

Title	子音強調による骨導デバイスの音声了解度の改善
Author(s)	朱, 雯雨
Citation	
Issue Date	2022-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17634
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

子音強調による骨導デバイスの音声了解度の改善

ZHU WENYU

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
(修士)

令和4年3月

Abstract

In recent years, computers have become an integral part of human life. They have evolved from the first large machines to the current cell phones and even smart watches. The size of computers is getting smaller and smaller, making them more convenient to be carried around. There are small computers such as smart watches and other wearable computers, sensors, and information presentation devices, and various efforts are being made. From these efforts, it can be seen that wearable computers (wearable devices) are becoming smaller than conventional computers. Therefore, wearable devices are considered to be an inevitable step in the development of computers.

It has been pointed out that there are two types of wearable devices: information wearable devices and environmental wearable devices. Among them, information wearable devices correspond to the five senses such as ears, eyes, etc. Bone conducted device is an information wearable device for the ears. Bone-conductive device is a device that can transmit sound by vibration of the skull or skin without blocking the ear. Since the bone-conductive device transmits sound by bone vibration, it has an advantage. That is users use bone-conducted device can hear the sound presented by bone vibration while air-conducted sound is heard. It is thought bone-conducted device can be useful for safe and secure communication. In addition to providing convenience and safety to people's lives, bone guidance devices can be effective in supporting the physically challenged and expanding the range of activities in their lives. Bone conduction devices can be used to assist people with conductive hearing loss who have problems in the ear canal or middle ear and have difficulty in hearing sounds by air-conducted headphone, or people with hearing impairment who have hematoma in ears, a condition in which the ear becomes blocked.

However, there is one problem when using bone conduction device. When speech is presented with a bone-conduction device, the sound quality and speech intelligibility are reduced compared to air-conduction speech. In particular, it was pointed out that the intelligibility of bone-conducted speech was reduced in noisy environment. There are two possible causes that may affect the intelligibility of bone-conducted speech. The first is that when users use a bone-conducted device, users can hear bone-conducted without covering their ears, then they can hear background sounds in the surrounding environment at the same time. In an environment with high levels of noise, the heard noise may affect the intelligibility of the bone-conducted speech. Second, the high-frequency component of bone-conducted speech is attenuated by the characteristic of bone-conducted transmission. The characteristic of bone-conducted transmission is that the high-frequency component of bone-conducted speech is attenuated compared to air-conducted speech.

If the speech intelligibility is reduced, there is a possibility that a discrepancy will occur when using bone-conducted device. This may cause a serious problem when communicating important instructions in the wearable information sharing space.

It is thought that the high-frequency attenuated of bone conduction affects the intelligibility of bone-conducted speech. Fujita focused on the high-frequency attenuated characteristic in the bone-conducted transmission and proposed a method to improve the intelligibility of bone-conducted speech. The method using high-frequency enhancement to compensate for the bone conduction transmission characteristic. This method improved the intelligibility of bone-conducted speech in noisy environment. On the other hand, consonants are concentrated at higher frequencies than vowels. So the high-frequency attenuated characteristic of bone-conduction may also affect the intelligibility of consonants. If we add processing to the consonants, we can expect further improvement in intelligibility.

The purpose of this paper is to improve the intelligibility of bone-conducted speech in noisy environment.

There are two important domains for speech: frequency domain and time domain. Since the method proposed by Fujita focused only on the frequency domain, it is expected to improve the intelligibility of bone-guided speech by focusing on the time domain as well. Focusing on the overall temporal structure of speech signals, Drullman pointed out that the temporal spectrum is important for speech intelligibility in the time domain. It was found that the intelligibility of bone-guided speech can be restored based on the modulation transfer function focusing on the time spectrum. The temporal structure of the speech signal leads to the phonological levels of consonants and vowels. The acoustic features of consonants are more complex than those of vowels in speech. Many consonants are concentrated in the high frequency range. In addition, the low power of consonants compared to vowels suggests that consonants are easily masked in noisy environments. Furthermore, the maintenance time of consonants is shorter than that of vowels. An important component to the perception of consonants is the formant transition from consonant to vowel. Due to the acoustic properties of consonants, it is thought that consonants are harder to identify than vowels in airborne speech. Hoshino proposed a method of consonant enhancement by emphasizing the consonant and formant transitions. Yasutake and Nakajima showed that consonant stress improved the intelligibility and comprehensibility of speech presented by loudspeakers in noisy environment.

Bone-conduction transmits sound by vibration of the skin and skull. The vibration of the skin and skull also transmits to the outer and middle ear. And the outer ear vibrates, the process of sound transmission from the outer ear to the middle ear and the characteristic of air-conducted sound are similar. Therefore,

it is thought that consonant enhancement is effective in air-conducted sounds and will also be effective in the bone conduction pathway. Therefore, the aim of this study is to improve the intelligibility of speech presented with a bone guide in a noisy environment by effectively emphasizing consonants on the time and frequency axes.

In this study, two methods were proposed that focus on the time domain. Consonant emphasize and modulation transfer function improvement methods were proposed. Among them, consonant emphasize was found to be able to improve speech intelligibility in noisy environments.

The proposed method in this study is to emphasize the consonants concentrated in the high frequency region which are easily affected by the bone conduction high-frequency attenuated characteristic. Other than that, the maintenance time of consonants was short, and it was difficult to obtain sufficient improvement effect only by emphasizing consonants. The formant transitions, which are important for the perception of consonants, should be emphasized together.

For improving the bone-conducted speech intelligibility well, the combination of the high-frequency emphasis proposed by Fujita and the consonant emphasis(CE) is proposed. Therefore, the first-order high-frequency emphasis (FOE) compensates for the transfer characteristic of regio temporalis(RT) vibration, which improved intelligibility well has been combined with consonant enhancement(RT-FOE+CE).

In order to confirm the improvement effect of the proposed method on bone-conducted speech intelligibility, speech intelligibility tests were conducted in noisy environment(55 dB, 75 dB). According to the results of tests, in noisy environment CE has significant difference in word correctness with No emphasis. And there was no significant difference between CE and RT-FOE. This indicates that CE has the same level of comprehension improvement as RT-FOE in Fujita. In addition, RT-FOE+CE significantly increased the correct response rate compared to CE and RT-FOE alone. This indicates that the RT-FOE+CE is most effective in improving the intelligibility of bone-conducted speech in noisy environment.

目次

第1章	序論	1
1.1	研究背景	1
1.2	研究目的	3
1.3	論文構成	4
第2章	関連研究	6
2.1	音響特性と音声知覚	6
2.2	骨導音声の音響特性	7
2.3	音声了解度改善に関する研究	10
2.4	骨導音声の了解度改善に関する研究	12
2.5	問題点	13
第3章	提案法	14
3.1	提案法のコンセプト	14
3.2	子音強調による骨導提示音声の了解度改善法	16
3.2.1	子音強調	16
3.2.2	子音強調と高域強調の組み合わせ	19
3.3	変調伝達関数に基づく骨導提示音声の了解度改善法	21
3.3.1	変調伝達関数の分析	21
3.3.2	フィルタ設計	26
3.3.3	変調伝達関数の補償強調	32
3.3.4	変調伝達関数改善法と高域強調の組み合わせ	33
第4章	評価	35
4.1	子音強調の評価	35
4.1.1	音声データ	35
4.1.2	実験条件	35
4.1.3	実験参加者	35
4.1.4	実験装置	36
4.1.5	実験手続き	38
4.1.6	実験結果	38
4.2	変調伝達関数改善法の評価	43
4.2.1	音声データ	43

4.2.2	実験条件	43
4.2.3	実験参加者	43
4.2.4	実験装置	43
4.2.5	実験手続き	44
4.2.6	実験結果	44
第 5 章	総合評価	49
5.1	三つの改善法の組合せ	49
5.2	音声データ	51
5.3	実験条件	51
5.4	実験参加者	51
5.5	実験装置	51
5.6	実験手続き	52
5.7	実験結果	52
第 6 章	全体考察	57
6.1	子音強調法の改善効果	57
6.2	変調伝達関数改善法の改善効果	58
6.3	三つの方法の組合せによる改善効果	61
6.4	音声了解度に対する最適な改善方法	61
第 7 章	結論	62
7.1	明らかにしたこと	62
7.2	残された課題	62

目 次

1.1	本論文の構成	5
2.1	気導・骨導音声の聴取経路（文献 [6] より引用）	8
2.2	（上）気導音声（下）骨導音声のスペクトログラム	9
3.1	音韻誤答率 (r)。子音（母音）の総数に占める間違えた子音（母音） の数の比 ($0 \leq r \leq 1$) を表す	15
3.2	提案法における子音強調処理の概略	18
3.3	子音強調と一次高域強調の組み合わせの概要	20
3.4	MTF の分析概要	23
3.5	骨導音声と気導音声の変調伝達関数	24
3.6	骨導音声と気導音声の変調伝達関数の逆特性	25
3.7	設計した理想的なフィルタ	27
3.8	設計したフィルタと理想的なフィルタの RMSE	30
3.9	MTF の逆特性に基づく理想的なフィルタと設計したフィルタ	31
3.10	変調伝達関数改善法と一次高域強調の組み合わせの概要	34
4.1	単語了解度試験の実験装置と実験環境	37
4.2	子音強調：雑音レベル 55 dB 環境下親密度別単語正答率	39
4.3	子音強調：雑音レベル 75 dB 環境下親密度別単語正答率	40
4.4	変調伝達関数：雑音レベル 55 dB 環境下親密度別単語正答率	45
4.5	変調伝達関数：雑音レベル 75 dB 環境下親密度別単語正答率	46
5.1	三つの改善法の組み合わせの概要	50
5.2	総合評価：雑音レベル 55 dB 環境下親密度別単語正答率	53
5.3	総合評価：雑音レベル 75 dB 環境下親密度別単語正答率	54
6.1	IMTF フィルタ処理前後の相互相関結果	59
6.2	IMTF フィルタ処理前後のパワー包絡線と子音強調後のパワー包絡線	60

表 目 次

4.1	子音強調：単語正答率	41
4.2	子音強調：有意差検定結果	42
4.3	変調伝達関数：単語正答率	47
4.4	変調伝達関数：有意差検定結果	48
5.1	総合評価：単語正答率	55
5.2	総合評価：有意差検定結果	56

第1章 序論

1.1 研究背景

現在、コンピュータは人間の生活に溶け込んでいる。最初の大型機械から現在の携帯電話、さらにはスマートウォッチまで発展して来た。コンピュータのサイズはますます小型化になり、持ち運びが便利になっている。スマートウォッチなどウェアラブルコンピュータといった小型のコンピュータやセンサや情報提示装置などがあり、さまざまな取り組みがなされている [1]。これらの取り組みからみると、従来のコンピュータよりウェアラブルコンピュータ（ウェアラブル装置）が小型化を求められている傾向と見受けられる。従って、ウェアラブル装置はコンピュータの開発の一環として避けられない段階と考えられる。

ウェアラブル装置は、一般的に体に装着できるウェアラブル物品の形態で存在している情報機械を指す [2]。また、あまり目立たないものの、健康、利便性、好奇心はウェアラブル技術の研究・開発を促進していると指摘されている [3]。ウェアラブル装置の使用は、ユーザがコメントを差し上げるときに、両手の操作を必要とせず、ジェスチャー、音声、または目の動きによって行われている [4]。モバイル・コンピュータなどのデバイスと比較すると、ユーザが意識的に装置を利用しようとしている面で考えれば、存在感が薄いと思われる。また、ウェアラブル装置は既存の装置よりも、ユーザの生活活動に影響を与えずにユーザを支援することができる。ウェアラブル装置を用いて構築できるウェアラブル情報共有空間は人々の生活に便利さ、安全さ、楽しさを提供すると考えられる [1]。ウェアラブル情報共有空間の中にある分散型ウェアラブル情報共有空間は、生活を見守ることができる。ユーザが随時、家電などを設定できることと家の情報を観察できることで、生活が便利になると想定される。遠隔にいる専門医師とコミュニケーションしながら医師の指示を受ける場合や、遠隔にいる消防、警察などが現場情報の取得と指示などの伝達を利用する場合などには、ウェアラブル装置で情報を交換することが必要だと考えられる。

ウェアラブル装置は情報ウェアラブル装置と環境ウェアラブル装置という二種類があると指摘されている [5]。そのうち、情報ウェアラブル装置は身につけられる、耳などの五感に照応している装置であり、耳に対する情報ウェアラブル装置は骨導スピーカ（骨導デバイス）があると知られている。骨導デバイスは耳を塞がずに、頭蓋骨、皮膚などの振動で音声を伝達できるデバイスである [6]。骨導デバイスは骨導振動で音を伝達するため、気導音声を聞きながら骨導振動で提示し

た音声も聞こえるというメリットがある。このような骨導デバイスを活用することで、上記で記した分散型ウェアラブル情報共有空間において、現場情報の取得や指示などを骨導デバイスで聴取することができ、それと同時に気導音で現場の背景音も聴取することができる。安心安全なコミュニケーションに役立つと考えられる。今の生活の中においても、骨導デバイスを装着しながら安心してジョギングなどができる、人々の生活に便利さ、安全さ、楽しさを提供することができるだけでなく、骨導デバイスは障害者支援、すなわち障害者の生活や活動範囲を普段よりも広げるのに有効だと考えられる。骨導デバイスは外耳道や中耳に問題があり、普通の気導で音を聴取しにくい伝音性難聴者 [7] や、耳介血腫などの原因で耳が塞がってしまう聴覚障害者を補助できる。昔から、骨伝導で音を聞いていたことがあり、骨伝導の特性を利用されていた [8]。

しかし、骨導デバイスを使う際に一つ問題がある。骨導デバイスによって提示される音声（骨導提示音声）は、気導音声と比べて、音質や音声了解度が低下してしまうことが知られている。特に高雑音環境下において、骨導提示音声の了解度が低下してしまうことが指摘された [9]。骨導提示音声の了解度に影響を与える可能性のある原因は2つあると考えられる。一つ目は、骨導デバイスを利用するときには耳を塞がずに、気導音も同時に聴取できるため、周囲環境の中にある背景音も気導音として聴取してしまうことである。特に、高いレベルの雑音がある環境下では、骨導提示音声は気導音で聴取した雑音にマスクされる可能性がある。このことは骨導提示音声の了解度に影響を与える可能性がある。二つ目は、骨導伝達の高域成分減衰特性だと考えられる。骨導伝達特性の影響を受けて、骨導音声では気導音声に比べ高域成分が減衰してしまうことがあると指摘されており [10]、骨導提示音声の了解度に影響を与えると考えられる。音声了解度が低下してしまうと、骨導デバイスを用いて音声を伝達する際に齟齬が起こる可能性がある。構築したウェアラブル情報共有空間内において、医療、消防、警察などの現場で重要な指示を伝える際に齟齬が起こってしまう場合には、重大な問題を起こしてしまう可能性がある。

雑音環境下において安心・安全なコミュニケーションを実現するためには、骨導提示音声の了解度を良好なレベルまで改善する必要がある。上記に骨導提示音声の了解度に影響を与える原因について、藤田は骨導提示音声の了解度が、雑音レベルや単語親密度に依存して大幅に低下してしまうことを明らかにした。さらに、骨導伝達における高域減衰に着目し、骨導伝達特性を補償する高域強調処理により骨導提示音声の了解度を改善する方法を提案した [11]。この方法によって雑音環境下で、骨導提示音声の了解度を改善できると認められた。しかし、高騒音環境下では、骨導提示音声の了解度は十分な改善効果が得られないという課題がある。

雑音環境下における安心安全なコミュニケーションを実現するためには、骨導提示音声の音声了解度をさらに改善する必要がある。そのため、骨導伝達特性を補償すること以外に、音声に対する重要な成分を補償することで、骨導提示音声の了解度を改善できる方法を提案することが重要であると考えられる。

1.2 研究目的

本研究の目的は、雑音環境下で骨導提示音声の了解度が低下することを防ぎ、低下した骨導提示音声の了解度を改善することである。

音声に対する重要な領域は周波数領域と時間領域2つがある。藤田による手法は、周波数領域に着目することで骨導提示音声の了解度を改善したが、時間領域には着目しなかった。本研究では、時間領域に着目することで、雑音環境下で骨導提示音声の了解度を改善できる方法を提案する。また、藤田による提案法がに時間領域の処理を追加すれば、了解度をより改善できることが期待される。本研究では、時間領域の情報として、音声の振幅包絡線情報に着目し、時間・周波数軸上で効果的に強調することで、雑音環境下で骨導提示音声の了解度の改善を目指す。また、藤田による提案した方法で強調した音声の了解度よりもさらに改善することを目指す。

1.3 論文構成

本論文は、7章で構成される。本論文の構成を示す図は図 1.1 である。

第1章

ウェアラブル装置と人間生活の関係と骨導デバイスの人間生活への利点について述べ、骨導デバイスで提示した音声の問題点と本研究の目的を明らかにする。

第2章

音声音響特性と人間の聴覚特性において重要な成分を説明する。聴覚特性に応じる音声了解度を改善した先行研究について述べる。骨導デバイスで提示した音声了解度の改善法と残っている問題点を述べる。

第3章

本研究の提案法について述べる。

第4章

提案法の効果を音声了解度試験で評価し、求めた結果を分析する。

第5章

提案法について総合評価を行い、最適な改善効果を音声了解度で評価する。

第6章

求めた結果を用いて、提案法について考察する。

第7章

結論として、本研究で明らかとなったことや残された課題について述べる。

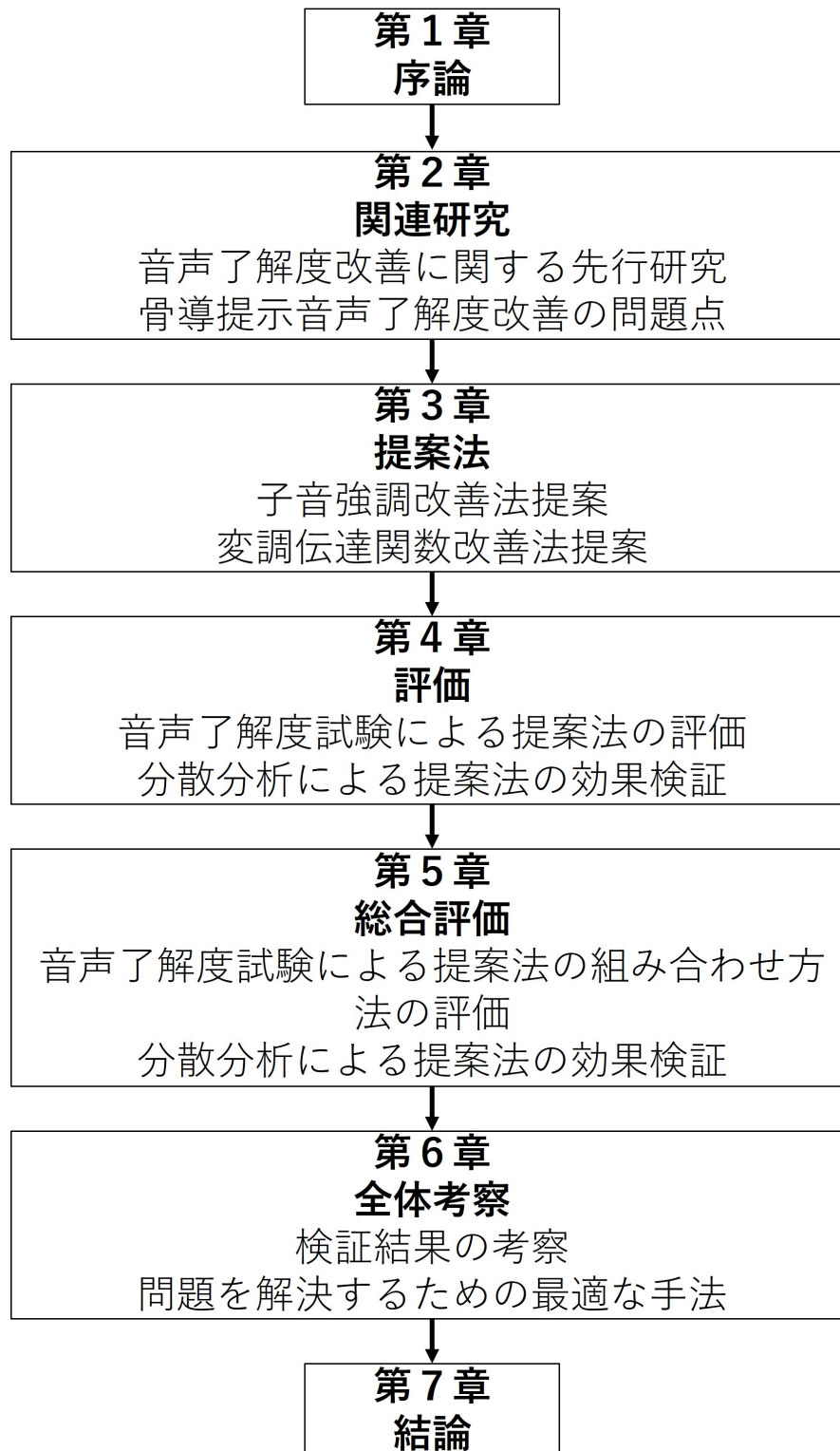


図 1.1: 本論文の構成

第2章 関連研究

2.1 音響特性と音声知覚

音は空気などの振動を経由して波という形で伝達する。波に関して、時間方向の波形と周波数方向の特性がある。音の特徴は大きさ、高さと音色である。大きさ、高さは周波数の変化により変わる。音色は周波数特性の変化と時間変化の影響を受ける。音声は音の一種類として、音の特徴と同じである。音声信号は人間のコミュニケーションに必要な情報を伝達する。波形と周波数特性を観察することで、人間は音声の聴取の理解に重要であると指摘された [12]。そのため、音声信号の重要な三つの特徴が時間的な波形と周波数特性を観察することから得られる。

人間の聴覚特徴からいうと、20 Hz から 20 kHz までの周波数範囲は可聴範囲である。可聴範囲の中でも、特に 3 kHz から 4 kHz の間の周波数成分が聞こえやすい。このことから、人間の音声知覚に対する、可聴範囲にかかわる周波数成分が重要であると考えられる。人間のコミュニケーションにとって、聞こえることだけではない、音を正しく聞くことが重要である。また、音を詳しく分ければ、子音と母音という音韻を分けている。音を聞くときに、単独な音韻を聞くのではなく [13]、子音と母音を一緒に構成した音節を正しく聞いて、音を知覚する。音声の母音と比べて子音の音響特徴は複雑である。母音は低周波領域に集中する。子音の音響特徴で分類すれば、破裂音、摩擦音、鼻音、流音とわたり音に分けられ、破裂音と摩擦音は有声音と無声音で分けられる。有声子音の周波数特性は高周波数領域に集中しているが、低周波数領域にもボイスバーがある。無声子音は高周波数領域に集中している [14]。鼻音と流音はほかの子音と比べ、低周波領域に集中している。また、母音と比べて子音はパワーが小さいため、雑音環境下で、子音が母音よりマスクされやすいと考えられる。さらには、子音の維持時間が母音より短い、例えば破裂音の時間は破裂した瞬間で維持するため、時間が短い。そのため、子音を単独で聞き取りにくい。通常、子音の後にある母音に変化する一部を聞くことで子音を聞き取る。その部分は、子音の知覚に対する重要な部分だと考えられる。フォルマント遷移は、子音から母音への遷移部分として定義される。破裂音の識別に対する重要な成分は破裂とフォルマント遷移部分である。摩擦音の識別に対する重要な成分は雑音部とフォルマント遷移部分であり。鼻音と流音に対しても遷移部分が重要な成分だと指摘された [14]。そのため、音韻を正しく聞き取るうえでフォルマント遷移は重要である。

それ以外に、音声了解度に関する重要な成分に関して、Drullman により、時間

領域の振幅包絡線情報が音声知覚に対して重要であると指摘された [15]。

つまり、音声信号の時間領域と周波数領域が音声了解度に対して重要であると考えられる。

2.2 骨導音声の音響特性

人間の聴覚は気導音だけを知覚するわけではない、頭蓋骨、皮膚などの振動で伝達する骨導音も知覚できる [12]。そのため、伝音性難聴者は、骨伝導を利用すれば、音声を聞くことができる。骨導音声の利点は、耳を塞がずに骨導伝導による音声を聞くことができるため、気導音声も同時に聴取できる点である。

骨導音声の伝達経路は気導音声のものと比べ複雑であり、頭部（頭蓋骨や軟組織）の振動が直接中耳や内耳に伝達される経路だけではなく、頭部の振動が外耳道内に空気振動として漏れ出る経路（外耳道内放射）も存在する [16]。骨導伝達の経路は図 2.1 に示す。図 2.1 に見せた気導・骨導の伝達過程の違いから、気導音声と骨導音声のスペクトル特徴は大きく異なる。骨導で音声を伝達するときに、皮膚と軟組織などの影響により、高い周波数成分が減衰するという特性がある [17]。図 2.2 には同時に収録した気導音声と骨導音声のスペクトログラムを示す。気導音声と骨導音声の内容が同じである。示した気導音声と骨導音声のスペクトログラムを見ると、スペクトログラムの色が濃ければ濃いほどパワーが強い。それより、骨導音声のパワーが気導音声より弱くなり、特に高域成分が大幅に減衰したことがわかった。そのことから、骨伝導はローパスフィルタのような低域通過特性があるとみえられる。骨導伝達の高域減衰特性による、骨導音声の高周波数帯域は気導音声より 10 dB 減衰したことが指摘された [18]。このことは骨導音声の了解度に影響すると考えられる。また、前節で述べた高域成分に集中している子音に対して大きく影響が及ぶと考えられる。それに関して、前田らは骨導音声と気導音声を比べ、音声の了解度が低下し、音韻レベルでの分析結果は、母音と子音の正答率が低下したことを示した [19]。

骨導音声は頭蓋骨などの振動で音声を伝導できるため、骨導で音声を聴取できるだけではなく、収録にも使える。骨導マイクロホンは雑音環境下でも、背景音の影響を受けないように、品質がよい音声を収録できることが期待される。しかし、高域成分が減衰することで、骨導で収録した音声の了解度を回復する必要がある。

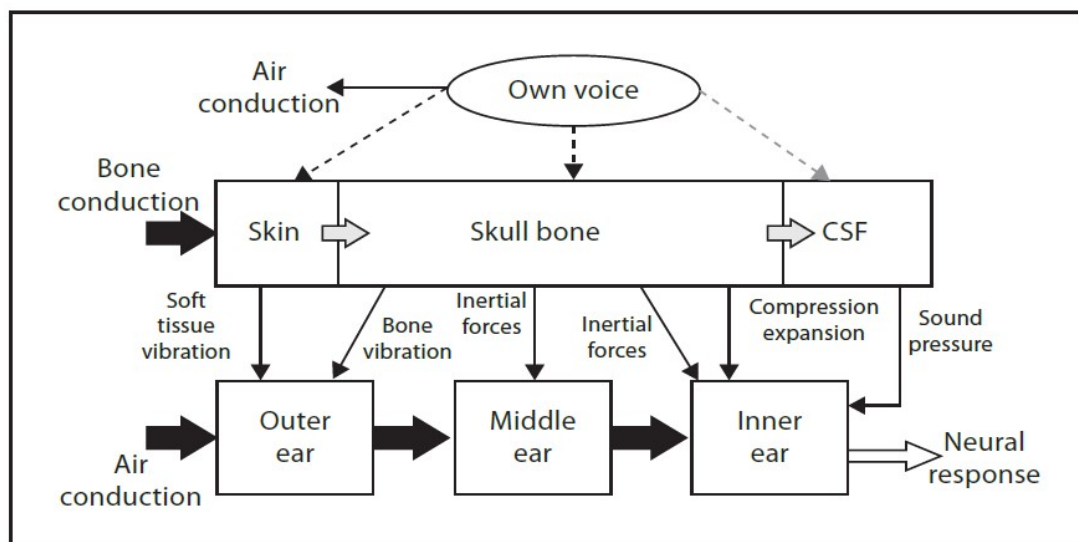


図 2.1: 気導・骨導音声の聴取経路（文献 [6] より引用）

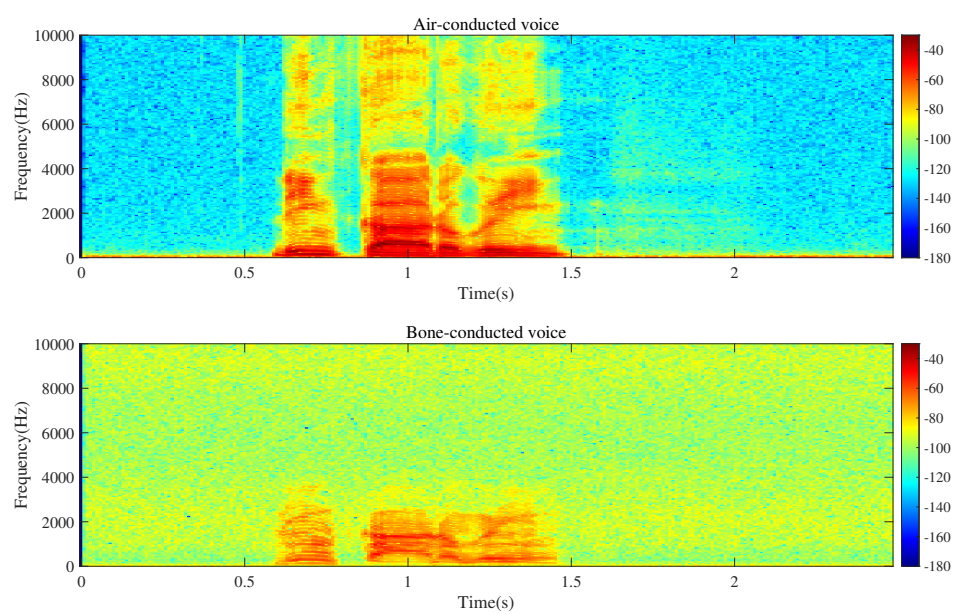


図 2.2: (上) 気導音声 (下) 骨導音声のスペクトログラム

2.3 音声了解度改善に関する研究

振幅包絡線情報を分析できる音声スペクトルに着目して、音声明瞭度を改善する先行研究について、川原らは音声スペクトルの時間変化強調方法を提案した [20]。音声の 500～4000 Hz の間の成分を強調することでスペクトルの時間変化を強調した。雑音環境下において音質評価値で強調効果を検証し、音声の明瞭性を改善できると証明された。Houtgast と Steeneken により提案された、音声スペクトルを用いて求めた変調伝達関数 (Modulation Transfer Function, MTF) に基づく音声明瞭度予測理論がある [21] ～ [23]。変調伝達関数は 2 種類の音声の変調スペクトル間の関係を表す関数である。音声信号の変調スペクトルは時間変化を周波数領域で表したものであると指摘された [24]。変調強調処理とは、伝達したい信号に基づき、信号に対する振幅や周波数に変化を与えることである。雑音環境下で音声了解度を改善するため、Koutsogianna と Styliou は音声スペクトルを変調強調する方法を提案した [25]。明瞭な音声には高い変調度があると証明された [26]。音声信号の変調度を変えるため、Koutsogianna と Styliou は音声信号の変調スペクトルに着目して、明瞭な音声のスペクトル特性に基づいて簡単な変調伝達関数を設計した。変調伝達関数を用いて音声時間スペクトルの変調度を変えることで、音声了解度を改善できると認められた。鶴木は変調伝達関数に基づいたパワーエンベロープ逆フィルタをかける処理で音声回復する方法を提案した [27]。

音韻レベルで見たとき、子音の音響特性が原因で、母音より子音の聞き取りが音声了解度に大きい影響をもたらす可能性がある。音声了解度を改善する研究の中に、子音を対象として改善法を提案した研究がたくさんある。その中に、星野は子音部とフォルマント遷移部分を強調することで、子音強調する方法を提案した [28]。星野は音声を帯域分割して、音声スペクトルのピークを強調した。ヘッドフォンで難聴者に音声を提示する了解度試験の結果、無声破裂音 /t/ と /k/ の明瞭度を改善でき、有声破裂音 /b/ と /d/ には改善効果が見られたが、強調した音声と原音声の間に有意差がなかった。破裂音以外の子音に対する顕著的な改善効果が見られなかった。星野の実験は、無響室で雑音のない環境下で行なわれた。騒音環境下における子音を強調する研究については、安武と中島は音声の子音部をリアルタイム強調する方法を提案した [29]。子音を強調することで、雑音環境下におけるスピーカなどで提示した音声の明瞭度・了解度の改善効果が認められた。また、DRC (Dynamic Range Compression) による処理は、雑音環境下で子音を強調することで音声了解度を向上できると証明された [30]～[31]。DRC によって、音声母音パワーの一部を子音に移して、パワーを再配置することで、原音声と比べて、相対的に子音を強調した。雑音環境下で音声了解度試験で DRC は音声了解度に改善効果がある。しかし、DRC で処理した音声と原音声を比べて子音は強調したが、母音も減衰した。本研究では母音を変えずに子音だけを変えることで、子音に強調処理を加えることができると期待される。

子音の時間が短いことに対して、安達らは子音部の持続時間を伸べることで子

音の音声明瞭度を改善できることを指摘した [32]。音声の中の子音部時間を伸長することで、鼻音・半母音の明瞭度を改善できた。

音声の子音を強調することで音声の振幅包絡線が変わるため、振幅包絡線が音声了解度に対する重要であると証明した。従って、音声の振幅包絡線情報は音声了解度改善に対する重要であると認められた。

2.4 骨導音声の了解度改善に関する研究

骨導音声の了解度を改善するため、骨導マイクによる収録した音声の了解度を改善する研究がある。Shin らは骨導マイクで収録した音声を強調することで了解度を改善する方法と骨導音声から気導音声へ回復する方法を述べた [33]。

衣笠 [34] は変調伝達関数に基づく骨導音声回復法を提案した。骨導音声と気導音声を帯域分割して、気導音声の変調スペクトルに対して骨導音声の変調スペクトルとの変調伝達関数を求めた。変調伝達関数の逆特性に基づく骨導音声のパワーエンベロープに減衰した部分を回復することで、骨導マイクで収録した音声の明瞭度を改善できた。

耳を塞がない骨導デバイスで音声を提示するときに、気導音による雑音も同時に聴取してしまうことが骨導提示音声の了解度に影響を与えると考えられる。雑音が骨導提示音声の了解度を低下に影響を与えることに関して、Fujimoto と Mori は、両耳同時に耳栓を装着した状態で、骨導デバイスを使用するときの音声了解度が高くなることを示した [35]。耳栓を装着することで、雑音の影響を受けないため、骨導提示音声の了解度が両耳を開放するときより高くなった。しかし、耳栓を装着すれば、骨導音声と気導音声を同時に聴取できるという骨導デバイスの利点が損なわれる。そのため、本研究では、外耳道を開放した状態で、高騒音環境下における音声了解度を改善できる方法を検討し、提案することが狙いである。

Toya らは、気導音声を基準とした側頭部振動 (RT) と外耳道内放射音 (EC) の骨導伝達特性を測定した。大域的には気導音声を基準として、骨導音声の高域成分が減衰する特性を有することを示した [36]。この高域減衰特性が骨導提示音声の了解度に影響を与えると考えられる。藤田による骨導伝達特性を補償する方法を提案し、外耳道を開放した状態で、高騒音環境下における骨導提示音声の了解度を改善できた [11]。藤田 [11] は骨導伝達における高域減衰に着目し、骨導伝達特性を補償する高域強調処理により骨導提示音声の了解度を改善する方法を提案した。藤田 [11] による提案法が直線的な高域減衰特性の補償と微細なスペクトル変形の補償の両方に着目するため、一次高域強調 (FOE) と高次高域強調 (HOE) という二種類の強調処理を適用する。その結果、骨導伝達特性の逆特性を用いて骨導デバイスに入力される音声信号の周波数特性を事前に補償する高域強調処理が骨導提示音声の了解度改善に有効であることを示した。特に、側頭部振動での骨導伝達特性を補償する高域強調によって骨導提示音声の了解度を効果的に改善できた。

藤田 [11] による提案した方法では骨導提示音声の了解度を改善したが、絶対的に良好なレベル (了解度で 90 % 以上) に至っていないわけではない。また、低親密度の音声に対して、十分な改善効果が得られないため、雑音環境下で骨導提示音声の了解度をさらに改善する必要があると考えられる。

2.5 問題点

藤田による提案した方法で骨導提示音声の改善効果が良好なレベルに至っていないことから、骨導伝達特性を補償することだけで強調できなかった部分があると考えられる。藤田による提案法は周波数領域に着目したが、音声了解度に対して重要な振幅包絡線に着目しなかった。

また、骨導伝達に伴う高域減衰特性の影響を受けることから、母音と比べてパワーがより高域に集中する子音は骨導デバイスで提示する際に減衰される可能性が高いと考えられ、骨導提示音声の子音の知覚が難しいと考えられる。気導音声の子音と比べて、骨導音声の子音のパワーがさらに弱くなり、さらに聴き取りにくくなる可能性がある。そのため、藤田による提案した手法は、骨導提示音声の子音は母音より異聴を起ししやすい問題に対応できなかったと考えられる。

第3章 提案法

3.1 提案法のコンセプト

藤田による提案法で着目されていなかった振幅包絡線に着目する骨導提示音声の了解度を改善する方法について説明する。

振幅包絡線の全体に着目する変調伝達関数について、衣笠は変調伝達関数の逆特性に基づいて、骨導デバイスで収録した骨導音声の了解度を回復できたため、同様に変調伝達関数の逆特性に基づく気導音声を強調できると考えられる。骨導提示する気導音声に変調伝達関数の逆特性で変調伝達関数を補償することによって、骨導提示するときに減衰する部分を強調できると考えられる。そのため、骨導伝達音声と気導伝達音声の MTF の逆特性に基づいて、骨導提示音声の了解度の改善効果があると期待される。

また、子音が骨導伝達特性の影響を受けて、母音より聞きにくくなると考えられる。気導音声と骨導音声の間に、骨導音声の子音の正答率が低下することが証明された [19]。藤田による提案した方法で処理した骨導提示音声の中に間違いやすい部分を解明するため、藤田の実験結果を用いて音韻誤答率を分析した。図 3.1 に音韻誤答率の分析した結果を示す。ここで、音韻誤答率は、子音（母音）の総数に占める間違えた子音（母音）の数の比を表す。どのような雑音環境下にも、子音は母音より誤答率が高い、齟齬を起ししやすい傾向が見つかった。特に、75 dB 雑音環境下に、子音誤答率は母音誤答率の 5 倍より高い（子音:0.196, 母音:0.038）と解明された。藤田による提案した手法は音声の時間領域全体にわたり一様に高域強調処理を施すことであるが、この手法では骨導提示音声の音韻レベルの子音が間違いやすさを改善できない。それに対して、子音を強調すれば、骨導提示音声の了解度を改善できると考えられる。子音強調に関する先行研究は気導音声を対象にしたものがほとんどであるが、骨導で音声を伝導するとき、皮膚と頭蓋骨振動し、皮膚と頭蓋骨の振動が外耳と中耳にも伝わる [6]。外耳振動すれば外耳から中耳に伝音する過程と気導音の特性が似ている。このことを踏まえると、子音強調が気導音において有効性があることから、骨導伝達通路のこの部分においても有効性があると考えられている。したがって、音声の時間構造に着目して子音のみを選択的に強調すれば、骨導提示音声の了解度をさらに向上させることができると考えた。子音強調は振幅包絡線を非線形強調する方法である。

本研究では、高騒音環境下における骨導提示音声の了解度改善に向けて、音声の振幅包絡線情報に着目し、時間・周波数軸上で効果的に強調する方法を提案する。

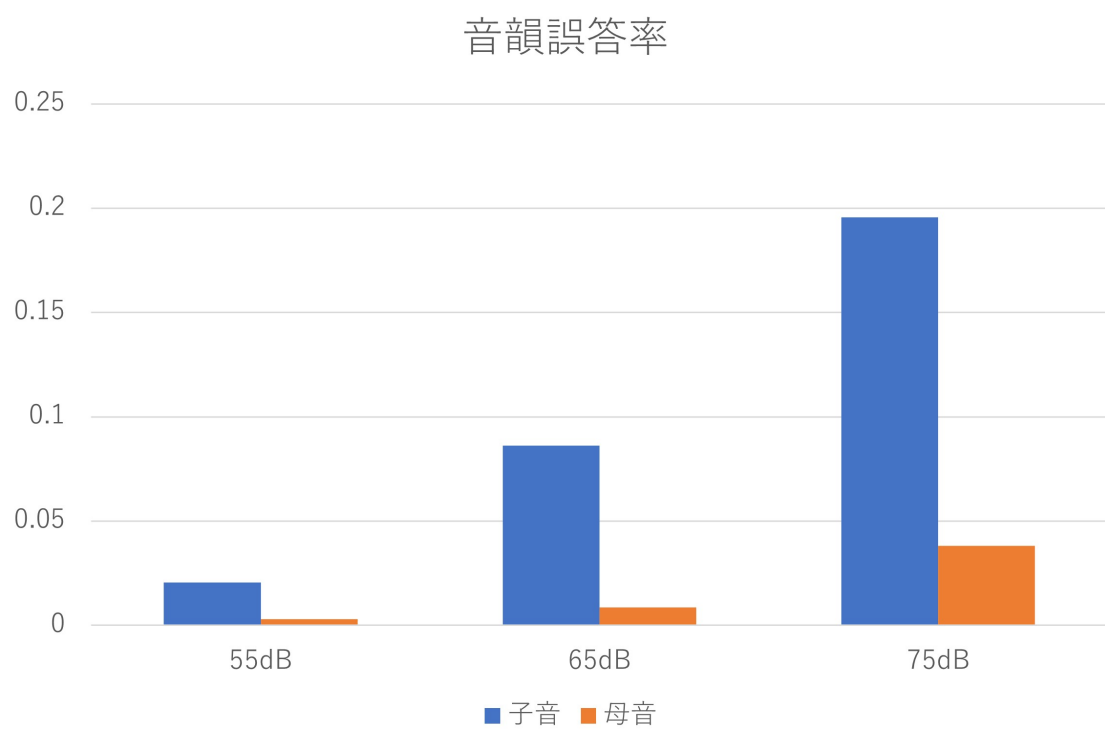


図 3.1: 音韻誤答率 (r)。子音 (母音) の総数に占める間違えた子音 (母音) の数の比 ($0 \leq r \leq 1$) を表す

3.2 子音強調による骨導提示音声の了解度改善法

3.2.1 子音強調

Toya らにより測定された骨導伝達特性の影響で、骨導提示音声の周波数成分の中に 5 kHz 以上の高域成分が減衰してしまう [36] ことがわかった。このような高域減衰は主に高周波数領域に集中している子音の成分を減衰して、子音の了解度低下の要因となっていると考えられる。

骨導伝達特性の影響を受けるため、骨導音声の周波数成分が気導音声の周波数成分よりパワーが小さくなると考えられるが、子音の音響特性にも依存し、特に無声子音は骨導伝達特性の影響をより受けやすいと考えられる。Kato と Shimamura は無声摩擦音がほかの子音より減衰しやすいと証明した [37]。そのため、本研究での提案法は高周波数領域に集中している骨導伝達特性の影響を受けやすい子音を強調することである。その以外に、子音の維持時間が短いために、子音を強調することだけでは十分な改善効果を得にくいと考えた。骨導提示音声の了解度をよく改善できるため、子音の知覚に対して重要なフォルマント遷移部分を一緒に強調することを提案した。

提案法による子音強調処理の概略を図 3.2 に示す。図 3.2 の A～D は下記の子音強調手順 A～D に対応する。ここでは、A～C は子音強調区間の決定、D は強調処理である。

A. 子音区間の検出：

音声の子音部の検出は、次の手順で行った。

- 1). 音声信号 $x(t)$ の全周波数帯域の振幅包絡線 $e_a(t)$ を、次式に従って求めた。

$$e_a(t) = \text{LPF}[\text{Hilbert}[x(t)]] \quad (3.1)$$

ここで、LPF はカットオフ周波数 100 Hz の 3 次の低域通過フィルタ処理、Hilbert はヒルベルト変換を表す。

- 2). 音声信号 $x(t)$ の 5 kHz 以上の高周波数帯域の振幅包絡線 $e_h(t)$ を、次式に従って求めた。

$$e_h(t) = \text{LPF}[\text{Hilbert}[\text{HPF}(x(t))]] \quad (3.2)$$

ここで、LPF と Hilbert は 1). と同様の処理を表し、HPF はカットオフ周波数 5 kHz の 9 次の高域通過フィルタ処理を表す。

- 3). $e_a(t)$ に対する $e_h(t)$ の相対レベル $P_r(t)$ を次式で求めた。

$$P_r(t) = 20 \log_{10} \frac{e_h(t)}{e_a(t)} \quad (3.3)$$

- 4). $P_r(t) > -12$ dB を満たす時間区間 t ($t_{s_k} \leq t \leq t_{e_k}$) を子音部とし、各子音の開始時点 t_{s_k} および終了時点 t_{e_k} を決定した。ただし、 $k = 1, 2, \dots, N$ (N : 音声信号に含まれる子音部の総数) である。

B. フォルマント遷移を考慮した強調区間拡張：

フォルマント遷移は子音の知覚にとっても重要なものであり、破裂音から母音への遷移時間は 50 ms 以内に完了すると言われている [14]。また、Furui はフォルマント遷移部分の時間長は子音の終了時点から約 10 ms であることを指摘した [38]。本研究では A で得られた t_{e_k} から $T_{\text{locus}} = 20$ ms の時間区間をフォルマント遷移部分として決定した。子音部に加えてフォルマント遷移部分も併せて強調した。

C. テーパー処理：

音声中の強調部分と非強調部分の間の急激な振幅変化を避けるため、フォルマント遷移部分の終端から $T_{\text{taper}} = 10$ ms にわたり振幅を緩やかに減衰させるようなコサインテーパー処理を施す時間区間を追加した。

D. 強調処理：

子音の開始時間 t_{s_k} からテーパー処理の終端 $t_{e_k} + T_{\text{locus}} + T_{\text{taper}}$ までの時間区間を強調区間とし、強調区間の振幅を +12 dB 増幅した。

以上の手順で子音強調（以下、CE と表記する）を施した。今回の提案法では無声破裂音、無声摩擦音と有声摩擦音を強調できる。

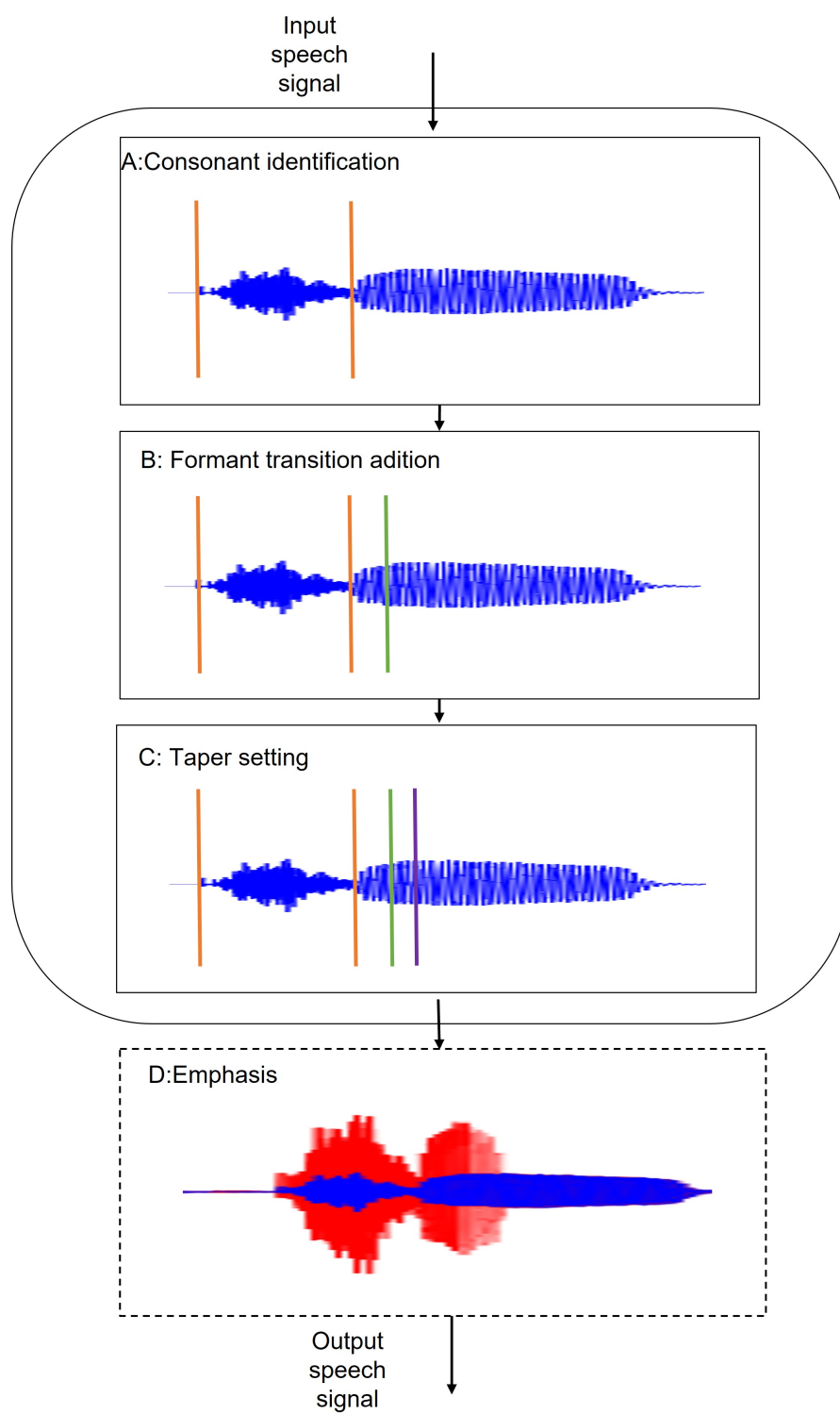


図 3.2: 提案法における子音強調処理の概略

3.2.2 子音強調と高域強調の組み合わせ

藤田 [11] による提案した高域強調と本稿での子音強調を組み合わせれば、高域強調単独での場合よりもさらに音声了解度を改善できると考えられる。そのため、藤田 [11] において改善効果の高かった、側頭部振動の伝達特性を補償する一次高域強調 (RT-FOE) と本研究での子音強調を、以下の手順で組み合わせた。子音強調と高域強調の組み合わせ方法の概略を図 3.3 に示す。図 3.3 の A~C は子音強調区間の決定、D は高域強調処理、E は子音強調である。

1. 音声信号 $x(t)$ を用いて 3.2.1 節の A~C により強調する時間区間を決定した。
2. 音声信号 $x(t)$ に RT-FOE の高域強調を施し信号 $x_e(t)$ を得た。
3. 音声信号 $x_e(t)$ において、1. で決定した時間区間に対して D の子音強調処理を施した。

上記の手順に従い、CE と RT-FOE のハイブリッド (RT-FOE+CE) により音声信号を事前に強調した。

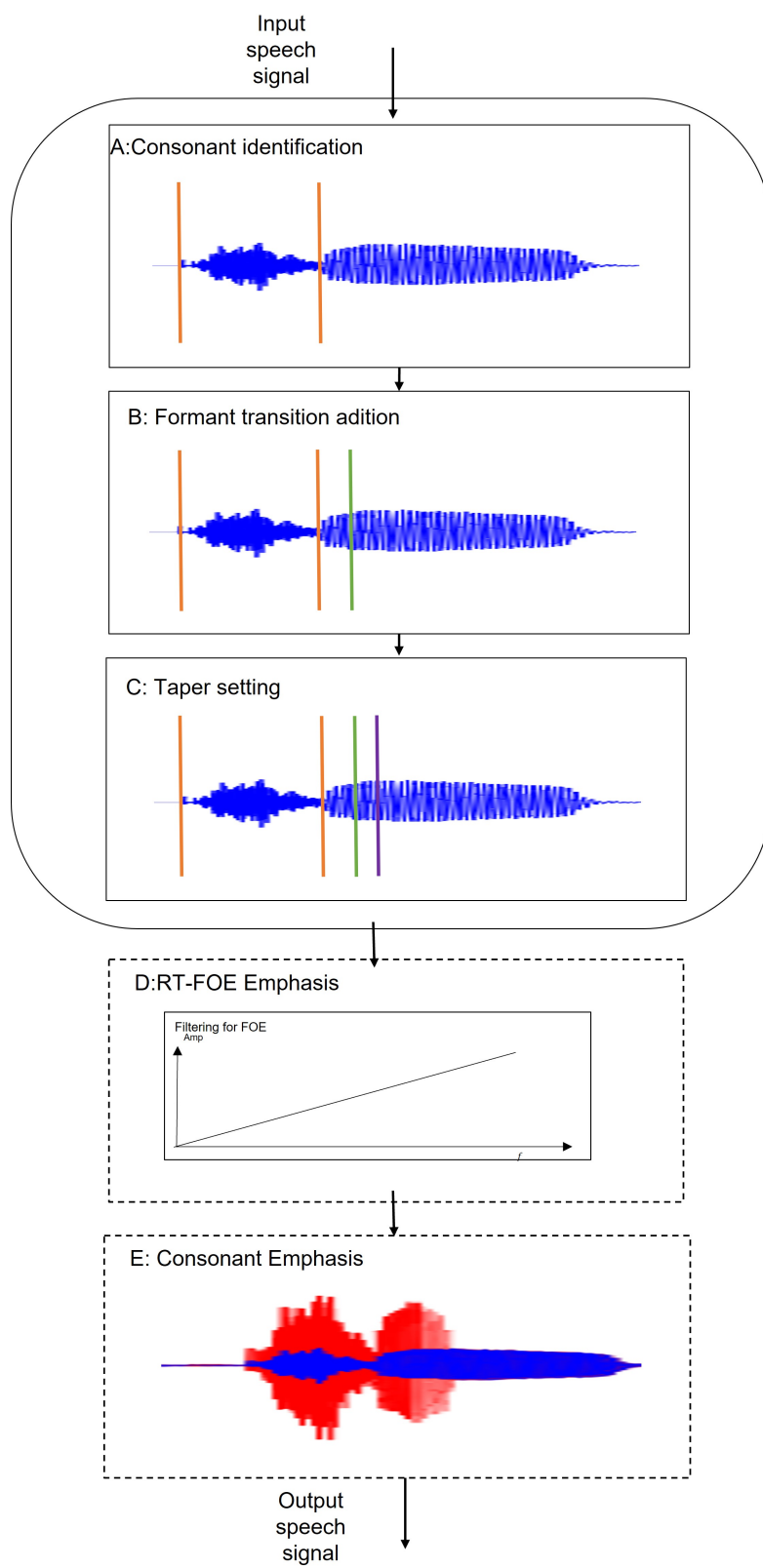


図 3.3: 子音強調と一次高域強調の組み合わせの概要

3.3 変調伝達関数に基づく骨導提示音声の了解度改善法

3.3.1 変調伝達関数の分析

骨導音声と気導音声の変調伝達関数 (Modulation Transfer Function, MTF) の逆特性に基づいて骨導提示音声の了解度に改善効果があると考えするため、骨導音声と気導音声の MTF を求める必要がある。そこで、MTF の定義に基づく骨導音声と気導音声の間の伝達特性を求めた。気導音声と骨導音声の変調スペクトルを求めてから伝達関数を明らかにした。気導音声と骨導音声の MTF を解明するため、同時に収録した気導音声と骨導音声が必要である。分析に使ったデータは気導音声（口唇付近）と骨導音声（側頭部）で同時に収録したデータベースである。データベースは本学で 15 人（男性 11 名、女性 4 名）による収録した音声 44 種類コマンドである（気導音声 $15 \times 44 = 660$ 個、骨導音声 $15 \times 44 = 660$ 個）。分析したデータの時間長が 1.5 秒である。MTF の分析概略を図 3.4 に示す。骨導デバイスでは音声にリアルタイムでの処理が求められると予想されるため、音声に MTF の逆特性を速く補償できることを求めた。そのため、音声を補償するときに処理速度に優れる帯域分割をしない状態で MTF を分析した。以下の手順で骨導音声と気導音声の MTF を求めた。

- 1). 気導音声信号 $v_a(t)$ と骨導音声信号 $v_b(t)$ の全周波数帯域の振幅包絡線 $s_a(t)$, $s_b(t)$ を、次式に従って求めた。

$$s(t) = \text{LPF}[\text{Hilbert}[v(t)]] \quad (3.4)$$

ここで、LPF はカットオフ周波数 60 Hz の 3 次の低域通過フィルタ処理、Hilbert はヒルベルト変換を表す。

- 2). 振幅包絡線 $s_a(t)$, $s_b(t)$ のパワーエンベロープの変調スペクトル ES_a , ES_b を次式に従って求めた。

$$ES = \text{FFT}[s(t)]^2 \quad (3.5)$$

ここで、FFT はフーリエ変換を表す。

- 3). 660 個音声信号 ES_a の平均値に対して 660 個音声信号 ES_b の平均値の伝達関数 F_{mtf} を次式で求めた。

$$F_{mtf} = \frac{\overline{ES_b}}{\overline{ES_a}} \quad (3.6)$$

以上の手順で求めた MTF は図 3.5 で示す。変調周波数成分において音声了解度に対する重要な成分は 1~16 Hz までである [39]。傾向を観察するため、30 Hz までを表す。

本研究での提案法は MTF の逆特性 (Inverse Modulation Transfer Function, IMTF) に基づく、気導音声を骨導デバイスで提示するときの減衰する部分を強調すると考えるため、IMTF を次式で求めた。求めた IMTF は図 3.6 で示す。

$$I_{mtf} = F_{mtf}^{-1} \quad (3.7)$$

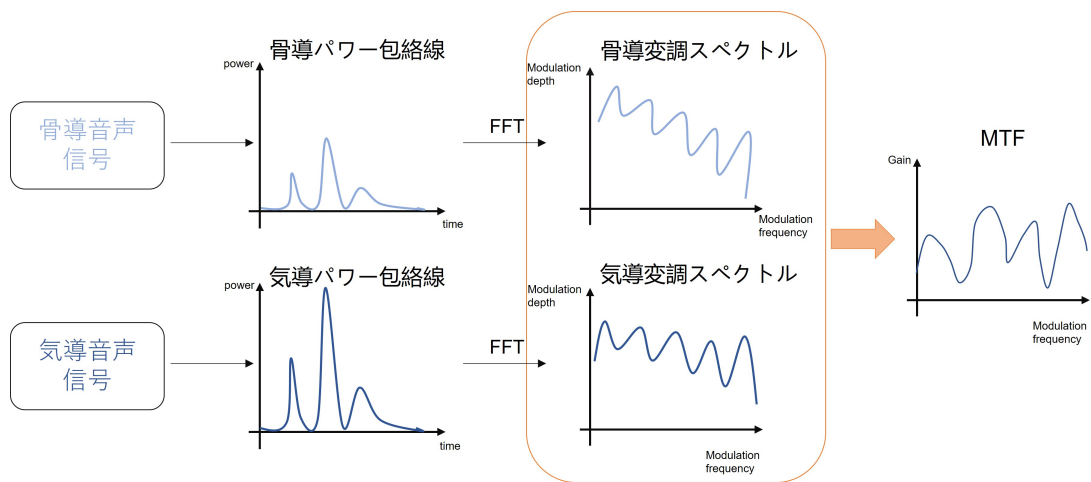


図 3.4: MTF の分析概要

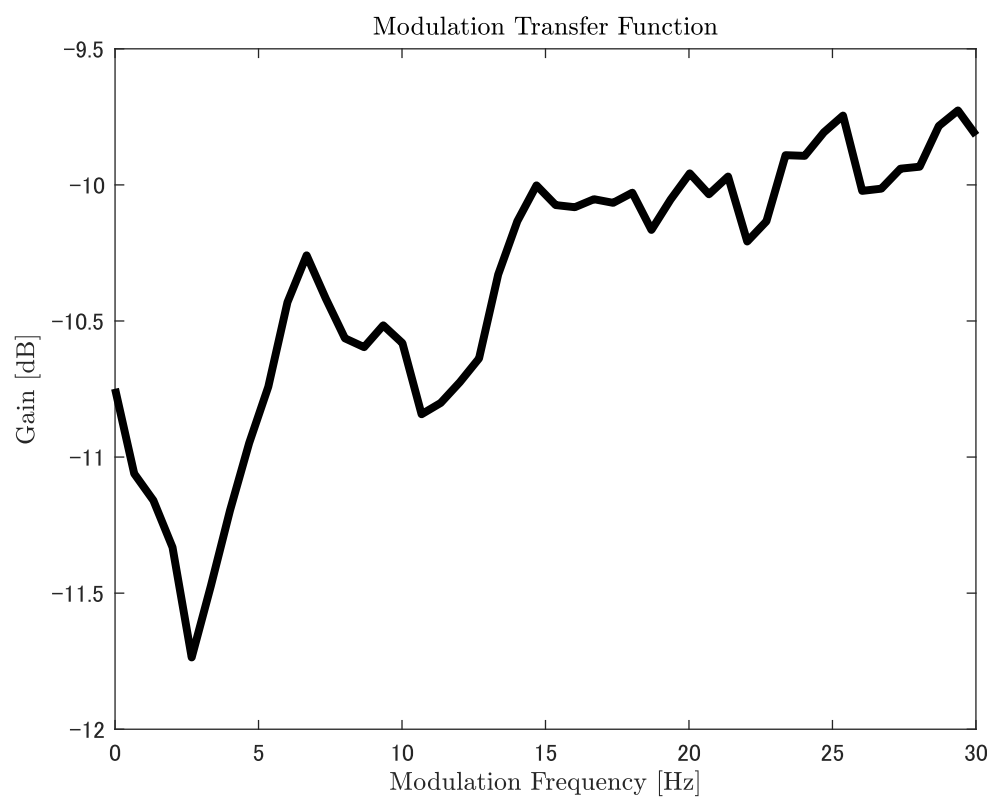


図 3.5: 骨導音声と気導音声の変調伝達関数

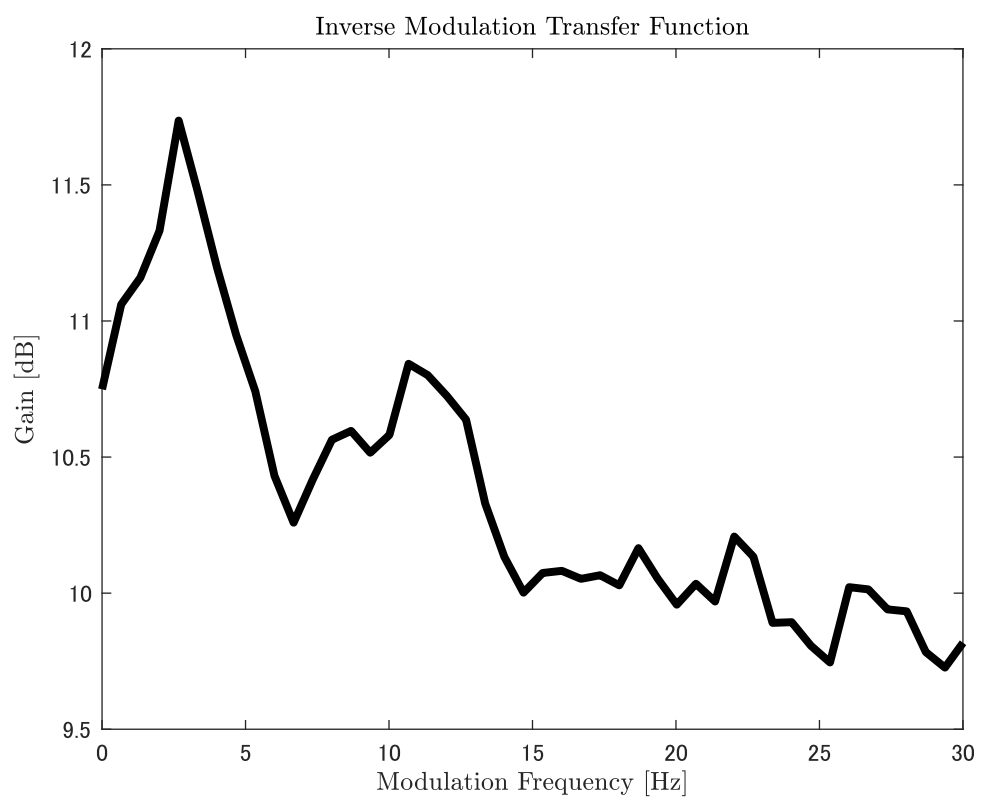


図 3.6: 骨導音声と気導音声の変調伝達関数の逆特性

3.3.2 フィルタ設計

骨導音声と気導音声の IMTF を利用して補償するフィルタを設計する。IMTF は図 3.6 で示す。IMTF に基づく設計したフィルタを気導音声にかける。これにより、骨導音声で減衰する部分を強調した気導音声を出力することができると考えられる。そのため、明らかにした MTF の逆特性を利用してフィルタを設計した。ただし、IMTF をそのまま使うことが音声の重要な成分に消極的な影響を与える可能性があることが知られている [40]。IMTF を全部使って、逆に音声了解度が悪化になる可能性があるため、IMTF の一部をもととしてフィルタを設計する必要があると考えられる。変調周波数成分の中に 4 Hz のところにあるピークは音声了解度に対して重要だと言われている [41]。今回分析した IMTF は 4 Hz の付近にピークがあるため、IMTF のピークがあるところを中心にして、理想的なフィルタを設計した。

図 3.6 に示すように、30 Hz までの範囲以内は、9 dB より高いと見える。その程度全体的な音圧レベルの増幅は物理的な処理で上げると考えているため、4 Hz 付近の強調倍数を中心に注目した。理想的なフィルタの設計は以下の手順で行った。

1. 緩やかな変化を示す 14 Hz から 30 Hz までの範囲の平均値を計算した。
2. 0 ～ 30Hz の値から 1. で求めた平均値を総体的に下げた。
3. ピークの終点 6 Hz 以降の音圧レベルから 1 dB まで緩やかに下げようようにした。
4. 3. の終わる時点から、その以降の値は 1 dB にした。

設計した理想的なフィルタを図 3.7 に示す。

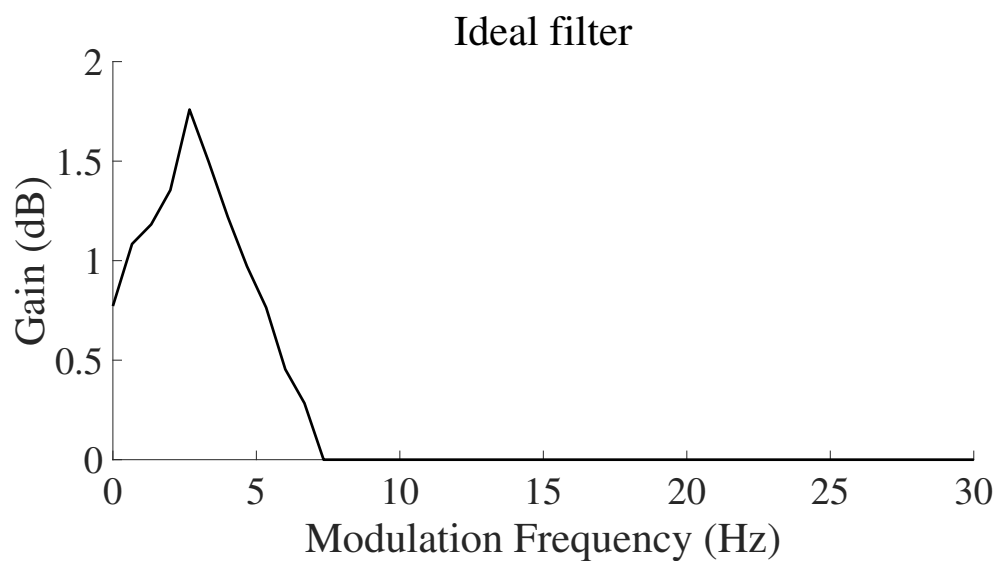


図 3.7: 設計した理想的なフィルタ

設計した理想的なフィルタを近似することで強調フィルタを用意した。骨導デバイスのリアルタイム処理を想定しているため、デジタルフィルタの中で処理速度が速い再帰型 Infinite Impulse Response (IIR) フィルタを設計した。一般に、IIR フィルタ特性 $H(z)$ は次式に従って定義される。

$$H(z) = \frac{B(z)}{A(z)} \quad (3.8)$$

ただし、上式の中にある $A(z)$, $B(z)$ は次式で表す。

$$A(z) = \sum_{k=0}^K a_k z^{-k} \quad (3.9)$$

$$B(z) = \sum_{k=1}^K b_k z^{-k} \quad (3.10)$$

ここで、 a_k ($a_0 = 1$) と b_k はフィルタ係数、 K はフィルタ次数である。分母 $A(z)$ および分子 $B(z)$ は、改良ユール・ウォーカー法 (Modified Yule-Walker Method) [42] で決めた。Modified Yule-walker では、設計したい理想的なフィルタの周波数応答に最小二乗近似を用いて、再帰型 IIR フィルタを設計できる。Yule-walker を用いて、以下の手順で分母 $A(z)$ および分子 $B(z)$ を決めた。

1. 設計した理想的なフィルタの周波数応答による、パワースペクトルを求めた。
2. 1. で求めたパワースペクトルを用いて逆 Fourier 変換により自己相関関数を求めた。
3. 自己相関関数から所望のフィルタ次数に即したフィルタの分母 $A(z)$ を算出した。
4. $A(z)$ を加法分解した。
5. スペクトル分解により、フィルタのインパルス応答を算出した。
6. 最小二乗法により、5. で求めたインパルス応答から分子 $B(z)$ を算出した。

理想的なフィルタと一番近似するフィルタを設計したいため、フィルタ次数が重要である。フィルタ次数が Modified Yule-walker を用いて設計したフィルタと理想的なフィルタの間の二乗平均平方根誤差 (Root Mean Square Error, RMSE) を用いて決めた。次数 1 から 52 までの間に設計したフィルタによる理想的なフィルタとの RMSE の変化を図 3.8 に示す。横軸はフィルタの次数、縦軸は Gain (dB) を示す。変化を観察しやすいため、 $Gain = 20 \log_{10} RMSE$ で表す。図 3.8 から、次数を高めると誤差が単調減少しているのがみえる。しかし、52 次以降からそうならないため、単調減少している範囲のなかで誤差が一番低い次数の 52 次を決めた。

52 次で設計したフィルタと理想的なフィルタの対照は図 3.9 に示す。図 3.9 の横軸は周波数 (Hz)，縦軸は Gain (dB) を表している。Ideal は近似元となった MTF の逆特性を表し，Postfilter は設計したフィルタを表す。

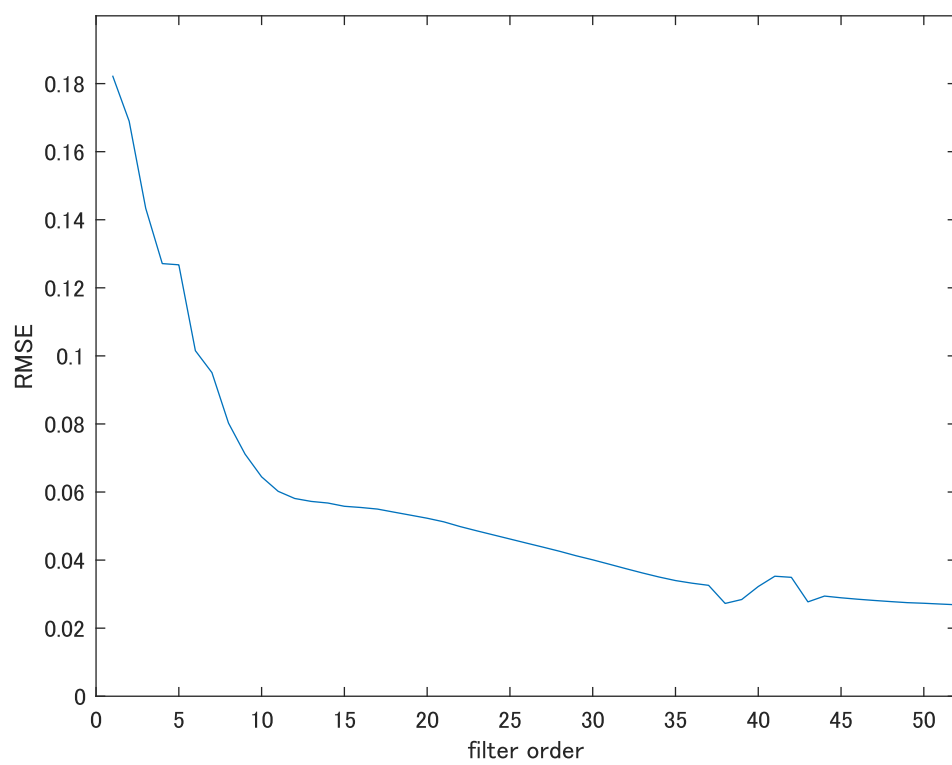


図 3.8: 設計したフィルタと理想的なフィルタの RMSE

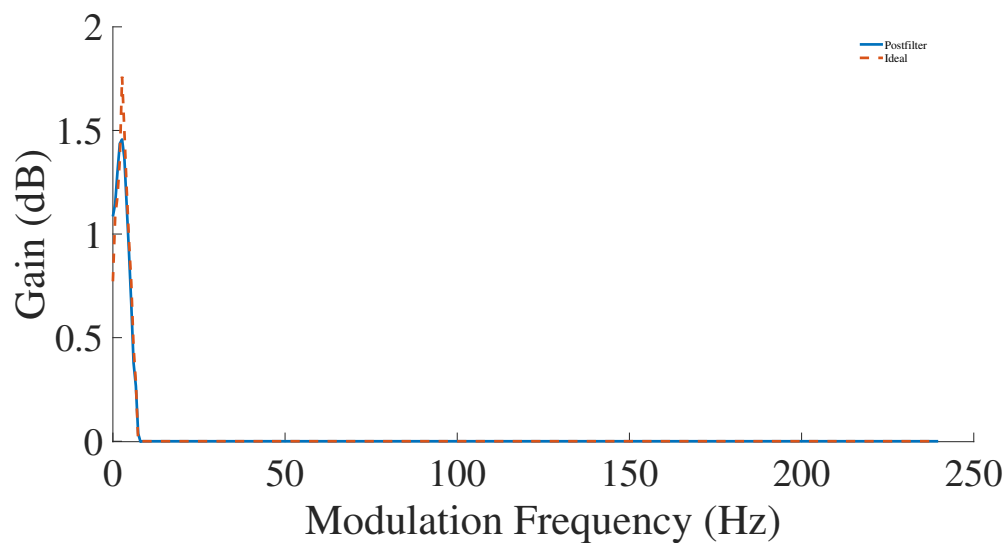


図 3.9: MTF の逆特性に基づく理想的なフィルタと設計したフィルタ

3.3.3 変調伝達関数の補償強調

変調伝達関数の補償強調処理は気導音声に設計したフィルタをかけることである。変調伝達関数は音声のパワー包絡線から求めたため、フィルタを同様に音声のパワー包絡線にかける。そのため、気導音声にフィルタをかける際にまずは音声の振幅包絡線とキャリアを分けた。変調伝達関数の補償強調は以下の手順で行う。

1. 気導音声信号 $x(t)$ の振幅包絡線 $y(t)$ を、次式に従って求めた。

$$y(t) = \text{LPF}|\text{Hilbert}[x(t)]| \quad (3.11)$$

ここで、LPF はカットオフ周波数 60 Hz の 3 次の低域通過フィルタ処理、Hilbert はヒルベルト変換を表す。

2. 気導音声信号 $x(t)$ の振幅包絡線 $y(t)$ を用いて、信号のキャリア $c(t)$ を次式に従って求めた。

$$c(t) = \frac{x(t)}{y(t)} \quad (3.12)$$

ここで、 $c(t)$ が無限大になってしまうことを防ぐため、 $y(t)$ の中に値が 0 のところを eps ($\text{eps} = 2.2204e^{-16}$) にした。

3. 振幅包絡線 $y(t)$ を二乗したパワー包絡線に前節で設計したフィルタをかけて、強調したパワー包絡線 $f_{y_p}(t)$ を生成した。
4. 強調した音声 $f_x(t)$ を次式で求めた。

$$f_x(t) = \sqrt{f_{y_p}(t)} \times c(t) \quad (3.13)$$

5. 強調した音声 $f_x(t)$ を原音声信号 $x(t)$ の RMS を一致するように調整した。

3.3.4 変調伝達関数改善法と高域強調の組み合わせ

藤田 [11] により提案された高域強調と本稿での変調伝達関数改善法を組み合わせれば、高域強調単独での場合よりもさらに音声了解度を改善できると考えられる。そのため、藤田 [11] における側頭部振動の伝達特性を補償する一次高域強調 (RT-FOE) と本研究の変調伝達関数改善法を、以下の手順で組み合わせた。子音強調変調伝達関数改善法と高域強調の組み合わせ方法の概略を図 3.10 に示す。図 3.10 の A~C は子音強調区間の決定、D は高域強調処理、E は子音強調である。

1. 音声信号 $x(t)$ を用いて 3.2.3 節の 1.~5. により変調伝達関数の補償を強調した信号 $x_m(t)$ を得た。
2. 音声信号 $x_m(t)$ に RT-FOE の高域強調を施し信号 $x_me(t)$ を得た。

上記の手順に従い、IMTF と RT-FOE のハイブリッド (IMTF+RT-FOE) により音声信号を事前に強調した。

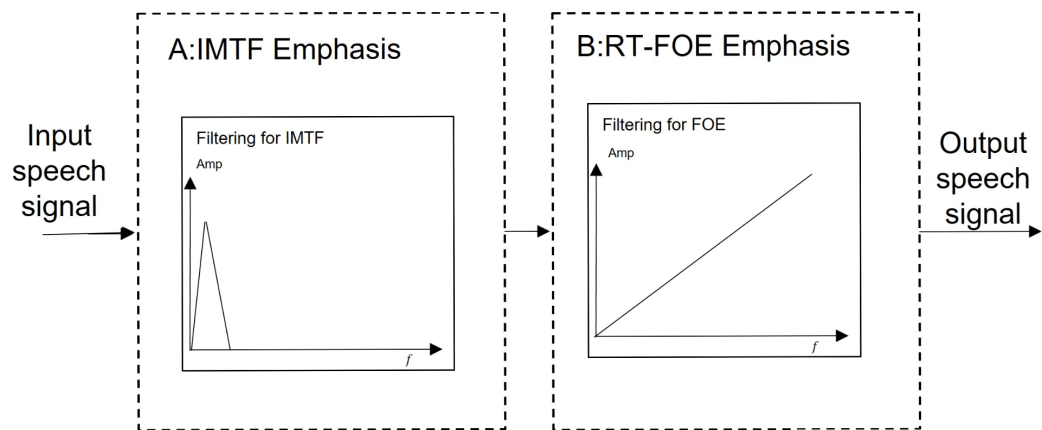


図 3.10: 変調伝達関数改善法と一次高域強調の組み合わせの概要

第4章 評価

雑音環境下で、本研究で提案した方法は骨導提示音声の了解度への改善効果を確認するため、2つ音声了解度試験を行った。

4.1 子音強調の評価

藤田[11]のRT-FOEによる強調処理および提案法の2種類の強調処理（CE, RT-FOE + CE）を施した際の骨導提示音声の了解度改善効果を比較するため、音声了解度試験を行った。

4.1.1 音声データ

親密度別単語了解度試験用音声データセット（FW07）[43]より、実験に用いる音声データを選定した。FW07は単語親密度の7段階評定値をもとに4つの親密度ランク（1: low（1.0～2.5）、2: lower-middle（2.5～4.0）、3: higher-middle（4.0～5.5）、4: high（5.5～7.0））に分けられた4モーラ単語から構成される。親密度が高い単語は日常生活の中に使う頻度が高い単語である。音声データのサンプリング周波数は48 kHz、量子化ビット数は16 bitsであった。

4.1.2 実験条件

3.2節で述べた2種類の強調処理（CE, RT-FOE+CE）を施した音声刺激、高域強調した音声刺激（RT-FOE）と、強調処理なし（No emphasis）の音声刺激を作成した。これら4種類の音声刺激には、4.1節で述べた4種類の親密度ランクの単語が含まれるようにした。雑音環境を模擬するため、2種類の提示音圧レベル（55 dB, 75 dB, 以後、雑音レベルと呼ぶ）のピンク雑音を用いた。条件数の合計は32（強調タイプ4種類×親密度4ランク×雑音レベル2種類）であった。

4.1.3 実験参加者

実験には、正常聴力を有する20代の日本語母語話者10名（男性6名、女性4名）が参加してもらった。

4.1.4 実験装置

実験は図 4.1 に示すように暗騒音レベル 15.2 dB の防音室で行った。音声刺激の提示には骨導トランスデューサ（株式会社テムコジャパン, KE08-01）とアンプ（audio-technica AT-HA5000）を用いた。音声をマスクするピンク雑音 [44] の提示にはスピーカ（ECLIPSE, TD508MK3）とパワーアンプ（Yamaha P4050）を用いた。スピーカは参加者の 70 cm 後方に設置した。音声・雑音刺激の提示制御は PC（LG Sharkoon, Windows 8）上のソフトウェア（MATLAB 2014a）および A/D コンバータ（Fireface UCX）を利用して行った。参加者の正面に、反応取得のためのスクリーンとキーボードを設置した。音声提示の際の骨導トランスデューサへの印加電圧は実効値で平均 0.368 V であり、このときのラウドネスと等価となる気導聴取での音圧レベルは約 60 dB であった。

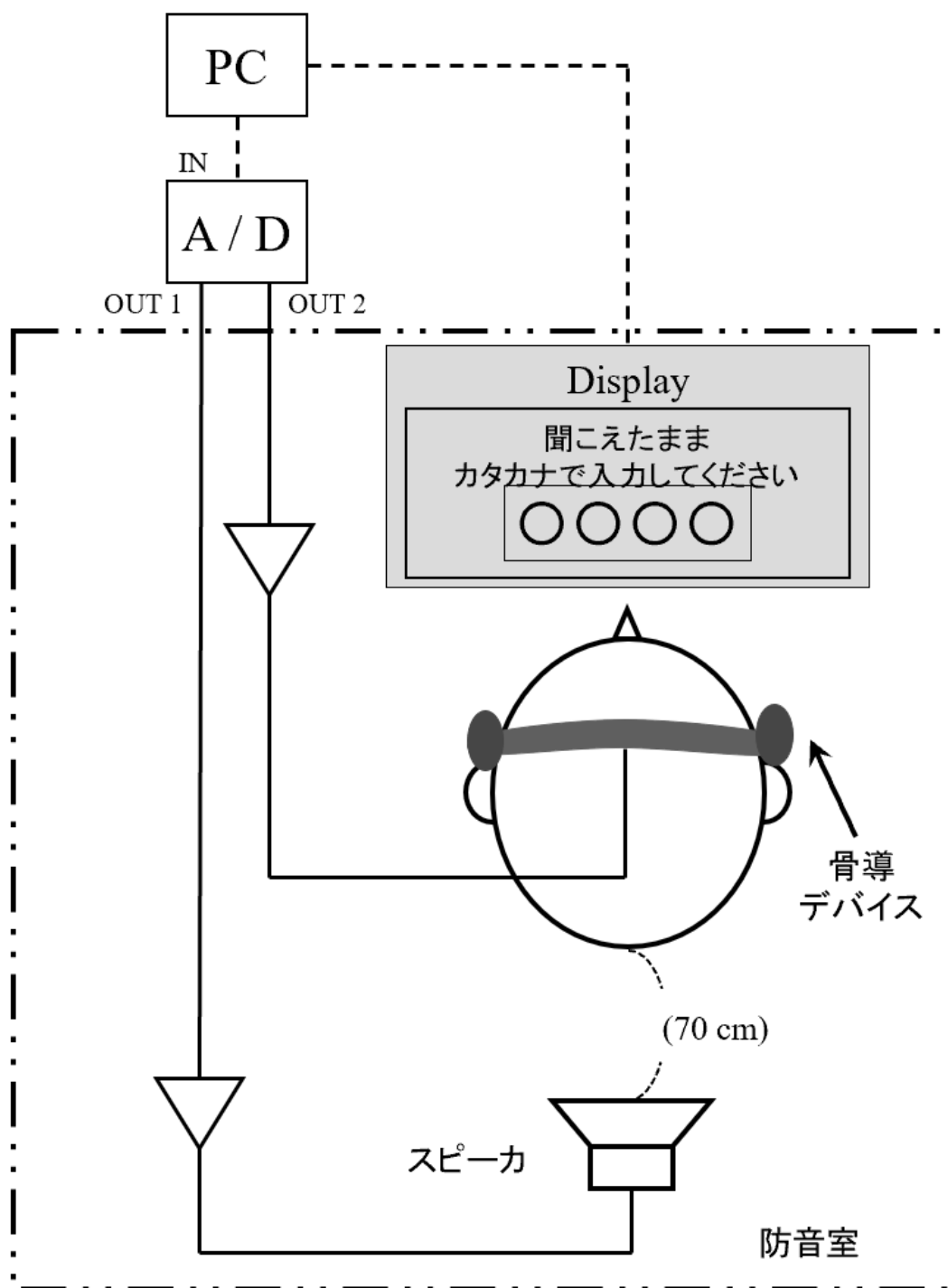


図 4.1: 単語了解度試験の実験装置と実験環境

4.1.5 実験手続き

実験参加者には、気導によるピンク雑音を提示しながら骨導デバイスから音声刺激を提示した。実験参加者には聴き取った単語をキーボードでカタカナ入力することを求めた。設問は1条件あたり20単語、計32条件640単語とした。含まれる単語が互いに異なる単語リストを4セット用意し、全参加者が同一の単語を聴取することを避けた。ある単語について、4モーラすべて正解したときにその単語を正答とした。雑音レベル、親密度ランク、強調タイプの種類（強調タイプ）ごとの単語正答率を算出し、比較評価に用いた。

4.1.6 実験結果

図4.2と図4.3はそれぞれ雑音レベル55 dB, 75 dBの条件下での単語正答率結果を表す。エラーバーは標準誤差を表す。図4.2と図4.3は、単語了解度試験で得られた強調タイプ・親密度ランクごとの単語正答率を示す。横軸は強調タイプに対して各親密度ランクを示し、縦軸である Word recognition rate は単語正答率を示している。

得られた単語正答率におけるANOVA4による分散分析を行った。ANOVA4では要因間の主効果、交互作用と単純主効果を同時に分析できる。

強調タイプ、親密度ランクと雑音レベル3要因の分散分析の結果は、単語正答率に対して強調タイプ [$F(3, 27) = 6.49, p < 0.01$], 親密度ランク [$F(3, 27) = 119.13, p < 0.01$], 雑音レベル [$F(1, 9) = 348.53, p < 0.01$] の主効果が認められた。また、強調タイプと雑音レベルの間に交互作用が認められたが [$F(3, 27) = 4.78, p < 0.01$], 強調タイプと親密度ランクの間には交互作用は認められなかった [$F(9, 81) = 0.56, p = 0.83$]。強調タイプと雑音レベルの交互作用に対して下位検定の結果、雑音レベル75 dBの下でフィルタタイプの単純主効果が認められた [$F(3, 54) = 11.25, p < 0.01$]。雑音レベル75 dBでの結果について、Holm-Bonferroni法による多重比較検定を行った。有意差が見られた強調タイプの組合せを、図4.2と図4.3に*印 ($p < 0.05$) および**印 ($p < 0.01$) で示す。CEにおける単語正答率はNo emphasisのものよりも有意に高かった。また、RT-FOE+CEは他の強調タイプよりも有意に正答率が高いこともわかった。

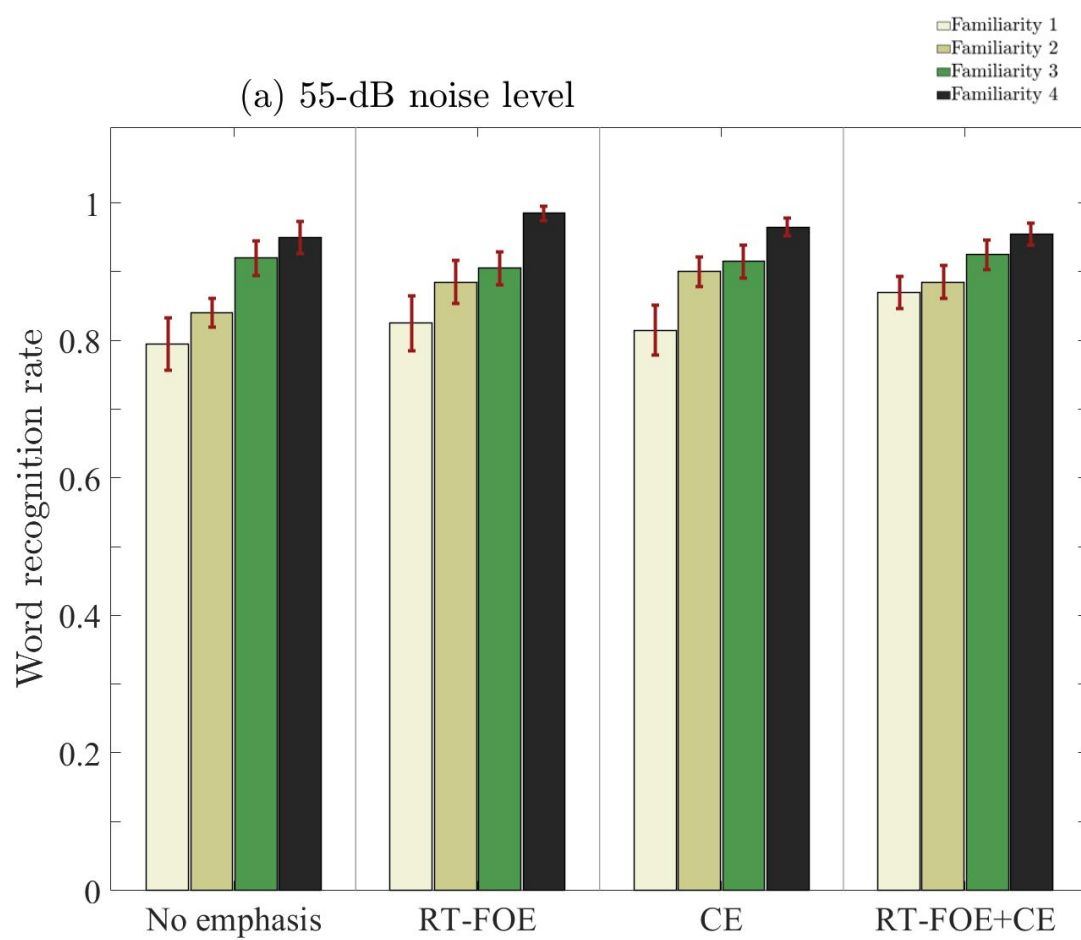


図 4.2: 子音強調：雑音レベル 55 dB 環境下親密度別単語正答率

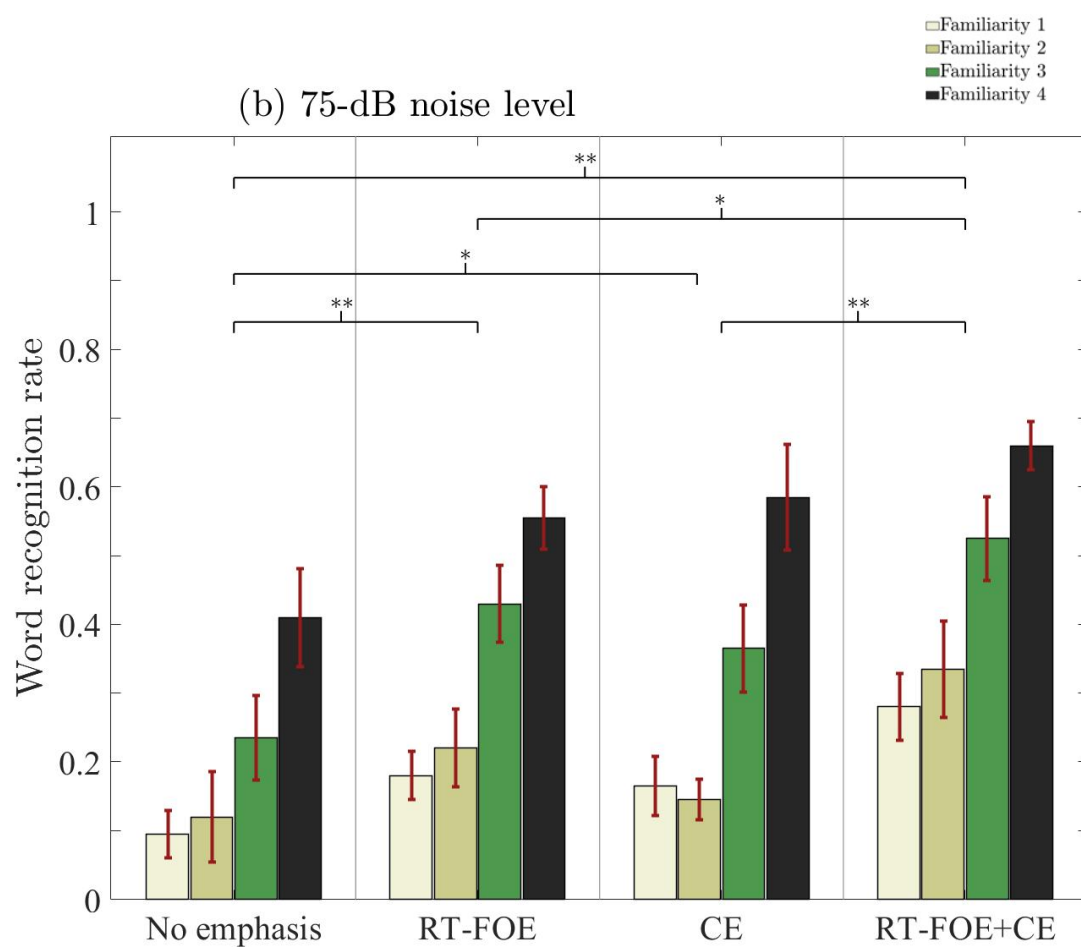


図 4.3: 子音強調：雑音レベル 75 dB 環境下親密度別単語正答率

表 4.1: 子音強調：単語正答率

音声条件	雑音レベル	親密度 1	親密度 2	親密度 3	親密度 4
No emphasis	55 dB	0.795	0.84	0.92	0.95
	75 dB	0.095	0.12	0.235	0.41
RT-FOE	55 dB	0.825	0.885	0.905	0.985
	75 dB	0.18	0.22	0.43	0.555
CE	55 dB	0.815	0.9	0.915	0.965
	75 dB	0.165	0.145	0.365	0.585
RT-FOE+CE	55 dB	0.87	0.885	0.925	0.955
	75 dB	0.28	0.335	0.525	0.66

表 4.2: 子音強調：有意差検定結果

source	df	F	p
A: 雑音レベル	1	348.532	0.0000
error[AS]	9		
B: 親密度ランク	3	119.132	0.0000
error[BS]	27		
C: 強調タイプ	3	6.494	0.0019
error[CS]	27		
AB	3	31.16	0.0000
error[ABS]	27		
AC	3	4.777	0.0085
error[ACS]	27		
BC	9	0.557	0.8282
error[BCS]	81		

4.2 変調伝達関数改善法の評価

藤田 [11] の RT-FOE による強調処理および提案法の 2 種類の強調処理 (IMTF, IMTF+RT-FOE) を施した際の骨導提示音声の了解度改善効果を比較するため、音声了解度試験を行った。

4.2.1 音声データ

親密度別単語了解度試験用音声データセット (FW07) [43] より、実験に用いる音声データを選定した。FW07 は単語親密度の 7 段階評定値をもとに 4 つの親密度ランク (1: low (1.0 ~ 2.5), 2: lower-middle (2.5 ~ 4.0), 3: higher-middle (4.0 ~ 5.5), 4: high (5.5 ~ 7.0)) に分けられた 4 モーラ単語から構成される。親密度が高い単語は日常生活の中に使う頻度が高い単語である。音声データのサンプリング周波数は 48 kHz, 量子化ビット数は 16 bits であった。

4.2.2 実験条件

音声データは子音強調評価の音声データを使った。3.4 節で述べた強調処理 (IMTF, IMTF+RT-FOE) を施した音声刺激, 高域強調した音声刺激 (RT-FOE) と, 強調処理なし (No emphasis) の音声刺激を作成した。これら 4 種類の音声刺激には, 4.1.1 節で述べた 4 種類の親密度ランクの単語が含まれるようにした。雑音環境を模擬するため, 2 種類の雑音レベル (55 dB, 75 dB) のピンク雑音を用いた。条件数の合計は 32 (強調タイプ 4 種類 × 親密度 4 ランク × 雑音レベル 2 種類) であった。

4.2.3 実験参加者

実験には, 正常聴力を有する 20 代の日本語母語話者 10 名 (男性 5 名, 女性 5 名) が参加した。

4.2.4 実験装置

実験は図 4.1 に示すように暗騒音レベル 15.2 dB の防音室で行った。音声刺激の提示には骨導トランスデューサ (株式会社テムコジャパン, KE08-01) とアンプ (audio-technica AT-HA5000) を用いた。音声をマスクするピンク雑音 [44] の提示にはスピーカ (ECLIPSE, TD508MK3) とパワーアンプ (Yamaha P4050) を用いた。スピーカは参加者の 70 cm 後方に設置した。音声・雑音刺激の提示制御は PC (LG Sharkoon, Windows 8) 上のソフトウェア (MATLAB 2014a) および A/D コ

ンバータ（Fireface UCX）を利用して行った。参加者の正面に、反応取得のためのスクリーンとキーボードを設置した。音声提示の際の骨導トランスデューサへの印加電圧は実効値で平均 0.368 V であり、このときのラウドネスと等価となる気導聴取での音圧レベルは約 60 dB であった。

4.2.5 実験手続き

実験参加者には、気導によるピンク雑音を提示しながら骨導デバイスから音声刺激を提示する。実験参加者には聴き取った単語をキーボードでカタカナ入力することを求めた。設問は 1 条件あたり 20 単語、計 32 条件 640 単語とした。含まれる単語が互いに異なる単語リストを 4 セット用意し、全参加者が同一の単語を聴取することを避けた。ある単語について、4 モーラすべて正解したときにその単語を正答とした。雑音レベル、親密度ランク、強調タイプの種類（強調タイプ）ごとの単語正答率を算出し、比較評価に用いた。

4.2.6 実験結果

図 4.4 と図 4.5 はそれぞれ雑音レベル 55 dB, 75 dB の条件下での単語正答率結果を表す。エラーバーは標準誤差を表す。図 4.4 と図 4.5 に、単語了解度試験で得られた強調タイプ・親密度ランクごとの単語正答率を示す。横軸は強調タイプに対して各親密度ランクを示し、縦軸である Word recognition rate は単語正答率を示している。

得られた単語正答率における ANOVA4 による分析を行った。強調タイプ、親密度ランクと雑音レベル 3 要因の分散分析の結果は、単語正答率に対して強調タイプ [$F(3, 27) = 6.94, p < 0.01$], 親密度ランク [$F(3, 27) = 83.54, p < 0.01$], 雑音レベル [$F(1, 9) = 1744.75, p < 0.01$] の主効果が認められた。また、強調タイプと雑音レベルの間に交互作用が認められたが [$F(3, 27) = 4.50, p < 0.05$], 強調タイプと親密度ランクの間には交互作用は認められなかった [$F(9, 81) = 0.44, p = 0.91$]。強調タイプと雑音レベルの交互作用に対して下位検定の結果、雑音レベル 75 dB の下でフィルタタイプの単純主効果が認められた [$F(3, 54) = 11.14, p < 0.01$]。雑音レベル 75 dB での結果について、Holm-Bonferroni 法による多重比較検定を行った。有意差が見られた強調タイプの組合せを、図 4.4 と図 4.5 に**印 ($p < 0.01$) で示す。IMTF における単語正答率は No emphasis のものよりも有意差が見られなかった。また、RT-FOE, IMTF+RT-FOE は No emphasis と IMTF のものよりも有意に正答率が高いこともわかった。

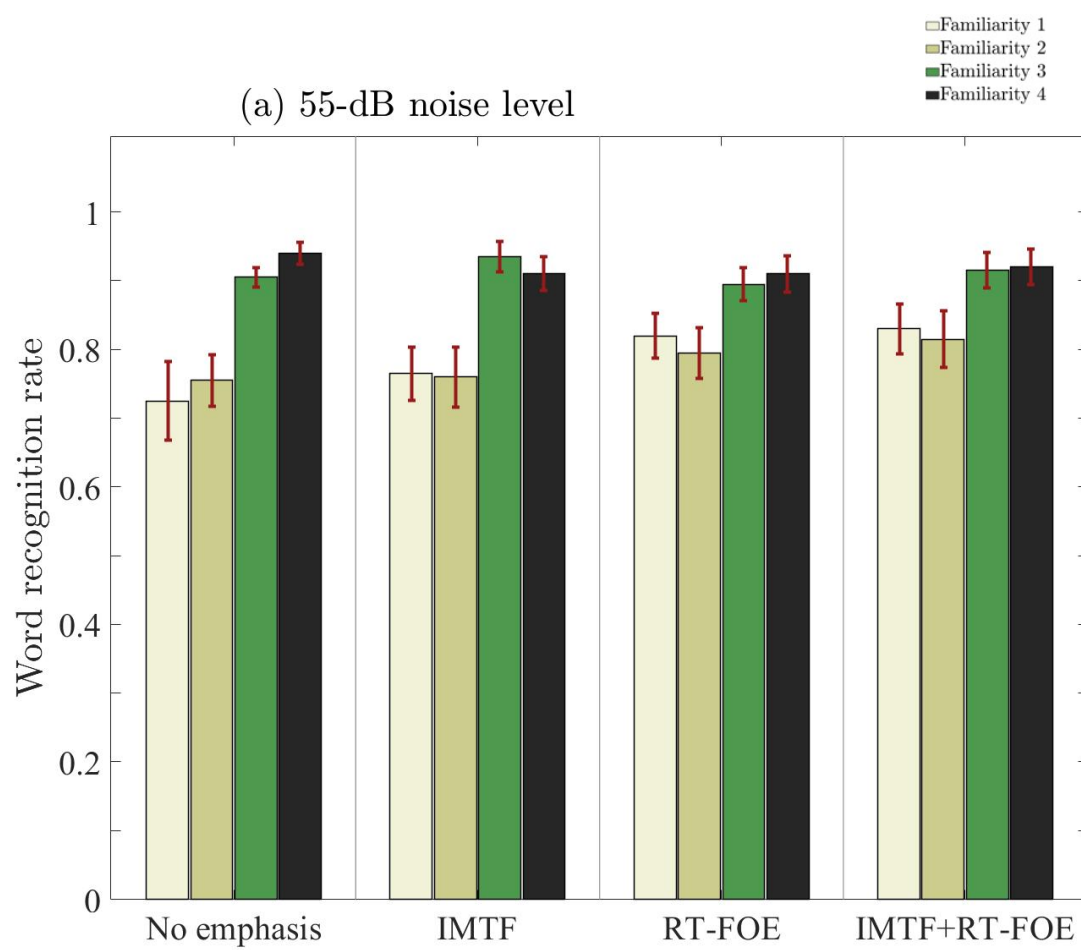


図 4.4: 変調伝達関数：雑音レベル 55 dB 環境下親密度別単語正答率

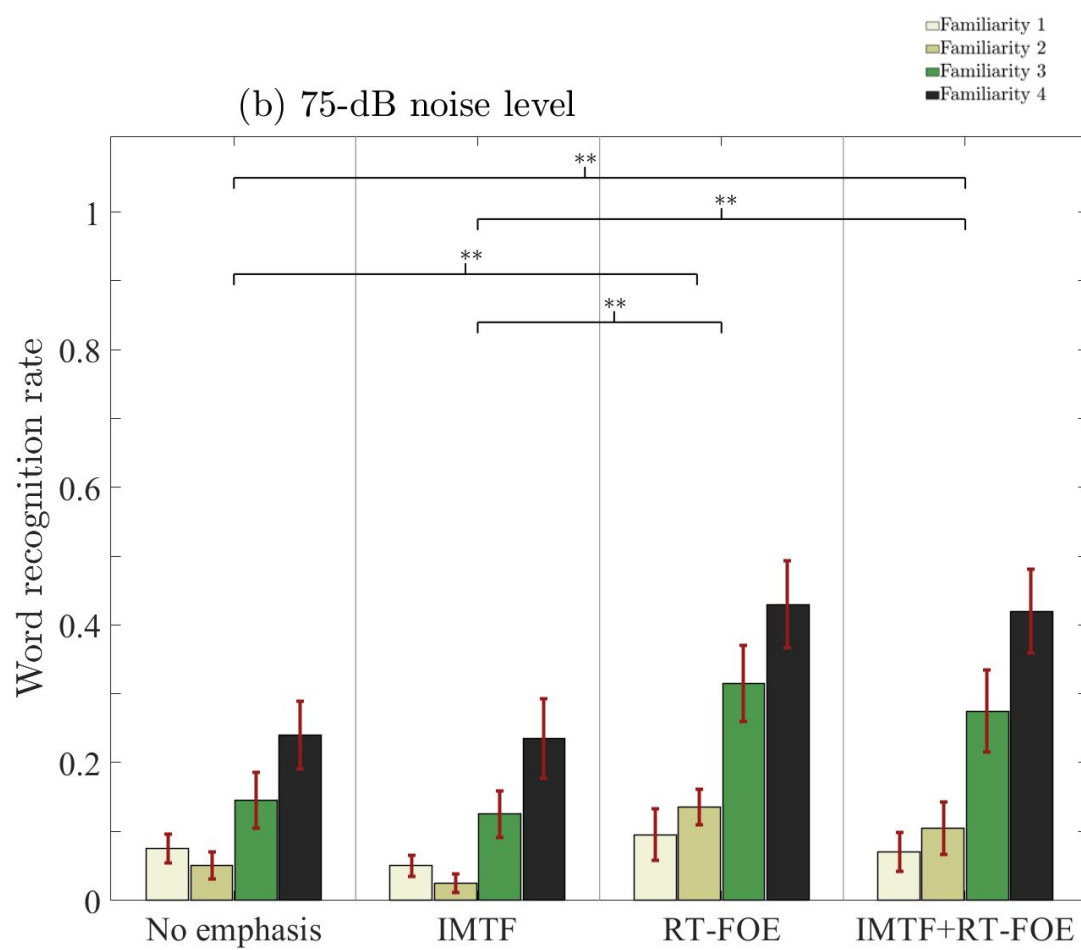


図 4.5: 変調伝達関数：雑音レベル 75 dB 環境下親密度別単語正答率

表 4.3: 変調伝達関数：単語正答率

音声条件	雑音レベル	親密度 1	親密度 2	親密度 3	親密度 4
No emphasis	55 dB	0.725	0.755	0.905	0.94
	75 dB	0.075	0.05	0.145	0.24
IMTF	55 dB	0.765	0.76	0.935	0.91
	75 dB	0.05	0.025	0.125	0.235
RT-FOE	55 dB	0.82	0.795	0.895	0.91
	75 dB	0.095	0.1135	0.315	0.43
IMTF+RT-FOE	55 dB	0.83	0.815	0.915	0.92
	75 dB	0.07	0.105	0.275	0.42

表 4.4: 変調伝達関数：有意差検定結果

source	df	F	p
A: 雑音レベル	1	1744.750	0.0000
error[AS]	9		
B: 親密度ランク	3	83.544	0.0000
error[BS]	27		
C: 強調タイプ	3	6.936	0.0013
error[CS]	27		
AB	3	4.391	0.0122
error[ABS]	27		
AC	3	4.500	0.0110
error[ACS]	24		
BC	9	0.444	0.9067
error[BCS]	81		

第5章 総合評価

本研究で提案した手法の組み合わせ方法の間で，骨導提示音声の了解度に対して改善効果が最も良い方法を検討するため，総合評価を行った。

5.1 三つの改善法の組合せ

藤田 [11] による提案した高域強調と本稿での子音強調，変調伝達関数改善法を組み合わせれば，高域強調単独の場合よりもさらに音声了解度を改善できることが期待される。そのため，藤田 [11] において改善効果の高かった，側頭部振動の伝達特性を補償する一次高域強調（RT-FOE）と本研究での子音強調，変調伝達関数改善法を，以下の手順で組み合わせた。変調伝達関数，子音強調，一次高域強調との組み合わせ方法の概略を図 3.3 に示す。図 3.3 の A～C は子音強調区間の決定，D は MTF の補償強調処理，E 高域強調処理，F は子音強調である。

1. 音声信号 $x(t)$ を用いて 3.1 節の A～C により強調する時間区間を決定した。
2. 音声信号 $x(t)$ に 3.2.3 節 MTF の補償強調を施し信号 $f_x(t)$ を得た。
3. 音声信号 $f_x(t)$ に RT-FOE の高域強調を施し信号 $h_x(t)$ を得た。
4. 音声信号 $h_x(t)$ を音声信号 $f_x(t)$ の RMS を一致するように調整した。
5. 音声信号 $h_x(t)$ において，1. で決定した時間区間に対して D の子音強調処理を施し信号 $s_x(t)$ を得た。

上記の手順に従い，CE，IMTF と RT-FOE のハイブリッド（IMTF+RT-FOE+CE）により音声信号を事前に強調した。

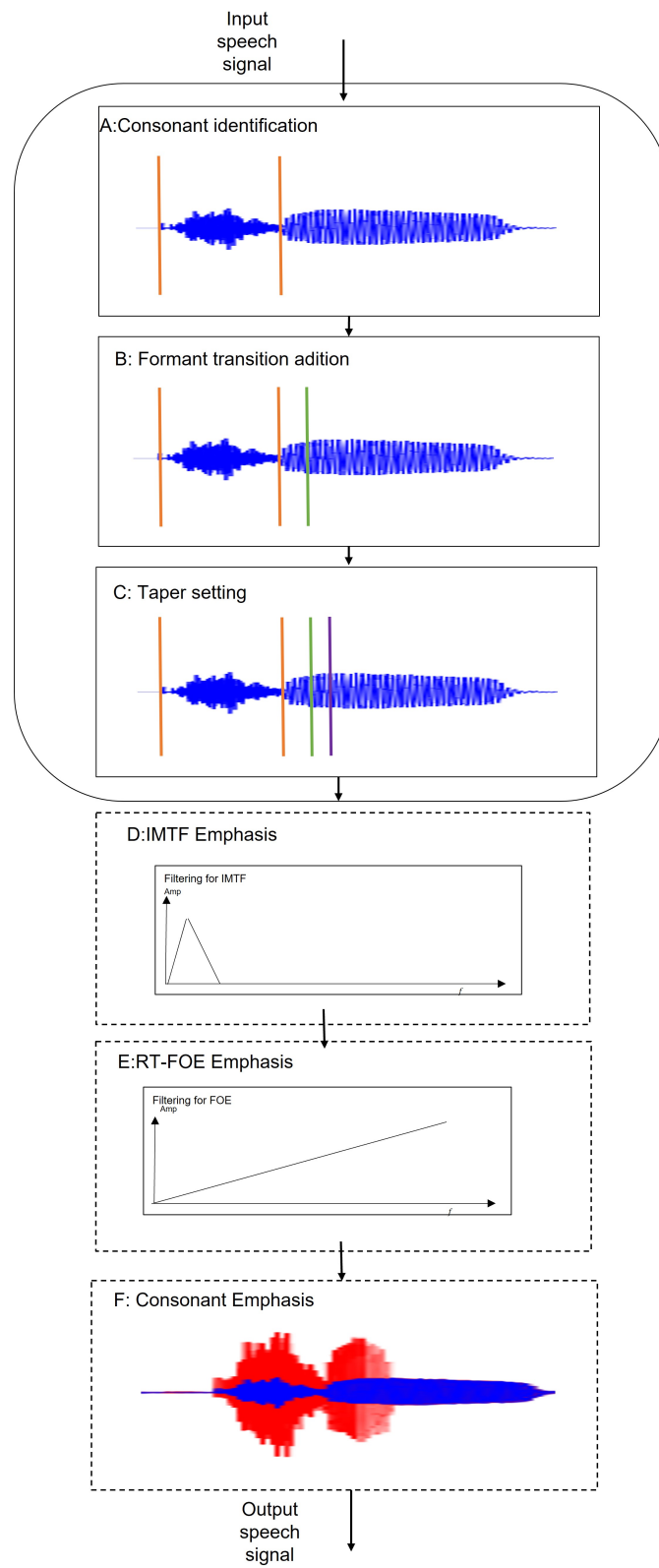


図 5.1: 三つの改善法の組み合わせの概要

5.2 音声データ

親密度別単語了解度試験用音声データセット (FW07) [43] より、実験に用いる音声データを選定した。FW07 は単語親密度の 7 段階評定値をもとに 4 つの親密度ランク (1: low (1.0 ~ 2.5), 2: lower-middle (2.5 ~ 4.0), 3: higher-middle (4.0 ~ 5.5), 4: high (5.5 ~ 7.0)) に分けられた 4 モーラ単語から構成される。親密度が高い単語は日常生活の中に使う頻度が高い単語である。音声データのサンプリング周波数は 48 kHz, 量子化ビット数は 16 bits であった。

5.3 実験条件

前述した 3 種類の組み合わせ強調処理 (RT-FOE+CE, IMTF+RT-FOE, IMTF+RT-FOE+CE) を施した音声刺激と、強調処理なし (No emphasis) の音声刺激を作成した。これら 4 種類の音声刺激には、4.1 節で述べた 4 種類の親密度ランクの単語が含まれるようにした。雑音環境を模擬するため、2 種類の雑音レベル (55 dB, 75 dB) のピンク雑音を用いた。条件数の合計は 32 (強調タイプ 4 種類 × 親密度 4 ランク × 雑音レベル 2 種類) であった。

5.4 実験参加者

実験には、正常聴力を有する 20 代の日本語母語話者 10 名 (男性 5 名, 女性 5 名) が参加した。

5.5 実験装置

実験は図 4.1 に示すように暗騒音レベル 15.2 dB の防音室で行った。音声刺激の提示には骨導トランスデューサ (株式会社テムコジャパン, KE08-01) とアンプ (audio-technica AT-HA5000) を用いた。音声をマスクするピンク雑音 [44] の提示にはスピーカ (ECLIPSE, TD508MK3) とパワーアンプ (Yamaha P4050) を用いた。スピーカは参加者の 70 cm 後方に設置した。音声・雑音刺激の提示制御は PC (LG Sharkoon, Windows 8) 上のソフトウェア (MATLAB 2014a) および A/D コンバータ (Fireface UCX) を利用して行った。参加者の正面に、反応取得のためのスクリーンとキーボードを設置した。音声提示の際の骨導トランスデューサへの印加電圧は実効値で平均 0.368 V であり、このときのラウドネスと等価となる気導聴取での音圧レベルは約 60 dB であった。

5.6 実験手続き

実験参加者には、気導によるピンク雑音を提示しながら骨導デバイスから音声刺激を提示する。実験参加者には聴き取った単語をキーボードでカタカナ入力することを求めた。設問は1条件あたり20単語、計32条件640単語とした。含まれる単語が互いに異なる単語リストを4セット用意し、全参加者が同一の単語を聴取することを避けた。ある単語について、4モラすべて正解したときにその単語を正答とした。雑音レベル、親密度ランク、強調タイプの種類（強調タイプ）ごとの単語正答率を算出し、比較評価に用いた。

5.7 実験結果

図5.2と図5.3はそれぞれ雑音レベル55 dB、75 dBの条件下での単語正答率の結果を表す。エラーバーは標準誤差を表す。図5.2と図5.3に、単語了解度試験で得られた強調タイプ・親密度ランクごとの単語正答率を示す。横軸は強調タイプに対して各親密度ランクを示し、縦軸である Word recognition rate は単語正答率を示している。

得られた単語正答率におけるANOVA4による分析を行った。強調タイプ、親密度ランクと雑音レベル3要因の分散分析の結果は、単語正答率に対して強調タイプ [$F(3, 27) = 9.46, p < 0.01$], 親密度ランク [$F(3, 27) = 77.02, p < 0.01$], 雑音レベル [$F(1, 9) = 702.95, p < 0.01$] の主効果が認められた。また、強調タイプと雑音レベルの間に交互作用が認められたが [$F(3, 27) = 4.38, p < 0.05$], 強調タイプと親密度ランクの間には交互作用は認められなかった [$F(9, 81) = 0.37, p = 0.95$]。強調タイプと雑音レベルの交互作用に対して下位検定の結果、雑音レベル75 dBの下でフィルタタイプの単純主効果が認められた [$F(3, 54) = 13.37, p < 0.01$]。雑音レベル75 dBでの結果について、Holm-Bonferroni法による多重比較検定を行った。有意差が見られた強調タイプの組合せを、図5.2と図5.3に*印 ($p < 0.05$)、**印 ($p < 0.01$) で示す。IMTF+RT-FOE, RT-FOE+CEと三つ方法の組み合わせ方法IMTF+RT-FOE+CEにおける単語正答率はNo emphasisのものよりも有意に高かったことがわかった。

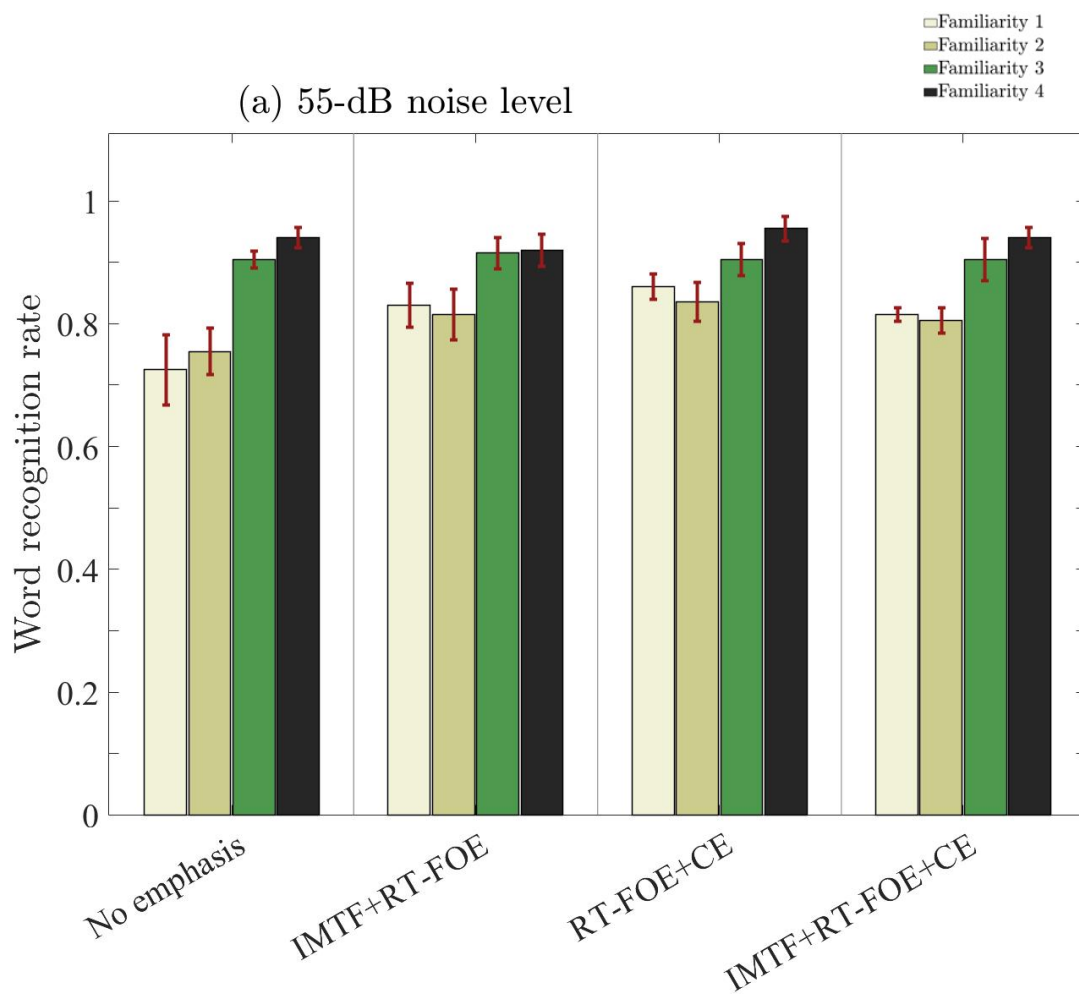


図 5.2: 総合評価：雑音レベル 55 dB 環境下親密度別単語正答率

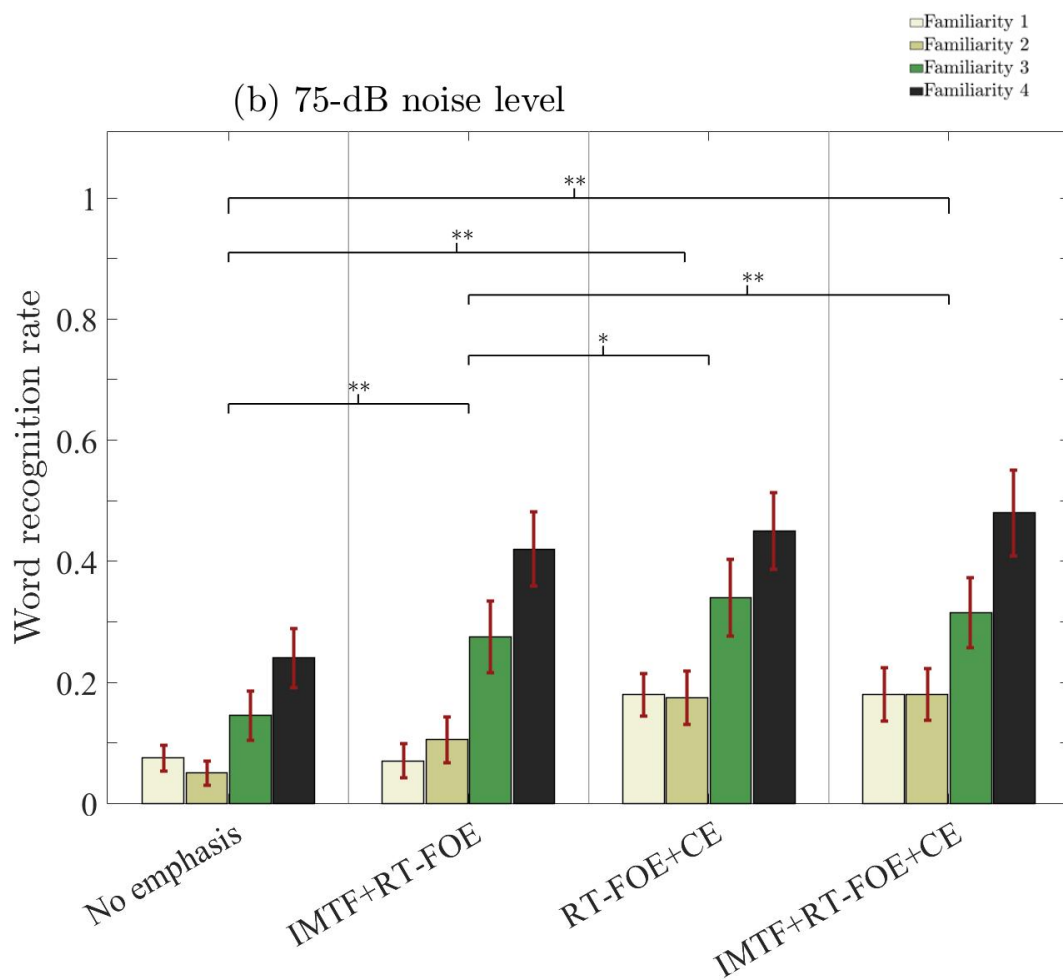


図 5.3: 総合評価：雑音レベル 75 dB 環境下親密度別単語正答率

表 5.1: 総合評価：単語正答率

音声条件	雑音レベル	親密度 1	親密度 2	親密度 3	親密度 4
No emphasis	55 dB	0.725	0.755	0.905	0.94
	75 dB	0.075	0.05	0.145	0.24
IMTF+RT-FOE	55 dB	0.83	0.815	0.915	0.92
	75 dB	0.07	0.105	0.275	0.42
RT-FOE+CE	55 dB	0.86	0.835	0.905	0.955
	75 dB	0.18	0.175	0.34	0.45
IMTF+RT-FOE+CE	55 dB	0.815	0.805	0.905	0.94
	75 dB	0.18	0.18	0.315	0.48

表 5.2: 総合評価：有意差検定結果

source	df	F	p
A: 雑音レベル	1	702.946	0.0000
error[AS]	9		
B: 親密度ランク	3	77.028	0.0000
error[BS]	27		
C: 強調タイプ	3	9.458	0.0002
error[CS]	27		
AB	3	6.249	0.0023
error[ABS]	27		
AC	3	4.380	0.0123
error[ACS]	27		
BC	9	0.365	0.0.9482
error[BCS]	81		

第6章 全体考察

6.1 子音強調法の改善効果

子音強調評価の結果から、高騒音環境下において、CE と No emphasis の間では単語正答率に有意差があったことがわかった。このことから、CE は骨導提示音声の了解度を改善することに有効であると言える。5 kHz 以上の高域成分は子音の了解度に重要であるが、骨導提示音声において高域成分が減衰してしまうことが音声の子音了解度に影響を与える。CE により高域に成分を集中する子音を選択的に強調したことで骨導提示音声の了解度が改善したと考えられる。

単語正答率に対して、強調タイプと雑音レベルの交互作用が認められたことから、CE と RT-FOE+CE は高騒音環境下でのみ了解度改善に有効であるといえる。また、単語正答率に対して強調タイプと親密度ランクの交互作用は認められなかったことから、CE と RT-FOE+CE は単語親密度に依らず有効であると考えられる。

CE と RT-FOE の間には、単語正答率に有意差が見られなかった。このことから、CE は藤田における RT-FOE と同程度の単語正答率に至るといえる。また、RT-FOE+CE では、CE および RT-FOE の単独の強調処理を施した場合に比べ、単語正答率が有意に増加したことを認められた。このことから、子音強調法においては CE と RT-FOE のハイブリットによる強調処理が、高騒音環境下での骨導提示音声の了解度改善に最も有効であったといえる。

6.2 変調伝達関数改善法の改善効果

変調伝達関数改善法評価の結果から、高騒音環境下において、IMTF と No emphasis の単語正答率の間では単語正答率に有意差がないことがわかった。IMTF+RT-FOE では RT-FOE の単独の強調処理を施した場合に比べ、単語正答率が有意に増加しなかった。このことから、本研究で提案した IMTF は骨導提示音声の了解度の改善に有効であると言えない。

単語正答率に対して強調タイプと雑音レベルの交互作用が認められたが、高騒音環境下で RT-FOE と IMTF+RT-FOE が骨導提示音声の了解度改善に有効であることを示した。RT-FOE では No emphasis および IMTF の単独の強調処理を施した場合に比べ、有意差が見られた。IMTF+RT-FOE では No emphasis および IMTF の単独の強調処理を施した場合に比べ、単語正答率が有意に増加した。IMTF+RT-FOE と RT-FOE の間には、単語正答率が有意差がなかった。このことから高騒音環境下で、本研究で提案した変調伝達関数改善法において、IMTF と RT-FOE のハイブリットによる強調処理は、藤田によって提案された RT-FOE による骨導提示音声の了解度をさらに改善するには至らなかったと考えられる。

本研究において IMTF フィルタで音声了解度に改善できなかった原因を探すため、IMTF フィルタを音声にかけた前後の位相を分析した。図 6.1 が相互相関の結果である。最大値が-25 のところにあるため、IMTF フィルタをかけた信号と原信号の間に位相差があることがわかった。位相がずれたことが音声了解度に影響を与える見込みがあると考えられる。IMTF で処理することを音声の子音に影響を与えることがあるため [40]、IMTF フィルタをかけた音声信号のパワー包絡線と原音声のパワー包絡線を比べた。子音に対して影響を見るため子音強調した音声のパワー包絡線を一緒に比べた。図 6.2 に三種類音声条件のパワー包絡線を示す。原音声と比べ、赤色で表す IMTF で処理した音声信号の子音部を低下することが分かる。このことによる骨導提示音声の子音を聞きにくくなる、骨導提示音声の了解度を改善できない原因の一つだと考えられる。さらに一つ可能的な原因は今回の IMTF フィルタでは最大 1.4 dB を増幅できるが、骨導提示音声の了解度を改善できる増幅レベルより低いと考えられる。また、今回の MTF を分析したときに、収録した音声信号の中には短い無声区間がある。660 個音声を用いて骨導音声と気導音声の MTF を求めるため、時間長を一致するようにした。そのため、ある音声信号の無声区間を一緒に含めて分析した。このことが今回の MTF を分析することに影響を与える可能性があると考えられる。この点を解決することで、IMTF を用いた了解度の改善効果が出てくる可能性があると考えている。

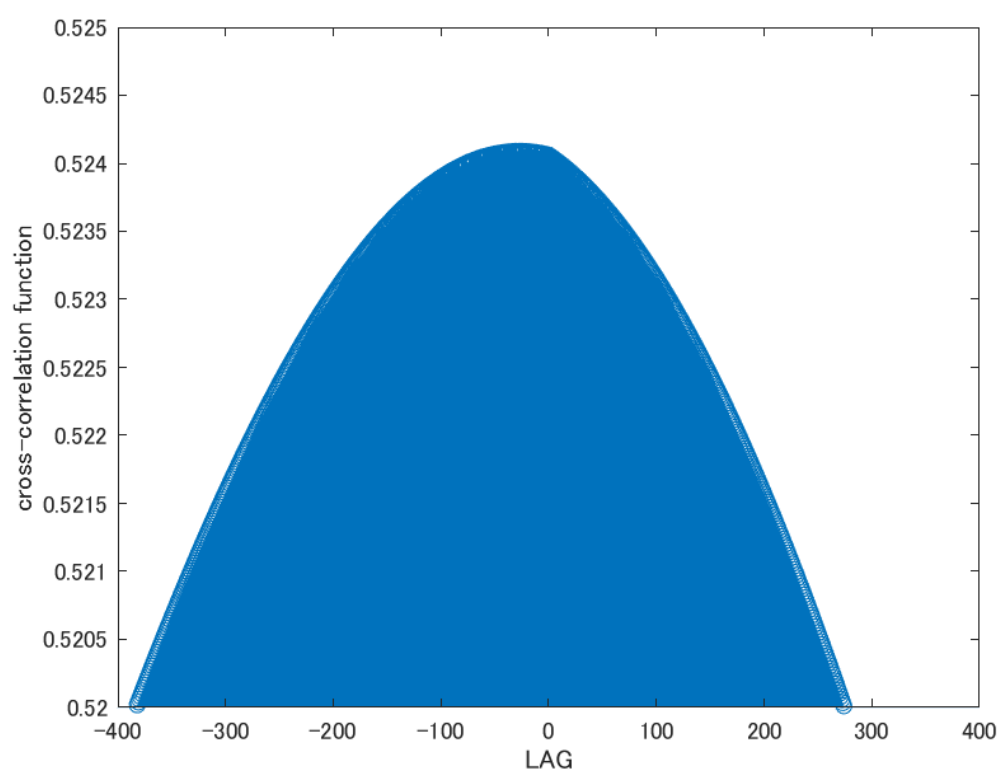


図 6.1: IMTF フィルタ処理前後の相互相関結果

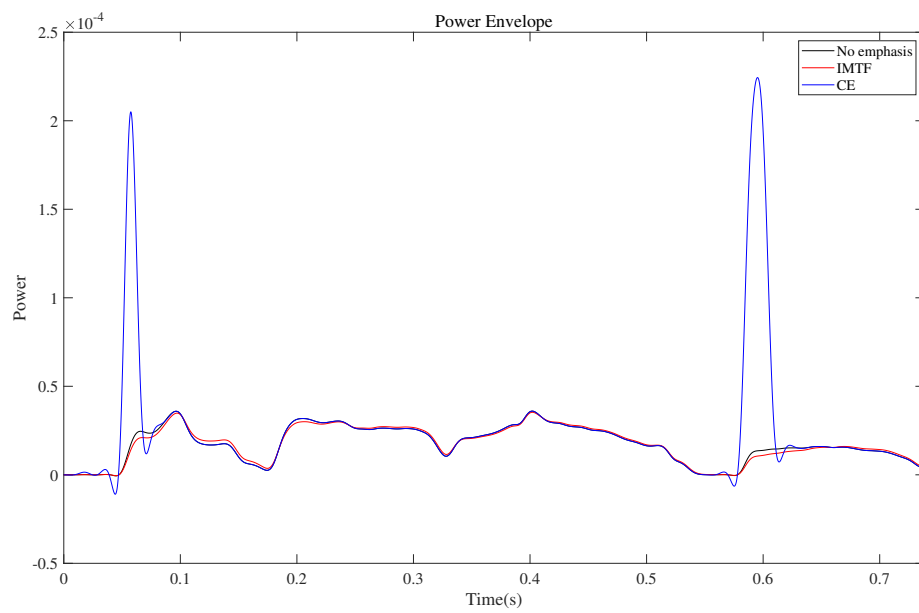


図 6.2: IMTF フィルタ処理前後のパワー包絡線と子音強調後のパワー包絡線

6.3 三つの方法の組合せによる改善効果

総合評価の結果から、高騒音環境下において、IMTF+RT-FOE、RT-FOE+CE、IMTF+RT-FOE+CE と No emphasis の間では単語正答率に有意差があることが示された。提案した3種類組み合わせ方法は、骨導提示音声の了解度に対して改善効果があると認められた。

単語正答率に対して強調タイプと雑音レベルの交互作用が認められたことから、提案した3種類組み合わせ方法は高騒音環境下でのみ了解度改善に有効であるといえる。また、単語正答率に対して強調タイプと親密度ランクの交互作用は認められなかったことから、提案した3種類組み合わせ方法は単語親密度に依らず有効であると考えられる。

3種類組み合わせ方法の間には、RT-FOE+CE では IMTF+RT-FOE の単独の強調処理を施した場合に比べ、単語正答率が有意に増加したことを認められた。IMTF+RT-FOE+CE では IMTF+RT-FOE の単独の強調処理を施した場合に比べ、単語正答率が有意に増加したことが認められた。IMTF+RT-FOE+CE と RT-FOE+CE の間には、単語正答率に有意差が見られなかった。このことから、IMTF+RT-FOE+CE は CE+RT-FOE と同程度の単語正答率に至るといえ、IMTF を IMTF を用いることは、骨導提示音声の了解度の十分な改善には至らなかったと考えられる。IMTF と比べて、CE と RT-FOE のハイブリットに基づいて、骨導提示音声の了解度をさらにの改善効果がよいと考えられる。

6.4 音声了解度に対する最適な改善方法

音声了解度試験の結果から、RT-FOE+CE と IMTF+RT-FOE の間には、有意水準が 5 % で有意差が認められた。IMTF+RT-FOE+CE では IMTF+RT-FOE の間には、有意水準が 1 % したときに有意差が認められた。このことから、RT-FOE+CE と IMTF+RT-FOE+CE は IMTF+RT-FOE より、雑音環境下で骨導提示音声の了解度に対して改善効果がよいと認められたと考えられる。RT-FOE+CE と IMTF+RT-FOE+CE の間には有意差が見られなかった。この結果から、RT-FOE+CE は IMTF+RT-FOE+CE と同程度の単語正答率に至るといえる。骨導デバイスを用いてリアルタイムコミュニケーションに役立つことから考えると、RT-FOE+CE は IMTF+RT-FOE+CE より処理速度が速いと考えため、本研究の提案法の中に、雑音環境下で骨導提示音声の了解度に対して最適な改善方法は RT-FOE+CE だと考えられる。

第7章 結論

7.1 明らかにしたこと

本研究では、音声の時間情報である振幅包絡線情報に着目し、時間・周波数軸上で効果的に強調することで、雑音環境下で骨導提示音声の了解度を改善することを目的とした。そこで、音声の子音部のみを強調する手法、また、変調伝達関数に基づく改善法で骨導提示音声の了解度を改善できるかどうかを検討した。また、本研究で提案した方法と、提案法に藤田によって提案された高域強調を組み合わせた方法で、骨導提示音声の了解度をさらに改善できるかどうかを評価した。その結果、明らかになったことを以下に示す。

1. 変調伝達関数に基づく改善法では、骨導提示音声の了解度を十分に改善することができなかった。
2. 子音強調は雑音環境下で骨導提示音声の了解度を効果的に改善できることが認められた。
3. 子音強調と高域強調のハイブリットによる方法では、高域強調よりさらに改善できることが認められた。
4. 子音強調と高域強調のハイブリットによる方法では、骨導提示音声の了解度改善効果が最も高いことがわかった。

これらの結果から、音声の振幅包絡線に着目し、子音部のみを強調することで、骨導提示音声の了解度を改善できることが明らかになった。また、音声信号の時間領域と周波数領域を同時に着目する方法は、音声了解度の改善効果が最も高いことが示された。

7.2 残された課題

本研究における今後の課題を以下に示す。

1. 本研究では、提案した骨導提示音声の了解度の改善法は話者個人性に影響を与えないように提案したが、いまの改善効果より、高騒音環境下で骨導提示音声の了解度をさらに改善する必要があると考えられる。特に、親密度の低

い音声の了解度は低いままであり、改善すべき重要な課題の一つといえる。この問題を解決できる方法の一つとして、音声の中に骨導伝達特性の影響を受けやすい高域成分を低域に集中させることが考えられる。高域成分を低域に集中させることにより、骨導伝達特性の影響を受けることを避けることができる。その方法では話者個人性も変わってしまう可能性があるが、骨導提示音声の了解度を改善できるかもしれないと考えている。

2. 本研究における提案した子音強調する方法について、高域周波数領域に集中している子音 (/p/,/t/,/k/,/s/,/z/,/j/) の正答率が改善したが、ほかの子音 (/b/,/g/,/m/,/n/,/r/,/y/,/w/,/f/) の正答率がまだ低い、改善余地があると考えられる。それは本研究の提案法で子音・母音の判定処理によって、高域周波数領域に集中している子音をしか識別できないと考える。しかし、多くの子音が母音よりパワーが小さいため、母音より雑音にマスクされやすいと考えられる。骨導提示音声の子音を全部強調すれば、雑音環境下で骨導提示音声の了解度をさらに改善できると考えている。
3. 本研究における子音強調方法の一番よい改善効果に至る増幅レベルについて検討しなかった。本研究では子音強調が骨導提示音声の了解度の改善に有効が認められたが、一番改善効果を得る増幅レベルについては明らかにしなかった。増幅レベルを明らかにすれば、骨導提示音声の了解度を改善できると考えられる。
4. 本研究で提案した変調伝達関数に基づく改善法について、無音区間がない骨導音声と気導音声のデータベースを使えば、求めた MTF の逆特性を用いて、骨導提示音声の了解度の改善効果がよくなるかもしれないと考えられる。

参考文献

- [1] 塚本昌彦, “社会の未来を拓くネットワーク情報共有空間:2. 情報共有空間のためのウェアラブルコンピューティング,” 情報処理, 128–134, 2007.
- [2] 廣瀬通孝, “ウェアラブル・コンピュータ: ウェアラブル・コンピュータの展開,” 情報処理, 40(9), 1999.
- [3] Beyond weiser, “ From ubiquitous to collective computing GD Abowd-Computer, 2016.
- [4] 福本雅朗, “実世界インタフェースの新たな展開: 8. ウェアラブルインタフェース&センシング,” 情報処理, 51(7), 812–818 2010.
- [5] 板生清, 駒澤真人, “ウェアラブルデバイスの応用と近未来の展開,” エレクトロニクス実装学会誌, 18(6), 384–389, 2015.
- [6] Stenfelt, S., “Acoustic and physiologic aspects of bone conduction hearing,” Implantable bone conduction hearing aids, 71, 10–21, 2011.
- [7] Siegert, R. and Kanderske, J., “A new semi-implantable transcutaneous bone conduction device: Clinical, surgical, and audioligic outcomes in patients with congenital ear canal atresia,” Otology & Neurotology, 34(5), 927–934, 2013.
- [8] Mudry, A., Tjellström, A., “Historical background of bone conduction hearing devices and bone conduction hearing aids,” Implantable Bone Conduction Hearing Aids, 71, 1–9, 2011.
- [9] Pellieux, L., Bouy, J. C., Blancard, C. and Guillaume, A., “Speech intelligibility with a bone vibrator,” Proc. RTO Human Factors and Medicine Panel (HFM) Symposium, 8, 1–20, 2005.
- [10] Rahman, M., Saha, A. and Shimamura, T., “Low frequency band noise suppression using bone conducted speech,” Proc. IEEE PacRim, 520–525, 2011.
- [11] 藤田周平, “骨導デバイスによる提示音声の音声了解度改善法の検討,” 北陸先端科学技術大学院大学修士論文, 2021.

- [12] 青木直史, “ゼロからはじめる音響学,” 講談社, 2014.
- [13] 日本音響学会, “音のなんでも小事典,” 講談社, 1996.
- [14] Kent, R. D., Read, C., (荒井隆行, 菅原勉 監訳), “音声の音響分析,” 海文堂出版株式会社, 1996.
- [15] Drullman R., “Temporal envelope and fine structure cues for speech intelligibility,” *The Journal of the Acoustical Society of America*, 97(1), 585–592, 1995.
- [16] Stenfelt, S., Wild, T., Hato, N. and Goode, R. L., “Factors contributing to bone conduction: the outer ear,” *J. Acoust. Soc. Am.*, 113(2), 902–913, 2003.
- [17] Munger, J. B., Thomson, S. L., “Frequency response of the skin on the head and neck during production of selected speech sounds,” *J Acoust Soc Am*, 124(6), 4001–4012, 2008.
- [18] 伊藤勲, 沖由香, 黒田英一, “骨伝導マイクイヤホン,” *テレビジョン学会誌*, 50(3), 351–357, 1996.
- [19] 前田秀彦, 西澤典子, 武市紀人, 本間明宏, 前田昌紀, 玉重詠子, 米本清, “骨固定型ピックアップから導出した直接骨導音の音響特性,” *音声言語医学*, 57(3), 294–304, 2016.
- [20] 川原竣介, 平川凜, 中藤良久, “音声スペクトルの時間変化強調による明瞭性改善手法の提案,” *産業応用工学会論文誌*, 6(1), 51–54, 2018.
- [21] Houtgast, T., Steeneken, H. J. M., “The modulation transfer function in room acoustics as a predictor of speech intelligibility,” *Acustica*, 28(1), 66–73, 1973.
- [22] Houtgast, T., Steeneken H. J. M., and Plomp R., “Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics,” *Acustica* 46(1), 60–72, 1980.
- [23] Houtgast, T., and Steeneken, H. J. M., “A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria,” *J. Acoust. Soc. Am.*, 77(3), 1069–1077, 1985.
- [24] Pek K., Arai, T., Kanedera, N., “Voice activity detection in noise using modulation spectrum of speech: Investigation of speech frequency and modulation frequency ranges,” *Acoustical Science and Technology*, 33(1), 33–44, 2012.

- [25] Koutsogiannaki, M., and Stylianou, Y., “ Modulation Enhancement of Temporal Envelopes for Increasing Speech Intelligibility in Noise,” *Interspeech*, 2508–2512, 2016.
- [26] Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F. G. , “ Clear speech perception in acoustic and electric hearing,” *The Journal of the Acoustical Society of America*, 116(4), 2374-2383, 2004.
- [27] 鶴木祐史, “変調伝達関数に基づく音声信号処理 (1) パワーエンベロープ逆フィルタ処理の原理とその応用について,” *Journal of Signal Processing* ,12(5), 339–348, 2008.
- [28] 星野聖, “スペクトルのローカル・ピーク強調による子音明瞭度の改善,” *Audiology Japan*, 37(1), 57–63, 1994.
- [29] 安武達朗, 中島祥好, “準実時間子音強調システム,” *信学技報*, 105(479), 79–84, 2005.
- [30] Schepker, H. F., Rennies, J., and Doclo, S., “Improving speech intelligibility in noise by SII-dependent preprocessing using frequency-dependent amplification and dynamic range compression,” In *INTERSPEECH*, 3577–3581, 2013.
- [31] Li, G., Hu, R., Ke, S., Zhang, R., Wang, X., and Gao, L., “Speech Intelligibility Enhancement Using Non-Parallel Speaking Style Conversion With Stargan And Dynamic Range Compression,” In *2020 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6, 2020.
- [32] 安達忠治, 小寺一興, 寺島邦男, 前川直子, 館野誠, “デジタル音声処理の子音明瞭度への影響,” *AUDIOLOGY JAPAN*, 47(3), 2004.
- [33] Shin, H. S., Kang, H.-G. and Fingscheidt, T., “ Survey of speech enhancement supported by a bone conduction microphones, *Speech Communication*,” 10, *ITG Symposium*, 2012.
- [34] 衣笠光太, “変調伝達関数に基づく骨導音声のブラインド回復法に関する研究,” 北陸先端科学技術大学院大学, 修士論文, 2009.
- [35] Fujimoto, T. and Mori, M., “Word intelligibility of bone conductive sound when wearing ear plugs,” *Proc. IEEE 4th Global Conference on Consumer Electronics (GCCE)*, 2015.
- [36] Toya, T., Birkholz, P., Unoki, M., “Estimates of Transmission Characteristics Related to Perception of Bone-Conducted Speech Using Real Utterances

- and Transcutaneous Vibration on Larynx,” *Speech and Computer. SPECOM*, 11658, 491–500, 2019.
- [37] Kato, T., and Shimamura, T., “Consonant behavior of bone conducted speech in Japanese,” 2012 International Symposium on Intelligent Signal Processing and Communications Systems. IEEE, 2012.
 - [38] Furui, S., “On the role of spectral transition for speech perception,” *The Journal of the Acoustical Society of America*. 80(4), 1016–1025, 1986.
 - [39] Kusumoto, A., Arai, T., Kinoshita, K., Hodoshima, N., and Vaughan, N., “Modulation enhancement of speech by a pre-processing algorithm for improving intelligibility in reverberant environments,” *Speech Commun.* 45 (2), 101–113, 2005.
 - [40] Ngo, T., Kubo, R., Akagi, M., “Increasing speech intelligibility and naturalness in noise based on concepts of modulation spectrum and modulation transfer function,” *Speech Communication*. 135, 11–24, 2021.
 - [41] Hermansky, H., “Modulation spectrum in speech processing,” *Signal Analysis and Prediction*. Birkhäuser, Boston, MA, 395–406, 1998.
 - [42] Friedlander, B. and Porat, B., “The modified yule-walker method of ARMA spectral estimation,” *Proc. 1984 IEEE Transactions on Aerospace Electronic Systems*, AES-20, 2, 158–173, 1984.
 - [43] 近藤公久, 坂本修一, 天野成昭, 鈴木陽一, “信号対雑音比調整による単語リスト間の単語了解度差補正：親密度別単語了解度試験用音声データセット (FW07) を用いた検証,” *日本音響学会誌*, 69(5), 224–231, 2013.
 - [44] 佐伯徹郎, 藤井健生, 山口静馬, 老松健成, “音声をマスクするための無意味定常雑音の選定,” *信学論 (A)* , J86-A(2), 187–191, 2003.

研究業績

国内発表

1. ZHU WENYU, 藤田周平, 鳥谷輝樹, 鵜木祐史, “日本語非母語話者を対象とした雑音環境下における骨導提示音声の了解度改善法の効果,” 電気・情報関係学会北陸支部連合大会, 2021.
2. ZHU WENYU, 藤田周平, 鳥谷輝樹, 小林まおり, 鵜木祐史, “骨導デバイスによる提示音声の音声了解度改善の検討,” 日本音響学会聴覚会研究会資料, 51(7), pp.491–496, 2021.
3. ZHU WENYU, 鳥谷輝樹, 中村健一, 鵜木祐史, “子音強調による骨導提示音声の了解度の改善,” 日本音響学会春季研究発表会講演資料, 2-4-6, 2022.

謝辞

本研究を進めるにあたり，多くのご指導，ご助言をくださった主指導教員である鵜木祐史教授に心から深く感謝致します。

研究に関する考察，研究の進め方，実験に関する助言などの様々なお助言をくださった赤木正人教授，木谷俊介助教，烏谷輝樹博士に深く感謝致します。

お忙しい中，実験に協力いただいた皆様に感謝致します。

本研究は，ウエストユニティス株式会社の共同研究開発の支援を受けて行われた。お礼を申し上げます。

大変忙しい中，様々なお助言をいただきました磯山さん，古田さん，郭さん，李さんに心から感謝申し上げます。同期として共に過ごした王さん，劉さん，田中さん，久保さん，市川さん，佐藤さんには研究から私生活に至るまで様々なところで助けられました。改めてお礼を申し上げます。

最後に，長きにわたる私の学生生活を支えてくださった家族の皆様に心から感謝致します。