修士論文

骨導提示音声の了解度改善のための子音強調処理の改良

WANG SICHENG

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５年３月

Abstract

The recognition of bone conduction hearing is old and was known at least in antiquity. In the 1st century, Pliny the Elder, a Roman scientist, remarked on the potential of sound conduction through solid bodies. This property enables the hearing impaired to hear through the bone medium. The first bone-anchored hearing device (BAHA) became widely commercially available in the 1980s. The BAHA system uses an Osseo integrated titanium implant to propagate sound directly to the inner ear through the skull, bypassing the impedance of the skin and subcutaneous tissues. BAHA is better than conventional bone-conduction hearing aids resulting in better sound quality. Bone conduction devices are used not only in medicine but also in military and industrial applications. In this case, wearing earplugs, can effectively reduce noise damage to the outer ear and can be used for hearing protection in the military in noisy environments. Bone conduction communication can work well both in high-noise and low-noise environments. These levels are promising for the broad implementation of bone conduction communication in industrial and military applications. And also, It is expected to maintain clear communication even in high-noise environments and improve work efficiency. It is thought that bone-conducted devices can be useful for safe and secure communication. Such as the conditions of medical, firefighting, police, and other emergencies that need to safely hear the background sound and important instructions at the same time.

It has been identified that five factors contribute to bone conduction hearing: 1) sound radiated into the external ear canal, 2) middle ear ossicle inertia, 3) inertia of the cochlear fluids, 4) compression of the cochlear walls, and 5) pressure transmission from the cerebrospinal fluid. However, different from air conduction, and bone conduction has different transmission characteristics from air conduction because the part of the sound will be absorbed by body tissues.

Therefore, using a bone conduction device has a drawback. The sound quality and clarity of speech are lower when using a bone-conduction device, compared to an air-conduction one. The commonly held explanation for this phenomenon is that, under high noise conditions, the bone-conducted sound is considered to be masked by the noise heard in the air-conducted sound. Also, the bone-conducted speech's high-frequency component is attenuated due to the nature of bone-conducted transmission.

It is believed that the attenuation of high-frequency sound in bone conduction negatively impacts the intelligibility of speech transmitted through bone conduction. Toya addressed this issue by proposing a method to enhance high-frequency sound to improve the intelligibility of speech transmitted through bone conduction in noisy environments (RT-FOE). On the other hand, Zhu investigated the

results of the experiment by Toya and found that the error rate of consonants is 5 times that of vowels. Zhu focused on the time domain instead of the method proposed by Toya, which focused only on the frequency domain. Because the low power of consonants compared to vowels suggests that consonants are easily masked in noisy environments, and also easily affected by the bone conduction high-frequency attenuated characteristic. Bone conduction transmits sound through vibrations in the skin and skull, which are also transferred to the outer and middle ears. Also, the sound is transmitted from the outer ear to the middle ear by vibrations, similar to air-conducted sound. Because consonant enhancement is effective in air-conducted sounds, it is thought to be effective in the bone conduction pathway as well. In addition, an important component of the perception of consonants is the formant transition from consonant to vowel. Therefore, Zhu proposed a method to improve the intelligibility of bone-conducted speech by emphasizing the maintenance time of consonants and the formant transitions (CE). Although the performance is at its best when both methods are used simultaneously, the improvement in speech intelligibility is affected by the accuracy of consonant detection.

Through the investigation of the consonant detection results by the CE, it is found that the accuracy in consonant detection is very low. This performance of consonant detection greatly reduces the effect of consonant emphasis. By labeling the consonant segment of the database, it is found that the CE is so poor for the detection of voiced consonant segments. This study aims to further improve the accuracy of accuracy in consonant detection, by considering the characteristics of unvoiced and voiced consonants. Concretely, consonants are divided into voiced consonants and unvoiced consonants according to their vocalization patterns. Based on the characteristics of voiced consonants and unvoiced consonants, it is designed to identify voiced consonants and unvoiced consonants through power ratio. Then, the consonant detection is judged through integrated processing. Finally, also consider that the detected consonants are the perception of consonants, which is the formant transition from consonant to vowel. The formant transitions, which are important for the perception of consonants, should be emphasized together. An improved method based on CE is proposed. (CE-IMP). The difference between CE-IMP and CE in terms of consonant emphasis is that CE does not perform taper processing on the parts before the emphasis segment, while CE-IMP performs taper processing on the parts before and after the emphasis for the naturalness of speech.

This study is based on the characteristics of the power ratio to judge the unvoiced consonants and voiced consonants. Unvoiced consonants have more power at high frequencies. Based on this feature, for unvoiced consonants, use the power

ratio of high-frequency power and overall power, and then compare the ratio with the threshold to judge unvoiced consonants. Voiced consonants have more power at low frequencies. Based on this feature, for voiced consonants, use the power ratio of low-frequency power and overall power, and then compare the ratio with the threshold to judge unvoiced consonants. For the thresholds related to unvoiced consonant detection and the boundary frequency of high-frequency, the best parameters are determined by the ROC curve of unvoiced consonant detection in the labeled database of unvoiced consonants and non-consonants. And also, For the thresholds related to voiced consonant detection and the boundary frequency of low-frequency, the best parameters are determined by the ROC curve of unvoiced consonant detection in the labeled database of voiced consonants and non-consonants.

To confirm the improvement effect of the proposed method on bone-conducted speech intelligibility, speech intelligibility tests were conducted in a noisy environment (55 dB, 75 dB). There are three test conditions, the first-order high-frequency emphasis compensates for the transfer characteristic of region temporalis vibration (RT-FOE), which was proposed by Fujita. consonant emphasis (CE) proposed by Zhu, and CE-IMP. According to the results of tests, in a noisy environment, CE-IMP has a significant difference in word correctness compared with other methods.

目 次

図 目 次

表 目 次

第1章　序論

1.1　研究背景

骨伝導聴覚は古代から知られていたと考えられており，その認識は古い歴史に遡ることができる．1 世紀に，ローマの科学者 Pliny the Elder は，固体を介した音伝導の可能性を述べていた [1]．この性質により，骨を媒介に聴覚障碍者が聴くことができる．1980 年代に埋め込み型骨導補聴器 (BAHA) が実用化された [2]．BAHA では皮膚や皮下組織のインピーダンスを回避し，頭蓋骨を通して内耳に直接音を伝播させる装置である．直接側頭骨に振動を与えるため, 皮膚の上から振動を伝導させる時と比べて良い音質の音を聴くことができる [3]．骨導提示デバイスは，医療だけでなく，軍隊，工業用途でも大いに活用されている [4]．骨を通して音を伝えることで，耳栓を付けて外耳への騒音被害を効果的に軽減できるため，雑音環境下で軍隊では聴覚保護に使用することができる [5]．骨伝導通信は，高ノイズ環境でも低ノイズ環境でもうまく機能する [6]．これらの通信レベルは，骨伝導通信を工場および軍事用途に幅広く導入するために有望である [7]．また，人々の生活に便利さ，楽しさ，安全さを提供することに役立つ [8]．さらに，骨導提示デバイスでは，医療，消防，警察などの緊急・安全を要するため，耳を塞がずに周囲環境の中にある背景音と重要な指示にある骨導音を同時に聴取できる利点がる [9]．

しかし，骨導提示デバイスで音声を聴取する際には，気導聴取時と比べて音質や音声了解度が低下することが知られている．特に高騒音環境下における骨導音声の音質劣化や骨導音声の了解度の低下が指摘されている [10]．それらの原因は高騒音環境下で，気導音で聴取した雑音が骨導提示音をマスクするためだと考えられる．また，骨導音声では気導音声と比べ高域成分が減衰することが指摘されており [11]，骨導伝達の高域減衰特性が音声の了解度に影響を与えると考えられる．

雑音環境下での音声聴取に骨導提示デバイスを応用するには，骨導音声の了解度の低下を防ぐための対策が必要である．鳥谷ら [12] は骨導伝達の高域減衰特性が主な要因と考え，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．また，朱ら [13] は子音部での異聴の起こしやすさに着目し，子音強調により，骨導提示音声の音声了解度を改善する対策法を提案した．いずれの方法，あるいは両者を組み合わせた方法により，雑音環境下で骨導提示音声の了解度を最大で約 30%改善できると認められた．しかし、音声了解度が 90%程度で十分な了解度であるとすると，高騒音環境下での骨導提示音声の了解度はそれには至っていない．

1.2 研究目的

本研究の目的は朱ら [13] の低下する子音区間検出性能を改善し，高騒音環境下での骨導提示音声の了解度を大幅に改善することである．

朱らによる手法では，子音部を選択的に強調することで，骨導提示音声の了解度をある程度改善したが，子音部の検出性能が低いため，了解度の改善効果が十分に至っていなかった．そのため，本研究では，子音が無声子音／有声子音に分けて，それぞれの音響特徴に基づき，子音部の検出性能を大幅に改善することで，朱らによる提案した手法で骨導提示音声の了解度をさらに改善する．

1.3　論文構成

本論文は，7 章で構成される．図 1.1 に本論文の構成を示す.

第 1 章

骨導提示デバイスの医療，軍隊，工業，生活への利点について述べ，骨導提示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

骨導音声の音響特性を説明する．音声了解度を改善した先行研究について述べる．骨導提示デバイスで提示した音声了解度が残っている問題点を述べる.

第 3 章

朱らによる子音強調法及び骨導提示音声の了解度の改善効果を説明し，朱らによる子音強調法の問題について述べる.

第 4 章

朱らによる子音強調法の問題に対して，改良法の全体構成と子音区間検出の改良について述べる.

第 5 章

子音区間検出法を評価し．改良法による骨導提示音声の了解度を評価する.

第 6 章

子音区間検出法と改良法による骨導提示音声の了解度の改善効果を考察する.

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.


第1章
序論
第2章
関連研究
骨導音声の音響特徴
骨導音声の了解度改善に関する研究
第3章
従来法
朱らによる子音強調法及び改善効果の説明
朱らによる子音強調法の問題
第4章
改良法
改良法の全体構成
子音区間検出の改良について
第5章
総合評価
子音区間検出法の評価
改良法による骨導提示音声の了解度の評価
第6章
全体考察
子音区間検出効果の考察
改良法による了解度の改善効果の考察
第7章
結論

図 1.1: 本論文の構成

第2章　関連研究

2.1　骨導音声の音響特性

ヒトの聴覚には２つの知覚経路として，気導音知覚と骨導音知覚である．気導音の知覚では，外耳・中耳・内耳を介して知覚される．空気の振動として外界から耳に到達した音は，外耳道を通過して鼓膜を振動させる．鼓膜の機械振動は，耳小骨という鼓膜につながった三つの骨（ツチ骨，キヌタ骨，アブミ骨）で増幅される．その後，内耳の蝸牛で機械振動が電気信号に変換される．最後に，電気信号が聴神経を介して脳に伝えられ，音として知覚される．一方，骨導音の知覚では，音が頭蓋骨，軟骨，皮膚や軟部組織，体液などを伝わって，複数の経路を通じて最終的に内耳の蝸牛に到達して知覚される [16].

骨導音知覚において、主要な伝達経路が 5 つの要因によって推測されている．1）外耳道内放射 ，2）中耳の耳小骨慣性振動（慣性骨導），3）内耳のリンパ液慣性振動，4）蝸牛壁の圧縮・伸長（圧縮骨導），5）脳脊髄液からの圧力の伝達である [17]．5 つの要因における骨導伝達の経路を図 2.1 に示す．骨を介して伝達されるため，骨導音の伝達特性は通常の音とは異なる．特に，生体組織での振動を吸収するため，高域成分に対する減衰が大きいという特性がある [18].

山田ら [19] は骨導マイクロホンによる録音の骨導音の周波数特性と気導音との比較して，1 kHz 以上の周波数帯域で骨導音の出力が約 20 dB 低下することを示した．この影響で，骨導音声の高周波数成分が十分に知覚できていないため，母音と子音の正答率が低くなる [20]．骨導提示デバイスを利用した音声知覚に関する研究があり，雑音環境下で，骨導音声が気導音よりも音声の了解度が低いことを示した [21].

骨導の伝達特性を調査した研究として鳥谷らの研究がある [11]．図 2.2 に鳥谷らによって，明らかにされた骨導の伝達特性を示す．図 2.2 の横軸は周波数（Hz），縦軸は振幅（dB）を表している．RT re AC が側頭部振動のみを考慮した場合の周波数特性で，EC re AC が側頭部振動に加えて，外耳道内放射の影響を含めて考慮した場合の周波数特性である．図 2.2 から骨導音の伝達特性として、高域成分の減衰（青の点線）が知られている．

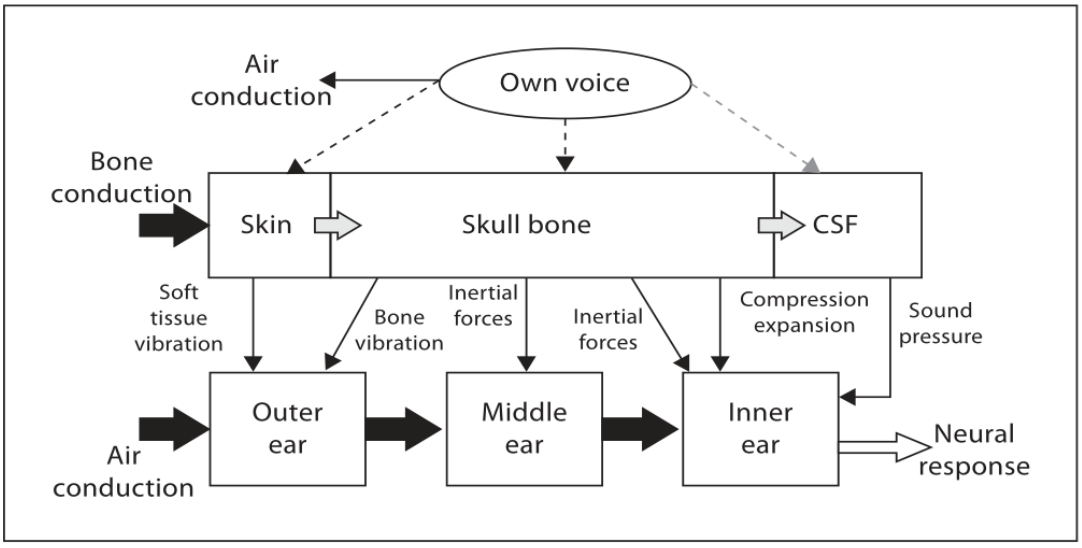

図 2.1: 気導・骨導音声の聴取経路 (文献 [16] より引用)

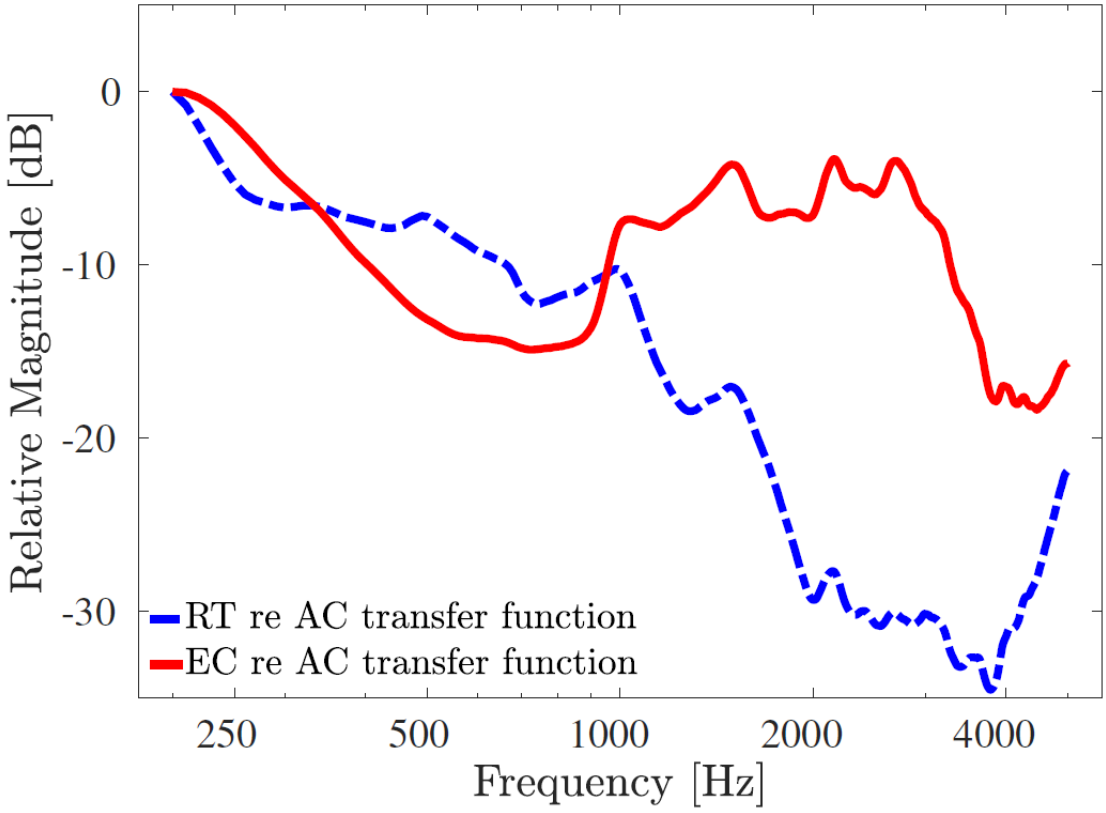

図 2.2: 骨導の伝達特性 (文献 [11] より引用)

2.2 骨導音声の了解度改善に関する研究

Fujimoto と Mori[22] は，骨導提示デバイスを用いて，雑音環境下での音声了解度を明らかにした．その結果，耳栓を装着した状態で高騒音環境下でも骨導提示音声の了解度が高いことを示した．Wang ら [21] は異なる骨導提示場所と耳の閉塞度の条件下で，骨導提示音声の了解度を調査した．その結果，骨導提示場所が関節丘（Condyle）である場合，最も優れた性能を発揮できると明らかにした．また，耳の閉塞度の高いほど，骨導提示音声の了解度が向上することを示した．

気導音だけでなく，骨導提示音を同時に聴取できるという骨導提示デバイスの利点を生かすために，本研究では，外耳道を開放した状態で，雑音環境下における音声了解度を改善できる方法を検討する．

鳥谷ら [12] は，骨導伝達の高域減衰特性に着目し，2 種類の強調処理（一次高域強調，高次高域強調）と 2 種類の骨導伝達特性（側頭部振動，外耳道内放射音）を組合せ，合計 4 種類の高域強調処理による骨導提示音声の了解度改善法を提案した．その結果，一次高域強調と側頭部振動から得られた骨導伝達特性を組合せた手法により，骨導提示音声の了解度を最も効果的に改善できることが分かった．一方で，一次高域強調処理は，骨導提示音声の時間領域全体にわたり一様に補償しているが，子音は母音より異聴を起こしやすい傾向があり，誤答率が高いことが指摘された [13]．

朱ら [13] は骨導伝達の高域減衰特性が音韻レベルの聴取にも影響すると考え，音声の時間構造にも着目し，音声の子音区間の検出・強調処理による骨導提示音声の了解度を改善する手法を提案した．単語了解度試験を実施することで，子音強調法および高域強調法による骨導提示音声の了解度改善効果を評価した．その結果，子音強調により骨導提示音声の了解度を効果的に改善できることが確認された．また，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分かった．

2.3 問題点

朱ら [13] と鳥谷ら [12] による提案した骨導提示音声の改善法に対して，いずれの改善法，あるいは両方を組合せた改善法により，提示音圧レベル 75 dB のピンク雑音環境下（以後，高騒音環境下と呼ぶ）で骨導提示音声の了解度を最大で約 30%改善できるが，静音環境での骨導提示音声の了解度と比べると十分には回復できていない．

特に，朱ら [13] の子音強調法（以後，従来法と呼ぶ）では，子音強調処理が核となることから，子音の検出精度が音声了解度の改善を左右してしまう．従来法では，一部の単語正答率の改善効果がみられないため，音声了解度を大幅に改善ができていない．これらは子音区間の誤検出・未検出が起因であると考えられる．

第3章　従来法

3.1　子音強調法

前田ら [20] は気導音声と骨導音声の間に，子音の正答率について調べた．結果から，子音の正答率が骨導音声の場合に低下することを明らかにした．朱らは子音が骨導伝達特性の影響を受け母音よりも聞き取りづらくなると考え，鳥谷ら [12] の実験結果の音韻誤答率を分析した．分析の結果から，どの雑音環境下にも，子音は母音より誤答率が高く，齟齬が生じやすい傾向が見られた．特に，75 dB 提示音圧レベル環境下において，子音の誤答率が母音よりも 5 倍高いことが指摘された．

子音強調にする多くの先行研究では，気道音声を対象とするものがあるが [23][24] が，骨導で音声を伝導するとき，皮膚と頭蓋骨の振動が外耳と中耳にも伝わる [16]．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とを踏まえると，子音強調が気導音声の了解度の改善に有効であることから，朱らは子音強調が骨導音声においても有効であると考えられている．

図 3.1 に，朱らの子音強調法 [13] のブロックダイアグラムを示す．ここでは，入力の音声信号に対し，子音区間検出法により子音区間が同定される．Kent[14] は破裂音から母音への遷移時間は 50 ms 以内に完了すると報告した．また，Furui[15] 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．これらの知見に基づき、従来法では子音の知覚に重要なフォルマント遷移部を考慮し，子音区間とその直後の 20 ms を強調処理区間とする．次に，音声中の強調部分と非強調部分の間の急激な振幅変化があるため，フォルマント遷移部分の終端から 10 ms 区間に対して，強調処理の前後で不連続が生じないようにテーパー処理を施す．最後に，子音の開始時間からテーパー処理の終端までの時間区間を強調区間とし，強調区間の振幅を +12 dB 強調する．

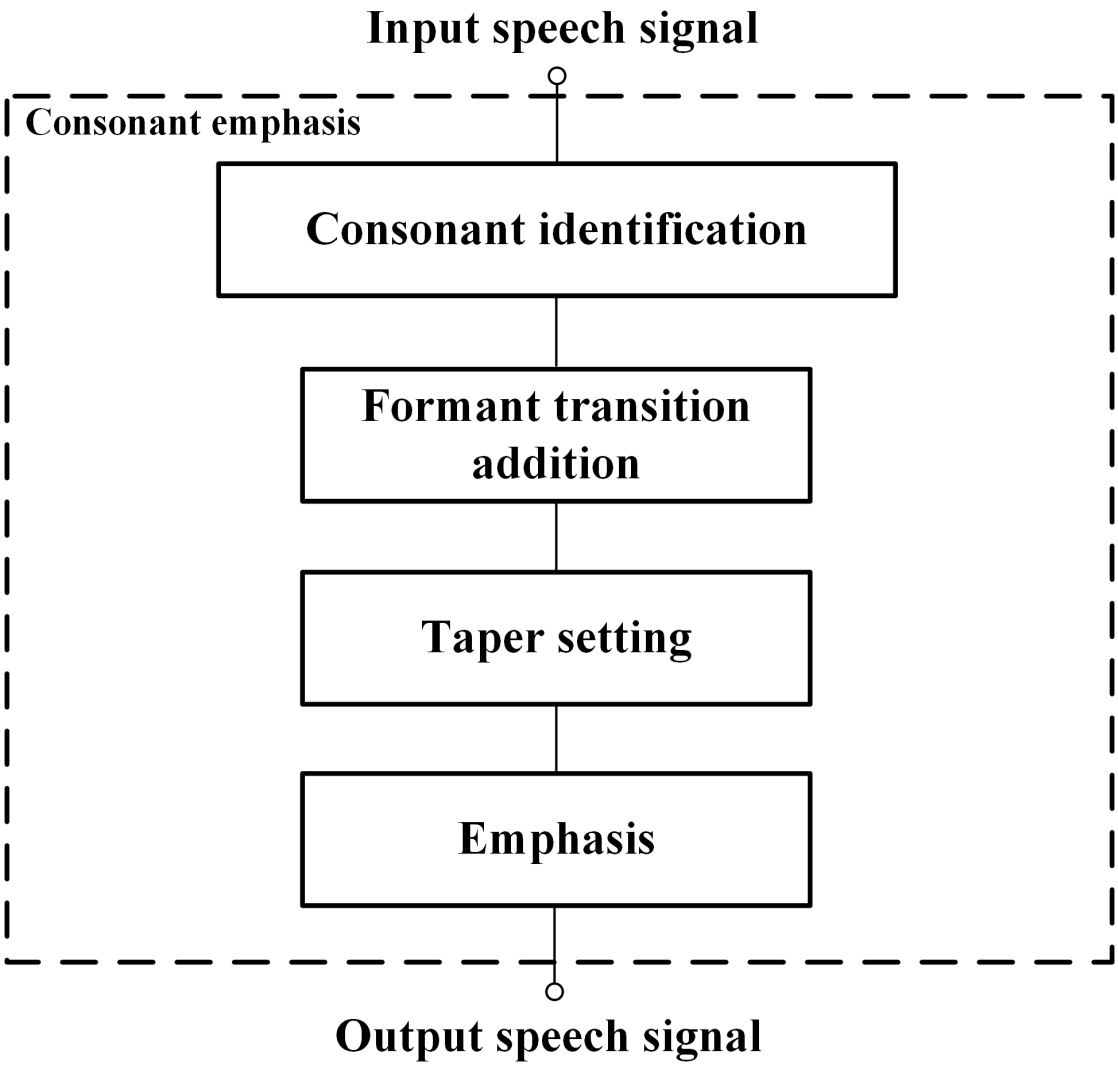

図 3.1: 従来法による子音強調処理のブロックダイアグラム

3.2　子音強調による骨導提示音声の了解度

朱ら [13] は，単語了解度試験を行うことで，子音強調を施した際の骨度音声の了解度改善効果を評価した．評価試験では，親密度別単語了解度試験用音声データセット（FW07）[25] を利用した．FW07 は，単語親密度の 7 段階評定値をもとに 4 種類の親密度ランク（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．音声データのサンプリング周波数は 48 kHz であり，量子化ビット数は 16bits であった．

実験では，強調法の条件（以後，強調タイプと呼ぶ）として，4 種類の音声刺激（No emphasis，RT-FOE，CE，CE+RT-FOE）を利用し，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．FW07 のうち音声データ 640 個を選定して利用した（各条件が 20 個音声データ）．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）に対して，ピンク雑音の気導提示と同時に音声刺激を骨導提示し，単語正答率を求めた．

図 3.2 と図 3.3 に従来法による雑音レベル 55 dB と 75 dB の単語正答率結果を示す．結果から，雑音レベル 75 dB での結果において，従来法 [13] による子音強調法（CE）と処理なし（No emphasis）の音声との間では単語正答率に有意差があることから，子音強調は音声了解度改善に有効であることが確認された．

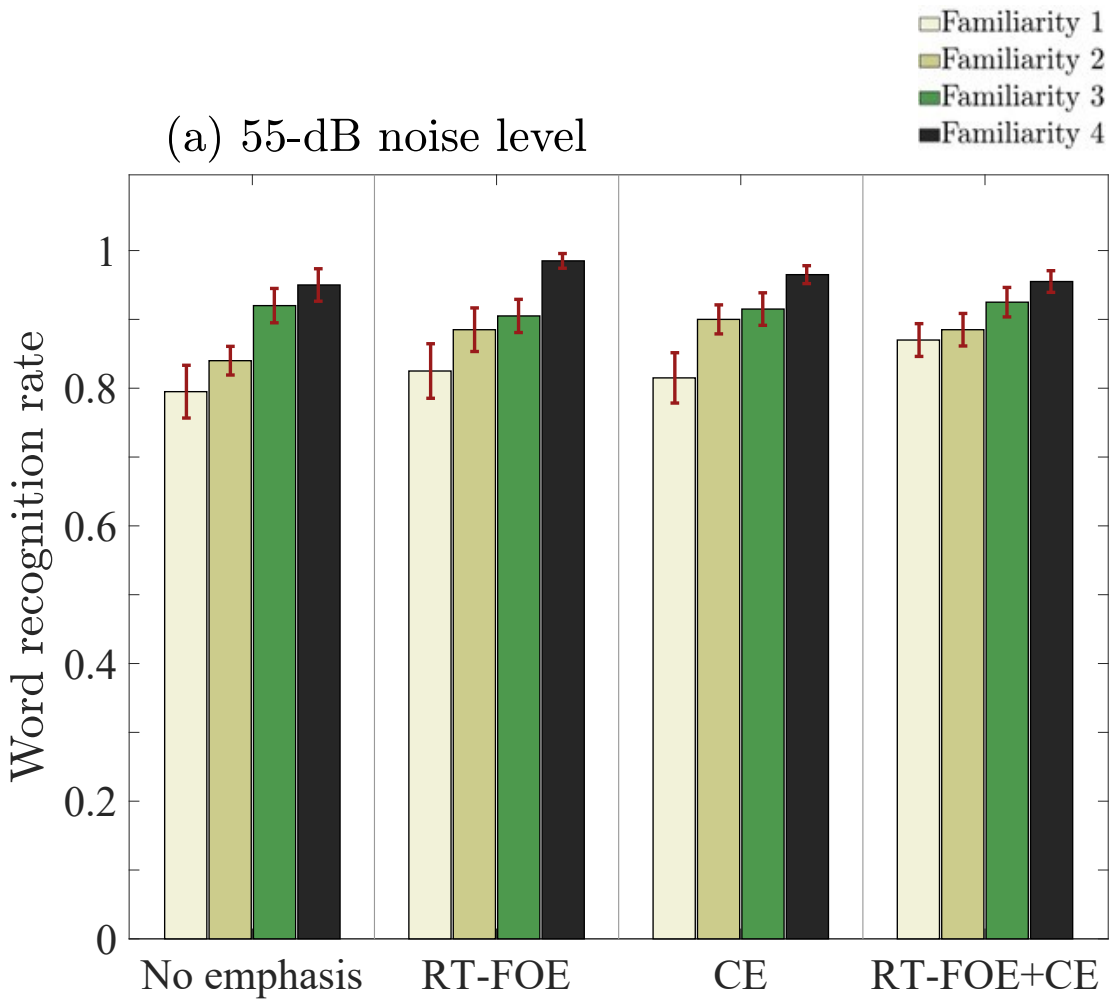

図 3.2: 従来法による雑音レベル 55 dB 環境下の強調タイプ・親密度別単語正答率

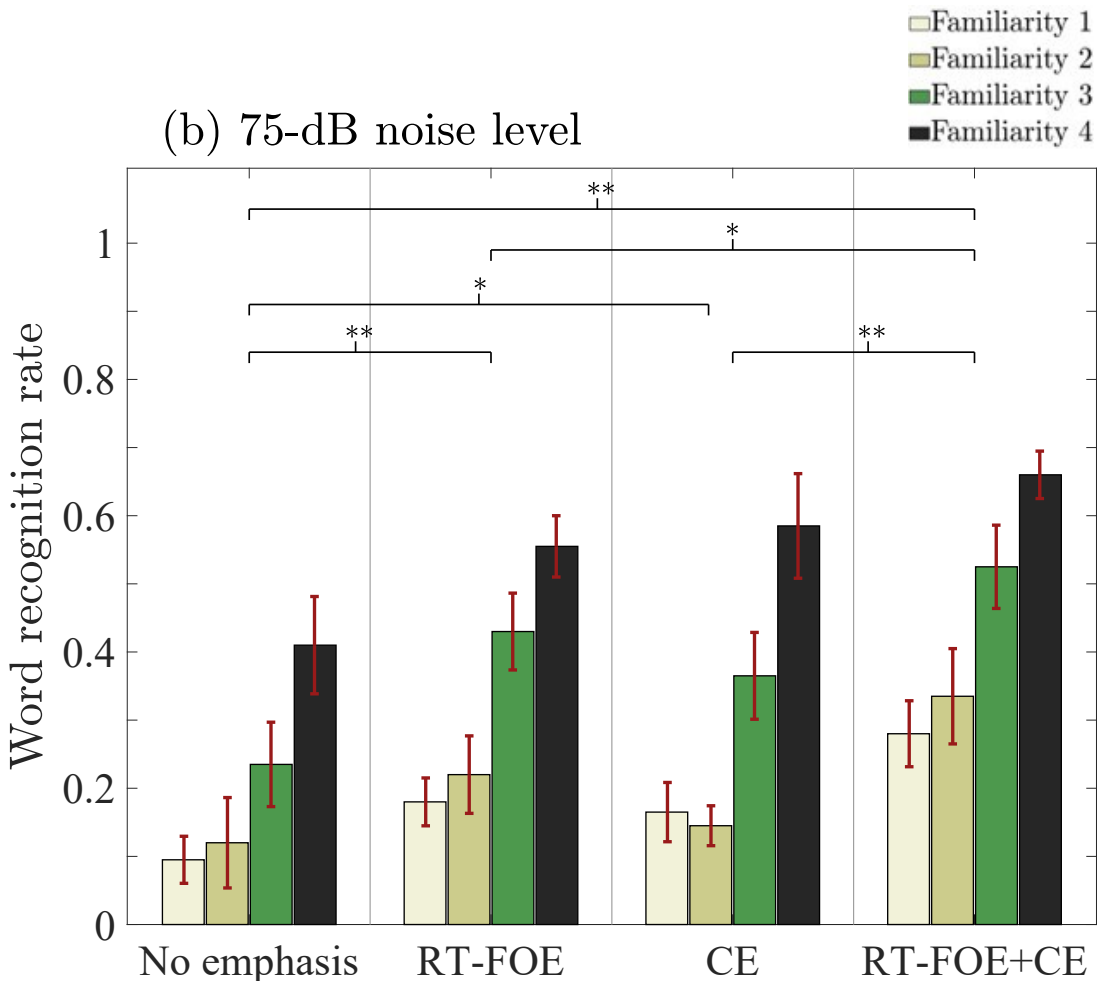

図 3.3: 従来法による雑音レベル 75 dB 環境下の強調タイプ・親密度別単語正答率

3.3　子音強調法の問題

従来法の子音区間検出処理では，主に，母音の周波数成分と比較して子音の周波数成分が高周波数帯域に集中していることに着目し，対象信号の全周波数帯域成分に対する高周波数帯域成分のパワー比を手がかりに，閾値処理を利用した子音区間検出処理が検討された [13]．図 3.4 に，従来法 [13] による子音区間検出処理のブロックダイアグラムを示す．

朱らの子音区間検出法の抱える問題を，事例をあげて説明する．まず，図 3.5 に音声信号/mihiraki/のパワースペクトログラムを示す．図中の黒の破線は各音素区間の正解を示し，5 kHz 以上の網掛け部は子音区間検出のための周波数帯域を示す．図 3.6 に音声に含まれている子音のパワースペクトルを示す．図中の黒の実線は子音区間検出のための境界周波数（5 kHz）を示す．まず，音声信号の帯域成分のパワーエンベロープが求められ，図 3.7(a) に示すようなパワー比（図中のピンクの実線）が得られる．次に，このパワー比が閾値（図中の黒の実線）を超えるかどうかで子音区間が求められる．最後に，図 3.7(b) に示すように，子音区間が得られる（図中の赤の実線）．

この結果から，多くの子音（子音/h/（図 3.6(a)）や子音/k/（図 3.6(b)）が高周波数帯域に成分をもち，このような子音に対して、従来法の子音区間検出法は正確に子音区間を検出していること確認された．その一方で，子音の周波数成分が高域側に集中していない（例えば，低域側にも相当あるような）子音/m/（図 3.6(c)）では一部のみ子音区間を検出しており，子音/r/（図 3.6(d)）ではまったく子音区間を検出していないことがわかった．

利用した音声データの子音区間検出の統計結果を図 3.8 に示す．従来法による子音区間のサンプル数が 2598068 であり，検出率が 32%しかない．この低い子音区間検出率が骨導提示音声の了解度の改善に妨げていると考えられる．

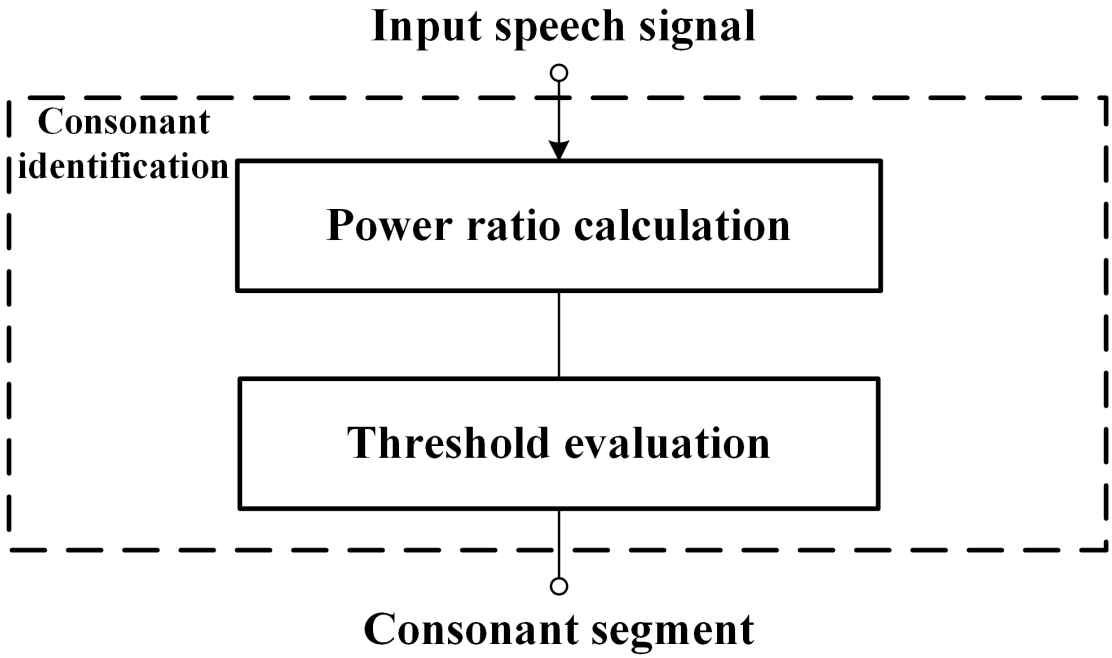

図 3.4: 従来法による子音区間検出処理のブロックダイアグラム

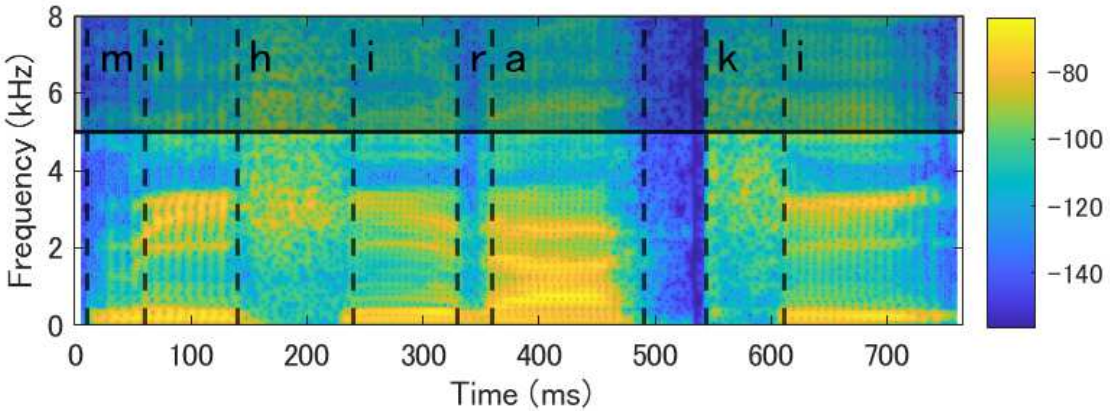

図 3.5: 音声信号/mihiraki/のパワースペクトログラム（境界周波数：5 kHz）

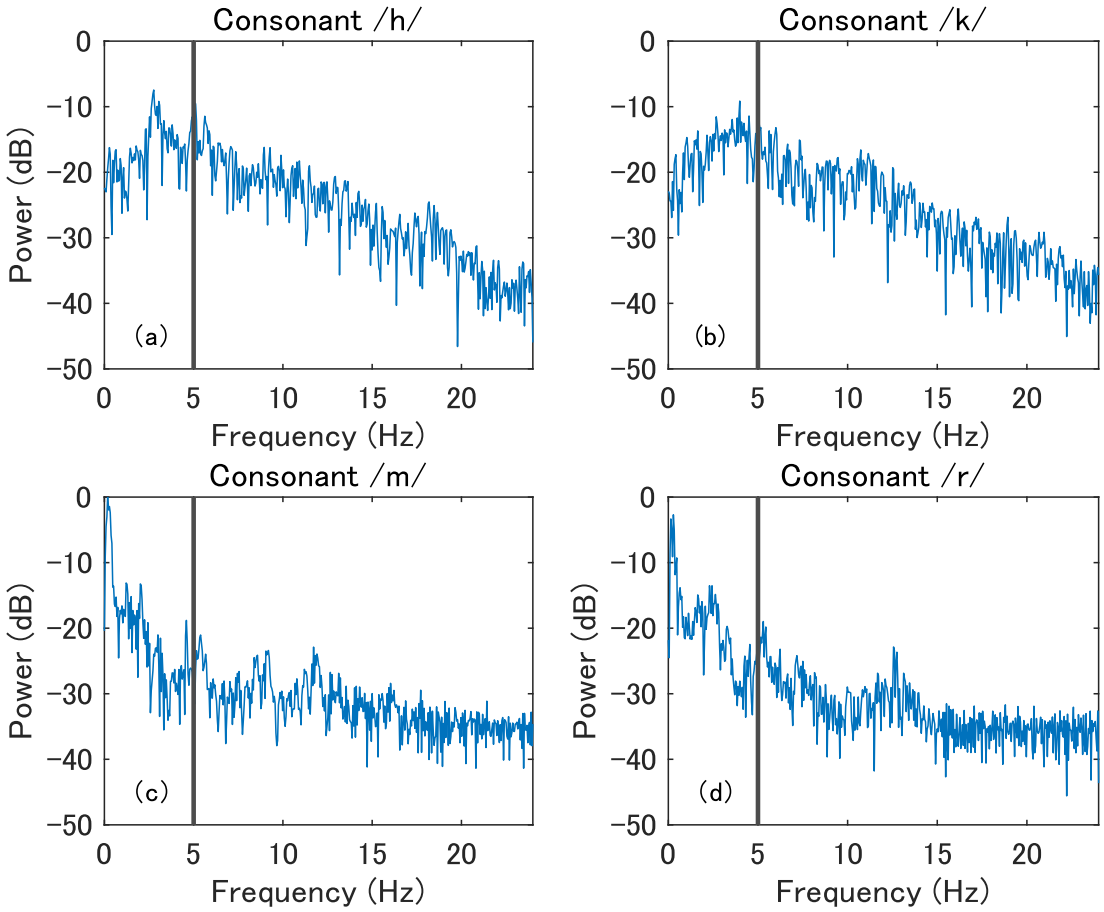

図 3.6: 子音のパワースペクトル：(a) 子音/h/，(b) 子音/k/，(c) 子音/m/，(d) 子音/r/

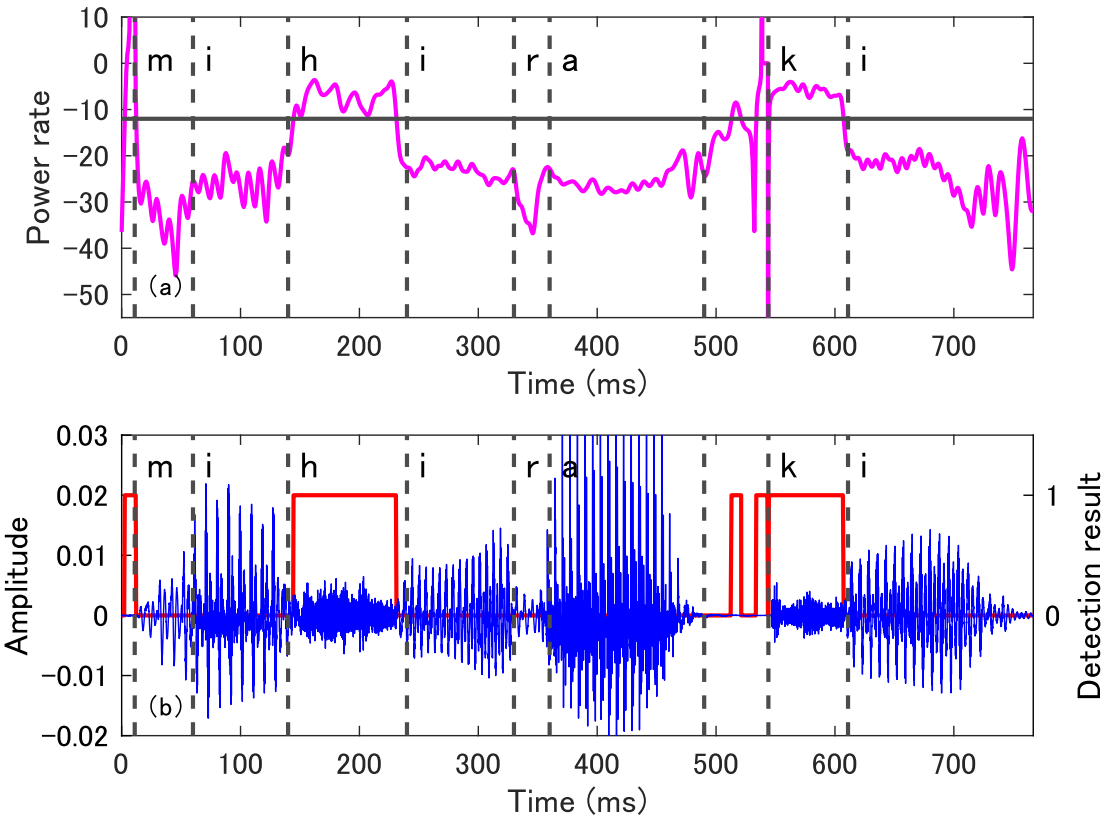

図 3.7: 従来法による子音区間検出：(a) パワー比と閾値，(b) 音声信号（青の実線）と子音区間の検出結果（赤の実線）

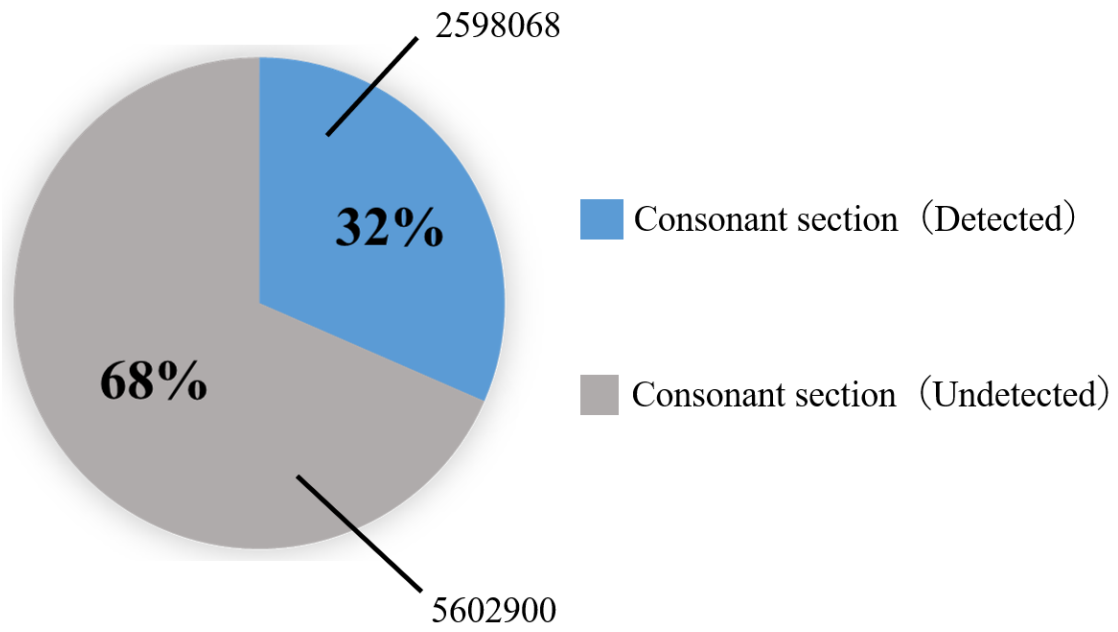

図 3.8: 従来法による子音区間の検出率

第4章　改良法

4.1　全体構成

従来法 [13] の問題を解決するためには，正確な子音区間検出が不可欠である．これらの検出精度の違い（誤検出）は，子音の発声様式（無声子音か有声子音）によって分類されたものに概ね一致した．つまり，従来法 [13] は無声子音に対して，高い精度で子音区間を検出することができるが，有声子音に対してほとんど子音区間を検出できないという子音区間検出性能の問題がある．

表 4.1 に示すように，子音は発声様式の違いにより，無声子音と有声子音に分類される [26]．無声子音と有声子音では音響的特徴が大きく異なり，無声子音（図 4.1）は高周波数側に，有声子音（図 4.2）は低周波数側に成分を多くもつことが知られている [14][27]．これらの性質を利用すれば，周波数範囲を決める境界周波数と子音を判別する閾値を適切に決めることで，無声子音／有声子音区間検出を実現できると考えられる．この考えに基づき，周波数帯域のパワー比を利用した子音区間検出法による子音強調の改良法を提案する．

この子音区間検出法では，無声子音／有声子音に対して，全周波数帯域に対する特定の周波数帯域のパワー比を閾値処理することで，無声子音／有声子音区間検出法を設計する．そして，両者の検出結果を統合処理することで，正確に子音区間検出を実現する．改良された子音強調処理（以後，改良法と呼ぶ）は，この子音区間検出法を子音強調法に組み込んで改良されたものである．改良法のブロックダイアグラムを図 4.3 に示す．

表 4.1: 発声様式に基づく子音分類

無声子音	/p/, /t/, /k/, /s/, /h/
有声子音	/b/, /d/, /g/, /m/, /n/, /z/, /r/, /j/, /w/

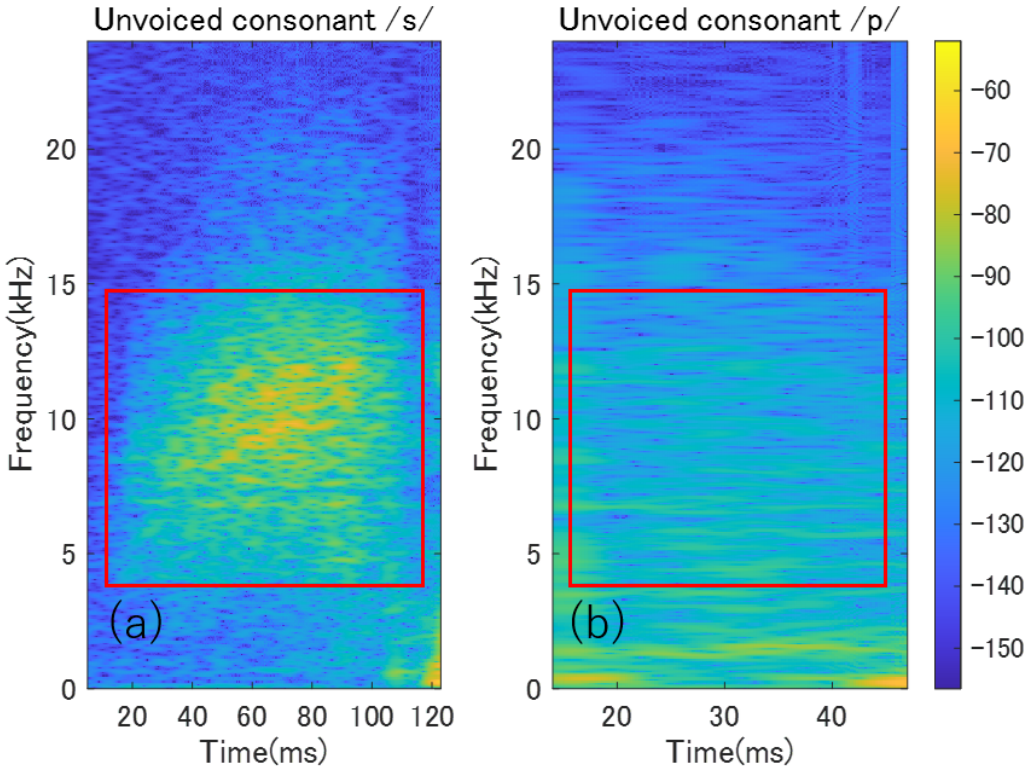

図 4.1: 無声子音のスペクトログラム：(a)/s/, (b)/p/

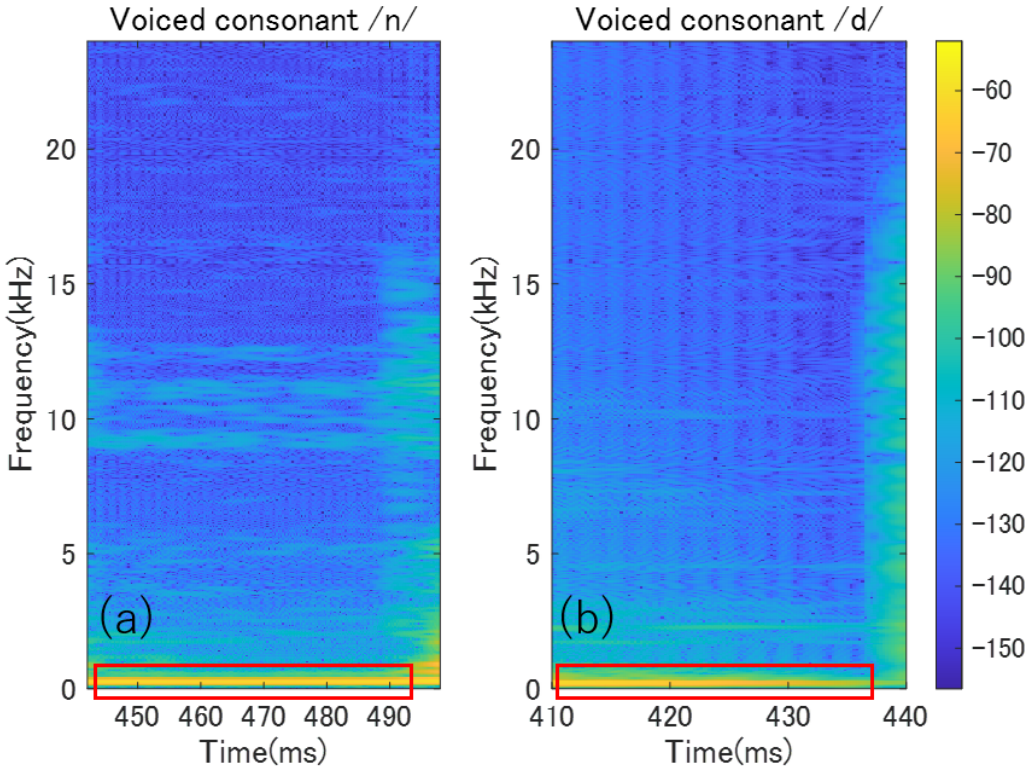

図 4.2: 有声子音のスペクトログラム：(a)/m/, (b)/d/

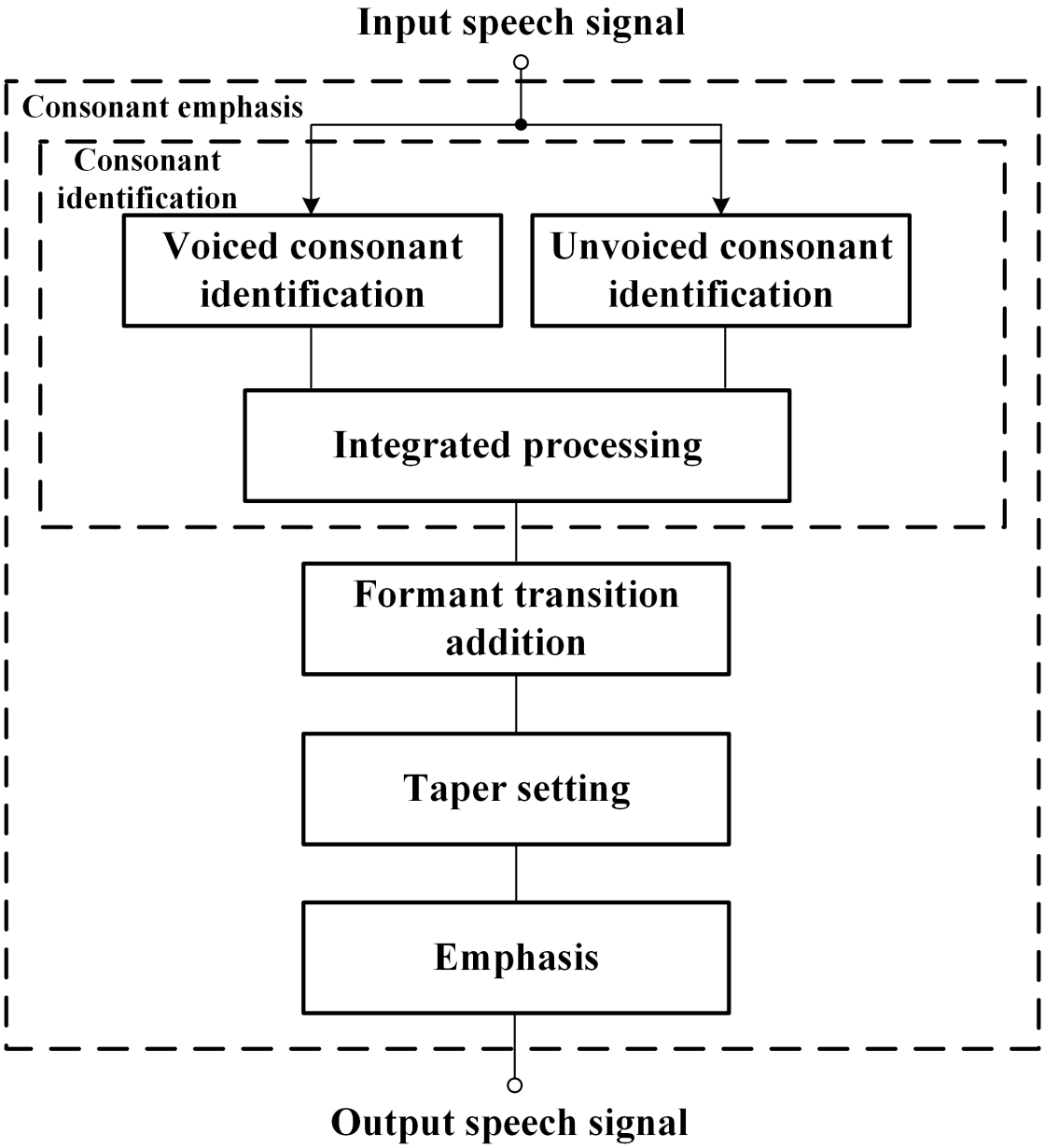

図 4.3: 改良法のブロックダイアグラム

4.2　子音区間検出の改良

従来法 [13] の子音区間検出性能が低いという問題を解決するためには，無声子音区間だけでなく有声子音区間も正確に検出する必要がある．そこで，無声子音区間ならびに有声子音区間の検出に特化した二つの方法を実現し，それらの結果を効果的に統合処理を行うことで，所望の子音区間検出法を実現できると考えられる．

改良法による子音区間検出法のブロックダイアグラムを図 4.4 に示す．子音区間検出法は，無声子音区間検出部，有声子音区間検出部，検出結果の統合処理部の三つで構成される．子音区間の検出は，次に手順で行った．

1. 音声入力：

   無声子音区間と有声子音区間の検出に特化した二つの方法があるため，音声信号は，無声子音区間検出部と有声子音区間検出部に同時に入力される．

2. 無声子音・有声子音区間の判定：

   無声子音区間は高周波数側に，有声子音区間は低周波数側に成分を多くもつ [14][27] ため，本研究では入力音声に対して，高周波数帯域が全周波数帯域のパワー比によって，無声子音区間／非無声子音区間を判定する．また，低周波数帯域が全周波数帯域のパワー比によって，有声子音区間／非有声子音区間を判定する．

3. 統合処理：

   無声子音区間と有声子音区間の検出結果から子音区間を判断するため，これらの検出結果を統合処理し，最終的に子音区間／非子音区間を判定する．

今回は従来法の子音区間検出性能が低いという問題に対して，無声子音/有声子音区間検出を判定し，それぞれの判定結果を統合処理することで，子音区間検出を改良する．

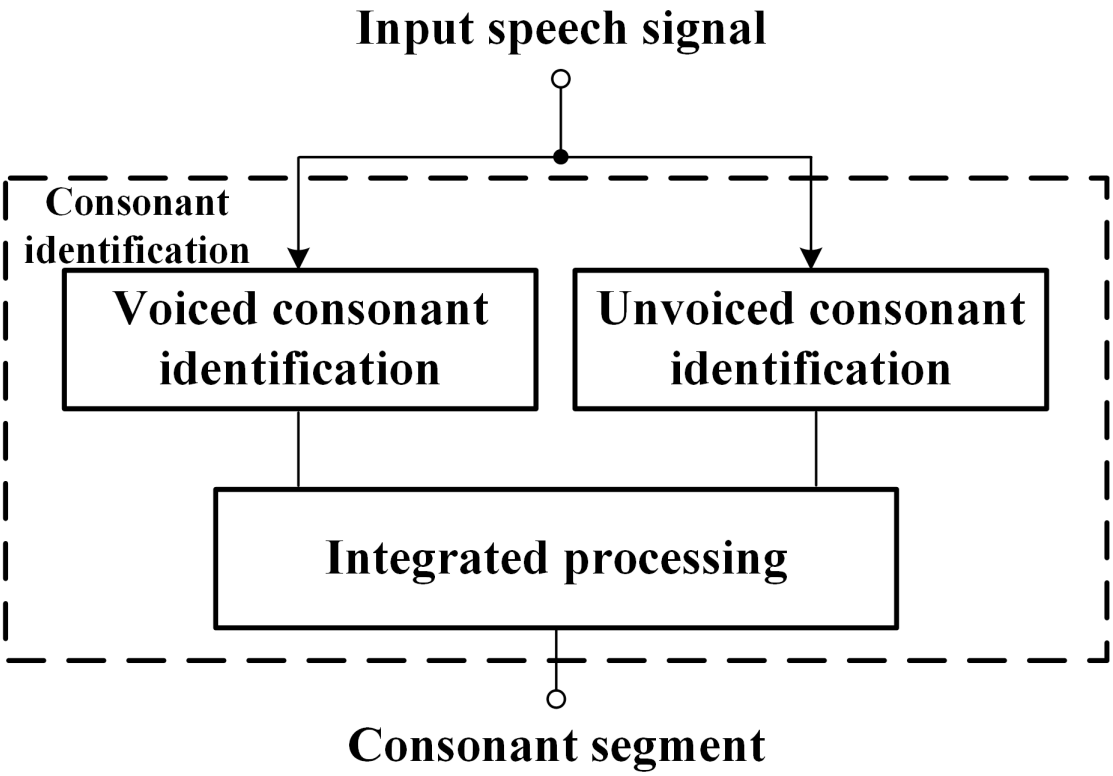

図 4.4: 改良法による子音区間検出法のブロックダイアグラム

4.2.1 無声子音区間検出

無声子音はパワースペクトルの高域成分にパワーが相対的に多く存在することから，音声信号の高周波数帯域と全周波数帯域のパワー比 $P_{\mathrm{UC}}(n)$ に対する閾値処理を行った．その結果により，無声子音区間 $D_{\mathrm{UC}}(n)$ か否かを判定する．

$$P_{\mathrm{UC}}(n) = 10\log_{10}\frac{e_{\mathrm{UC}}^2(n)}{e_{\mathrm{All}}^2(n)} \tag{4.1}$$

$$D_{\mathrm{UC}}(n) = \begin{cases} 1, & P_{\mathrm{UC}}(n) \geq \theta_{\mathrm{UC}} \\ 0, & P_{\mathrm{UC}}(n) < \theta_{\mathrm{UC}} \end{cases} \tag{4.2}$$

ここで，$e_{\mathrm{UC}}(n)$ は高周波数帯域成分で構成される信号を時間領域において求められた振幅包絡線，$e_{\mathrm{All}}(n)$ は全周波数帯域成分で構成される信号を時間領域において求められた振幅包絡線であり，θ_{UC} は無声子音区間を判別するための閾値である．振幅包絡線 $e_{\mathrm{UC}}(n)$ および $e_{\mathrm{All}}(n)$ は次式により求められる．

$$e_{\mathrm{UC}}(n) = \mathrm{LPF}\{|\mathrm{Hilbert}[\mathrm{HPF}_{\mathrm{UC}}x(n)]|\} \tag{4.3}$$

$$e_{\mathrm{All}}(n) = \mathrm{LPF}\{|\mathrm{Hilbert}[x(n)]|\} \tag{4.4}$$

ここで，LPF はカットオフ周波数を 100 Hz とした低域通過フィルタ，Hilbert はヒルベルト変換，$x(n)$ は入力音声信号を表す．高周波数帯域の下限（境界周波数）は f_{UC} kHz であり，$\mathrm{HPF}_{\mathrm{UC}}$ はカットオフ周波数を f_{UC} kHz とした高域通過フィルタである．

4.2.2 有声子音区間検出

有声子音はパワースペクトルの低域成分にパワーが相対的に集中していることから，音声信号の低周波数帯域と全周波数帯域のパワー比 $P_{\mathrm{VC}}(n)$ に対する閾値処理を行った，その結果により，有声子音区間 $D_{\mathrm{VC}}(n)$ か否かを判定する．

$$P_{\mathrm{VC}}(n) = 10\log_{10}\frac{e_{\mathrm{VC}}^2(n)}{e_{\mathrm{All}}^2(n)} \tag{4.5}$$

$$D_{\mathrm{VC}}(n) = \begin{cases} 1, & P_{\mathrm{VC}}(n) \geq \theta_{\mathrm{VC}} \\ 0, & P_{\mathrm{VC}}(n) < \theta_{\mathrm{VC}} \end{cases} \tag{4.6}$$

ここで，$e_{\mathrm{VC}}(n)$ は低周波数帯域成分で構成される信号を時間領域において求められた振幅包絡線，$e_{\mathrm{All}}(n)$ は全周波数帯域成分で構成される信号を時間領域において求められた振幅包絡線であり，θ_{VC} は有声子音区間を判別するための閾値である．振幅包絡線 $e_{\mathrm{VC}}(n)$ は次式により求められる．

$$e_{\mathrm{VC}}(n) = \mathrm{LPF}\{|\mathrm{Hilbert}[\mathrm{LPF}_{\mathrm{VC}}x(n)]|\} \tag{4.7}$$

ここで，低周波数帯域の上限（境界周波数）は f_{VC} kHz であり，$\mathrm{LPF}_{\mathrm{VC}}$ はカットオフ周波数を f_{VC} kHz とした低域通過フィルタである．

4.2.3 統合処理

統合処理では，無声子音／有声子音区間検出法によって、求められた子音区間の検出結果に対し，論理和を取ることで子音区間かどうかを決定した（$D_{\mathrm{C}}(n)$=1：子音区間，$D_{\mathrm{C}}(n)$=0：非子音区間）．ここで，次式により子音区間 $D_{\mathrm{C}}(n)$ が求められた．

$$D_{\mathrm{C}}(n) = D_{\mathrm{UC}}(n) \cup D_{\mathrm{VC}}(n) \tag{4.8}$$

4.2.4 パラメータの最適化

無声子音／有声子音区間検出法のパラメータについて，境界周波数とパワー比の閾値を最適化することにより，これらの方法の性能を最大化する．

本稿では，評価用音声データとして，3.2 節と同じ親密度別単語了解度試験用音声データセット（FW07）[25]（4 つの親密度ランクから 4 モーラ単語の音声データを 160 個ずつ選んだ合計 640 単語）を利用した．評価用音声データには，Julius 音素セグメンテーションキットを利用して正解値の音素区間をラベリングした．ここでは,「研究用日本語音声データベース」[32] を参考に，Julius による音素ラベリング結果を微修正し，子音区間を決定した．

無声子音／有声子音区間検出法の性能が最大化になるように，これらの方法のパラメータである境界周波数ならびにパワー比の閾値を，次の手順で最適化した．

まず，無声子音／有声子音区間検出法で利用する境界周波数とパワー比の閾値の範囲を設定した．無声子音区間検出法の境界周波数 f_{UC} kHz は $3, 3.5, \cdots, 6$ kHz の合計 7 個，パワー比の閾値 θ_{UC} は $0, -4, \cdots, -44$ dB の合計 12 個とした．有声子音区間検出法の境界周波数 f_{VC} kHz は $0.7, 0.8, \cdots, 1.3$ kHz の合計 7 個，パワー比の閾値 θ_{UC} は $-0.02, -0.04, \cdots, -0.3, -0.5, -1, \cdots, -6$ dB の合計 27 個とした．

次に，これらの範囲内で，任意の境界周波数，任意の閾値を利用して，無声子音区間ならびに有声子音区間検出を行った．さらに，表 4.2 と表 4.3 に示す無声／有声子音区間検出の評価尺度に基づいて，無声子音／有声子音区間の検出結果から，FP（False Positive），FN（False Negative），TP（True Positive），TN（True Negative）を求めた．その次に，TPR（True Positive Rate）と FPR（False Positive Rate）を次式で求めた．

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{4.9}$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4.10}$$

最後に，任意の境界周波数の条件ごとに，閾値を変化させながら (TPR, FPR) を結ぶことで，ROC 曲線を得た．ここで，(TPR, FPR) = (1, 0) が理想的な解で

あることから，この点と ROC 曲線上の点との距離 d_{ROC} が最小となる点が “最適解” になる [28]．d_{ROC} は次式より得られる．

$$d_{\mathrm{ROC}} = \sqrt{\mathrm{FPR}^2 + (1 - \mathrm{TPR})^2} \tag{4.11}$$

図 4.5 と図 4.6 に，無声子音／有声子音区間検出法に対する ROC 曲線を示す．また，表 4.4 に “最適解 ” における境界周波数とパワー比の閾値，d_{ROC} を示す．

表 4.2: 無声子音区間検出の評価尺度

		真の結果	
		無声子音区間	非子音区間
検出結果	無声子音区間	True Positive (TP)	False Positive (FP)
	非子音区間	False Negative (FN)	True Negative (TN)

表 4.3: 有声子音区間検出の評価尺度

		真の結果	
		有声子音区間	非子音区間
検出結果	有声子音区間	True Positive (TP)	False Positive (FP)
	非子音区間	False Negative (FN)	True Negative (TN)

表 4.4: 各区間検出法における最適パラメータおよび d_{ROC}

区間検出法	周波数 (kHz)	閾値 (dB)	d_{ROC}
無声子音	$f_{\mathrm{UC}} = 4$	$\theta_{\mathrm{UC}} = -16$	0.166
有声子音	$f_{\mathrm{VC}} = 0.9$	$\theta_{\mathrm{VC}} = -0.12$	0.349

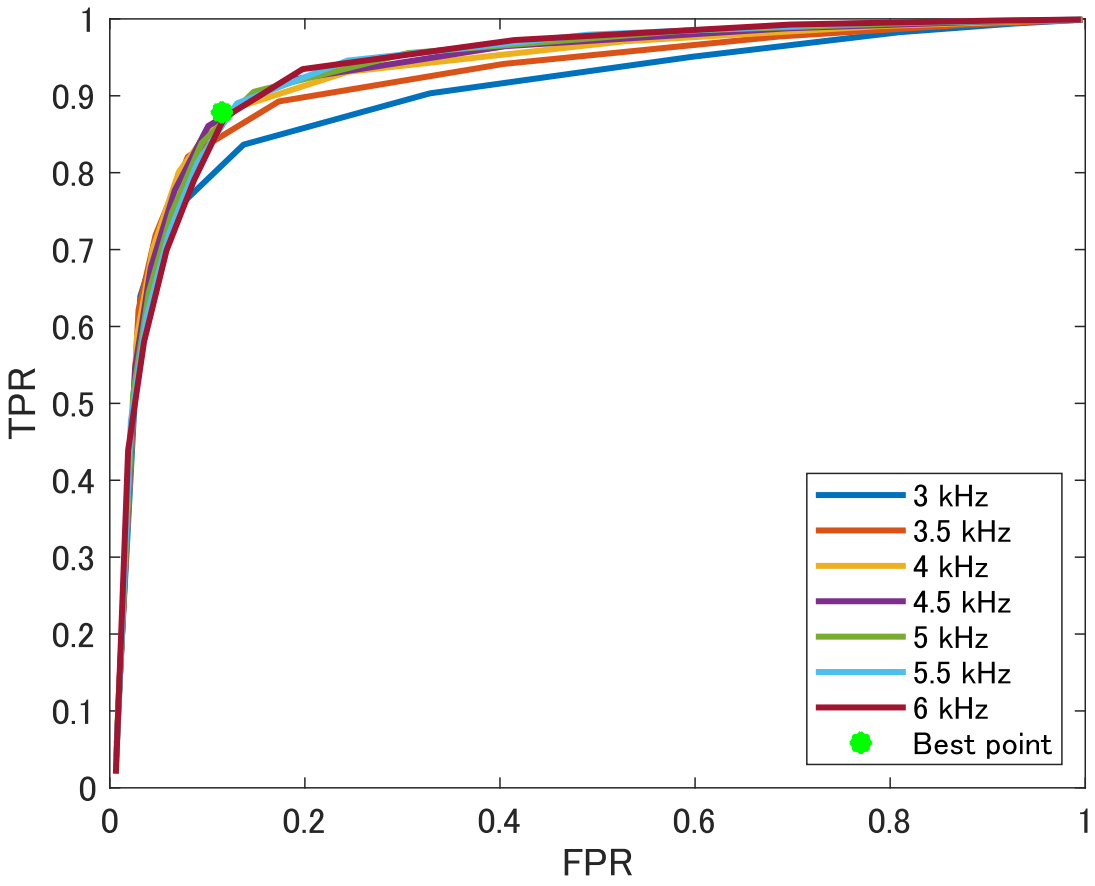

図 4.5: 無声子音区間検出に対する ROC 曲線

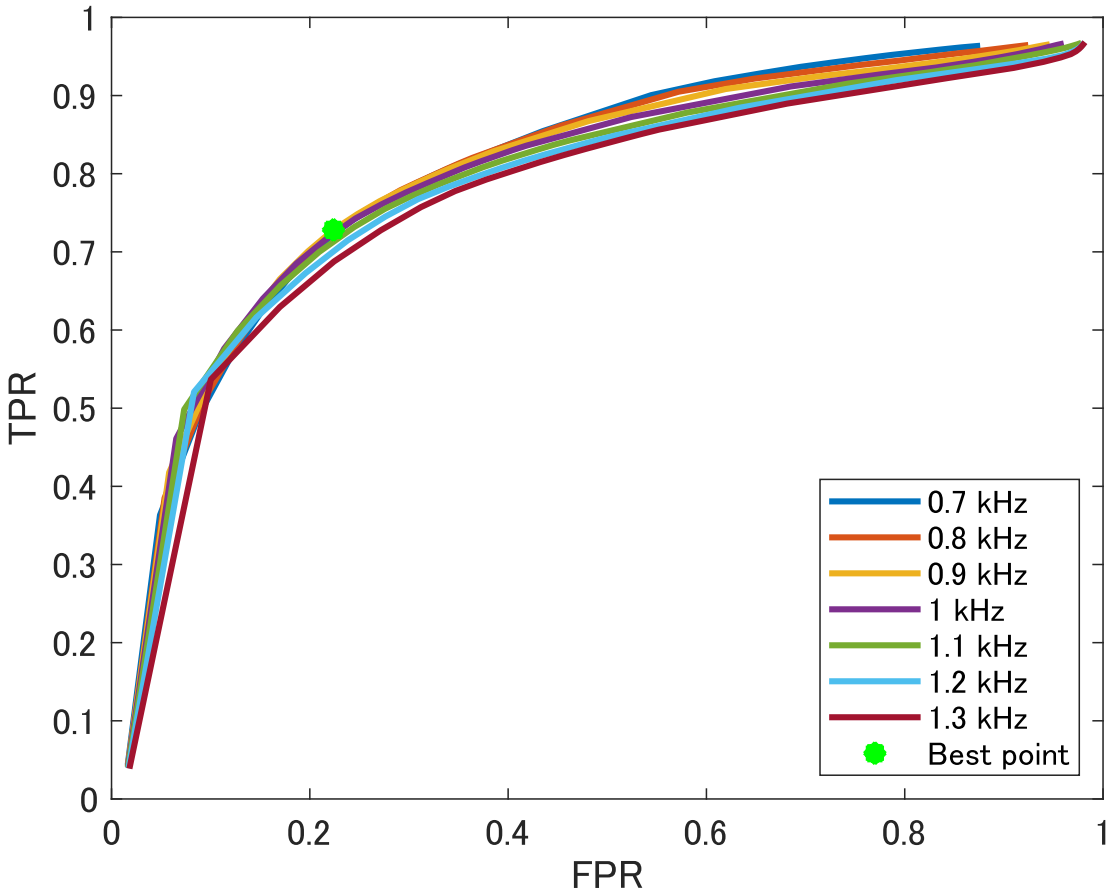

図 4.6: 有声子音区間検出に対する ROC 曲線

4.3 子音強調

朱ら [13] は，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で，骨導提示音声の了解度の改善に有効であることを明らかにした．したがって，従来法の考えに基づき，検出された子音区間に強調処理を施す．従来法と異なる点は，検出された子音区間の冒頭部分にテーパー処理が施される．(1)〜(4) の処理手順に従って，強調処理が施される．本研究の子音強調処理の概略を図 4.7 に示す．

1. 子音区間検出：

   本研究が提案した子音区間検出法の結果により，子音区間を判定する．

2. 子音の強調処理区間の拡張：

   子音の知覚に対して重要なフォルマント遷移部分を強調するために，(1) で検出された子音区間の終了時点から 20 ms の時間区間を併せて強調する．

3. テーパー処理：

   強調の有無で急激な振幅変化が生じることを避けるため，強調処理の前後区間 10 ms にわたり，振幅を緩やかに減衰させるような cos ランプ関数の荷重を提示音声に施す．

4. 強調処理：

   子音の開始時間からテーパー処理の終端まで（子音区間+30 ms），子音強調区間とし，振幅を +12 dB 増幅させる．

音声/mihiraki/の音声波形を図 4.8(a)，子音区間の検出結果を図 4.8(b)，子音区間の強調結果を図 4.8(c) に示す．図中の青の実線が音声信号であり，赤の実線が強調する部分である．子音区間の正確な検出によって，音声に含まれる子音部をすべて強調できたことがわかった．

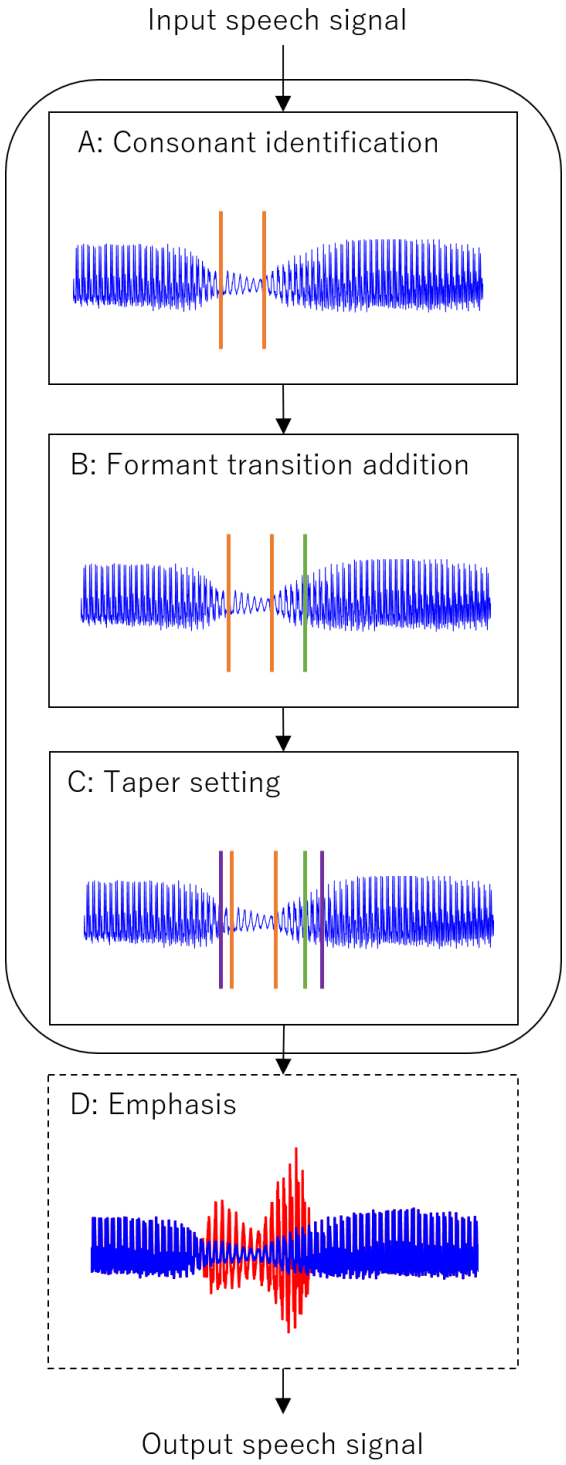

図 4.7: 子音強調処理の概略

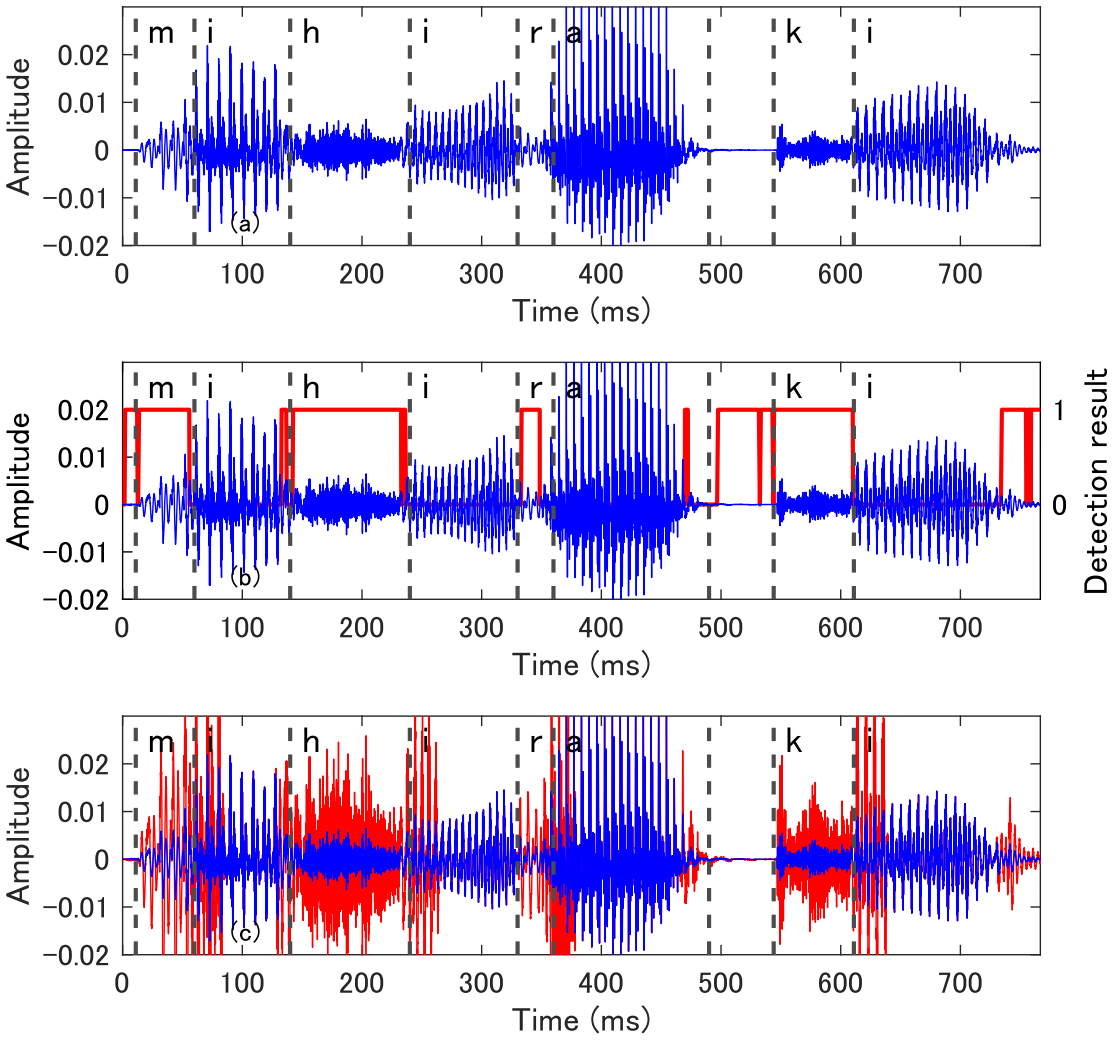

図 4.8: 子音強調処理の例：(a) 音声信号（青の実線），(b) 子音区間検出結果（赤の実線），(c) 強調処理前の音声信号（青の実線）と強調処理後の音声信号（赤の実線）

第5章　総合評価

5.1　子音区間検出法の評価

総合評価では，3.2節と同じ親密度別単語了解度試験用音声データセット（FW07）[23] の640個単語を総合評価として利用した．比較対象は，従来法，無声子音区間検出，有声子音区間検出，子音区間検出法の四つの子音区間検出法である．ここで，無声子音／有声子音区間検出ならびに子音区間検出法は，表4.4で示したパラメータの最適値を利用した．評価指標はF-scoreとd_{ROC}を用いた．F-scoreは0〜1，d_{ROC}は0〜$\sqrt{2}$の範囲を取り，F-scoreは1に近いほど，d_{ROC}は0に近いほど，子音区間検出の性能が高いことを示す．表5.1に示す，子音区間検出の評価尺度に基づいて，子音区間の検出結果から，FP，FN，TP，TNを求めた．その次に，PrecisionとRecallを求めた．最後に，F-scoreを次式で求めた．

$$\text{F-score} = \frac{2}{1/\text{Precision} + 1/\text{Recall}} \tag{5.1}$$

ただし，

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5.2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5.3}$$

表5.2に，子音区間検出法の比較評価の結果を示す．F-scoreに注目すると，従来法では0.438，無声子音区間検出法のみの場合では0.471，有声子音区間検出法のみの場合では0.498となり，無声子音／有声子音区間検出法をそれぞれ単独で用いた場合では，従来法と同程度の性能を持っていることが分かった．これに対し，子音区間検出法のF-scoreは0.678であり，従来法と比較して0.24改善された．また，子音区間検出法のd_{ROC}は0.356であり，従来法と比較して0.33改善された．

ROC曲線により無声子音／有声子音区間の検出性能が最大となるパラメータを利用して，無声子音／有声子音区間を検出した．図5.1に示す4 kHz以上の網掛け部は，無声子音区間検出のための周波数帯域を表す．図5.2に，音声に含まれている無声子音のパワースペクトルを示す．これらの図中の赤の実線は，子音区間検出のための境界周波数（4 kHz）を示す．図5.3(a)に示すようにパワー比（図中のピンクの実線）が得られ，図5.3(b)に示すように，このパワー比が閾値（−16 dB）を超えるかどうかで無声子音区間が判定される．図5.4に示す0.9 kHz以下の網

掛け部は，有声子音区間検出のための周波数帯域を表す．図 5.5 に，音声に含まれている有声子音のパワースペクトルを示す．これらの図中の赤の実線は子音区間検出のための境界周波数（0.9 kHz）を示す．図 5.6(a) に示すようにパワー比（図中のピンクの実線）が得られ，図 5.6(b) に示すようにのパワー比が閾値（−0.12 dB）を超えるかどうかで有声子音区間が判定される．

従来法（図 3.7(b)）と改良した子音区間検出法（図 5.7(d)）の子音区間検出結果を比較すると，無声子音/h/と/k/の区間について，改良した子音区間検出法の精度が高いことが分かった．加えて，従来法では有声子音/m/と/r/の区間を検出していないのに対し，改良した子音区間検出法では検出していることがわかった．

子音区間検出の統計結果を図 5.8 に示す．従来法の子音区間の検出率（図 3.8）と比較すると，サンプル数が 7048149 であり，検出率が 86%に改善できた．

表 5.1: 子音区間検出の評価尺度

		真の結果	
		子音 区間	非子音 区間
検出 結果	子音 区間	True Positive (TP)	False Positive (FP)
	非子音 区間	False Negative (FN)	True Negative (TN)

表 5.2: 子音区間検出法の比較評価

	Precision	Recall	F-score	FPR	TPR	d_{ROC}
従来法 [13]	0.708	0.317	0.438	0.063	0.317	0.686
無声子音区間検出法	0.618	0.381	0.471	0.115	0.381	0.630
有声子音区間検出法	0.513	0.483	0.498	0.223	0.483	0.563
子音区間検出法	**0.560**	**0.859**	**0.678**	**0.328**	**0.859**	**0.356**

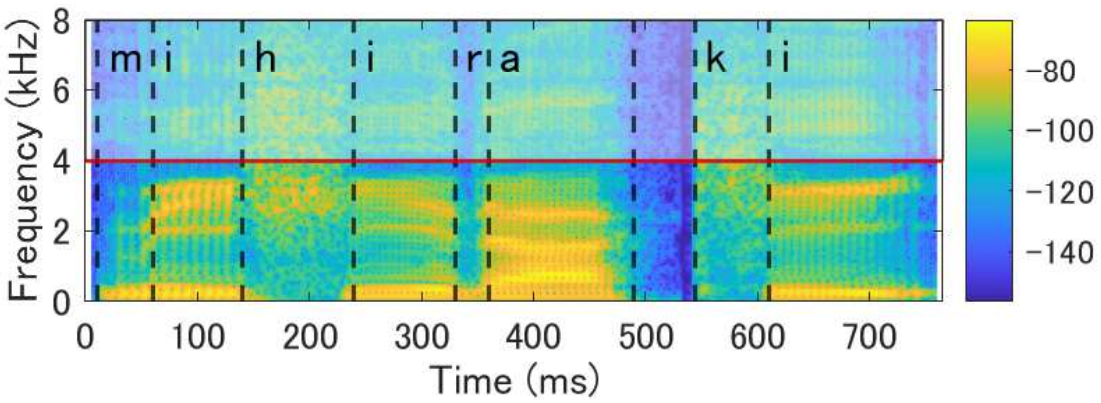

図 5.1: 音声信号/mihiraki/のパワースペクトログラム（境界周波数：4 kHz）

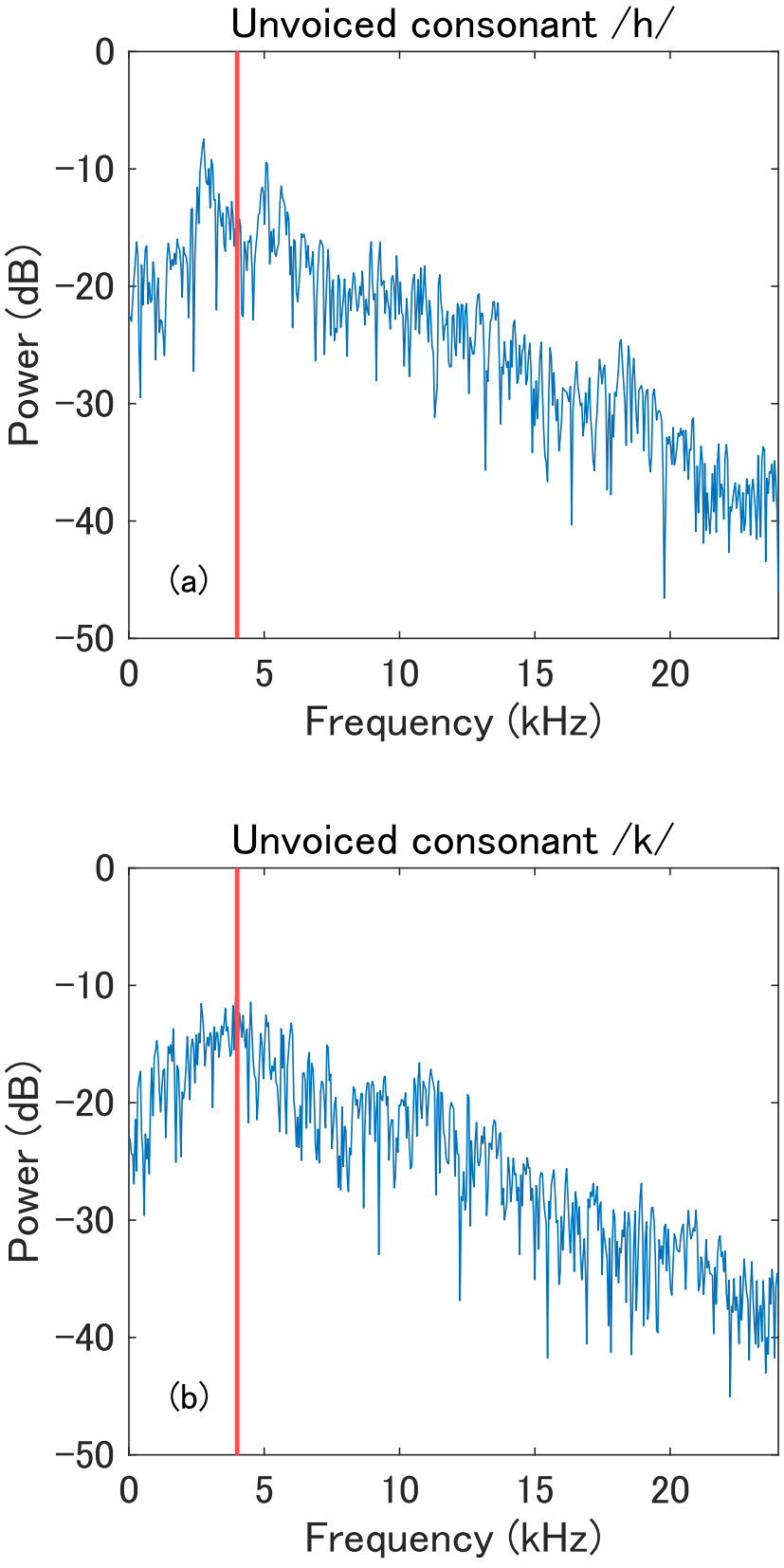

図 5.2: 有声子音のパワースペクトル：(a) 有声子音/h/，(b) 有声子音/k/

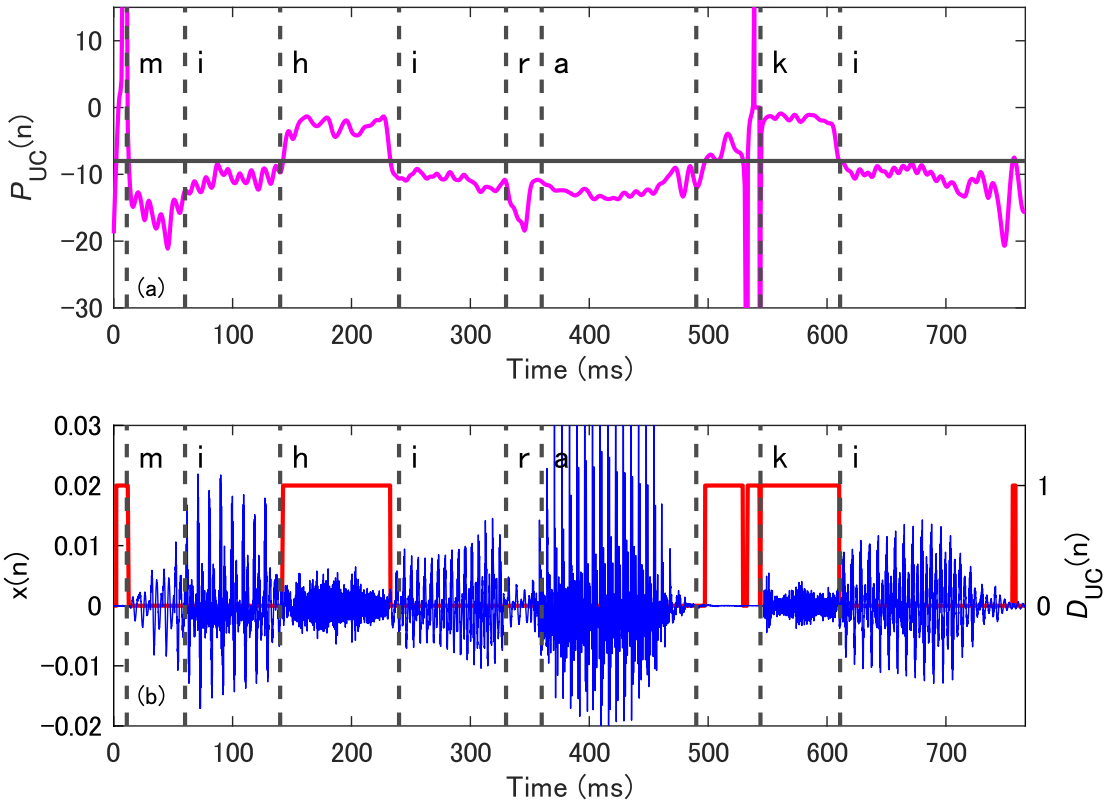

図 5.3: 無声子音区間検出による：(a) パワー比 $P_{\mathrm{VC}}(n)$ と閾値 θ_{VC}，(b) 音声信号（青の実線）と子音区間の検出結果 $D_{\mathrm{VC}}(n)$（赤の実線）

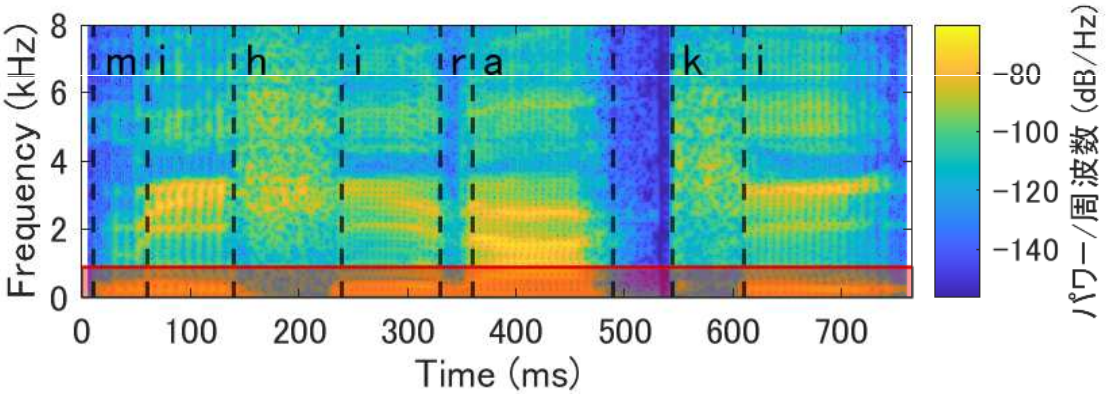

図 5.4: 音声信号/mihiraki/のパワースペクトログラム（境界周波数：0.9 kHz）

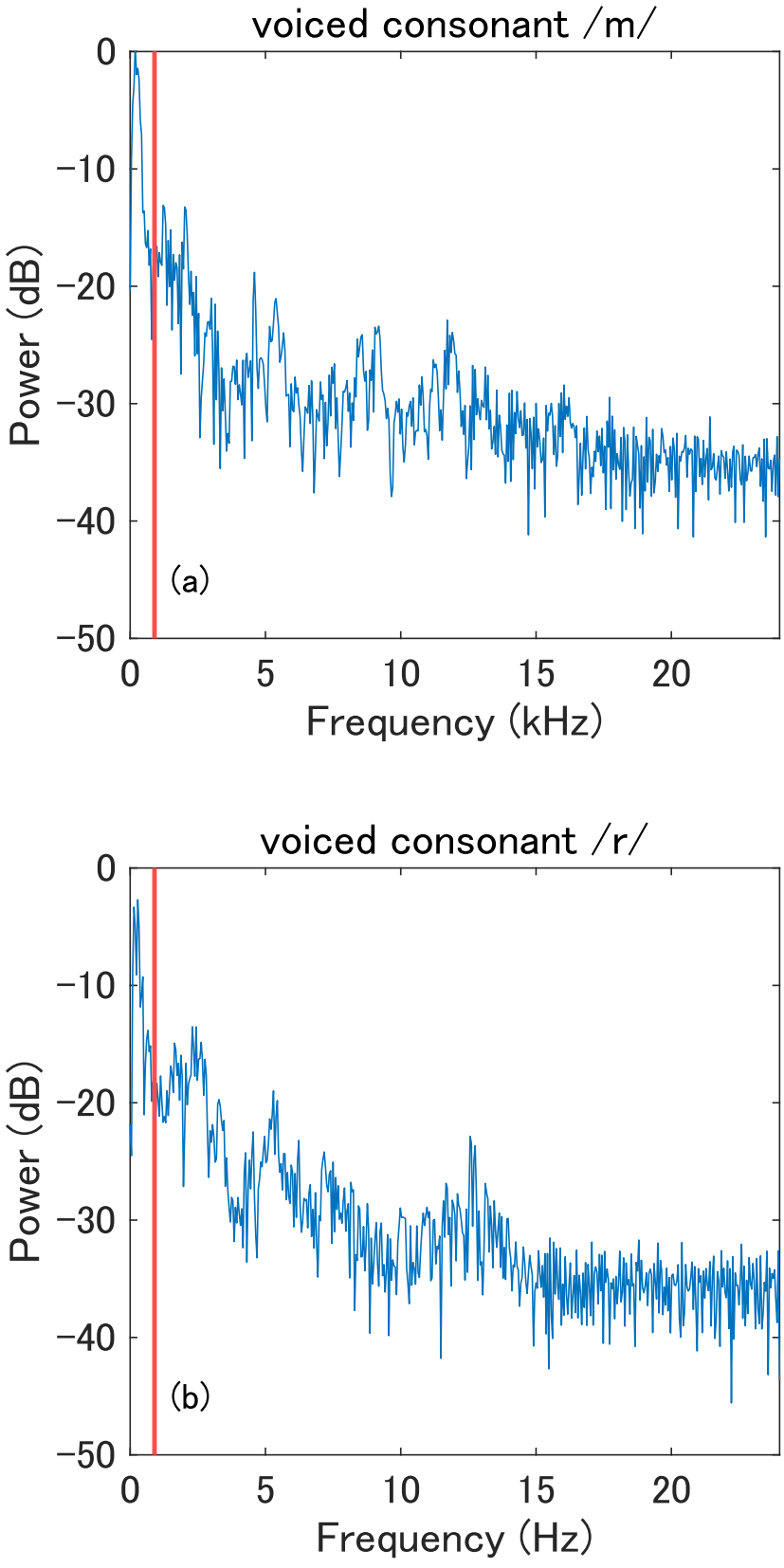

図 5.5: 有声子音のパワースペクトル：(a) 有声子音/m/，(b) 有声子音/r/

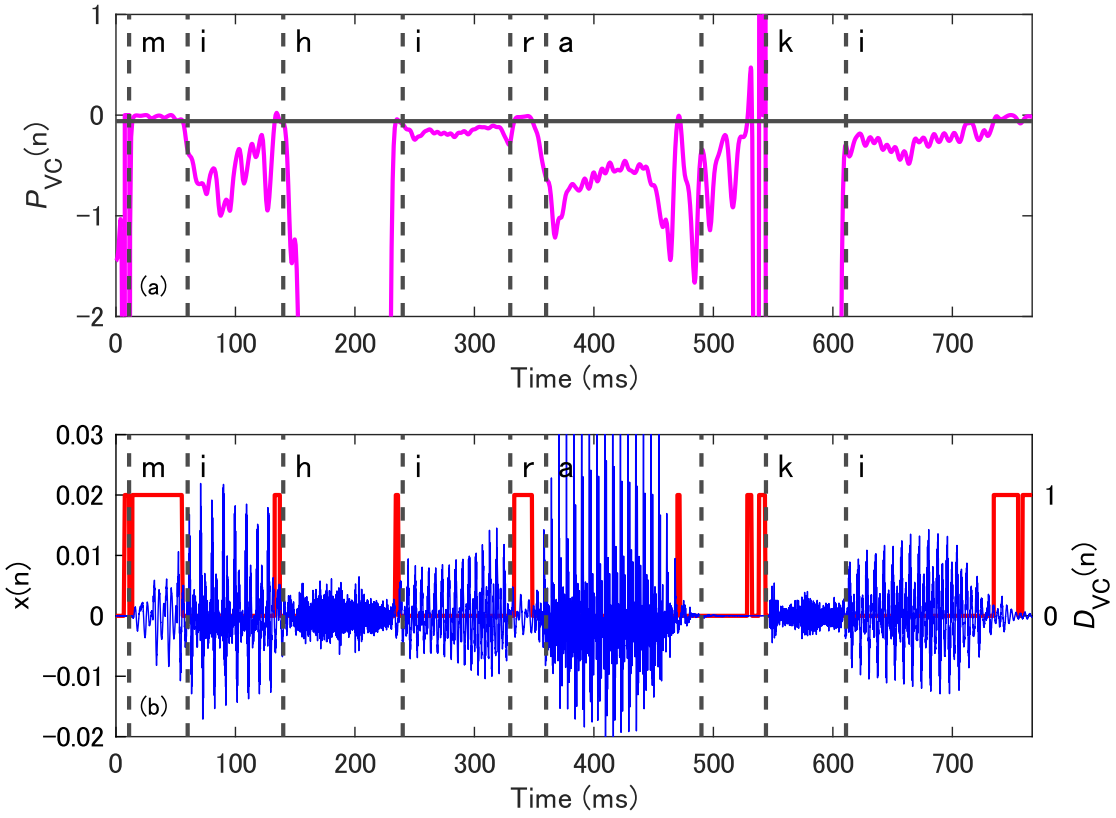

図 5.6: 有声子音区間検出による：(a) パワー比 $P_{\mathrm{VC}}(n)$ と閾値 θ_{VC}，(b) 音声信号（青の実線）と子音区間の検出結果 $D_{\mathrm{VC}}(n)$（赤の実線）

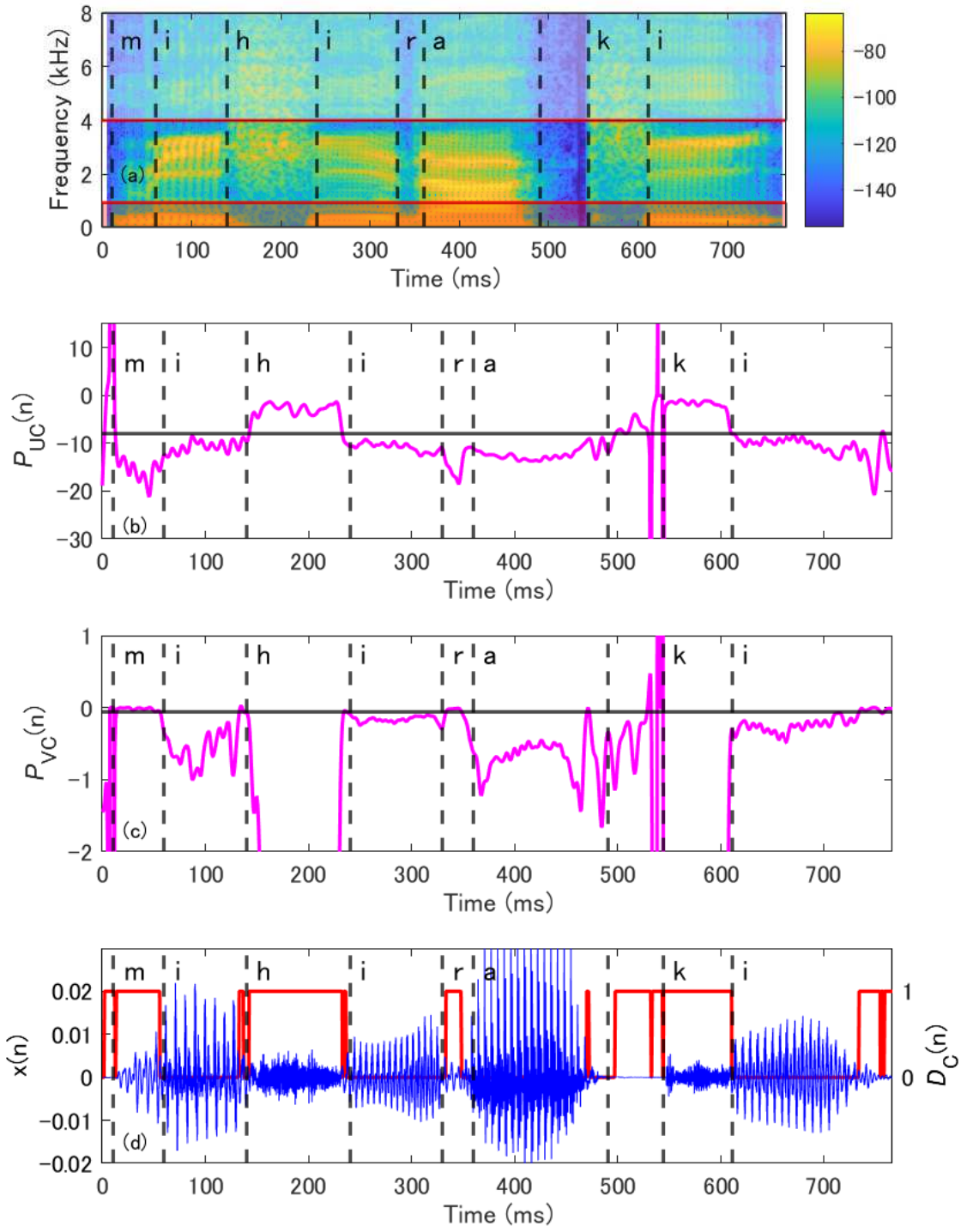

図 5.7: 子音区間検出法による子音区間検出：(a) 音声信号/mihiraki/のパワースペクトログラム（境界周波数：4 kHz と 0.9 kHz），(b) 子音区間の検出結果 $D_{\mathrm{UC}}(n)$（無声子音区間検出法），(c) 子音区間の検出結果 $D_{\mathrm{VC}}(n)$（有声子音区間検出法），(d) 子音区間の検出結果 $D_{\mathrm{C}}(n)$（子音区間検出法）

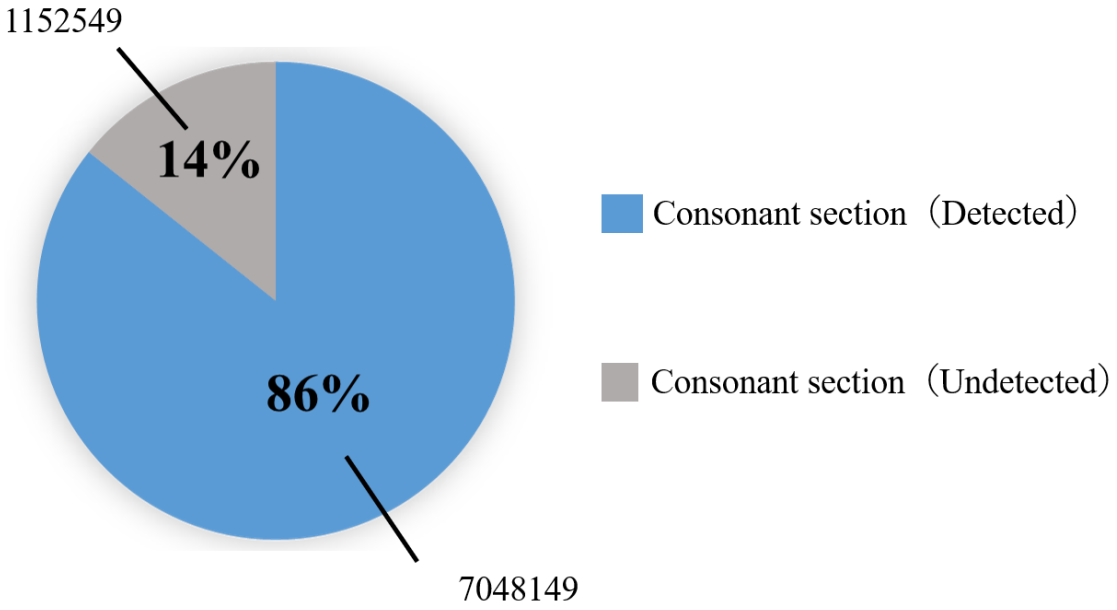

図 5.8: 改良法による子音区間の検出率

5.2 改良法による骨導提示音声の了解度の評価

改良した子音強調法（CE-IMP）を利用し，骨導提示音声の了解度試験を行った．強調タイプとして，CE-IMP の他，強調処理なし（No emphasis），従来法 [13] の音声子音強調法（CE），鳥谷ら [12] の一次高域強調法（RT-FOE）を利用した．

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．評価用音声データは子音区間検出法の評価に利用した 640 刺激を利用した．いずれの強調タイプの音声刺激も，4 種類の親密度ランクの単語が含まれていった．4 モーラ単語すべての入力を正解したとき，その単語を正答と判定した．雑音環境を模擬するため，2 種類の雑音レベル（55 dB，75 dB）のピンク雑音を利用した．そのため，実験の条件数は 32（強調タイプ 4 種×親密度 4 ランク×雑音レベル 2 種）であった．

図 5.9 に，単語了解度試験の実験装置と実験環境の概略を示す．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，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 に設定した．

図 5.10 と図 5.11 はそれぞれ改良法による雑音レベル 55 dB，75 dB での条件下での単語正答率結果を表す．図 5.10 と図 5.11 に，単語了解度試験により求められた単語正答率結果を強調タイプ・親密度ランクごとに示す．図中のエラーバーは標準誤差を表す．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．

求められた単語正答率における ANOVA4 による分析をした．強調タイプ，親密度ランクと雑音レベル 3 要因分散分析の結果，単語正答率に対して強調タイプ $[F(3,27) = 30.13,\ p < 0.01]$，親密度ランク $[F(3,27) = 177.29,\ p < 0.01]$，雑音レベル $[F(1,9) = 127.08,\ p < 0.01]$ 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 $[F(3,27) = 30.06,\ p < 0.01]$．強調タイプと雑音レベルの交互作用に対する下位検定の結果，雑音レベル 75 dB の下で強調タイプの単純主効果が認められた $[F(3,54) = 59.10,\ p < 0.01]$．雑音レベル 75 dB での結果について，Holm 法による多重比較検定の結果により，強調タイプ間で有意差が見られた（図 5.11 参照，*は $p < 0.05$，**は $p < 0.01$ の結果を示す）．

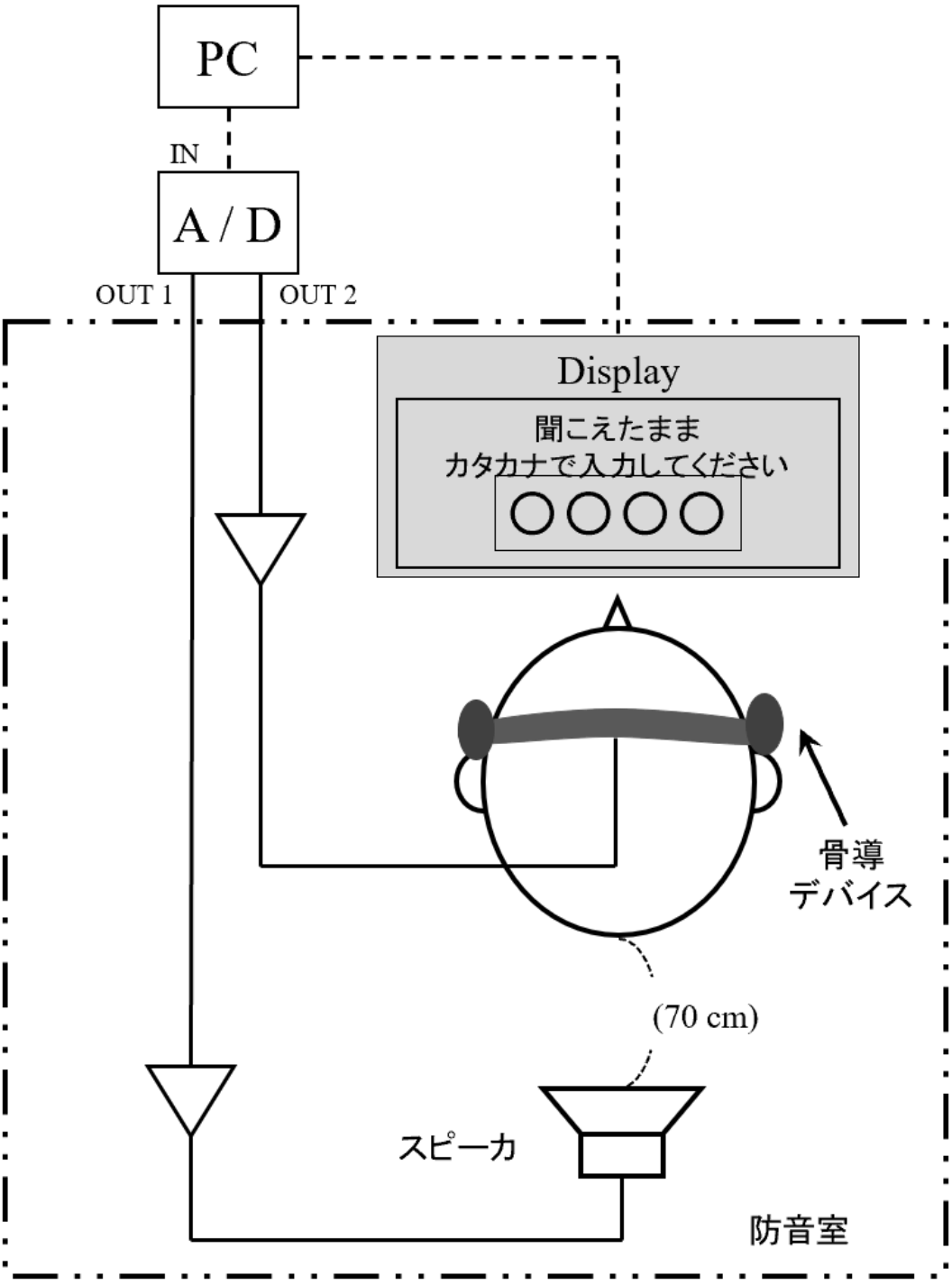
PC
IN
A / D
OUT 1
OUT 2
Display
聞こえたまま
カタカナで入力してください
骨導
デバイス
(70 cm)
スピーカ
防音室

図 5.9: 実験装置

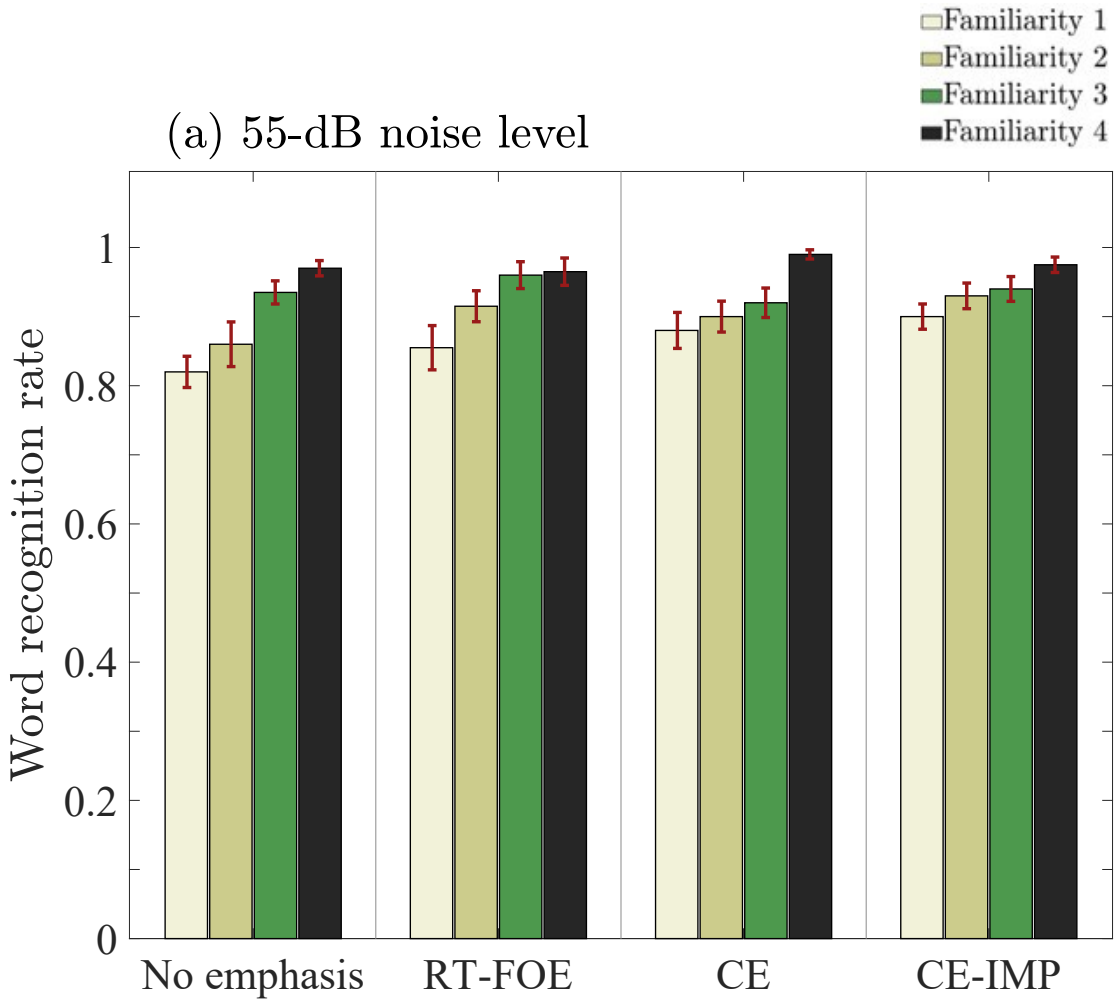

図 5.10: 改良法による雑音レベル 55 dB 環境下の強調タイプ・親密度別単語正答率

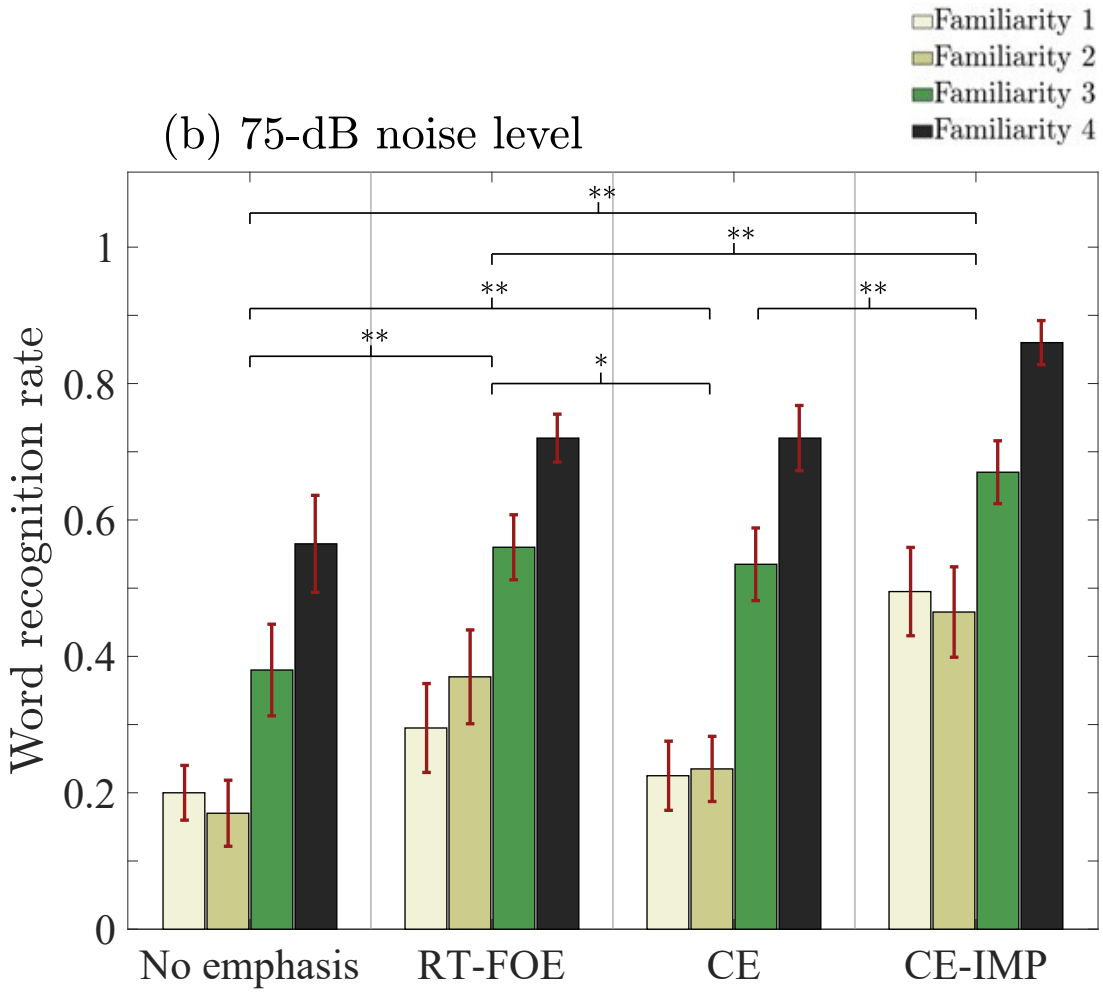

図 5.11: 従来法による雑音レベル 75 dB 環境下の強調タイプ・親密度別単語正答率

表 5.3: 改良法による子音強調：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	0.82	0.86	0.935	0.97
	75 dB	0.2	0.17	0.38	0.565
RT-FOE	55 dB	0.855	0.915	0.96	0.965
	75 dB	0.295	0.37	0.56	0.72
CE	55 dB	0.88	0.9	0.92	0.99
	75 dB	0.225	0.235	0.535	0.72
CE-IMP	55 dB	0.9	0.93	0.94	0.975
	75 dB	0.495	0.465	0.67	0.86

表 5.4: 改良法による子音強調：有意差検定結果

source	df	F	p
A: 雑音レベル	1	127.082	0.0000
error[AS]	9		
B: 親密度ランク	3	177.285	0.0000
error[BS]	27		
C: 強調タイプ	3	30.134	0.0000
error[CS]	27		
AB	3	65.935	0.0000
error[ABS]	27		
AC	3	30.055	0.0000
error[ACS]	27		
BC	9	1.721	0.0975
error[BCS]	81		

第6章　全体考察

6.1　子音区間検出の効果

子音区間の検出について，無声子音／有声子音の音響特徴に基づいて，無声子音／有声子音区間検出法を設計し，それぞれの結果を統合処理するように子音区間検出法を改良した．改良法による子音区間検出性能は，従来法 [13] と比べて 32%から 86%に改善できたことが分かった．従来法と改良法による各子音区間の検出結果を図 6.1 と図 6.2 に示す．横軸が各子音，縦軸が子音区間（サンプル数），バーの上に各子音区間の検出率を表している．その結果，従来法と比較して，各子音の検出性能が全体的に改善できた．特に，/m/や/n/など有声子音区間を 90%検出している．この結果は，相対的に低周波数帯域に集中している子音を考慮したことに由来すると考えられる．以上のことから，従来法と比較して提案法は，子音区間を高い精度で検出が可能であることが明らかになった．

しかし，改良法の子音区間検出における/w/と/r/の検出率が依然として低いままである．その理由として，/w/と/r/の音響特性について，低周波数帯域 (0〜0.9 kHz) と全周波数帯域 (0〜24 kHz) のパワー比が相対的に大きくないためだと考えられる．

一方で，ROC 曲線の概形と d_{ROC} に着目すると，有声子音区間検出法では無声子音区間検出法と比べて，(TPR，FPR) = (1，0) の点から離れていることが確認された．また，図 5.7(d) の結果では，子音区間の完全な検出には至っておらず，誤検出と未検出の区間が見られる．これは，子音の音響特徴が複雑であることに起因していると考えられる．加えて、子音区間と非子音区間を完全に分ける方法についても深く検討する必要がある．

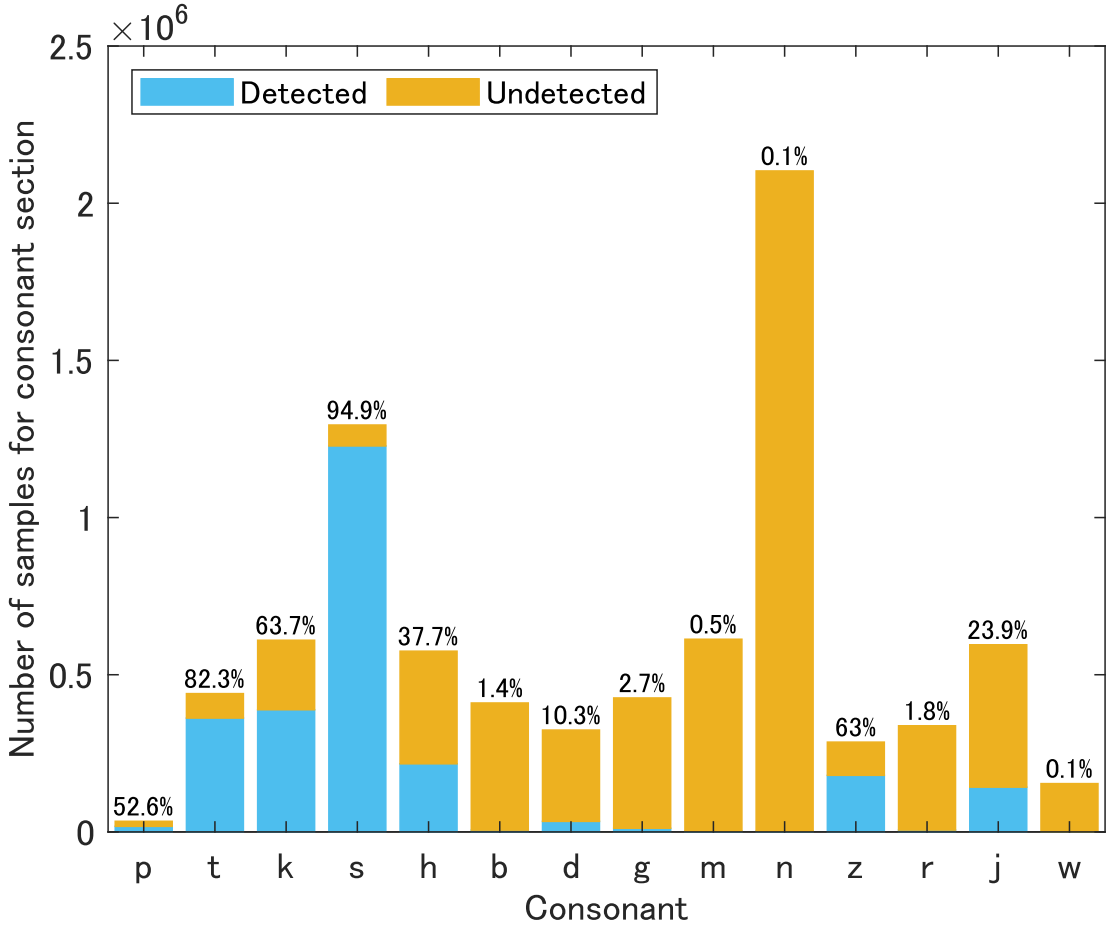

図 6.1: 従来法による各子音の区間検出結果

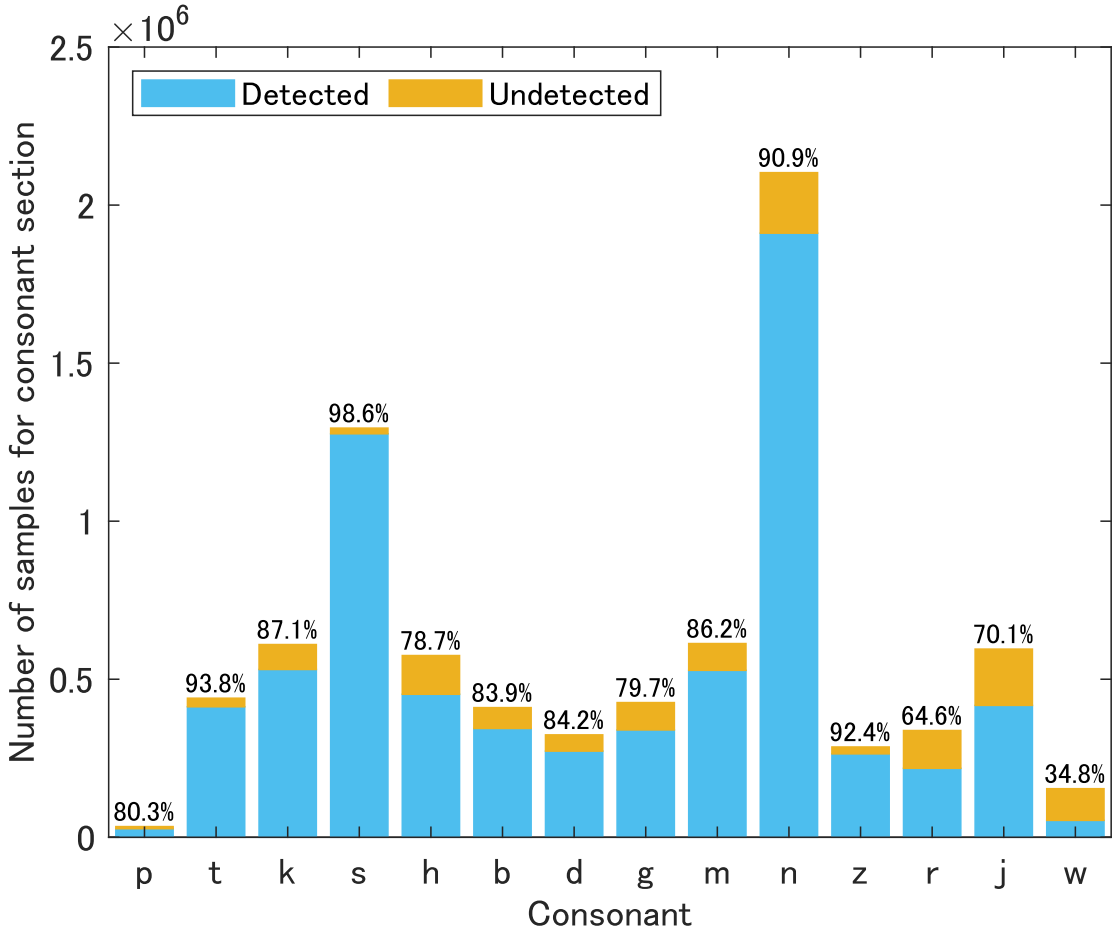

図 6.2: 改良法による各子音の区間検出結果

6.2　改良法による了解度の改善効果

高騒音環境下において，CE-IMP と RT-FOE の間では単語正答率に有意差があったことがわかった．この結果から，時間全体にわたり一貫して高域強調するよりも，子音全体を強調するほうが，高騒音環境下における骨導提示音声の了解度改善に有効であると考えられる．また，CE-IMP と CE 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，子音区間の検出性能の違いに起因すると考えられる．つまり，相対的に高域にパワーが集中する無声子音だけではなく，低域にパワーが集中する有声子音も含め，改良法では子音区間全体をより正しく検出し強調することで，他の強調手法と比較して，骨導提示音声の了解度がさらに改善されたと考えられる．

単語の正答率に対して強調処理と雑音レベルの交互作用が認められた．このことから，CE-IMP は高騒音環境下での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処理と親密度ランクの交互作用は認められなかった．この結果から，CE-IMP は単語親密度によらず有効であると考えられる．

第7章　結論

7.1　明らかにしたこと

本研究では，従来法による子音部の検出性能が低いため，雑音環境下で骨導提示音声の了解度の改善効果が十分に至っていなかったという問題に対して，子音部の検出性能を大幅に改善することで，雑音環境下で骨導提示音声の了解度をさらに改善することを目指した．そこで，無声子音／有声子音の音響特徴に基づき，無声子音／有声子音区間検出法を設計し，それぞれの結果の統合処理による子音区間検出法を検討した．また，子音区間検出法を子音強調に組み込むことで，改良法による骨導提示音声の了解度改善効果を調べた．その結果から，明らかになったことを下記に示す．

1. 子音の発声様式に基づく子音分類に着目して，無声子音／有声子音区間検出法を設計し，それぞれの検出結果を統合処理することで，従来法と比べると子音区間の検出性能が大幅に改善できることが明らかにした．

2. 子音区間検出法を改良した子音強調処理を利用することで，高騒音環境下で骨導提示音声の了解度を大幅に改善できたことが明らかにした．

7.2　残された課題

本研究では，残された課題を下記に示す．

1. 本研究の改良法による骨導提示音声の了解度の結果から，75 dB 雑音レベル環境下で，了解度が 90%以上に回復できていない．特に親密度ランク 1〜3 の単語の了解度が 70%以下であるため，高騒音環境下における骨導提示音声の了解度の改善が必要である．また，母音を子音として誤判定し，子音を未検出することがある．特に，子音/r/と/w/の区間検出率が 70%未満である．そのため，子音区間検出の性能を向上すれば，子音区間検出の性能をさらに改善するかもしれないと考えられる．

2. 本研究では子音強調処理を改良することで，骨導提示音声の了解度の改善に有効が認められたが，最も良い子音強調による了解度の改善効果がまだ実現できていない．各子音の区間検出率と骨導提示音声の了解度の改善効果を明らかにすれば，骨導提示音声の了解度をさらに改善できると考えられる．

3. 従来法による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を示している [13]．本研究による子音強調は，従来法の子音強調よりも骨導提示音声の了解度を改善できることが明らかにした．そのため，骨導提示音声の了解度改善に，改良された子音強調と高域強調を組み合わせた方法が，最も役立つか確認の必要がある．

4. 子音強調処理と高域強調処理では骨導伝達特性の影響に対して，高域成分の補償・子音区間の強調処理を行った．それ以外に，雑音環境下で骨導提示音声の了解度を改善するために，高域減衰という骨導伝達特性を考慮した上で，音声知覚に重要な高域成分を低域に集約させることで，骨導伝達特性の影響を避けられると考えられる．周波数スペクトルの伸長圧縮処理によって，音声言語知覚で重要な周波数帯域を骨導伝達帯域に集約することで，話者個人性が壊れ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．

参考文献

[1] A. Mudry, and A. Tjellström, “Historical background of bone conduction hearing devices and bone conduction hearing aids, ” Implantable bone conduction hearing aids. Karger Publishers, Vol. 7, pp. 1–9, 2011.

[2] S. E. Ellsperman, E. M. Nairn, and E. Z. Stucken, “Review of bone conduction hearing devices, ” Audiology Research, Vol. 11, No. 2, pp. 207–219, 2021.

[3] A. Hagr, “BAHA: bone-anchored hearing aid, ” International journal of health sciences, Vol. 1, No. 2, pp. 265, 2007.

[4] C. Manning, T. Mermagen, and A. Mermagen, “The effect of sensorineural hearing loss and tinnitus on speech recognition over air and bone conduction military communications headsets, ” Hearing Research, Vol. 349, pp. 67–75, 2017.

[5] E. H. Berger, R. W. Kieper, and D. Gauger, “Hearing protection: Surpassing the limits to attenuation imposed by the bone-conduction pathways, ” The Journal of the Acoustical Society of America, Vol. 114, No. 4 pp. 1955–1967, 2003.

[6] T. Letowski, P. Henry, and T. Henry, “Use of bone conduction transmission for communication with mounted and dismounted soldiers, ” Proceedings of the ASNE Human System Integration Symposium on Enhancing Combat Effectiveness through Warfighter Performance. Arlington, VA, 2005.

[7] T. Letowski, T. Mermagen, N. Vause, and P. Henry, “Bone conduction communication in noise: a preliminary study, ” Proceedings of the XXI International Congress on Sound and Vibrations. Vol. 37, pp. 3037–3044, 2004.

[8] B. N. Walker, and J. Lindsay, “Navigation performance in a virtual environment with bonephones, ” Georgia Institute of Technology, 2005.

[9] Z. J. Lim, and J. Claydon, “The use of bone conduction headsets to improve communication during the COVID‐19 pandemic, ” Emergency Medicine Australasia, 2020.

[10] 鵜木祐史,“骨導音の考え方とその応用事例,”騒音制御, Vol. 46, No. 2, pp. 53–58, 2022.

[11] T. Toya, P. Birkholz, and M. Unoki,“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.

[12] 鳥谷輝樹, 小林まおり, 中村健一, 鵜木祐史,“骨導デバイスによる提示音声の了解度改善法,”音講論（春）, pp. 689–692, 2-4-5, 2022.

[13] シュブンウ, 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,”音講論（春）, pp. 693–696, 2-4-6, 2022.

[14] R. D. Kent, and C. Read, (荒井隆行, 菅原勉　監訳), 音声の音響分析, 海文堂出版, 東京, 1996.

[15] S. Furui, “On the role of spectral transition for speech perception,”The Journal of the Acoustical Society of America. Vol. 80, No. 4, pp. 1016–1025, 1986.

[16] S. Stenfelt, “Acoustic and physiologic aspects of bone conduction hearing, ” Implantable bone conduction hearing aids, Vol. 71, pp. 10–21, 2011.

[17] S. Stenfelt, and R. L. Goode, “Bone-conducted sound: physiological and clinical aspects, ” Otology & Neurotology, Vol. 26, No. 6, pp. 1245–1261, 2005.

[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, Vol. 50, No. 3, pp. 351–357, 1996.

[19] 山田芳靖, 土方啓暢, 川原伸章, 藤坂洋一, 中川誠司,“骨伝導音による音声認識の検討,”電気学会論文誌 E (センサ・マイクロマシン部門誌), , Vol. 124, No. 8, pp. 272–277, 2004.

[20] 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, Vol. 57, No. 3, pp. 294–304, 2016.

[21] J. Wang, S. Stenfelt, S. Wu, Z. Yan, J. Sang, C. Zheng, and X. Li, “The Effect of Stimulation Position and Ear Canal Occlusion on Perception of Bone Conducted Sound, ” Trends in Hearing, Vol. 26, 23312165221130185, 2022.

[22] T. Fujimoto, and M. Mori, “Word intelligibility of bone conductive sound when wearing ear plugs, ”2015 IEEE 4th Global Conference on Consumer Electronics (GCCE). IEEE, pp. 38–39, 2015.

[23] 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, Vol. 37, No.1, pp. 57–63, 1994.

[24] 安武達朗, 中島祥好, “準実時間子音強調システム, ”信学技報, Vol. 105, No. 479, pp. 79–84, 2005.

[25] 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, ” 音学誌, Vol. 69, No. 5, pp. 224–231, 2013.

[26] 日本音響学会, “音響用語辞典, ” コロナ社, 2013.

[27] R. Lawrence, and J. Biing-Hwang, （古井貞熙　監訳）, 音声認識の基礎（上）, NTT アドバステクノロジ株式会社出版, 東京, 1995.

[28] 荒木雅弘, フリーソフトではじめる機械学習入門 Python/Weka で実践する理論とアルゴリズム (第 2 版), 森北出版, 東京, 2020.

[29] 石塚健太郎, 藤本雅清, 中谷智広, “音声区間検出技術の最近の研究動向” 音学誌, Vol. 65, No. 10, pp. 537–543, 2009.

[30] Y. Qi, and B. R. Hunt, “Voiced-unvoiced-silence classifications of speech using hybrid features and a network classifier, ” *IEEE Transactions on speech and audio processing*, Vol. 1, No. 2, pp. 250–255, 1993.

[31] Julius development team, “Julius 音素セグメンテーションキット” *Julius now on GitHub*, https://julius.osdn.jp/index.php?q=ouyoukit.html, (参照 2022-10-21)

[32] 武田一哉, 匂坂芳典, 片桐滋, 桑原尚夫, “研究用日本語音声データベースの構築” 音学誌, Vol. 44, No. 10, pp. 747–754, 1998.

研究業績

国内発表

1. 王思成, 上江洲安史, 鳥谷輝樹, 鵜木祐史, “骨導提示音声の了解度改善のための子音強調処理の改良,” 日本音響学会聴覚会研究会資料, Vol. 52, No. 7, pp. 543–548, 2022.

2. 王思成, 上江洲安史, 鵜木祐史, “周波数帯域のパワー比に基づいた子音区間検出法の検討,” 第 37 回信号処理シンポジウム, pp. 320–325, 2022.

3. 王思成, 上江洲安史, 鳥谷輝樹, 鵜木祐史, “子音強調処理の改良による骨導提示音声の了解度改善,” 日本音響学会聴覚会研究会資料, 2-4-4, 2022.

謝辞

研究活動から私生活，社会の一般常識にいたるまで，多くのご指導，ご助言をいただいた，主指導教員である鵜木祐史教授には，心から深く感謝致します．研究の進め方，研究に関する考察，様々なご助言を頂いた上江洲安史特任助教に心から感謝申し上げます．また，研究室会議やミーテイングなどの場合において，様々なご助言をくださった木谷俊介助教，大田さん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ご支援をいただきました．お礼を申し上げます．大変忙しい中，様々なご助言をいただきました，磯山さん，鳥谷さん，郭さん，李さんに感謝申し上げます．同期として共に過ごした井上さん，市川さん，宮家さん，Wang さん，Benita 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，研究と生活に支えてくださった朱さん，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．