

時間および周波数特性変更による音声情報の聞き取りやすさ向上 Improvement of Speech Information Listening Performance by Time and Frequency Characteristics Modification

精密工学専攻 14号 江口 和希
Kazuki Eguchi

1. はじめに

近年、東京オリンピックの開催や高齢化により、音声情報の聞き取りやすさへの関心が高まると考えられる。これまでの研究により、音声情報の聞き取り性能を評価する研究が多数行われている⁽¹⁻³⁾。また、避難警報音をはじめとした拡声音は様々な人が同時に聞き取るひとつであり、聞き取り性能が非常に重要なものであり、拡声音の基準化⁽⁴⁾など様々な人の同一環境での音声情報の聞き取りやすさへの関心も高まった。しかし、聴取する人全員が満足するような音声に関する研究は少なく同一環境で音声情報を聞く場合、難聴者に合わせ音を大きくする音響特性の変更処理を行うと同一環境で聞く人はうるさすぎるなど⁽⁵⁾の問題点がある。同一環境で音声情報を聞き取る聴取者全員に対して、聞き取り性能が高く、不快に感じない音声情報を提供する音環境の構築が求められている。

そこで本研究では、同一環境下で音声を取り取ることを想定し、難聴者および健常者が聞き取る拡声音に着目する。まずは難聴者に対し、音声情報の音響特性を変更することで聞き取り性能の高い音声情報を検討する。その結果より、音声情報の周波数毎の聞き取りやすさを変更し、健常者に対しても同様な傾向が得られるか検証し、聞き取りやすい音響特性の推定を行う。推定された音響特性に補正する音声情報の加工方法を提案する。その提案音声の聞き取りやすさの印象および聞き取り性能に対する効果の妥当性を検討する。

2. 難聴者に対する音質変更による音声情報の聞き取り性能評価

本章では、難聴者に対し、音声情報の聞き取りやすさを把握するため、再生機器変更による音質変更や音源加工による音質変更による単音節の聞き取り性能評価実験を行う。

2.1 再生機器変更による聞き取り性能評価

Fig. 1(a), (b)に本実験で用いるスピーカ2種(Speaker A, Speaker B)に440 Hzの正弦波を入力した際の高調波歪特性を示す。Speaker Bは高調波歪を多く出力する特性を持つ。この特性により、出力される各単音節の周波数特性に変化が加わり、聞き取り性能が変化する。単音節の聞き取り性能評価実験は67-S語表内⁽⁶⁾の単音節20個をランダムに並び替えたものを8セット再生し、聞こえた単音節を書き取ってもらい、その正答率を聞き取り性能とした。聞き取り性能の正答率の求め方の例をFig. 2に示す。これにより各単音節の母音(V)・子音(C)・全体(T)の正答率をそれぞれ算出する。

Fig. 3に得られた各平均正答率を示す。Speaker A出力による結果をSA、Speaker B出力による結果をSBとする。母音正答率の差は100%に近くあまり差がないが、子音正答率がSAでは約74%、SBでは約84%とSBの方が10%高い。それに伴い、全体の正答率もSBが高くなっている。全単音節のSBにおける子音と母音の振幅差CVR(Consonant Vowel Rate)とSAとのCVRの差の平均値が3.69となっており、SBはSAに対し、子音部が強調されている。次節では子音部の音圧の大きさの変化と正答率の変化に着目し、検討を行う。

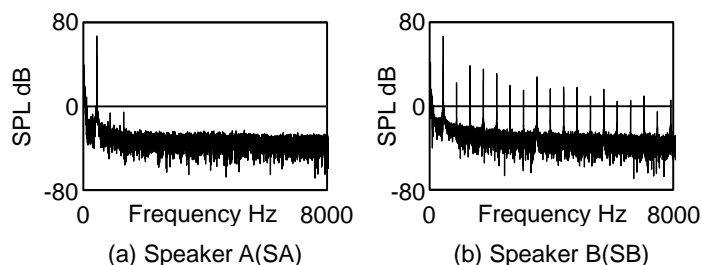


Fig. 1 Speaker's harmonics characteristics

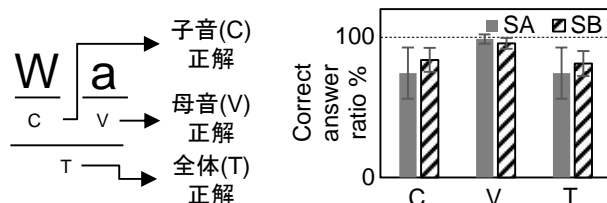


Fig. 2 Scoring example

Fig. 3 Correct answer rate

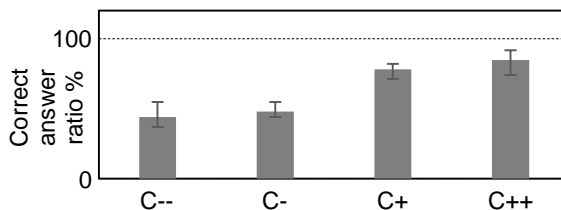


Fig. 4 Correct answer rate (Total)

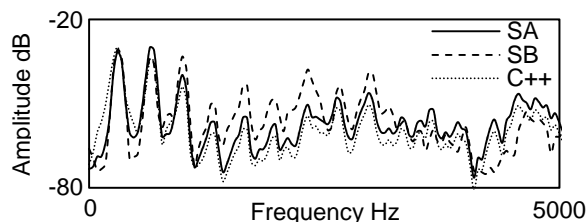


Fig. 5 Spectrum (Ne)

2.2 子音部音圧変更音による聞き取り性能評価

本節では67-S語表内の子音を部もつ単音節17音節の子音部の音圧を変更させた音源を作成し、2.1節と同じく聞き取り性能評価実験を行う。各単音節の子音部を目視にて定め、音圧を段階に分け4つ作成する。作成した音源とSAのCVRの差の平均値はC++では10.7、C+では5.1、C-では-4.6、C--では-8.4となっている。作成した音源をSpeaker Aより出力し、2.1節と同様の実験を行う。

Fig. 4に作成した音源毎の全体の平均正答率を示す。子音部の音圧が大きくなるにつれ、正答率は向上する傾向が見られた。これにより、子音部の音圧を大きくすると音声情報の聞き取り性能が向上すると推測される。

ここで2.1節にて行った実験との比較を行う。正答率の違いが特に大きかった単音節Neでは正答率がC++では50%であったが、SBでは100%と正答率の違いがある。Fig. 5に単

音節 Ne の SA, SB, C++それぞれのスペクトラムを示す。e 行の母音部の第二フォルマントの帯域が存在する中高音帯域の音圧が SB では大きくなっていることがわかる。そのため中高音帯域における聞き取りのヒントがより聞き取りやすくなり、正答率が向上したと考えられる。よって子音部を聞き取りやすくするだけでなく、単音節の周波数特性を変化させることで聞き取り性能が向上すると推測される。

2.3 帯域別音圧圧縮加工による聞き取り性能評価

本節では、子音部強調及び高域部強調を実現するため、帯域別音圧圧縮(マルチバンドコンプレッサ)を用い、聞き取り性能評価を行う。Fig. 6に今回用いる帯域別音圧圧縮加工の概要を示す。これにより、音声をもつ低域および母音部を持つ大きい音圧を中高域および子音部に分配することができる。この加工を67-S語表内の全単音節に施し、評価を行う。加工後の音源による結果をSA2とする。また作成した音源とSAのCVRの差の平均値は5.9となっており、子音部が強調されていることもわかる。作成した音源をSpeaker Aより出力し、2.1節と同様の実験を行う。

Fig. 7に得られた平均正答率を示す。作成した音源 SA2の平均正答率が約74%から約94%とSAに比べ向上したことがわかる。以上より、音声情報の音圧バランスを変更することで難聴者に対し、聞き取りやすい音声情報を出力できると推測できる。

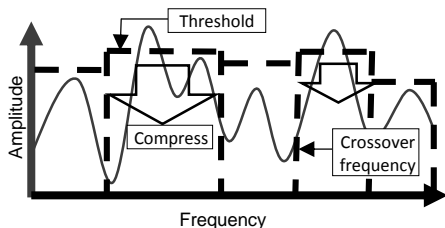


Fig. 6 Overview of multi band compression

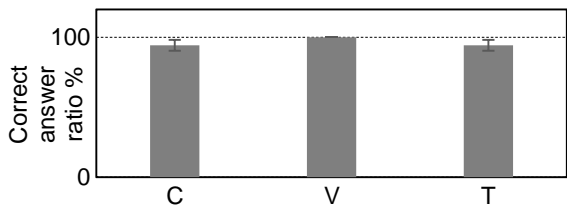


Fig. 7 Correct answer rate(SA2)

3. 騒音下における音声情報の聞き取り性能評価

本章では、3章にて示唆された単音節の周波数毎の音圧バランスの違いが健常者に対し、音声情報の聞き取り性能に及ぼす影響を把握する。

3.1 騒音の違いによる聞き取り性能評価

健常者に対し、単音節と同時に提示する騒音の変化による音声情報の聞き取り性能の変化を把握する。

実験に使用する騒音は、ホワイトノイズに250-8000 Hzのバンドパスフィルタを掛けた騒音(BN1)、BN1に対しカットオフ周波数250 Hz、-6 dB/Octのローパスフィルタを掛けた騒音(BN2)、BN1に対しカットオフ周波数250 Hz、-12 dB/Octのローパスフィルタを掛けた騒音(BN3)である。Fig. 8に作成した各騒音のスペクトラムを示す。これらの騒音のマスクングにより、周波数毎の聞き取りやすさが変化すると考えられる。音声情報は女性による発話であり、SP1とする。作成した騒音を65 dB、SP1の各単音節を60 dBでヘッドホンよりそれぞれ提示し、評価を行う。被験者は20代8名である。

実験により得られた各正答率および後続母音毎に分けた正答率をFig. 9(a), (b)に示す。図より提示する騒音がBN2

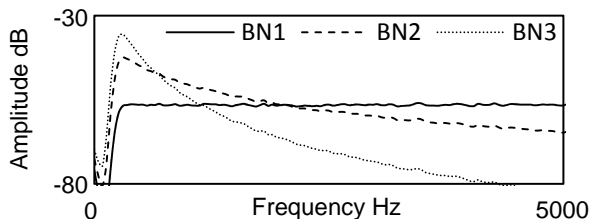


Fig. 8 Spectrum of noise

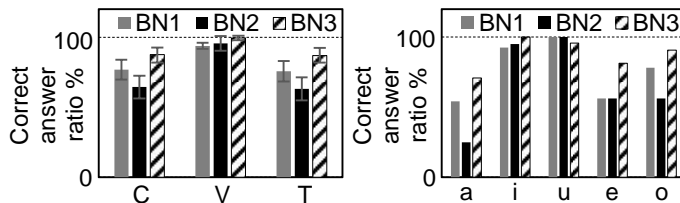


Fig. 9 Correct answer rate(SP1)

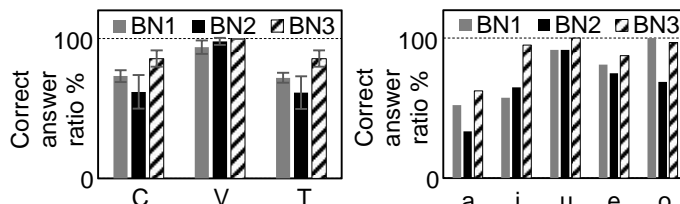


Fig. 10 Correct answer rate(SP2)

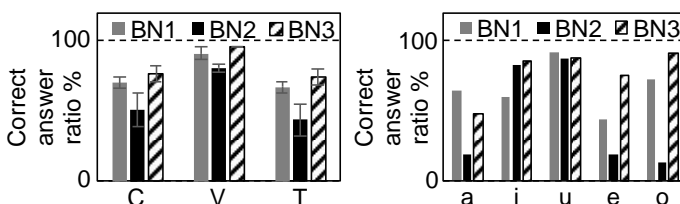


Fig. 11 Correct answer rate

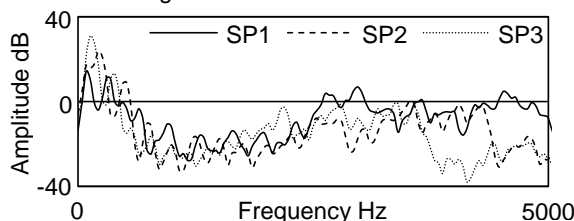


Fig. 12 Spectrum difference(Ri_BN1)

の際、聞き取り性能が大きく低下することがわかる。それぞれのノイズの特徴として第一フォルマントを含む低音域のマスクング量はBN3 > BN2 > BN1であり、中高音域のマスクング量はBN1 > BN2 > BN3である。BN2により第一フォルマントが多く、同時に第二フォルマント以降の帯域もマスクングされ、両方が聞き取りにくくなったと考えられる。

一方、BN3で最も正答率が高いのは、第一フォルマント帯域のマスクング量が多いが、それ以降の帯域のマスクング量が少なく、聞き取りやすくなっているため正答率が向上したと考えられる。

3.2 話者違いによる聞き取り性能評価

3.1節で行った実験を別の話者による発話においても行った。使用する音源はプロアナウンサの男女それぞれ1名による発話である。女性による発話をSP2、男性による発話をSP3とし、3.1節と同様の条件で実験を行う。

実験により得られた SP2, SP3 の各正答率および後続母音

毎に分けた正答率を Fig. 10(a), (b) および Fig. 11(a), (b) に示す. 3.1 節と同様に提示する騒音が BN2 の際, 正答率が一番低く, BN3 の際, 最も正答率が高い結果が得られた. 話者毎で比較すると後続母音が i の単音節に差が見られた. BN1 の際の正答率が SP1 では 100%, SP2 では 0%, SP3 では, 50% となっている単音節 Ri に着目する. Fig. 12 にそれぞれ話者により単音節 Ri のスペクトラムと BN1 のスペクトラムの差の図を示す. この図より第一フォルマント帯域の聞こえは SP3 > SP2 > SP1 の順で聞こえやすいと考えられるが第二フォルマント以降は SP1 > SP2 ≒ SP3 である. これらより, 聞き取り性能には, 音圧が大きい第一フォルマントを含む帯域も重要であるが, 第二フォルマント以降を含む帯域がより重要であると示唆された.

4. フォルマント間音圧バランス変更による音声情報の聞き取り性能評価

本章では, 2.3 章にて示唆されたフォルマント間の音圧バランスを帯域別音圧圧縮で変更し, 音声情報の聞き取りやすさ及び聞き取り性能に及ぼす影響を把握する.

4.1 加工音の聞き取り印象評価

3 章でみられた結果より, 帯域別音圧圧縮により, 各フォルマント箇所にあわせた周波数域値を設け, 音圧が大きい低次フォルマント帯域は圧縮し, 全体の音圧バランスを揃えることで相対的に音圧の小さい高次フォルマント帯域の音圧を大きくし, フォルマント毎の音圧バランスを変更する加工を施した. 本実験では 67-S 語表内の単音節 Wa, Ni, Ku, Te, Yo に着目し, SP1 内のこれらの単音節に加工を加え, 評価を行う. 低次フォルマント帯域の圧縮量が大きく, フォルマント間の音圧差が小さい順に MS, Small, 逆に高次フォルマントを圧縮するように加工し, フォルマント間の音圧差を大きくした音源を MB とする. 評価は無響室で行い, 評価音源をヘッドホンより提示する. 評価方法はシェッフェの対比較法(中屋の変法), 評価項目は意識しやすさ, 聞き取りやすさ, 被験者は 20 代 8 名である.

評価結果を Fig. 13(a), (b) に示す. 単音節においてフォルマント毎の音圧差が小さいほど意識しやすく, 聞き取りやすいという傾向が得られた. フォルマント間の音圧差が小さくなることで相対的に中高音域が大きくなり, 目立つようになるため意識しやすかったという印象を得たと考えられる. 聞き取りやすさに着目すると, 単音節 Ni では Small と MS の差はかなり小さい. 今回の加工音は, 二段階の音圧差で実験を行ったが, フォルマント間の音圧差を小さくしすぎること, 聞き取りやすさの印象が悪くなると予想される.

4.2 騒音下に対する加工音の聞き取り印象評価

4.1 節で作成した加工音を 3 章にて作成した騒音下において, 印象変化を把握する. 使用する単音節, 騒音および評価項目は同様である. それぞれの騒音に対する評価結果を Fig. 14(a)~(c) に示す. BN1 の際は, 加工前後で評価結果の差が小さい. BN2 の際は, Ref に対し, 印象の差が小さいものもあるが MB に対してはどれも MS は聞き取りやすいという結果が得られた. BN3 の際は, どの単音節においても Ref, MB に対し, MS は聞き取りやすいという評価が得られた. 聞き取りやすいという印象を得る音声は, 第二フォルマント以降の帯域が十分に聞こえることが重要であり, 第一フォルマント帯域の聞き取りやすさの印象に対する影響は小さいことが示唆された.

4.3 騒音下の加工音の聞き取り性能評価

SP1 による発話の全単音節に MS の加工を適用し, 騒音下での聞き取り性能の変化を把握する. 評価条件は, ヘッドホンにて各単音節を 60 dB, 各騒音を 65 dB で提示し評価を行う.

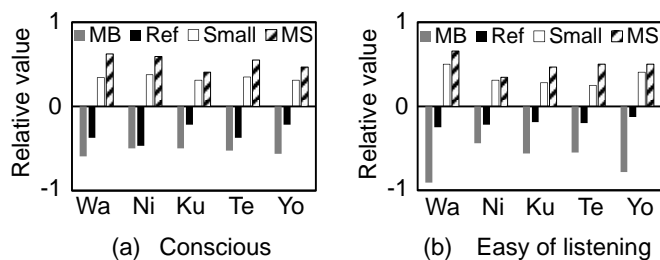


Fig. 13 Evaluation result

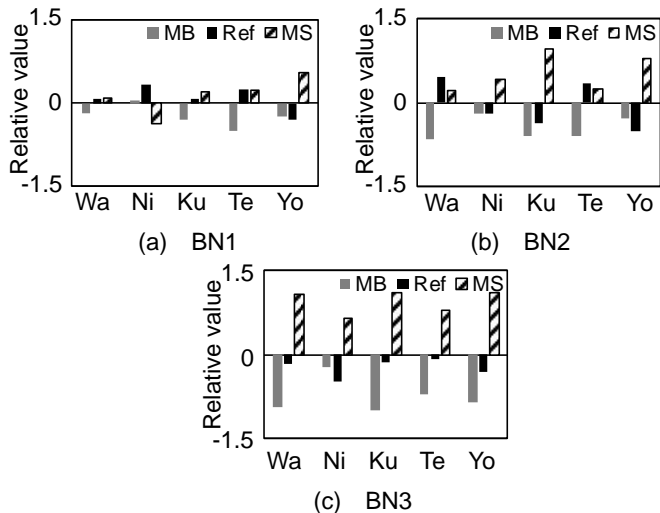


Fig. 14 Evaluation result

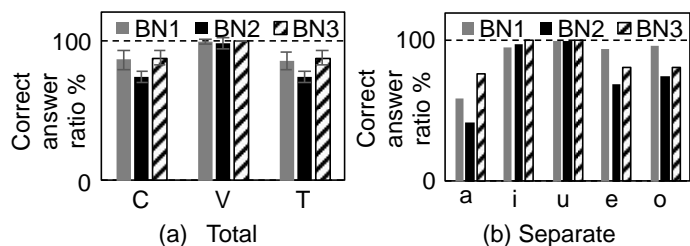


Fig. 15 Correct answer rate(SP1-MS)

各正答率および後続母音毎に分けた正答率を Fig. 15(a), (b) に示す. Fig. 9 と比較して加工音は騒音が BN3 の際は正答率がほぼ同等であり, BN1 および BN2 の際は約 10% 以上正答率が向上した. 第一フォルマントを含む低域の聞こえは, 第二フォルマント以降を含む中高域に比べ, 聞き取り性能への影響は小さいと考えられる. そのため, 今回の加工において第一フォルマント帯域の音圧を第二フォルマント以降の帯域である中高域に分配することで BN1, BN2 に対し, 第一フォルマント帯域である低域と, それ以降の帯域を含む中高域における音声情報をバランスよく聞き取ることができ, 加工後の正答率が向上したと考えられる.

また, 4.2 節の聞き取りやすさの印象変化と比較すると, BN3 の際には加工前後で印象が大きく変化したが聞き取り性能の正答率の変化は小さい, 逆に BN1, BN2 の際には印象変化の差が小さいが聞き取り性能は大きく変化していることがわかる. このことより聞き取りやすい印象を得る音声情報が, 必ずしも聞き取り性能が高い音声情報とはいえない可能性が示唆された.

4.4 別話者への適用

3.3 節で用いた SP2, SP3 に MS の加工を施し, 騒音下での聞き取り性能の変化を把握する. 実験条件は 4.3 節と同様である. Fig. 16(a), (b) に SP2-MS, Fig. 17(a), (b) に SP3-MS の各正答率および後続母音毎に分けた正答率を示す. Fig. 10,

Fig. 11と比較するとSP2, SP3共にBN1, BN2の際に正答率が上昇していることが確認でき、話者によらず提案した加工方法にて聞き取り性能を向上させられることが確認できる。

全話者の後続母音毎での正答率を比較すると後続母音がaの際、加工後においても、他と比較すると正答率が低い傾向であった。原因としてフォルマント間の音圧バランスの崩れが考えられる。Fig. 18に加工前後の単音節Baのスペクトラムを示す。aは第一フォルマント帯域と第二フォルマント帯域の音圧差が小さく、フォルマント周波数も近い。今回提案する加工により第一フォルマント帯域を大きく圧縮し、第二フォルマント以降の帯域へ分配したことで第二フォルマント帯域との音圧差が逆に大きくなり、バランスよく聞き取ることができず、後続母音がaの単音節では正答率が低かったと考えられる。

またBN3の際、平均正答率が加工前後でほぼ同等であった。後続母音に着目すると特にo行の単音節で低下していた。SP3による発話に着目すると、特に単音節Moでの低下が大きい。

Fig. 19にSP3における加工前後での単音節MoのスペクトラムとBN3のスペクトラムの差分の図を示す。加工前後で特に第一フォルマント帯域で差が大きい。これはa行と同様に第一フォルマント帯域を大きく圧縮するようにMSを作成してしまい、BN3による低域マスキングに加わることで、第一フォルマント帯域が大きくマスキングされ、聞き取りが悪くなったと考えられる。中高音域は加工前より判別できるレベルで聞き取れ、その箇所が強調されているが、聞き取りの結果には影響を及ぼさなかった。これらより、低域より中高域での聞き取り判別のヒントは多くあるが、低域の聞こえが悪くなりすぎると逆に聞き取り性能が低下する恐れが示唆された。特に後続母音がa, oの場合、第一フォルマント帯域に対する聞き取り性能の重要性が他の後続母音より大きく、過度にフォルマント毎の音圧バランスを変更すると聞き取り性能が低下してしまうことが示唆された。

また、SP2, SP3でBN1, BN2の際、加工前後で正答率の変化が大きいi行の単音節に着目する。特に加工前では正答率が低かったSP2の単音節Riの加工前後のスペクトラムとBN1のスペクトラムとの差分のグラフをFig. 20に示す。今加工により、スペクトラムとの差分が大きかった第一フォルマント帯域の音圧が圧縮され、その音圧が第二フォルマント以降の帯域に分配され大きくなっている。結果として各フォルマント帯域でバランスよく聞くことができ、正答率が向上したと考えられる。

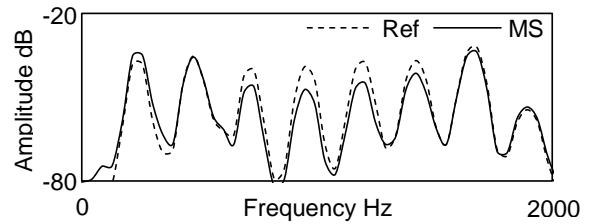


Fig. 18 Spectrum(Ba_SP2)

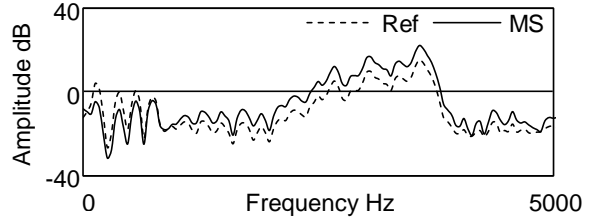


Fig. 19 Spectrum difference(Mo_BN3_SP3)

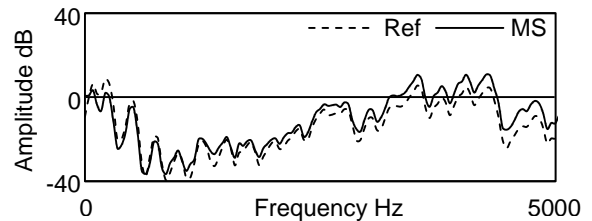


Fig. 20 Spectrum difference(Ri_BN2_SP2)

5. 研究成果

- (1) 難聴者に対する再生機器および音質変更により、子音部の音圧および周波数のバランス変更により聞き取り性能が変化することを確認した。
- (2) 健常者に対して、騒音を用い音声の周波数毎における聞き取りやすさを変更することで、聞き取り性能が変化する事を把握した。
- (3) 健常者に対し、周波数毎の聞き取りやすさの変更により、音声情報の聞き取りやすい印象でも聞き取り性能は必ずしも高いといえない可能性あることが示唆された。
- (4) フォルマント毎の音圧バランスを変更することにより、聞き取り性能が向上することを確認できた。

参考文献

- (1) 江口和希, 戸井武司, 桑原和男, 藤澤孝史, 感音性難聴に対し音質変化が単音節の聞き取りに及ぼす影響評価, 日本機械学会環境工学総合シンポジウム, No. 17-9 (2017) pp. 128-131.
- (2) 江口和希, 戸井武司, 周波数毎の音圧バランス変更による騒音環境における音声の聞き取りやすさ向上, 日本音響学会講演論文集(春), (2018) 1-13-12.
- (3) 中家諒, 栗栖清浩, 坂本真一, 難聴者向けスピーカ及び難聴者支援システムの評価, 日本音響学会講演論文集(秋), (2016) pp. 1123-1124.
- (4) 佐藤逸人, 栗栖清浩, 荻木禎史, 鈴木陽一, 災害等非常時屋外拡声システム性能確保のためのASJ技術基準(第1版), 日本音響学会講演論文集(秋), (2017) pp. 1521-1522.
- (5) 永田徳, 私の感じる音環境-難聴者予備軍の1人として, 騒音制御, Vol. 29 No. 5 (2005) pp. 6-7.
- (6) 立木孝, 小田恂, 「聴覚検査法(1990)2. 語音聴覚検査法(2000)(案)」について, Audiology Japan, Vol. 43 No. 6 (2000) pp. 668-683.

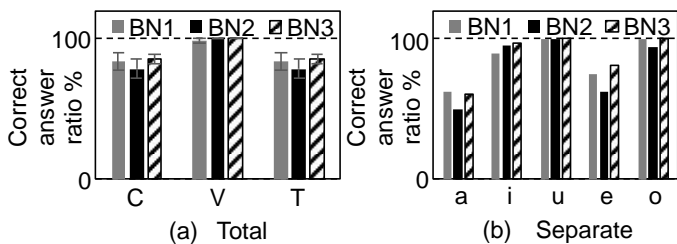


Fig. 16 Correct answer rate(SP2-MS)

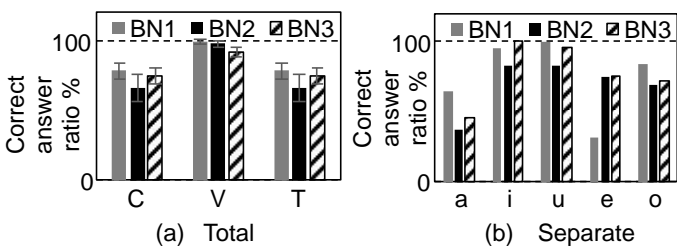


Fig. 17 Correct answer rate(SP3-MS)