\hat{e}_k(n) = \sum_{m=1}^M e_{k,m}(n) \quad (3.4)$$

処理 IV では、時間伸長圧縮 (TSC) 処理といった TAE の時間領域における処理を行う。この処理の詳細は、次節で述べる。TAE の時間反転処理に関しては、本研究では処理 IV で行わずに、音声信号の入力時に、原音声事前に時間反転させることで対応した。図 3.4 から図 3.7 に時間反転した TAE 及び IMCs の一例を示す。

最後に、処理 V では、狭帯域雑音 (Narrow-band Noise: NBN) $W_k(n)$ に $e_k(n)$, $\hat{e}_k(n)$, または時間伸長圧縮処理を経て作成された TAE $E_k(n)$ のいずれかの TAE を乗じて得た帯域信号を、全周波数帯域で足し合わせることで NVS $v(n)$ を作成する。 $W_k(n)$ は白色ガウス雑音を AF で帯域分割することで作成される。NVS $v(n)$ は以下の式で表される。

$$v(n) = \sum_{k=1}^K \hat{e}_k(n) W_k(n) \quad (3.5)$$

3.4 時間伸長圧縮処理

処理 IV での TAE の時間伸長圧縮処理の詳細について述べる。まず、変調周波数帯域が制限された TAE $\hat{e}_k(n)$ の IMF $f_k(n)$ を抽出する。TAE の瞬時位相 $p_k(n)$ 及び IMF $f_k(n)$ は次式で表される。

$$p_k(n) = \arctan \left(\frac{\mathcal{H}[\hat{e}_k(n)]}{\hat{e}_k(n)} \right) - 2\pi f_c t_k(n) \quad (3.6)$$

$$f_k(n) = \begin{cases} 0 & n = 0 \\ \frac{(p_k(n) - p_k(n-1))f_s}{2\pi} + f_c & \text{otherwise} \end{cases} \quad (3.7)$$

ただし、 $\mathcal{H}(\cdot)$ は Hilbert 変換を、 $t_k(n)$ は時間を、 f_c は TAE $\hat{e}_k(n)$ の中心周波数を表す。

表 3.1: 緊迫感知覚実験における評価カテゴリ

カテゴリ	評価
かなり緊迫している	+2
緊迫している	+1
同程度	0
緊迫していない	-1
かなり緊迫していない	-2

表 3.2: 実験で利用した音声刺激の呼称と作成条件

ラベル	TAE の変調周波数領域の処理	TAE の時間領域の処理
A, B, C, D (原音声)	–	–
a ₁ , b ₁ , c ₁ , d ₁	–	–
a ₂ , b ₂ , c ₂ , d ₂	MF #1, #2, #3, #4, #5, #6	–
a ₃ , b ₃ , c ₃ , d ₃	MF #3, #4	–
a ₄ , b ₄ , c ₄ , d ₄	MF #3, #4	時間反転
a _{5C} , b _{5C} , c _{5C} , d _{5C}	MF #3, #4	圧縮のみ
a _{5S} , b _{5S} , c _{5S} , d _{5S}	MF #3, #4	伸長のみ
a _{5SC} , b _{5SC} , c _{5SC} , d _{5SC}	MF #3, #4	伸長圧縮

表 3.3: 各 MF の帯域幅

MF (ch)	帯域幅 (Hz)
#1	0 ~2
#2	2 ~4
#3	4 ~8
#4	8 ~16
#5	16 ~32
#6	32 ~64

時間変動成分 $\hat{f}_k(n)$ は、TAE に含まれる瞬時変調周波 (IMF) $f_k(n)$ の瞬時周波数として計算される。IMF の時間変動成分を導出するための瞬時位相 $\hat{p}_k(n)$ は次式で表される。

$$\hat{p}_k(n) = \arctan\left(\frac{\mathcal{H}[f_k(n)]}{f_k(n)}\right) - 2\pi\hat{f}_c t_k(n) \quad (3.8)$$

ただし、 $\hat{f}_c$ は IMF $f_k(n)$ の中心周波数である。IMF の時間変動成分 $\hat{f}_k(n)$ は、先に導出した瞬時位相から次式を利用して求める。

$$\hat{f}_k(n) = \begin{cases} 0 & n = 0 \\ \frac{(\hat{p}_k(n) - \hat{p}_k(n-1))f_s}{2\pi} + \hat{f}_c & \text{otherwise} \end{cases} \quad (3.9)$$

次に、時間伸長圧縮処理で利用する時間伸長圧縮係数 (時間伸長圧縮率) $\alpha(n)$ を求めるために、次式を利用して、時間変動成分の傾き a_k と時間重心 $c_{t,k}$ を求める。

$$a_k = \frac{N \sum_{n=1}^N t_k(n) \hat{f}_k(n) - \sum_{n=1}^N t_k(n) \sum_{n=1}^N \hat{f}_k(n)}{N \sum_{n=1}^N t_k^2(n) - \left(\sum_{n=1}^N t_k(n)\right)^2} \quad (3.10)$$

$$c_{t,k} = \frac{\sum_{n=1}^N t_k(n) \hat{f}_k(n)}{\sum_{n=1}^N \hat{f}_k(n)} \quad (3.11)$$

これらを利用して、次式から $\alpha(n)$ を求める。

$$\alpha(n) = 1 + \left(\frac{1}{3}\right)^{p_l} \bar{a}(t(n) - \bar{c}_t) > 0 \quad (3.12)$$

ここで、 $\bar{a}$ 及び $\bar{c}_t$ は、AF #7, #8, #9 から算出された傾きの平均値と重心の平均値をそれぞれ表す。時間伸長圧縮係数 $\alpha(n)$ は正の値である必要があるが、傾き $\bar{a}$ や重心 $\bar{c}_t$ の値によっては、負の値をとる場合がある。そのため、傾きを抑制するための係数として、傾き抑制係数 $(\frac{1}{3})^{p_l}$ を導入する。傾き抑制係数 $(\frac{1}{3})^{p_l}$ の指数項である p_l は次式で表される。

$$p(n) = \begin{cases} 0 & 1 + \bar{a}(t(n) - \bar{c}_t) > 0 \\ 1 & 1 + \bar{a}(t(n) - \bar{c}_t) = 0 \\ \lceil \log_{\frac{1}{3}}\left(\frac{-1}{\bar{a}(t(n) - \bar{c}_t)}\right) \rceil & 1 + \bar{a}(t(n) - \bar{c}_t) < 0 \end{cases} \quad (3.13)$$

$$p_l = \max(p(n)) \quad (3.14)$$

処理条件による各 α を、図 3.8 から図 3.11 に示す。

最後に TAE $\hat{e}_k(n)$ に対して時間伸長圧縮を行うことで、間接的に IMF の時間変動成分を制御する。時間伸長圧縮では、 $\alpha(n)$ の値が 1 以上の場合 $\hat{e}_k(n)$ は伸長さ

れ、 $\alpha(n)$ の値が 1 未満の場合 $\hat{e}_k(n)$ は圧縮される．時間伸長圧縮の過程においては、TAE の時間ビンが不連続になるため、それを一定に保つために補間処理を行う．時間伸長圧縮された TAE $E_k(t(n))$ は次式によって表される．

$$\hat{t}(n) = \begin{cases} 0 & n = 0 \\ \hat{t}(n-1) + \Delta t(n)\alpha(n) & \text{otherwise} \end{cases} \quad (3.15)$$

$$\Delta t(n) = t(n) - t(n-1) \quad (3.16)$$

$$E_k(t(n)) = \text{Interp}[\hat{e}_k(\hat{t}(n))] \quad (3.17)$$

ただし、 $\hat{t}(n)$ は、時間伸長圧縮された時間のサンプル点を表し、 $\text{Interp}[\cdot]$ は補間処理を表す．各処理条件合わせて作成された TAE の一例を、図 3.12 から図 3.23 に示す．

T

0/0

Subject

Select the section

Start

Aに比べてBはどの程度緊迫している,あるいはしていないか
(Bの音声について教えてください)

0.0

0.1

0.2

0.3

0.4

0.5

0.6

0.7

0.8

0.9

1.0

かなり緊迫していない

緊迫していない

同程度

緊迫している

かなり緊迫している

Next

図 3.2: 操作画面

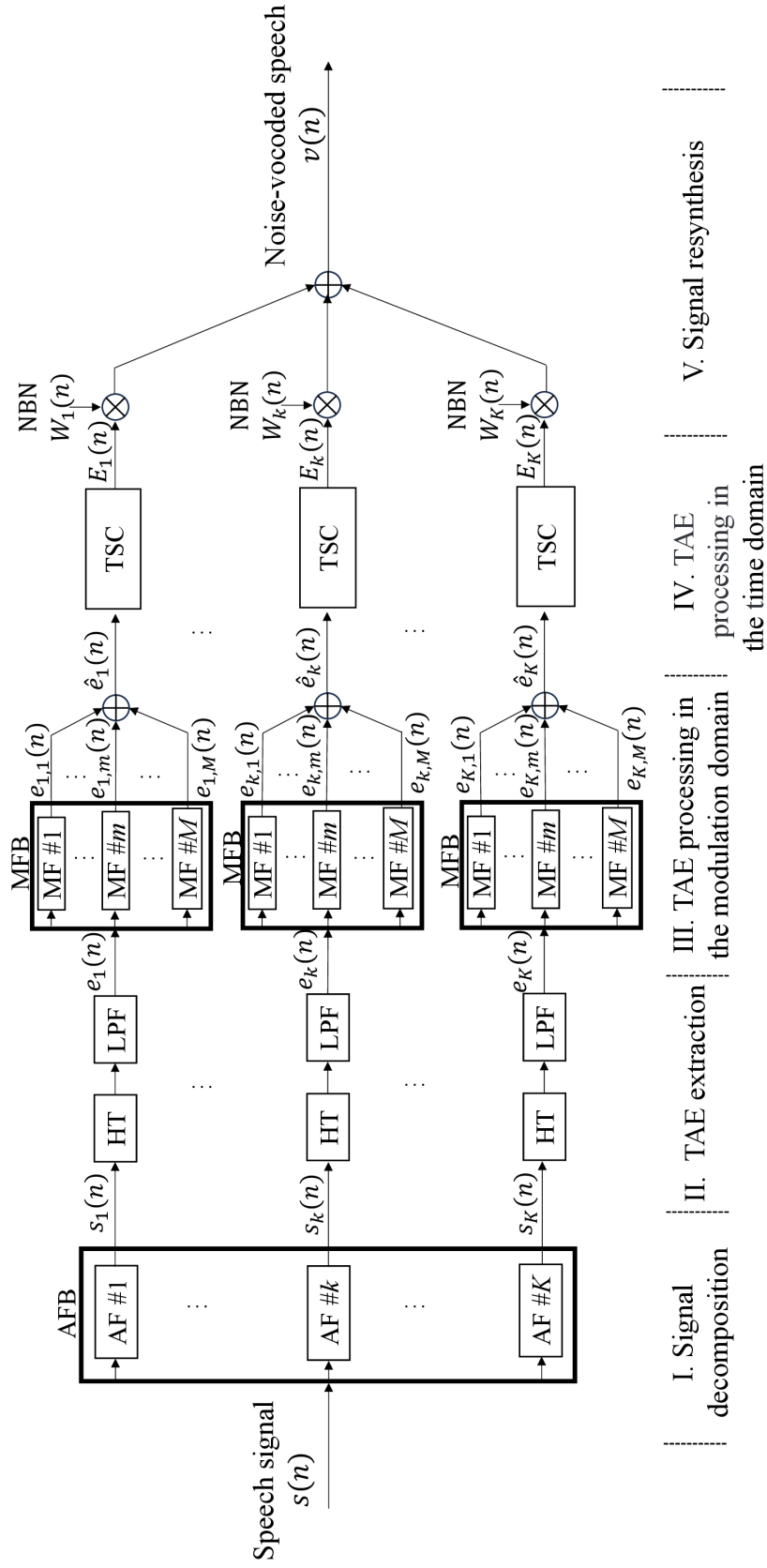
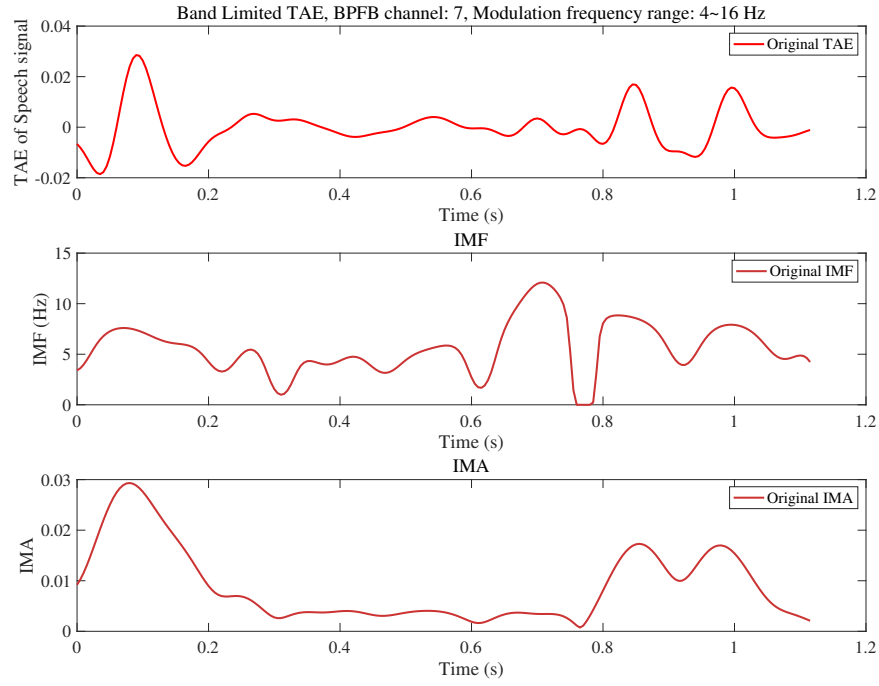
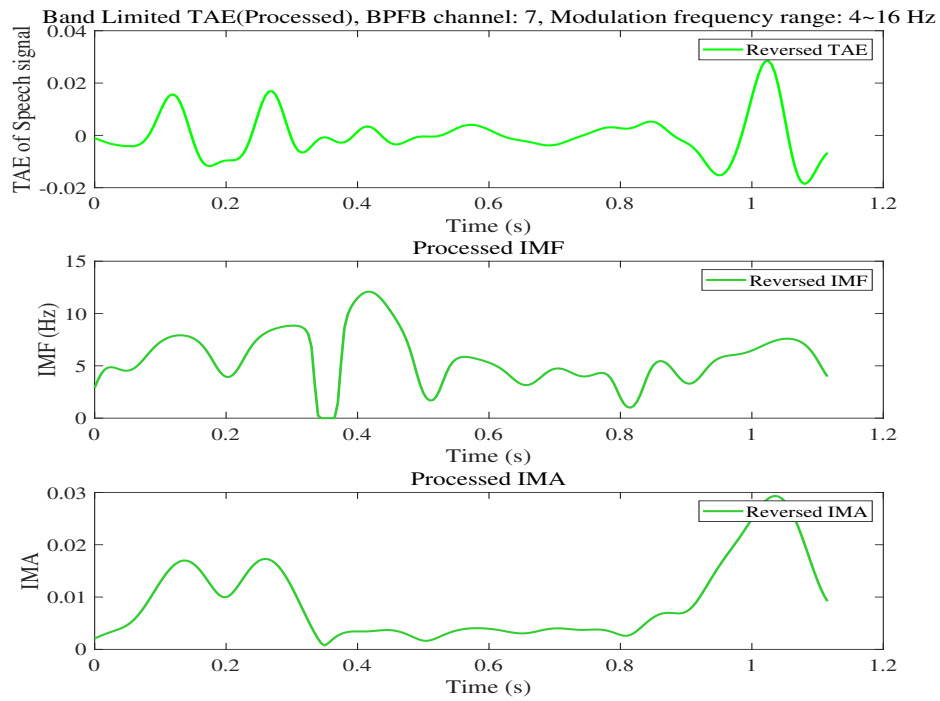


図 3.3: NVS の作成方法の概略

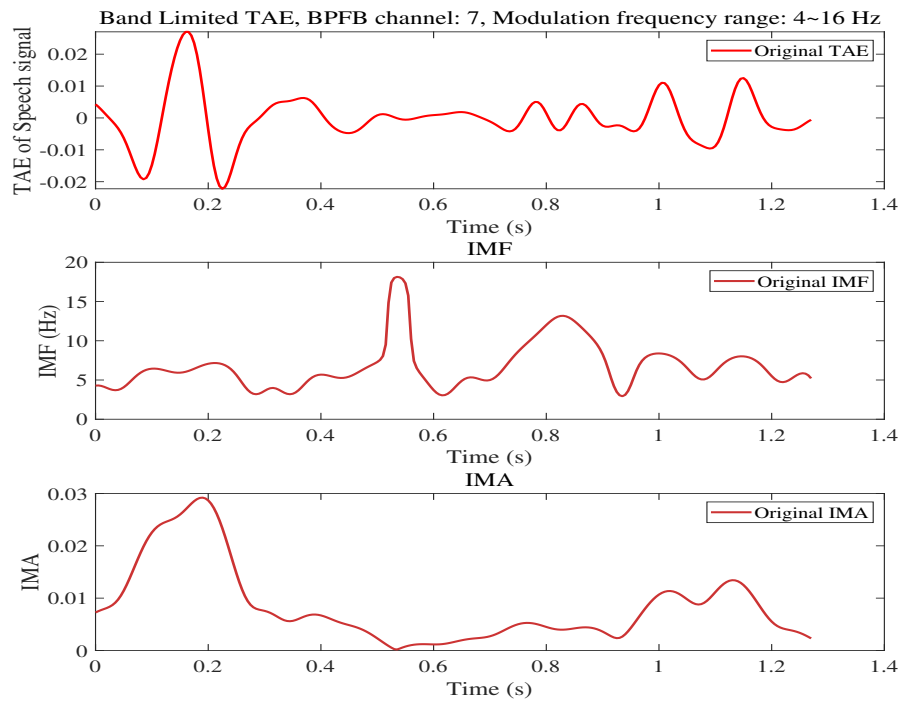


(a) 元の音声

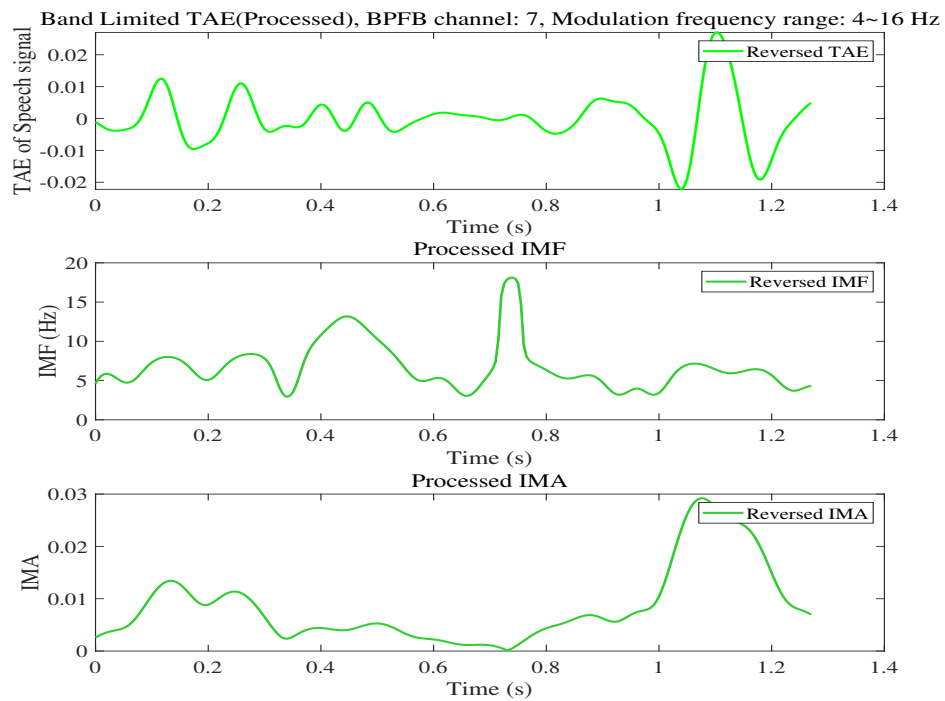


(b) 時間反転処理

図 3.4: TAE と IMCs (音声 A)

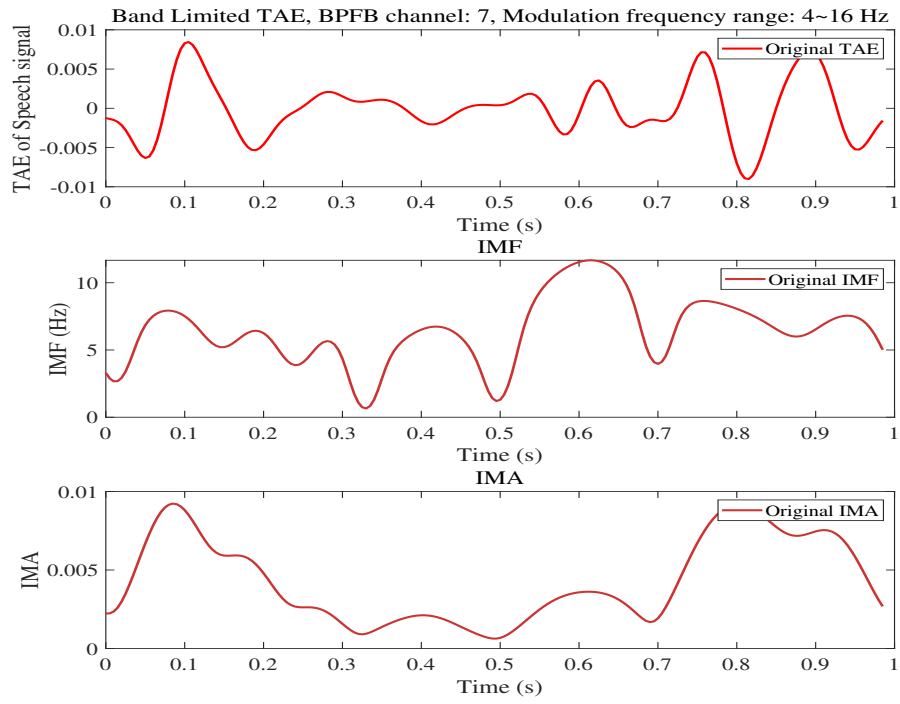


(a) 元の音声

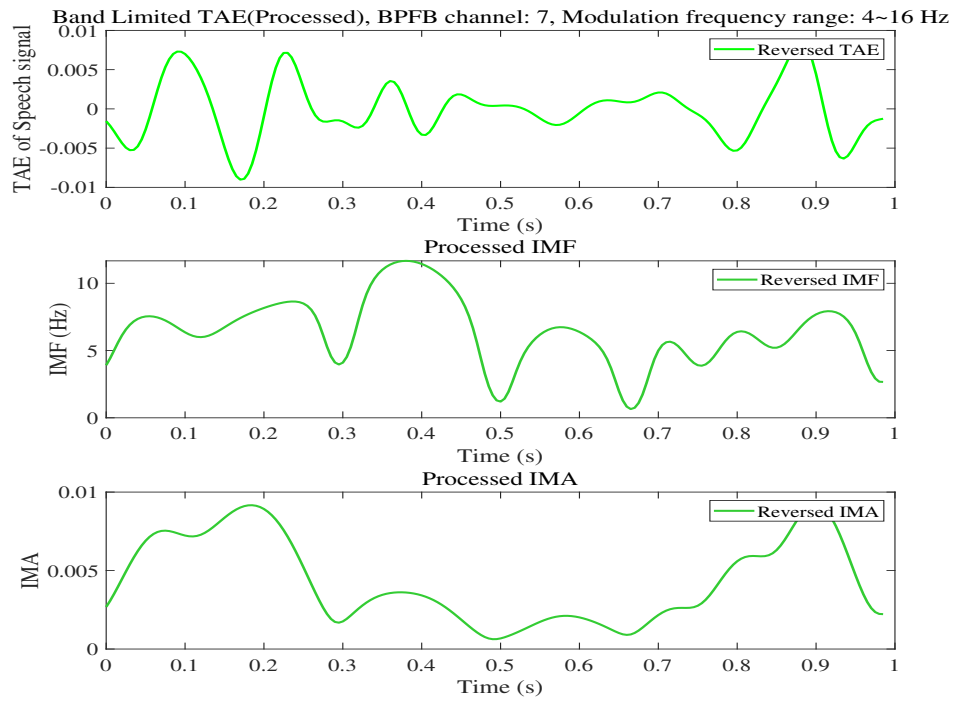


(b) 時間反転処理

図 3.5: TAE と IMCs (音声 B)

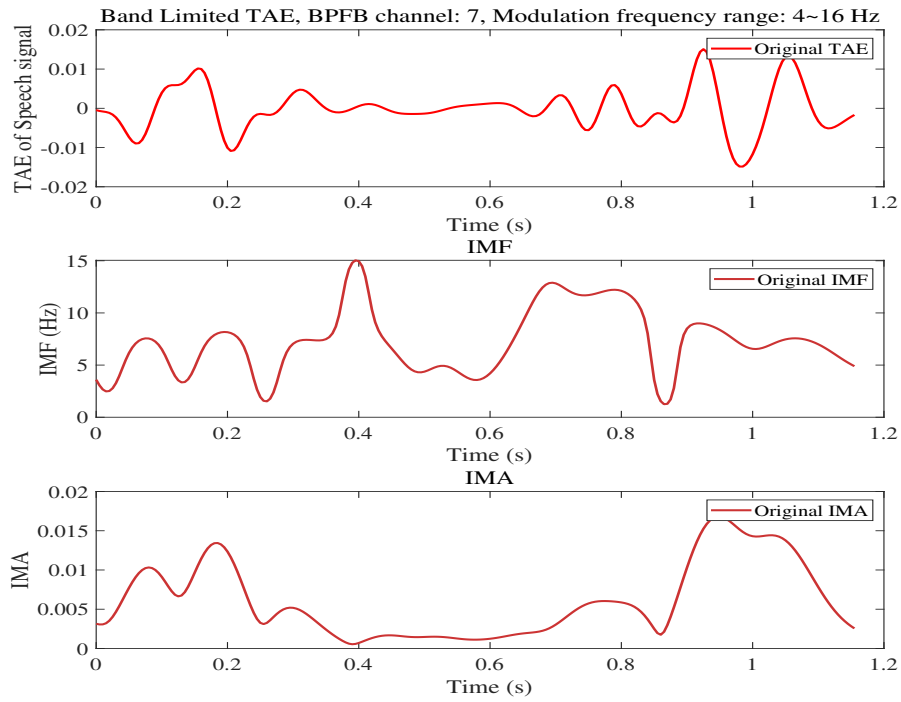


(a) 元の音声

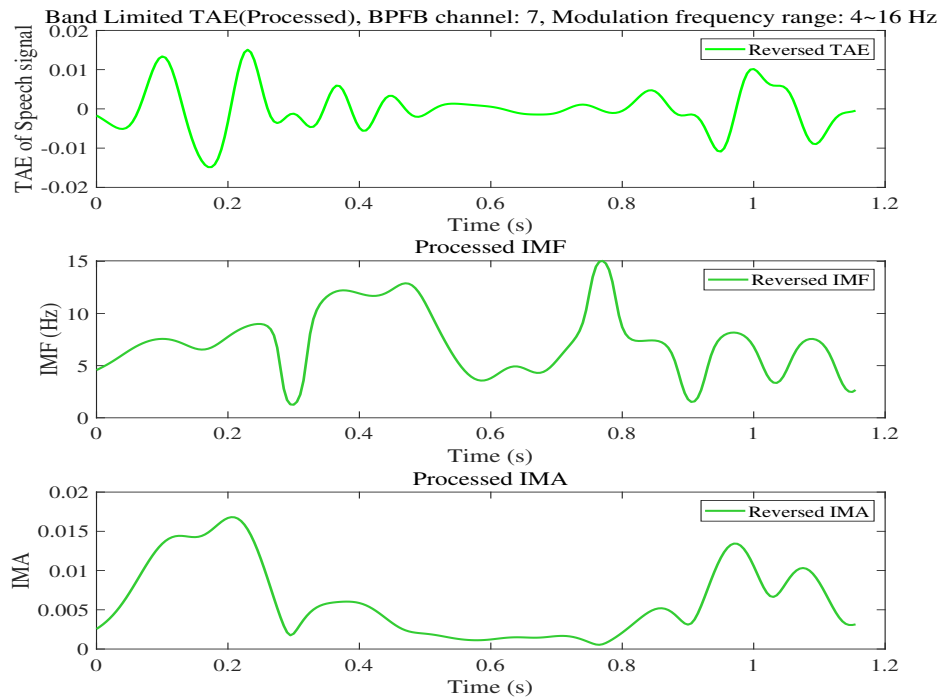


(b) 時間反転処理

図 3.6: TAE と IMCs (音声 C)

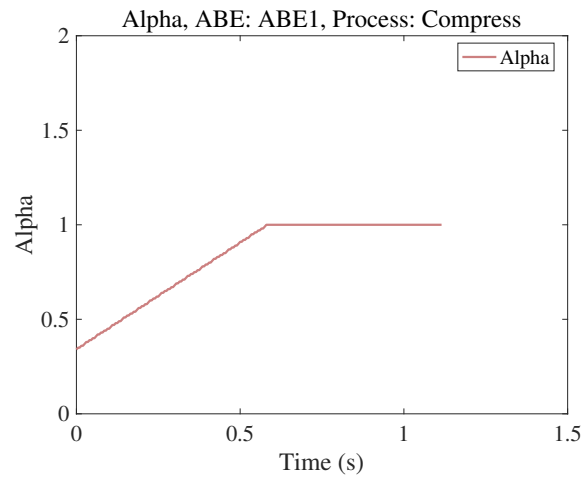


(a) 元の音声

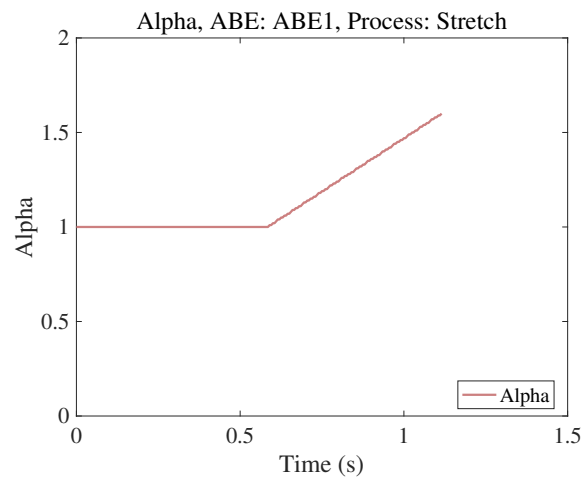


(b) 時間反転処理

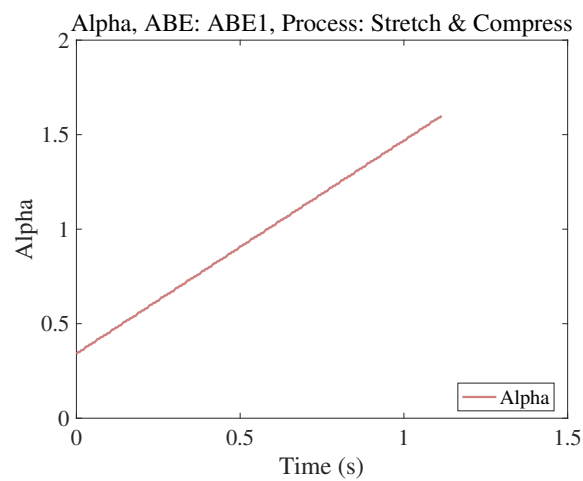
図 3.7: TAE と IMCs (音声 D)



(a) Alpha (圧縮のみ)

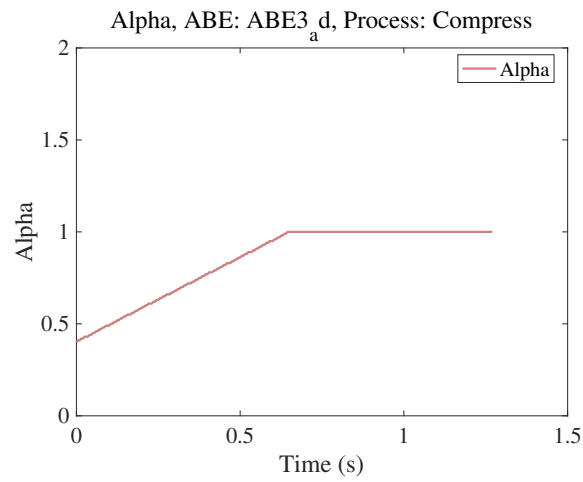


(b) Alpha (伸長のみ)

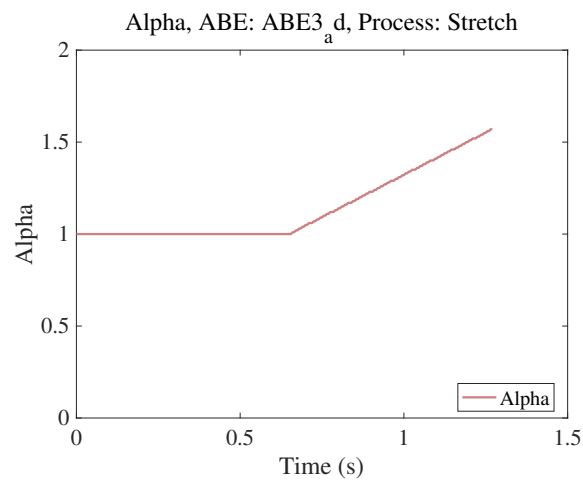


(c) Alpha (伸長圧縮)

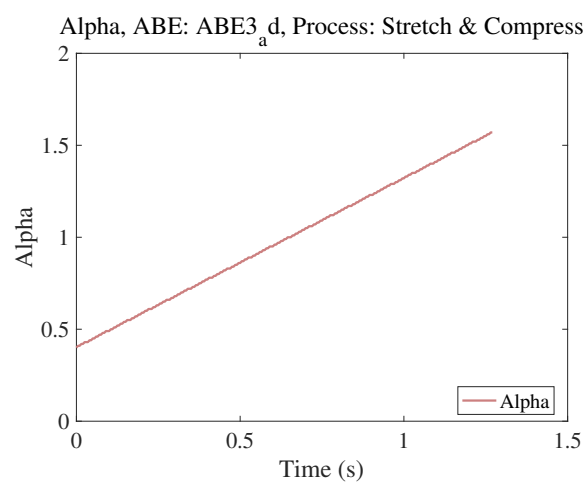
図 3.8: Alpha (音声 A)



(a) Alpha (圧縮のみ)

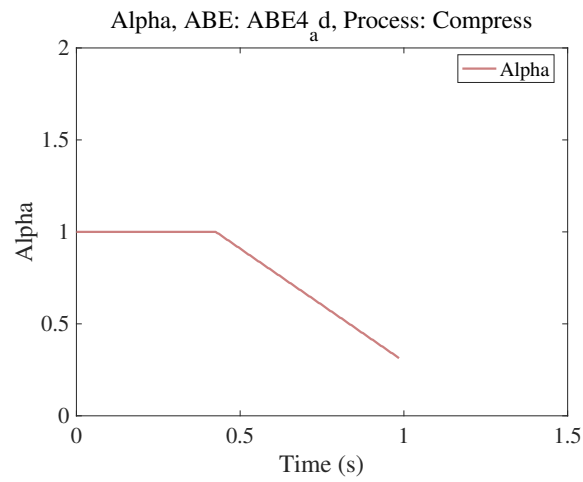


(b) Alpha (伸長のみ)

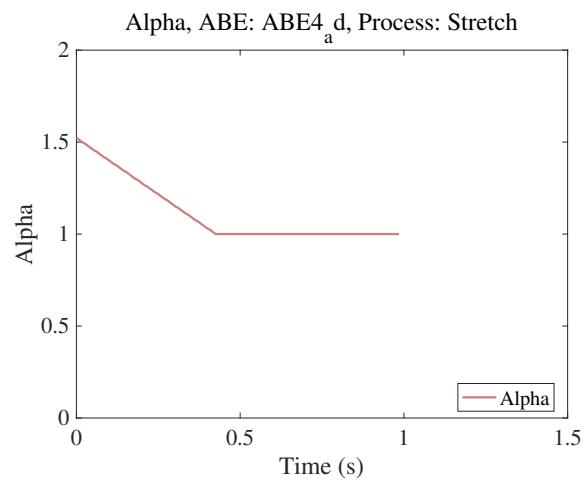


(c) Alpha (伸長圧縮)

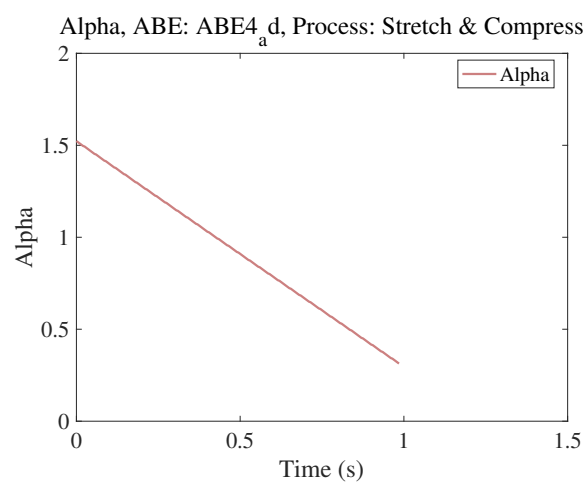
図 3.9: Alpha (音声 B)



(a) Alpha (圧縮のみ)

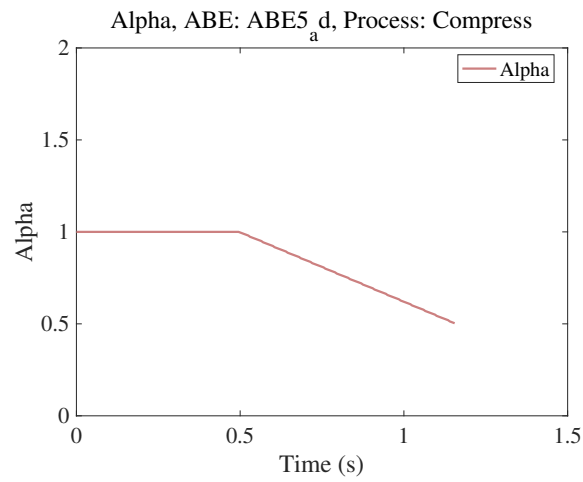


(b) Alpha (伸長のみ)

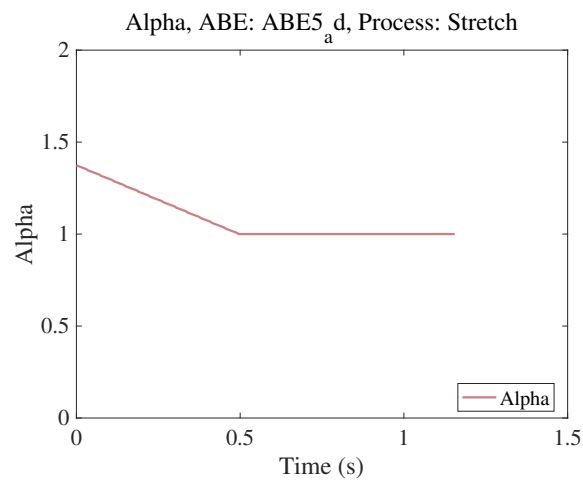


(c) Alpha (伸長圧縮)

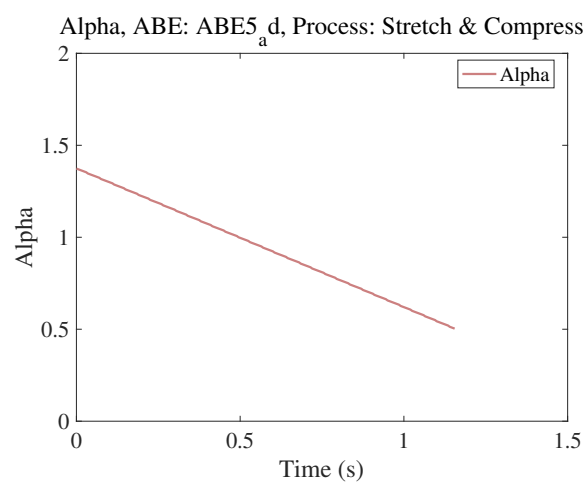
図 3.10: Alpha (音声 C)



(a) Alpha (圧縮のみ)



(b) Alpha (伸長のみ)



(c) Alpha (伸長圧縮)

図 3.11: Alpha (音声 D)

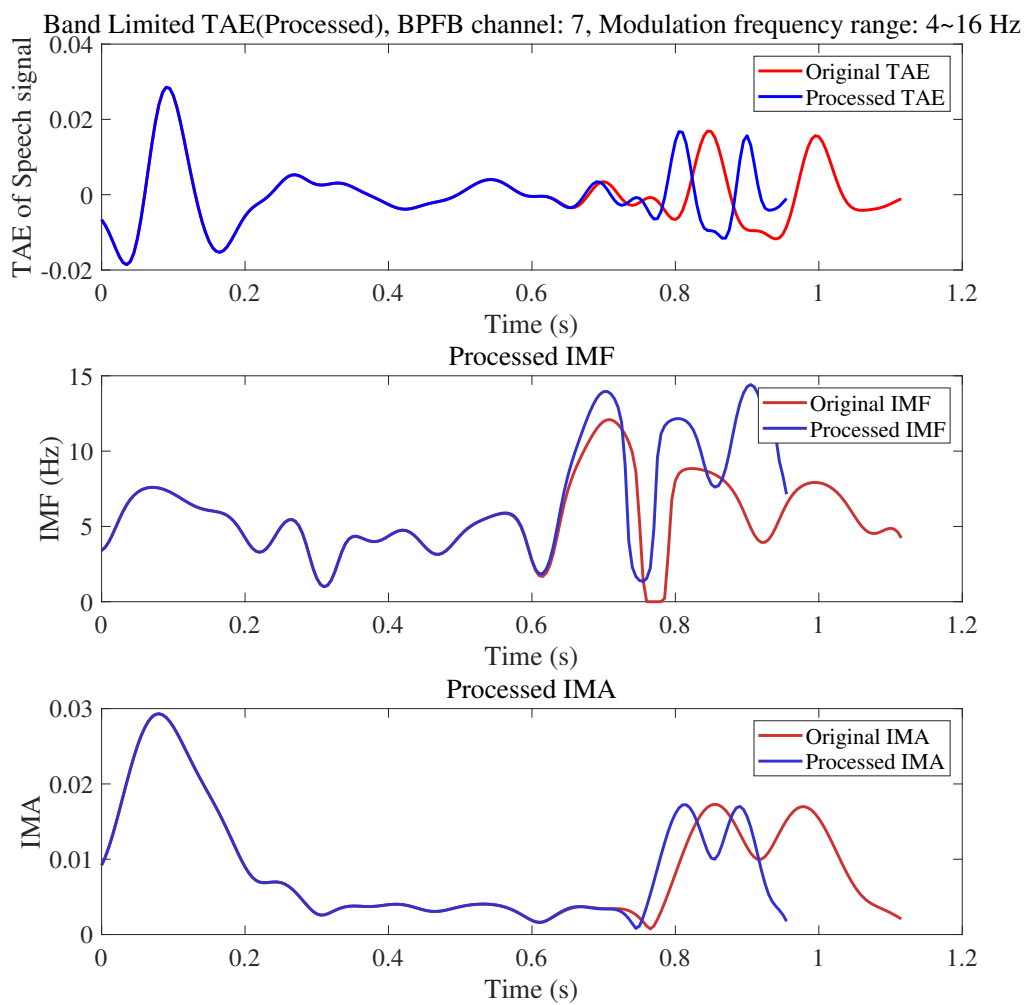


図 3.12: 時間圧縮処理した TAE と IMCs (音声 A)

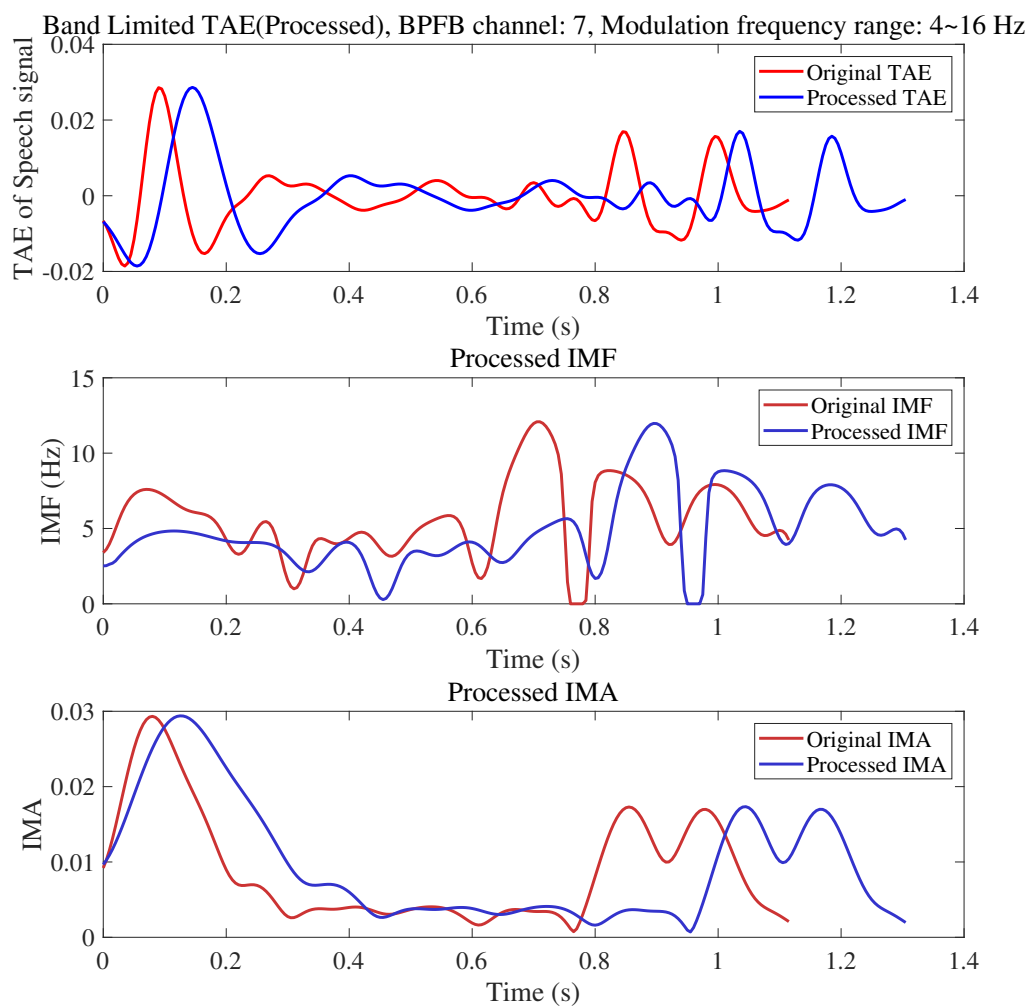


図 3.13: 時間伸長処理した TAE と IMCs (音声 A)

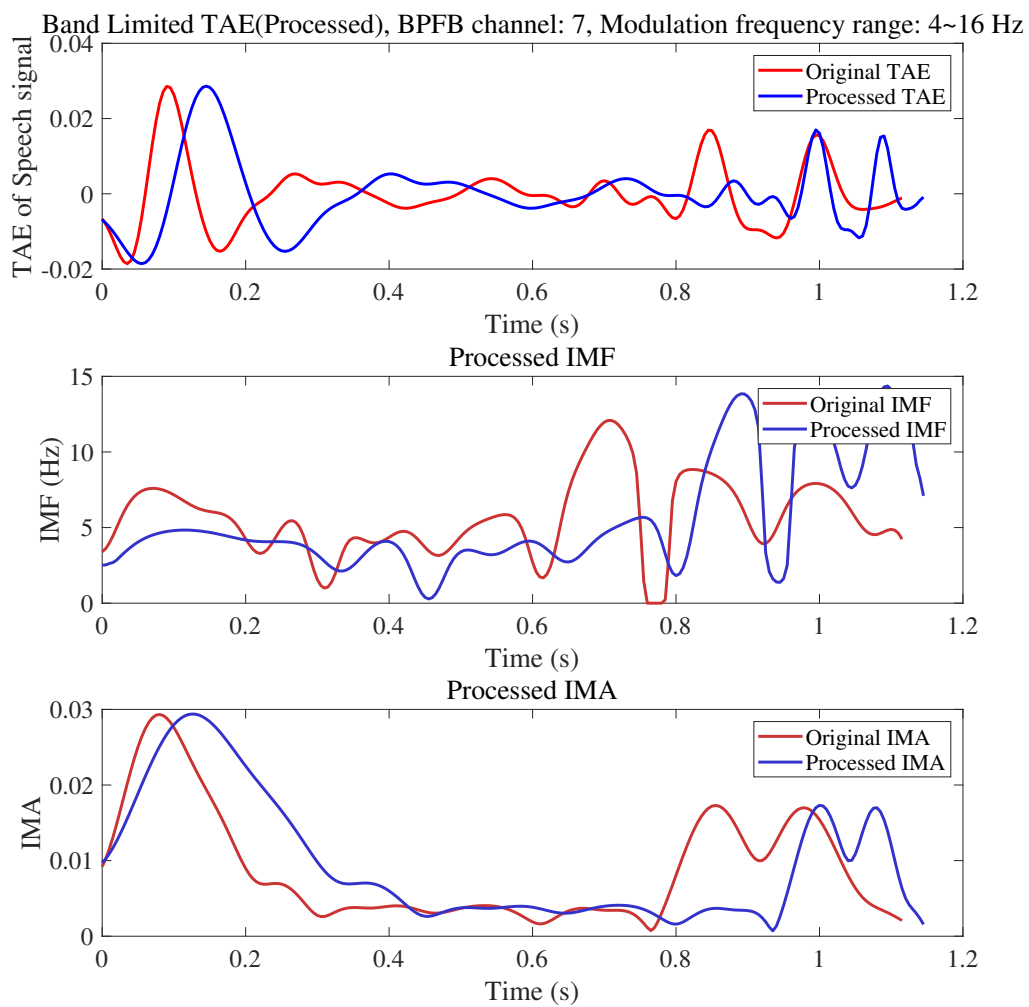


図 3.14: 時間伸長圧縮処理した TAE と IMCs (音声 A)

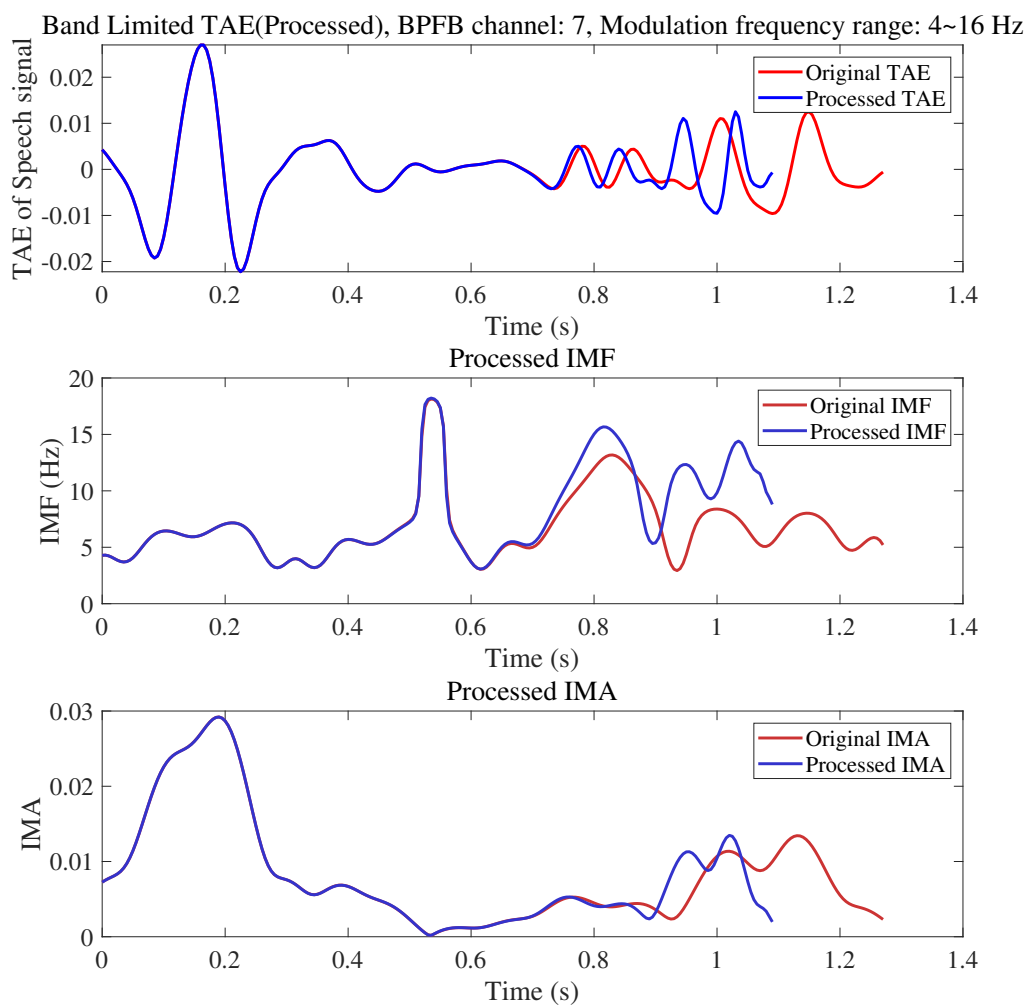


図 3.15: 時間圧縮処理した TAE と IMCs (音声 B)

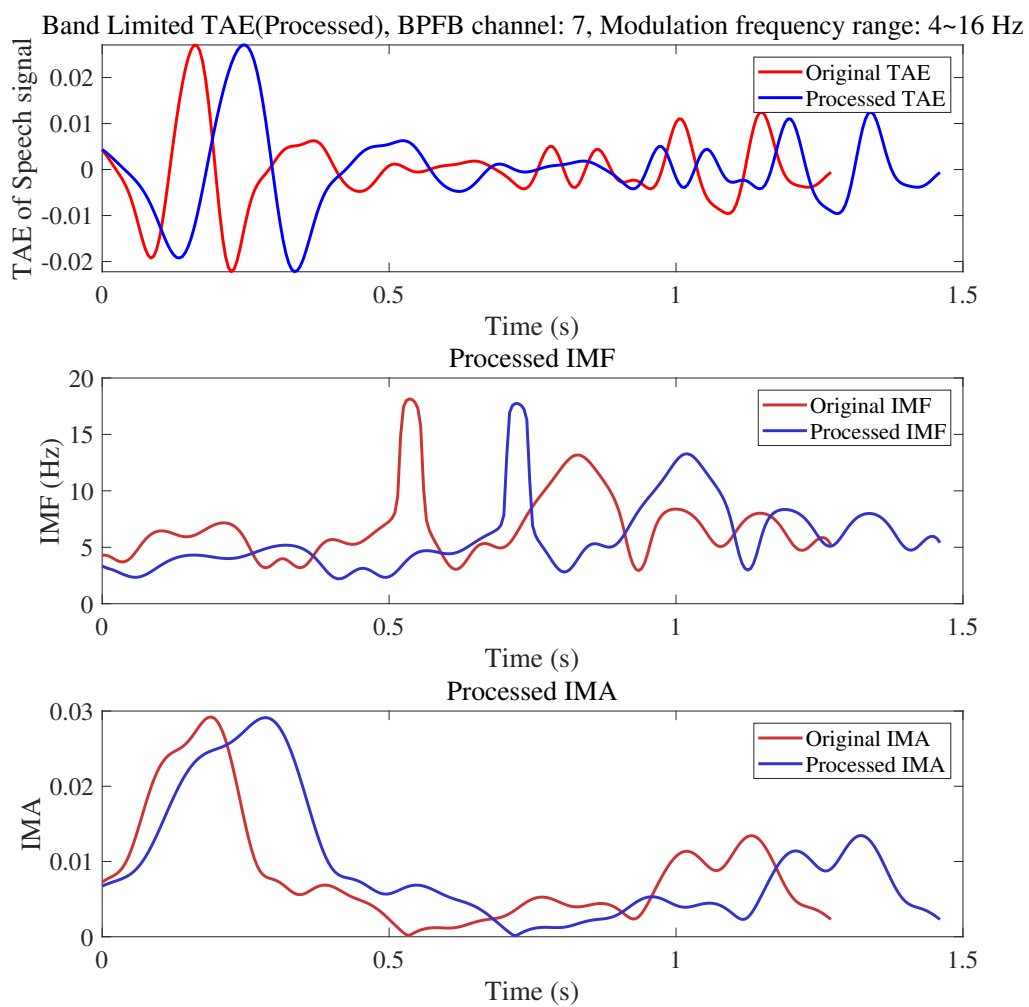


図 3.16: 時間伸長処理した TAE と IMCs (音声 B)

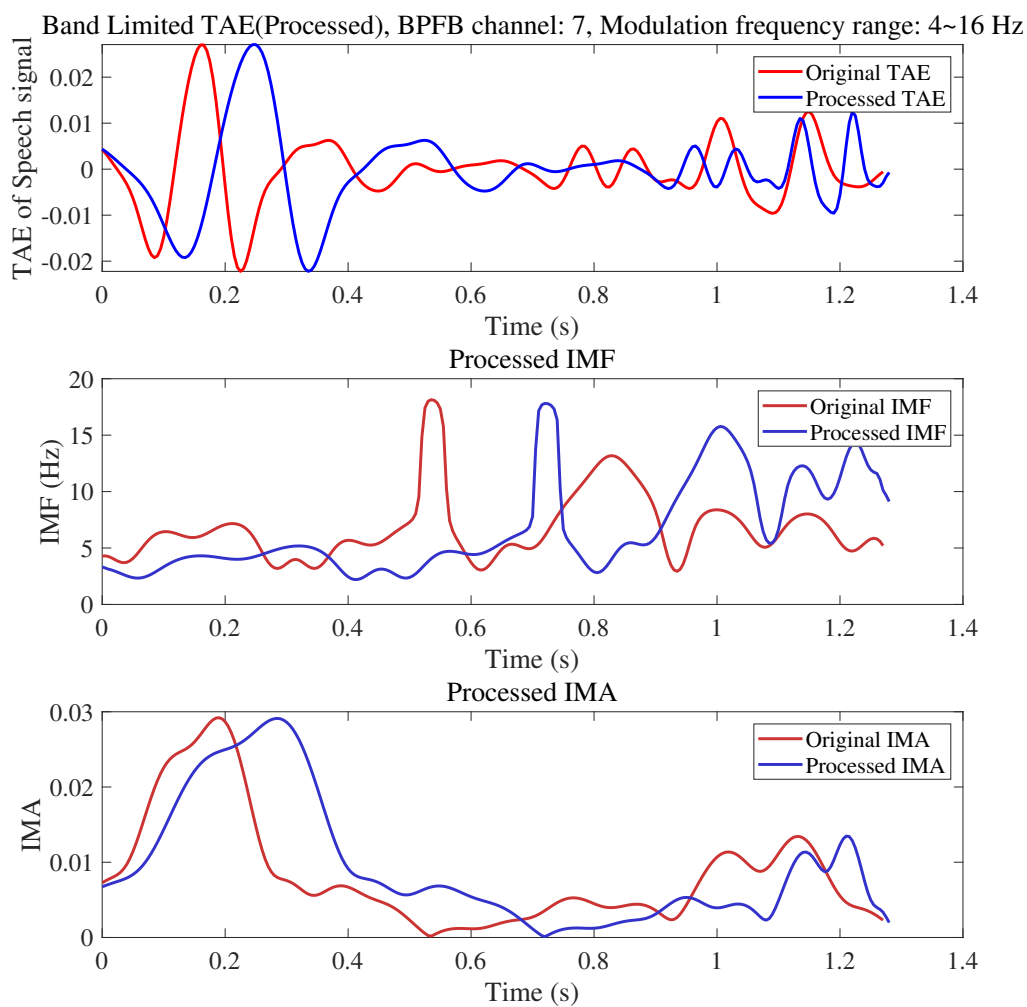


図 3.17: 時間伸長圧縮処理した TAE と IMCs (音声 B)

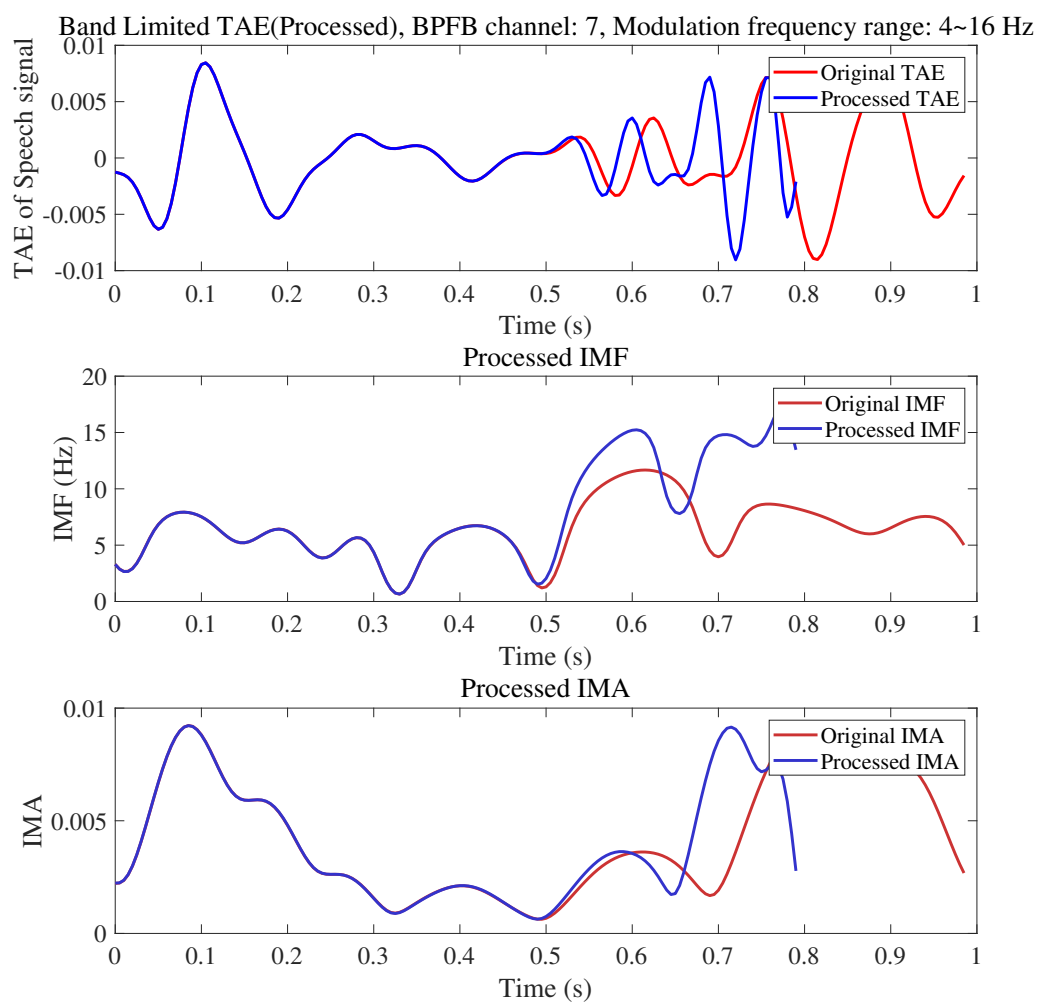


図 3.18: 時間圧縮処理した TAE と IMCs (音声 C)

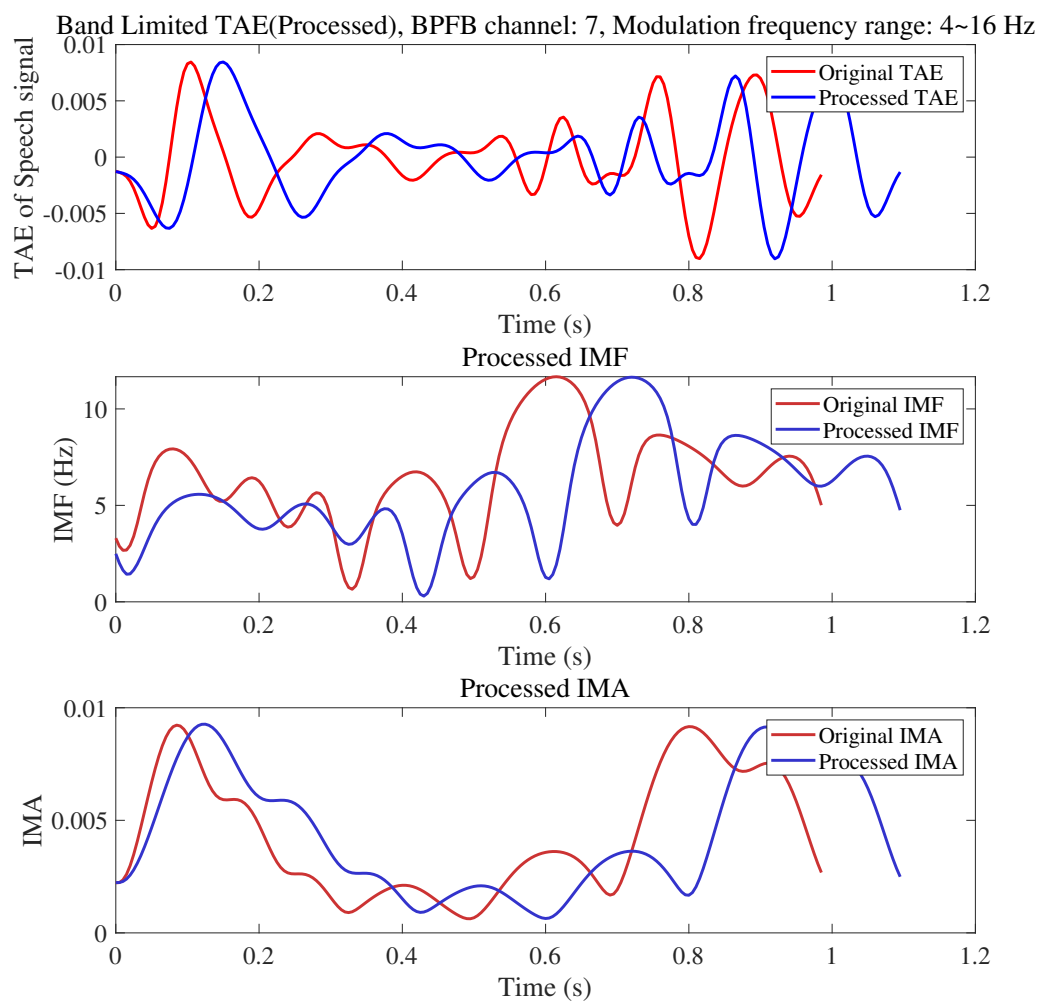


図 3.19: 時間伸長処理した TAE と IMCs (音声 C)

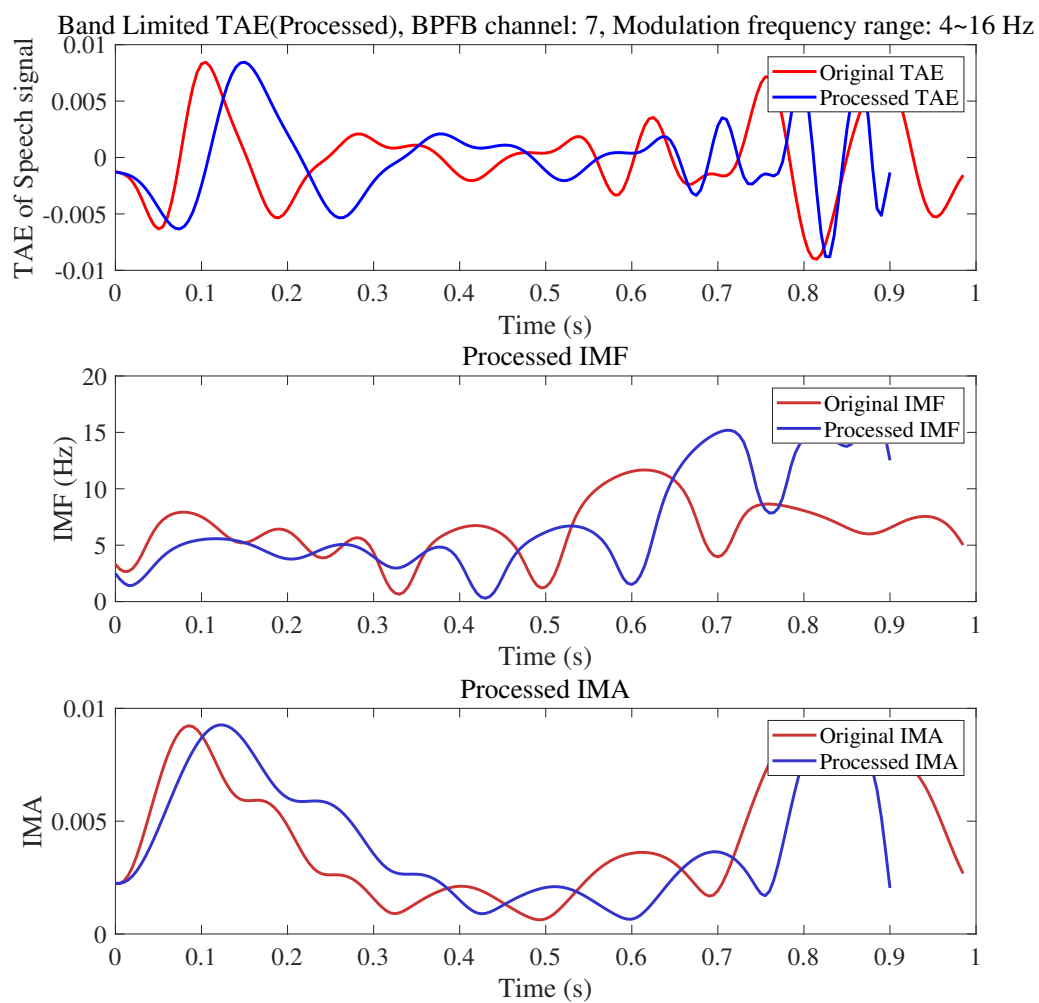


図 3.20: 時間伸長圧縮処理した TAE と IMCs (音声 C)

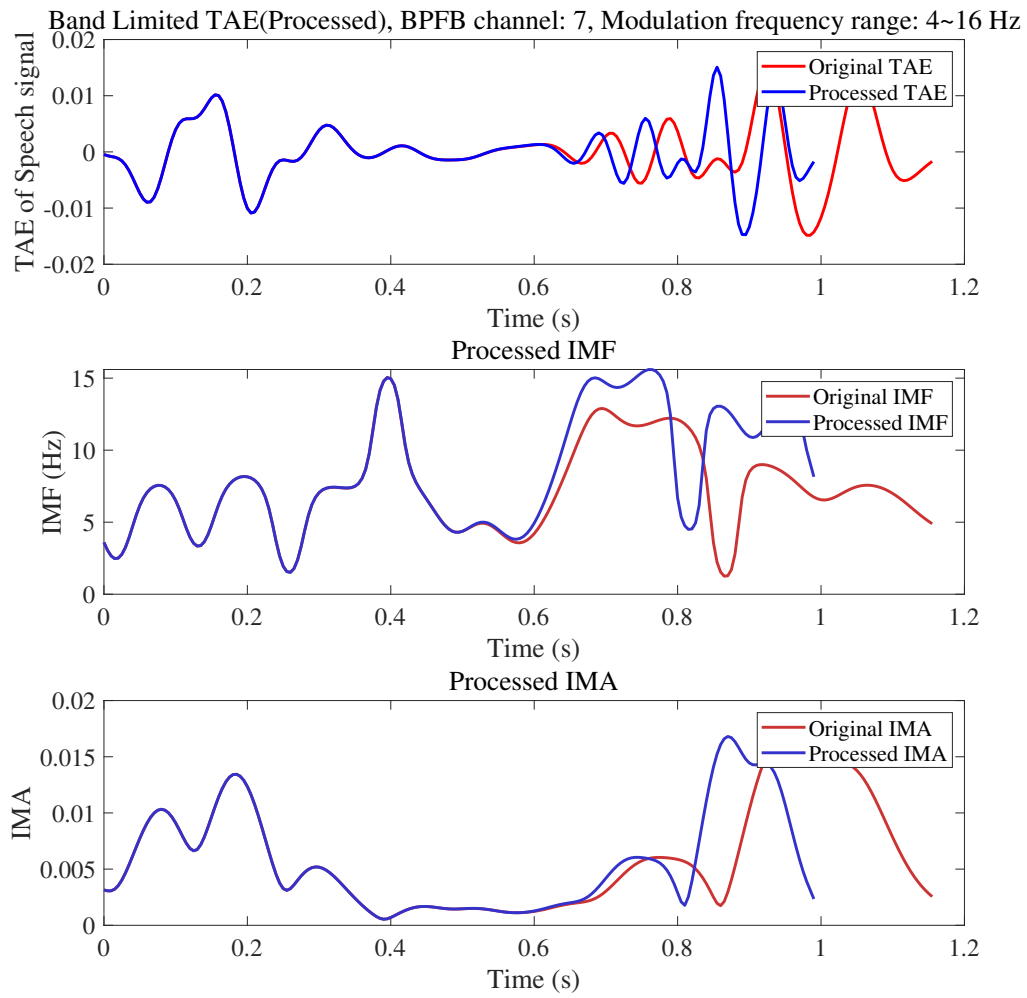


図 3.21: 時間圧縮処理した TAE と IMCs (音声 D)

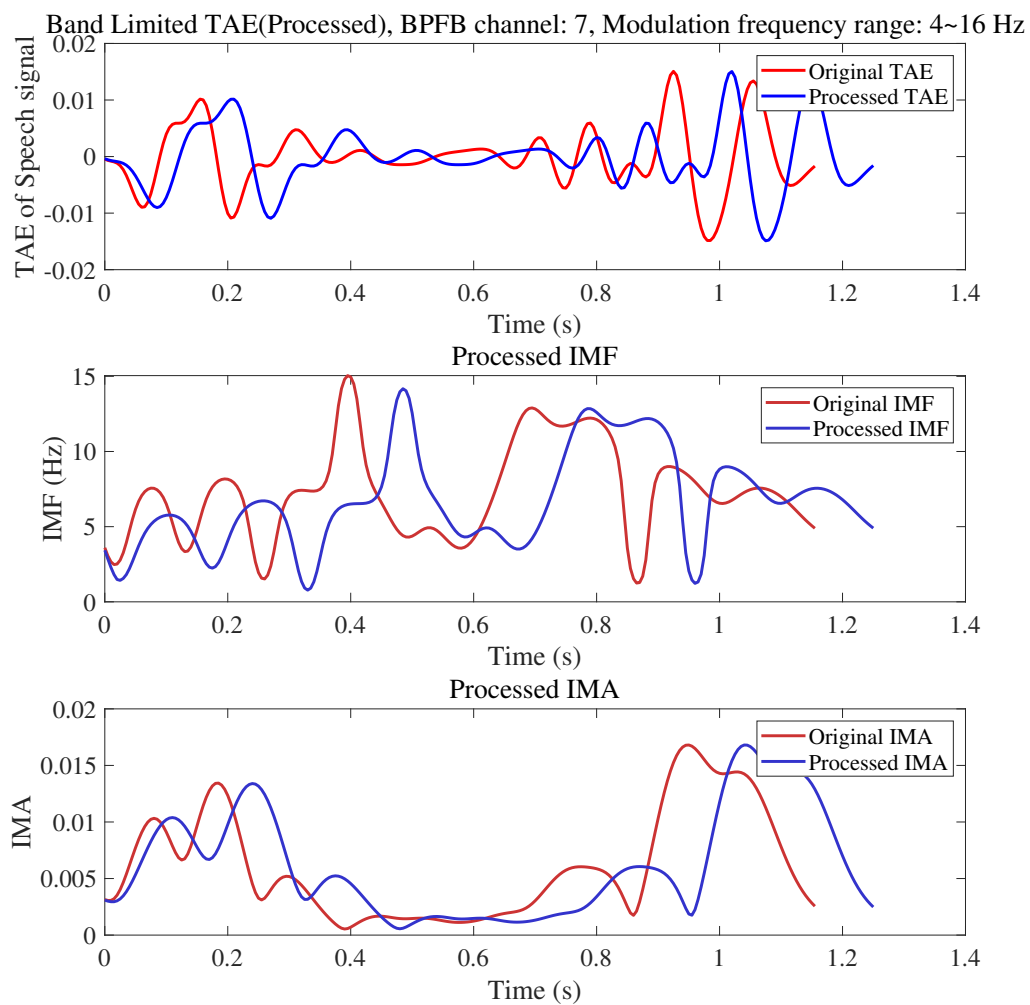


図 3.22: 時間伸長処理した TAE と IMCs (音声 D)

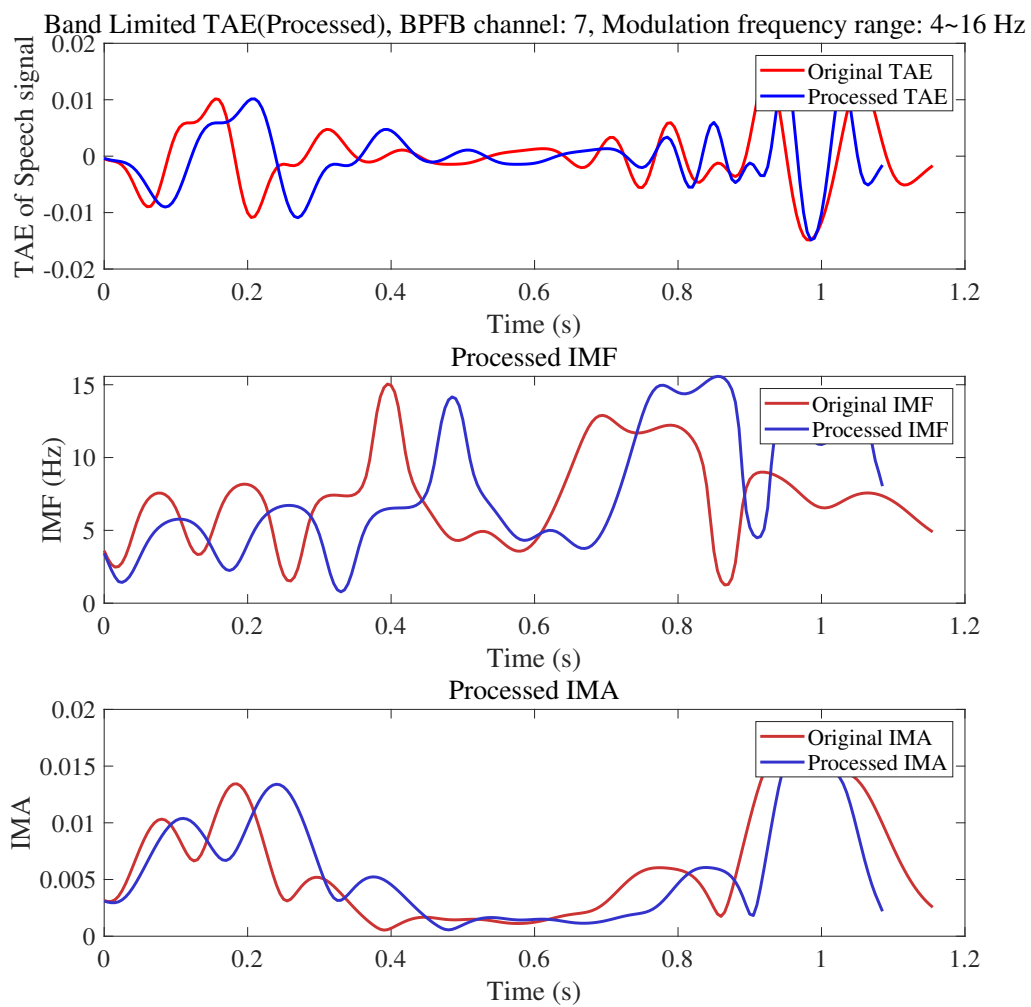


図 3.23: 時間伸長圧縮処理した TAE と IMCs (音声 D)

第4章 時間反転処理による緊迫感の変化

元の NVS 4 種 (a_3, b_3, c_3, d_3) と時間反転した TAE を持つ NVS 4 種 (a_4, b_4, c_4, d_4), 計 8 個の NVS を利用して実験 4 を行った. 実験に用いた音声刺激対は 56 対, 実験の実施時間は約 10 分である.

4.1 仮説

Unoki らの研究 [27] では, 時間反転した TAE を持つ NVS 同士の緊迫感知覚について検討された. 元の NVS では, 明確に緊迫感の程度に差が見られたが, 時間反転した TAE を持つ NVS 同士では緊迫感の程度の差がなくなった. この変化が, 時間反転処理によって, TAE に含まれる緊迫感知覚の手がかりが減少したことによる変化だと仮定すると, 時間反転した TAE を持つ NVS は元の NVS に比べて緊迫感が低減すると考えられる. 逆に, 時間反転処理によって, TAE に含まれる緊迫感知覚の手がかりが増強されたことによる変化だと仮定すると, 時間反転した TAE を持つ NVS は元の NVS に比べて緊迫感が増加すると考えられる. 以上の仮説を踏まえて, 実験 4 を行った.

4.2 実験結果

表 4.1 及び図 4.1 に実験 4 の結果を示す. 次の図の横軸は緊迫感の程度を示し, 値が大きいほど緊迫感が相対的に高く知覚されていることを表す. 結果から, 緊迫感の程度が高い方から $d_3, c_3, a_3, d_4, b_3, b_4, a_4, c_4$ の順に並ぶことがわかった. TAE を要因とする 1 要因 8 水準の分散分析の結果, 音声刺激に対する主効果が認められた ($F(7, 651) = 30.04, p < 0.01$). 時間反転した TAE を持つ NVS は元の NVS と比べて緊迫感の程度が低く, b_3 に寄った.

4.3 考察

実験 4 の結果より, 時間反転した TAE を持つ NVS の緊迫感度は元の NVS より低減することが示された. 時間反転した TAE を持つ NVS 同士の緊迫感の程度の差

がなくなったのは、TAEに含まれる緊迫感知覚の手がかりが減少したことが原因だと考えられる。これより、緊迫感知覚の手がかりとして、時間反転処理によって変化した成分も重要であることが示された。

表 4.1: 実験 4 の結果

ラベル	緊迫感の程度
d_3	0.4598
c_3	0.2232
a_3	0.0357
d_4	-0.0357
b_3	-0.0491
b_4	-0.1964
a_4	-0.2009
c_4	-0.2366

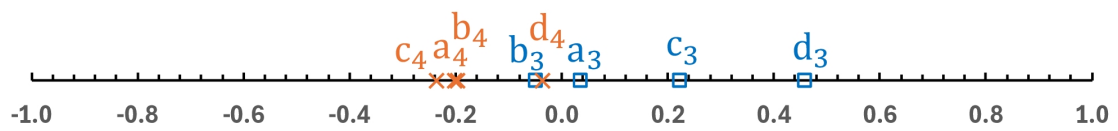


図 4.1: 実験 4 の結果

表 4.2: 実験 4 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	89.86	7	12.84	30.04	0.0000
主効果 × 個人	286.0	91	3.143	7.354	0.0000
組み合わせ効果	7.571	21	0.3605	0.8436	0.6659
順序効果	14.33	1	14.33	33.53	0.0000
順序 × 個人	35.99	13	2.768	6.478	0.0000
誤差	278.2	651	0.4274		
全体	712.0	784			

第5章 時間伸長圧縮処理による緊迫感の変化

実験4で利用した NVS 8 種 ($a_3, b_3, c_3, d_3, a_4, b_4, c_4, d_4$) と時間圧縮処理のみされた TAE をもつ NVS 4 種 ($a_{5C}, b_{5C}, c_{5C}, d_{5C}$), 時間伸長処理のみされた TAE をもつ NVS 4 種 ($a_{5S}, b_{5S}, c_{5S}, d_{5S}$), 時間伸長圧縮処理された TAE をもつ NVS 4 種 ($a_{5SC}, b_{5SC}, c_{5SC}, d_{5SC}$), 計 20 個の NVS を用いて実験5を行った。実験に用いた音声刺激対は 380 対, 実験の実施時間は約 90 分である。

5.1 仮説

TAE の時間伸長圧縮処理を利用して, 時間反転した TAE の IMCs を模擬した TAE を持つ NVS を作成した。実験4の結果が, TAE の IMCs の違いによるものだと仮定すると, 時間反転した TAE の IMCs を模擬した TAE を持つ NVS の緊迫感の程度は, 時間反転した TAE を持つ NVS と同様に低減すると考えられる。そのため, 時間伸長圧縮処理した TAE を持つ NVS の緊迫感は低減すると考えられる。以上の仮説を踏まえて, 実験5を行った。

5.2 実験結果

次のように実験結果を整理して図示する。まず, 図 5.1 に, 全 NVS の結果を示す。次に, 表 5.2 及び図 5.2 に, a_4 から d_4 , a_{5C} から d_{5C} の NVS 8 種の結果について抜粋したものを示す。更に, 表 5.3 及び図 5.3 に, a_4 から d_4 , a_{5S} から d_{5S} の NVS 8 種の結果について抜粋したものを示す。最後に, 表 5.4 及び図 5.4 に, a_4 から d_4 , a_{5SC} から d_{5SC} の NVS 8 種の結果について抜粋したものを示す。これらの図は, 時間圧縮のみ, 時間伸長のみ, 時間伸長圧縮を施したときの結果に対応する。

TAE を要因とする 1 要因 20 水準の分散分析を行った。その結果, 音声刺激に対する主効果が認められた ($F(19, 4869) = 138.4, p < 0.01$)。 a_4 から d_4 は, 実験4と同様に, 元の NVS (a_3 から d_3) より, 緊迫感の程度が低い。また, 下位検定の結果, a_3, b_3 に対し, $a_{5S}, b_{5S}, a_{5SC}, b_{5SC}$ は緊迫感の程度が有意に低い ($p < 0.01$)。

表 5.1: 実験 5 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	1193	19	62.79	138.4	0.0000
主効果 × 個人	1172	247	4.746	10.46	0.0000
組み合わせ効果	80.22	171	0.4691	1.034	0.3676
順序効果	7.823	1	7.823	17.24	0.0000
順序 × 個人	85.37	13	6.567	14.47	0.0000
誤差	2209	4869	0.4538		
全体	4748	5320			

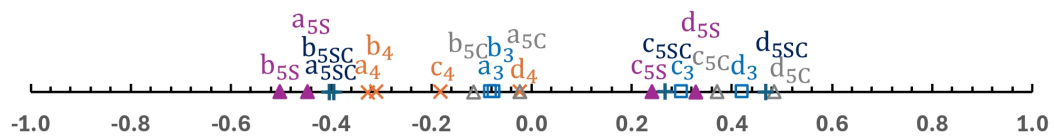


図 5.1: 実験 5 の結果

表 5.2: 実験 5 の結果 (時間反転処理と時間圧縮処理の比較)

ラベル	緊迫感の程度
d _{5C}	0.4857
c _{5C}	0.3714
d ₄	-0.0232
a _{5C}	-0.0232
b _{5C}	-0.1161
c ₄	-0.1821
b ₄	-0.3089
a ₄	-0.3268

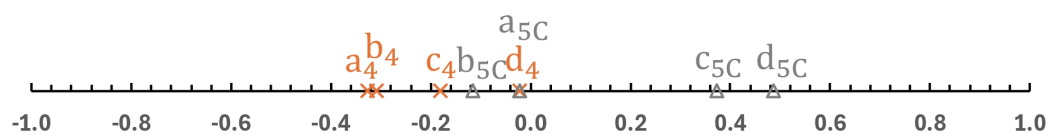


図 5.2: 実験 5 の結果 (時間反転処理と時間圧縮処理の比較)

表 5.3: 実験 5 の結果 (時間反転処理と時間伸長処理の比較)

ラベル	緊迫感の程度
d_{5S}	0.3286
c_{5S}	0.2411
d_4	-0.0232
c_4	-0.1821
b_4	-0.3089
a_4	-0.3268
a_{5S}	-0.4464
b_{5S}	-0.5018

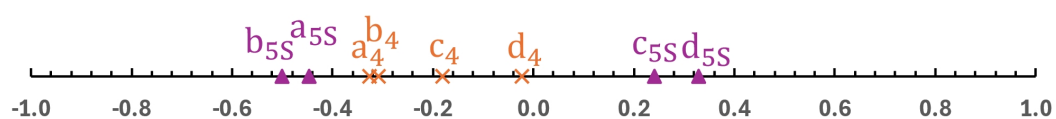


図 5.3: 実験 5 の結果 (時間反転処理と時間伸長処理の比較)

表 5.4: 実験 5 の結果 (時間反転処理と時間伸長圧縮処理の比較)

ラベル	緊迫感の程度
d_{5SC}	0.4679
c_{5SC}	0.2679
d_4	-0.0232
c_4	-0.1821
b_4	-0.3089
a_4	-0.3268
a_{5SC}	-0.3946
b_{5SC}	-0.4036

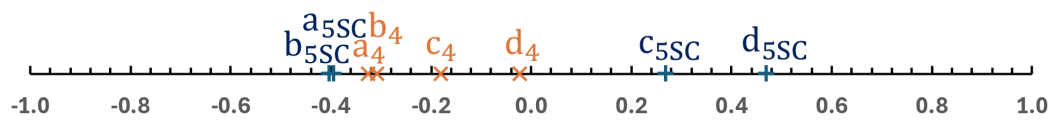


図 5.4: 実験 5 の結果 (時間反転処理と時間伸長圧縮処理の比較)

5.3 考察

実験5の結果より，時間伸長圧縮処理したTAEを持つNVSの一部が，時間反転したTAEを持つNVSと同程度に低い緊迫感の程度を持つことがわかった．時間反転したTAEのIMCsに近づける処理によって，時間反転したTAEを持つNVSと同様の緊迫感の変化が起きたことから，TAEのIMCsが緊迫感知覚の直接的な手がかりであることが示された．ただし，時間伸長圧縮処理によって明確な緊迫感の低減が見られなかったNVSもあった．この点に関する考察は，次章で述べる．

第6章 全体考察

時間反転した TAE を持つ NVS の緊迫感知覚に関する実験を通して、TAE に対する時間反転処理による、NVS の緊迫感の増減の変化について検討した。時間反転した TAE は、緊迫感知覚に重要な変調周波数帯域を含む TAE であるにもかかわらず、元の TAE よりも手がかりが失われていた。これより、緊迫感知覚の直接的な手がかりとして、変調周波数帯域といった長時間平均的な変調周波数成分だけではなく、瞬時的な変調周波数成分も重要な働きをすることが示された。

時間伸長圧縮した TAE を持つ NVS の緊迫感知覚に関する実験を通して、TAE に対する時間伸長圧縮処理によって、実験 4 で見られた時間反転処理による緊迫感の低減を説明できるかどうか検討した。対象の TAE の IMCs に近づけるように操作することで、元の NVS を対象の NVS の緊迫感の程度に近づけることができた。このことから、時間反転処理による NVS の緊迫感の変化は、TAE の IMCs の変化に起因することが示唆された。

明確な緊迫感の低減が見られなかった NVS について、主に c, d とラベリングされた音声の変化が確認できなかった。これらの音声の時間伸長圧縮処理において、IMF の時間変動成分の傾きが大きいことから、時間伸長圧縮係数 $\alpha(n)$ の傾きが、傾き抑制係数 $(\frac{1}{3})^n$ によって抑制されていた。以上より、今回の時間伸長圧縮係数 $\alpha(n)$ の算出方法では、IMCs の時間変化の急峻さに対応できず、時間反転した TAE の IMCs に十分に近づけることができなかったと考えられる。

第7章 結論

7.1 本研究で明らかになったこと

本研究では、TAE の IMCs の制御による NVS の緊迫感知覚の変化について検討した。TAE の変調周波数領域及び時間領域を操作した NVS を作成し、聴取実験を通して各音声の緊迫感の程度を比較した。実験の結果、次のことがわかった。

- (i) TAE は音声の緊迫感知覚の手がかりを含む。
- (ii) 緊迫感知覚に重要な手がかりとなる変調周波数成分は、4 Hz から 16 Hz までの変調周波数帯域である。
- (iii) TAE の時間反転処理によって NVS の緊迫感は低減する。
- (iv) TAE の時間伸長圧縮処理を用いて IMCs を制御することで、一部の結果を除き、時間反転処理による NVS の緊迫感の変化を再現できる。

以上の結果から、音声の緊迫感知覚において TAE の IMCs が重要なはたらきをすると考えられる。

7.2 残された課題

本研究では、TAE の IMCs が音声の緊迫感知覚に影響を与えることが明らかになった。ただし、今回は TAE の IMCs の操作による緊迫感の低減にのみ注目した検討となった。TAE の IMCs と音声の緊迫感知覚の関係について詳細に検討するためには、緊迫感知覚の増加も視野に入れた検討も必要である。また、感情や緊迫感以外の非言語情報知覚の検討にも拡張する事で、TAE 情報と非言語情報知覚の関係について更に知見が深まる。

付 録 A 雑音駆動音声の緊迫感知覚

実験 1-1 として原音声 4 種 (A, B, C, D), 実験 1-2 として NVS 4 種 (a_1, b_1, c_1, d_1) を用いて聴取実験を行った。

実験参加者を 7 名ずつのグループに分け、片方のグループは実験 1-1, 実験 1-2 の順で行い、もう片方のグループはその反対の順で行う。これにより、順序効果による原音声及び NVS の緊迫感の程度の評価への影響を抑える。実験に用いた音声刺激対はそれぞれ 12 対、実験の実施時間は約 10 分である。

A.1 仮説

先行研究より、原音声と NVS は、緊迫感の程度の大きさに基づき音声刺激を並べた際に、同様の並びが見られることがわかった [27]。この結果に基づくと、原音声と NVS の緊迫感の程度の並びは同じになると考えられる。以上を踏まえて、実験 1 を行った。

A.2 実験結果

表 A.1 及び図 A.1 に実験 1-1 (原音声) の結果を、表 A.2 及び図 A.2 に実験 1-2 (NVS) の結果を示す。原音声は緊迫感の程度が刺激ラベル C, D, B, A の順に高いことがわかった。TAE を要因とする 1 要因 4 水準の分散分析の結果、音声刺激に対する主効果が認められた ($F(3, 109) = 184.2, p < 0.01$)。これに対して、NVS は緊迫感の程度が d_1, c_1, b_1, a_1 の順に高いことがわかった。1 要因 4 水準の分散分析の結果、音声刺激に対する主効果が認められた ($F(3, 109) = 36.73, p < 0.01$)。

表 A.1: 実験 1-1 (原音声) の結果

ラベル	緊迫感の程度
C	0.7232
D	0.4286
B	-0.2857
A	-0.8661

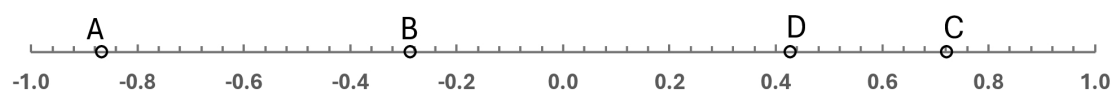


図 A.1: 実験 1-1 (原音声) の結果

表 A.2: 実験 1-2 (雑音駆動音声) の結果

ラベル	緊迫感の程度
d_1	0.3661
c_1	0.2589
b_1	-0.2500
a_1	-0.3750

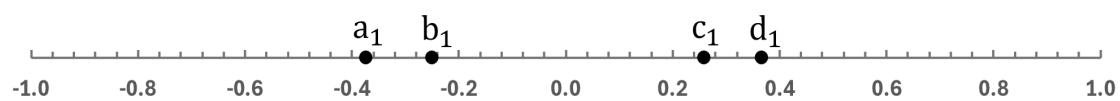


図 A.2: 実験 1-2 (雑音駆動音声) の結果

表 A.3: 実験 1-1 (原音声) 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	172.3	3	57.43	184.2	0.0000
主効果 × 個人	19.70	39	0.5050	1.619	0.0271
組み合わせ効果	1.589	3	0.5298	1.699	0.1716
順序効果	0.7202	1	0.7202	2.309	0.1315
順序 × 個人	2.696	13	0.2074	0.6651	0.7928
誤差	33.99	109	0.3119		
全体	231.0	168			

表 A.4: 実験 1-2 (雑音駆動音声) 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	45.27	3	15.09	36.73	0.0000
主効果 × 個人	51.98	39	1.333	3.244	0.0000
組み合わせ効果	1.304	3	0.4345	1.058	0.3702
順序効果	2.381	1	2.381	5.796	0.0177
順序 × 個人	10.29	13	0.7912	1.926	0.0346
誤差	44.78	109	0.4108		
全体	156.0	168			

A.3 考察

実験 1-1 及び実験 1-2 の結果より，NVS の緊迫感の程度の並びは，原音声と同様の様相を見せることが示された．具体的には，a, b とラベリングされた音声の緊迫感の程度が低く，c, d とラベリングされた音声の緊迫感の程度が高い．以上より，NVS に保存された TAE は緊迫感知覚の手がかりを含むと考えられる．また，この結果は，Unoki らの研究の結果を支持する．

ただし，Unoki らの研究との相違点として，NVS と原音声で c と d の緊迫感の程度が入れ替わる様子が見られたことと，図の数直線上において NVS の緊迫感の程度が原音声に比べて幾分狭まっていることが挙げられる．本実験では，半分の実験参加者が原音声を聞く前に NVS の評価を行った．この参加者らは，NVS 聴取の訓練が十分でなく，かつ原音声から得られる詳細な手がかりを得ていない．原音声から得られるメッセージ形式や基本周波数といった音響特徴は，音声の主観的な緊急性の評価に影響を与える．そのため，この順序効果による影響を抑えた措置が，緊迫感の程度の並びやその幅に影響を与えたと考えられる．

付 録 B 緊迫感知覚に重要な変調周波数成分

B.1 変調フィルタバンクの影響

MFB を利用しないで作成した NVS 4 種 (a_1, b_1, c_1, d_1) と MFB を利用して作成した NVS 4 種 (a_2, b_2, c_2, d_2), 計 8 個の NVS を用いて実験 2 を行った. 実験に用いた音声刺激対は 56 対, 実験の実施時間は約 10 分である.

B.1.1 仮説

先行研究では, MFB の有無によって NVS の緊迫感の程度とその並びに違いは確認できなかった [28]. この結果及び実験 1 の結果に基づくと, MFB を利用して作成した NVS は, MFB を利用しないで作成した NVS と同様に, a, b とラベリングされた NVS の緊迫感が低く, c, d とラベリングされた NVS の緊迫感が高く知覚されると考えられる. また, 同じ原音声から作成された NVS (同じアルファベットの NVS) の緊迫感の程度は同程度になると考えられる. 以上を踏まえて, 実験 2 を行った.

B.1.2 実験結果

表 B.1 及び図 B.1 に実験 2 の結果を示す. 緊迫感の程度が高い方から $d_1, c_1, c_2, d_2, b_2, b_1, a_2, a_1$ の順に並ぶことがわかった. TAE を要因とする 1 要因 8 水準の分散分析の結果, 音声刺激に対する主効果が認められた ($F(7, 651) = 78.66, p < 0.01$). また, 同じ原音声から作成された NVS 同士 (a_1 と a_2 , b_1 と b_2 , c_1 と c_2 , d_1 と d_2) の間に有意差はなかった.

B.1.3 考察

実験 2 の結果より, MFB の有無に関わらず, どちらの NVS の緊迫感の程度とその並びについて概ね同様の様相が見られた. MFB を利用したことによる NVS の緊迫感知覚の変化が見られなかったため, TAE に対する MFB 処理は NVS の緊迫

感知覚に影響を与えないと考えられる。また、この結果は、木谷らの研究の結果を支持する。

表 B.1: 実験 2 の結果

ラベル	緊迫感の程度
d_1	0.4062
c_1	0.3750
c_2	0.2991
d_2	0.2500
b_2	-0.2411
b_1	-0.2455
a_2	-0.3661
a_1	-0.4777

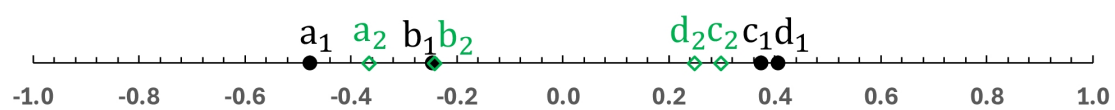


図 B.1: 実験 2 の結果

表 B.2: 実験 2 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	210.2	7	30.02	78.66	0.0000
主効果 × 個人	171.6	91	1.886	4.940	0.0000
組み合わせ効果	11.59	21	0.5519	1.446	0.0898
順序効果	11.51	1	11.51	30.16	0.0000
順序 × 個人	19.68	13	1.514	3.967	0.0000
誤差	248.5	651	0.3817		
全体	673.0	784			

B.2 重要な変調周波数帯域について

0 Hz から 64 Hz の変調周波数帯域を持つ NVS4 種 (a_2, b_2, c_2, d_2) と 4 Hz から 16 Hz の変調周波数帯域を持つ NVS 4 種 (a_3, b_3, c_3, d_3), 計 8 個の NVS を用いて実験 3 を行った. 実験に用いた音声刺激対は 56 対, 実験の実施時間は約 10 分である.

B.2.1 仮説

先行研究では, MFB を用いて変調周波数帯域を制御した TAE を持つ NVS を作成し, その緊迫感を比較することで, 緊迫感知覚に重要な変調周波数帯域について検討した [28]. その結果, 全ての変調周波数帯域 (0 Hz から 64 Hz) を含む NVS と 4 Hz から 16 Hz の変調周波数帯域を持つ NVS が, 同様の緊迫感の程度と並びを持つことが示された. この結果及び実験 2 の結果に基づくと, 4 Hz から 16 Hz の変調周波数帯域を持つ NVS は, 0 Hz から 64 Hz の変調周波数帯域を持つ NVS と同様に, a, b とラベリングされた NVS の緊迫感が低く, c, d とラベリングされた NVS の緊迫感が高く知覚されると考えられる. また, 同じ原音声から作成された NVS 同士の緊迫感の程度は同程度になると考えられる. 以上を踏まえて, 実験 3 を行った.

B.2.2 実験結果

表 B.3 及び図 B.2 に実験 3 の結果を示す. 緊迫感の程度が高い方から $c_2, d_2, c_3, d_3, b_2, a_2, a_3, b_3$ の順に並ぶことがわかった. TAE を要因とする 1 要因 8 水準の分散分析の結果, 音声刺激に対する主効果が認められた ($F(7, 651) = 51.96, p < 0.01$). また, a_2, a_3, b_2, b_3 からなるすべての組み合わせの間に有意差はなかった. 同様に c_2, c_3, d_2, d_3 からなるすべての組み合わせの間に有意差はなかった.

B.2.3 考察

実験 3 の結果より, 変調周波数帯域の違いによらず, どちらの NVS の緊迫感の程度とその並びについて概ね同様の様相が見られることがわかった. 変調周波数帯域が制限された上で, 全ての帯域を含む NVS と同じ緊迫感を持つことから, その変調周波数帯域に緊迫感知覚の重要な手がかりが含まれていると考えられる. 以上より, 緊迫感知覚に重要な変調周波数成分は, 4 Hz から 16 Hz の変調周波数帯域だと考えられる. またこの結果は, 木谷らの研究の結果を支持する.

表 B.3: 実験 3 の結果

ラベル	緊迫感の程度
c_2	0.4018
d_2	0.3482
c_3	0.2098
d_3	0.1830
b_2	-0.1696
a_2	-0.3214
a_3	-0.3214
b_3	-0.3304

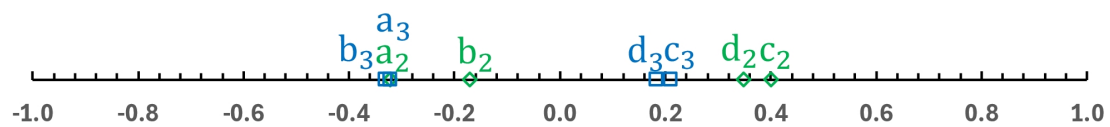


図 B.2: 実験 3 の結果

表 B.4: 実験 3 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	157.9	7	22.55	51.96	0.0000
主効果 × 個人	199.0	91	2.187	5.039	0.0000
組み合わせ効果	9.705	21	0.4622	1.065	0.3822
順序効果	7.367	1	7.367	16.97	0.0000
順序 × 個人	23.49	13	1.807	4.163	0.0000
誤差	282.6	651	0.4340		
全体	680.0	784			

謝辞

本研究を行うにあたって，多大なる御指導及び御助言頂いた鶴木祐史教授に心から御礼申し上げます。研究及びに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郭太陽さん，磯山拓都先生，木谷俊介講師に深く感謝申し上げます。実験に際し，貴重なお時間を割いてくださった参加者の皆様に心よりお礼申し上げます。苦楽の時間を共に過ごしてくださった鶴木研究室の皆様に感謝申し上げます。最後に，本学での研究を支え，暖かい応援をくれた家族に心から最大の感謝を申し上げます。

参考文献

- [1] 森大毅, 前川喜久雄, 粕谷英樹, 音声は何を伝えているか—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, 2014.
- [2] 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.
- [3] Loizou, P.C., “Mimicking the human ear,” *IEEE Signal Processing Magazine*, Vol.15, No.5, pp.101–130, 1998.
- [4] 平原達也, “音を聴く聴覚の仕組み,” 日本音響学会誌, Vol.66, No.9, pp.458–465, 2010.
- [5] 長南浩人, 齋藤佐和, “人工内耳を装用した聴覚障害児の音韻意識の発達,” 特殊教育研究, Vol.44, No.5, pp.283–290, 2007.
- [6] 城間将江, 松永達男, “人工内耳装用者の聴取に関する考察,” *IRYO*, Vol.58, No.9, pp.522–527, 2004.
- [7] 坂本圭, 小渕千絵, 城間将江, 松田帆, 関 恵美子, 荒木隆一郎, 池園哲郎, “人工内耳装用者の聴覚的時間情報処理に関する研究—倍速音声に対する統語修復の効果—,” *Audiology Japan*, Vol.57, No.1, pp.92–98, 2014.
- [8] 福田友美子, 城間将江, 船坂宗太郎, “人工内耳埋め込み患者の音声の知覚—聴覚と視覚の併用について—,” 音声言語医学 30, Vol.30, No.4, pp.334–339, 1989.
- [9] Barone, P., Chambaudie, L., Strelnikov, K., Fraysse, B., Marx, M., Belin, P., Deguine, O., “Crossmodal interactions during non-linguistic auditory processing in cochlear-implanted deaf patients,” *Cortex*, Vol.83, pp.259–270, 2016.
- [10] Paquette, S., Ahmed, G.D., Goffi-Gomez, M.V., Hoshino, A.C.H., Peretz, I., Lehmann, A., “Musical and vocal emotion perception for cochlear implants users,” *Hearing Research*, Vol.370, pp.272–282, 2018.
- [11] Jiam, N.T., Caldwell, M., Deroche, M.L., Chatterjee, M., Limb, C.J., “Voice emotion perception and production in cochlear implant users,” *Hearing Research*, Vol.352, pp.30–39, 2017.

- [12] Wiefferink, C. H., Rieffe, C., Ketelaar, L., Raeve, L. D., Frijns, J. H. M., “Emotion Understanding in Deaf Children with a Cochlear Implant,” *Journal of Deaf Studies and Deaf Education*, Vol.18, No.2, pp.175–186, 2013.
- [13] Xin, L., Qian-Jie, F., Galvin III, J. J., “Vocal Emotion Recognition by Normal-Hearing Listeners and Cochlear Implant Users,” *Trends in Amplification*, Vol.11, No.4, pp.301–315, 2007.
- [14] Massida, Z., Marx, M., Belin, P., James, C., Fraysse, B., Barone, P., Deguine, O., “Gender Categorization in Cochlear Implant Users,” *Journal of Speech, Language, and Hearing Research*, Vol.56, No.5, pp.1389–1401, 2013.
- [15] 佐藤友貴, 小渕千絵, 籠宮隆之, 大金さや香, 城間将江, 野口佳裕, 加我君孝, “人工内耳装用児の話者の男女識別に関する検討,” *Audiology Japan*, Vol.63, No.3, pp.181–188, 2020.
- [16] 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.
- [17] 小林まおり, 濱田康弘, 赤木正人, “音声の緊迫感に影響する音響特徴の検討,” *電子情報通信学会技術研究報告*, Vol.118, No.426, SP2018-53, pp.9–14, 2019.
- [18] Ofuji, K., Ogasawara, Naomi., “Verbal disaster warnings and perceived intelligibility, reliability, and urgency: The effects of voice gender, fundamental frequency, and speaking rate,” *The Acoustical Society of Japan*, Vol.39, No.2, pp.56–65, 2018.
- [19] Phil-Sik, J., “Designing acoustic and non-acoustic parameters of synthesized speech warnings to control perceived urgency,” *International Journal of Industrial Ergonomics*, Vol.37, pp.213–223, 2007.
- [20] Li, X., Thompson, C. S., Pfingst, B. E., “Relative contributions of spectral and temporal cues for phoneme recognition,” *The Journal of the Acoustical Society of America*, Vol.117, No.5, pp.3255–3267, 2005.
- [21] Massidaa, Z., Belinc, P., Jamesd, C., Rougera, J., Frayssee, B., Baronea, P., Deguine, O., “Voice discrimination in cochlear-implanted deaf subject,” *Hearing Research*, Vol.275, pp.120–129, 2011.
- [22] Loebach, T. L., Wickesberg, R. E., “The psychoacoustics of noise vocoded speech: A physiological means to a perceptual end,” *Hearing Research*, Vol.241, pp.87–96, 2008.

- [23] Roman, A. S., Pisoni, D. B., Kronenberger, W. G., Faulkner, K. F., “Some Neurocognitive Correlates of Noise-Vocoded Speech Perception in Children with Normal Hearing: A Replication and Extension of Eisenberg et al., 2002,” *Ear and Hearing*, Vol.38, No.3, pp.344–356, 2017.
- [24] Shannon, R.V., Zeng, F., Kamath, V., Wygonski, J., Ekelid, M., “Speech Recognition with Primarily Temporal Cues,” *Science*, Vol.270, No.5234, pp.303–304, 1995.
- [25] Drullman, R., “Temporal envelope and fine structure cues for speech intelligibility,” *The Journal of the Acoustical Society of America*, Vol.97, No.1, pp.585–592, 1995.
- [26] Zhu, Z., Miyauchi, R., Araki, Y., Unoki, M., “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.
- [27] Unoki, M., Kawamura, M., Kobayasi, M., Kidani, S., Li, J., Akagi, M., “Contributions of Temporal Modulation Cues in Temporal Amplitude Envelope of Speech to Urgency Perception,” *Applied Science*, Vol.13, pp.6239, 2023.
- [28] 木谷俊介, 劉小婷, 郭太陽, 磯山拓都, 李軍鋒, 赤木正人, 鵜木祐史, “音声に含まれる緊迫感の変調周波数帯域の検討,” 日本音響学会音声研究会資料, Vol.3, No.1, pp.35–38, 2023.
- [29] Guo, T., Isoyama, T., Kidani, S., Unoki, M., “Contributions of Instantaneous Modulation Components in Temporal Amplitude Envelope to Vocal Emotion Perception,” *Proceedings of Spring Meet of Acoustical Science and Technology*, 1-3-4, pp.10–11, 2024.
- [30] 古川茂人, “聴覚の時間情報処理,” *Audiology Japan*, Vol.59, No.6, pp.615–622, 2016.
- [31] 小畑宣久, 力丸祐, “帯域雑音により合成された日本語音声の了解度 –聴覚中枢神経の機能を利用したスピーチプロセッサを目指して–,” 日本音響学会聴覚研究会資料 H-2000-3, 2000.
- [32] Tachibana, R., Sasaki, Y., Riquimaroux, H., “Relative contributions of spectral and temporal resolutions to the perception of syllables, words, and sentences in noise-vocoded speech,” *The Acoustical Society of Japan*, Vol.34, No.4, pp.263–270, 2013.

- [33] Drullman, R., Festen, J. M., Plomp, R., “Effect of temporal envelope smearing on speech reception,” *The Journal of the Acoustical Society of America*, Vol.95, No.2, pp.1053–1064, 1994.
- [34] Zhu, Z., Nishino, Y., Miyauchi, R., Unoki, M., “Study on linguistic information and speaker individuality contained in temporal envelope of speech,” *Acoustical Science and Technology*, Vol.37, No.5, pp.258–261, 2016.
- [35] Oganian, Y., Chang, E. F., “A speech envelope landmark for syllable encoding in human superior temporal gyrus,” *Science Advances*, Vol.5, No.11, pp.1–13, 2019.
- [36] Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D., Schalk, G., “The Tracking of Speech Envelope in the Human Cortex,” *PLoS ONE*, Vol.8, No.1, e53398, 2013.
- [37] Qian, J. F., Chinchilla, S., Galvin, J. J., “The Role of Spectral and Temporal Cues in Voice Gender Discrimination by Normal-Hearing Listeners and Cochlear Implant Users,” *Journal of the Association for Research in Otolaryngology*, Vol.5, pp.253–260, 2004.
- [38] Yosida, K., Kazama, M., Toyama, M., “Pitch and speech-rate conversion using envelope modulation modeling,” *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol.4, pp.1-425–1-428, 2002.
- [39] 風間道子, 東山三樹夫, 山崎芳男, “狭帯域音声波形包絡線の帯域間相関行列に現れる話者情報,” *電子情報通信学会論文誌 A*, Vol.J92-A, No.4, pp. 205–215, 2009.
- [40] 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.
- [41] Moore, B. C. J., Glasberg, B. R., “Suggested formulae for calculating auditory-filter bandwidths and excitation patterns,” *The Journal of the Acoustical Society of America*, Vol.74, No.750, pp. 50–753, 1983.

研究業績

(口頭，査読無)

1. 房野早希，木谷俊介，鵜木祐史，時間振幅包絡線の時間変化の違いによる緊迫感知覚の変化，2024年度電気・情報関係学会北陸支部連合大会，H1-3，石川，2024年9月
2. 房野早希，郭太陽，磯山拓都，木谷俊介，鵜木祐史，音声の時間振幅包絡線の時間伸長圧縮操作による緊迫感知覚の変化，日本音響学会聴覚研究会資料，Vol.54, No.6, pp.569-574, 石川，2024年11月
3. 房野早希，郭太陽，磯山拓都，木谷俊介，鵜木祐史，時間振幅包絡線の瞬時変調成分を制御した雑音駆動音声の緊迫感知覚の変化，日本音響学会春季研究発表会，3-4-6, 埼玉，2025年3月（発表予定）