技 術 報 告 43.71.Es, Gv; 43.55.Hy; 43.60.Qv

残響環境における音声明瞭度改善を目的とした子音強調・母音抑圧による前処理*

辻 美咲[*1] 荒井隆行[*1] 安 啓一[*1,†]

［要旨］ 残響時間が長い環境では情報伝達が困難となることがある。そのため，残響による影響を軽減させ，正しく情報伝達することが求められている。本研究は，残響下での聴き取りを改善するため，子音強調及び母音抑圧を施す前処理手法を提案する。処理音声に残響を畳み込んだ刺激を用いて単語了解度試験を行い，処理の効果を検討した。その結果，各子音部の最大振幅を 1.0 とし，各母音部の最大振幅を 0.4～1.0 とした場合に関しては有意差はなかった。一方，各母音部の最大振幅を 0.2 とした場合に関しては有意に了解度が低下した。

キーワード 子音強調，母音抑圧，残響，単語了解度，定常部抑圧処理
Consonant emphasis, Vowel suppression, Reverberation, Word intelligibility, Steady-state suppression

1. はじめに

日常生活において，残響とは至る所に存在する。コンサートホールで音楽鑑賞をしているときは，残響による効果で，心地よい音楽を楽しむことができるという利点がある一方，欠点も存在する。例えば地下鉄のホームで，残響により電車の案内アナウンスが聴き取りづらくなったり，非常時の緊急アナウンスが残響により聴き取れず，逃げ遅れたりというようなことである。後者のような欠点を防ぐため，最近の研究では残響による影響を軽減させ，正しく情報を伝達する手法の開発が求められている。

上記のように，残響によって音声の聴き取りが困難になる原因の一つとして，overlap-masking [1] が挙げられる。これは，ある音素に残響が畳み込まれ，その残響の尾が後続する音素をマスクするという現象である。

荒井ら [2, 3] はこの overlap-masking による聴き取りの低下を防ぐため，定常部抑圧処理を提案した。遷移部は音声知覚に関する情報を多く有しているのに対し，定常部はエネルギーが相対的に大きいものの比較的その情報が少ないとされている [4]。そのため，定常部を抑圧することにより，それに後続する遷移部にかかる overlap-masking 量を減少させることが可能となり，音声の明瞭性を改善することができると考えられている。また，定常部は主に母音に存在し，逆に遷移部は主に子音やその前後に存在する。先行研究 [2, 3, 5] では，定常部抑圧処理を用いて単音節明瞭度試験が行われ，特定の条件下でその効果が確認されている。しかし，文章了解度試験では，定常部抑圧処理の音声明瞭度改善に対する明確な効果が確認されていない。

また，千葉ら [6] は単語了解度試験において，音声明瞭度を改善することを目的とした母音定常部抑圧処理を行った。これは定常部と判断された区間の内，子音部と見なされるものは原音声のまま保持し，母音部と見なされるものは従来どおり抑圧を行うという手法である。この手法を用いて音声に処理を施し，実験参加者に正解率と聴き取りにくさ [7] について評価を行ってもらった。その結果，正解率・聴き取りにくさのどちらにおいてもこの手法の有意な効果は確認されなかった。

以上のことから，残響環境下における，単語了解度を向上させるためには，手法の改善が必要であると考えられる。そこで本研究では，その改善として子音強調に着目した。音声知覚に関する情報が多く含まれる子音部を強調することで，残響が付加された後にも明瞭な音声になると考えられる。よって本研究では，母音定常部抑圧処理に加え，子音強調を音声に施し，残響環境下での単語了解度を向上させることを目的とした。

2. 処理方法

本研究では，千葉ら [6] の提案した母音定常部抑圧処理に加え，子音強調を行った。従来の定常部抑圧処

---

* Preprocessing using consonant emphasis and vowel suppression for improving speech intelligibility in reverberant environments,
by Misaki Tsuji, Takayuki Arai and Keiichi Yasu.

*1 上智大学理工学部
† 現在，パイオニア㈱
（問合先：荒井隆行 e-mail: arai@sophia.ac.jp）

（2012 年 4 月 11 日受付，2012 年 10 月 15 日採録決定）

理や母音定常部抑圧処理では，処理後も母音部の振幅が子音部の振幅よりも大きくなることが多かった。そこで本実験では，更にその効果を最大限高めるため，各子音・母音区間における最大振幅によって区間内振幅を正規化することを試みた，処理後の振幅が母音よりも子音の方が一律に大きくなるような処理を提案する。具体的な処理方法は以下のとおりである。

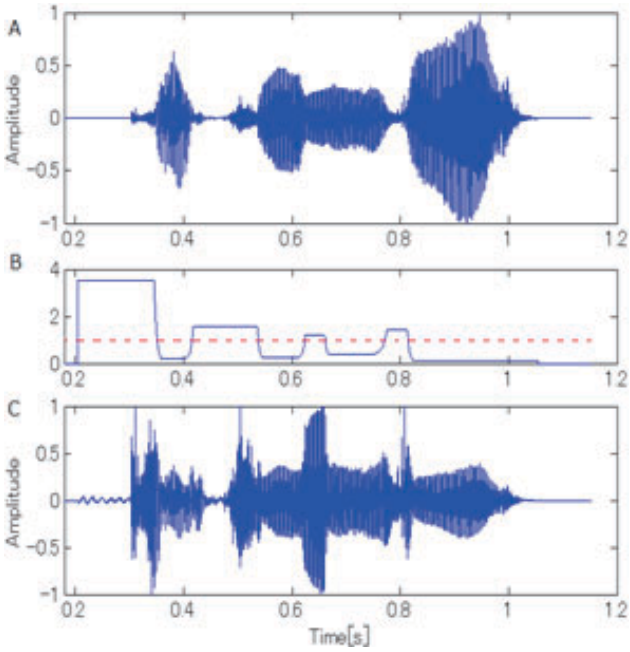
まず初めに，Praat [8] を用いて，音声に対し時間波形やスペクトログラム上におけるフォルマントなどの時間変化や周波数の特徴から，手動でセグメンテーションを行った。手動で行った理由は，本研究ではセグメンテーションを自動的に行うことによって生じるエラーの影響を排除するためである。そしてその結果から，音声を子音部・母音定常部・母音遷移部の三つに分けた。その後その境界の位置情報を Matlab に取り込み，子音区間ごとに処理後の最大振幅が 1.0 となるように増幅/抑圧を行った。母音区間においては，母音区間ごとに処理後の最大振幅が 0.2～1.0 の間で定められた規定値となるように増幅/抑圧を行った。更に母音遷移部においては，なだらかに増幅/抑圧が行われるよう，指数関数を適用した。ここで，子音から母音への遷移部には式 (1) を，母音から子音への遷移部には式 (2) を用いた。式 (1) の a_{c} は遷移部の直前にある子音部の増幅率，a_{v} は遷移部の直後にある母音定常部の増幅率を表す。式 (2) の a_{v} は遷移部の直前にある母音定常部の増幅率，a_{c} は遷移部の直後にある子音部の増幅率を表す。更に n は遷移部の開始点を基準としたときのサンプル番号を表す。r は変数であるが，複数の単語に対して r を変化させた結果，処理後の最大振幅が 1 を超えないように $r = 25$ とした。今後，処理後の波形における子音部の最大振幅を 1.0 としたときの母音定常部の最大振幅を CV 比（本研究では 0.2～1.0）として表現する。**図–1** に原音声と処理音声の時間波形，及び増幅/抑圧関数を表した。B のグラフにおいて，1 よりも値が大きい箇所が増幅，1 よりも小さい箇所が抑圧であることを示している。B のグラフの破線は，値が 1 である基準線である。処理音声は，原音声に増幅/抑圧関数を掛け合わせたものである。また，以降は上記の手法を CE と呼ぶこととする。

$$\frac{1}{r-1}\cdot\left((a_{\mathrm{c}}-a_{\mathrm{v}})\cdot r^{1-n}+r\cdot a_{\mathrm{v}}-a_{\mathrm{c}}\right) \qquad (1)$$

$$\frac{1}{r-1}\cdot\left((a_{\mathrm{c}}-a_{\mathrm{v}})\cdot r^{n}+r\cdot a_{\mathrm{v}}-a_{\mathrm{c}}\right) \qquad (2)$$

3. 実　　験

本研究では，処理の効果を評価するものとして，単語了解度試験を行った。



図–1　音声波形比較「かちにげ」
A：原音声，B：増幅/抑圧関数，C：処理音声，C:V=1:0.4

3.1 原　音　声

原音声は，NTT-AT 親密度別単語了解度試験用音声データベース（FW03）[9] から，単語親密度 5.5～4.0 の日本語 4 モーラ語を 42 語選出した。これらに対し，アルカディア社製の音声合成ソフトウェア「SPeeCAN SFT5」を用いて女性話者による合成音声を作成した。なお，音声合成ソフトを使用した理由は，セグメンテーションが容易だからである。予備実験として 3 名に対し 62 語を対象とした了解度試験を行い，その結果，了解度が 100%となったものから刺激として用いる 42 語を選び出した。

3.2 処 理 条 件

処理音声として，CV 比が 0.2, 0.4, 0.6, 0.8, 1.0 となるように，原音声に処理を施したものを用いた。なお，これ以降，それぞれの CV 比に対応して，これらの処理を CE2, CE4, CE6, CE8, CE10 と呼ぶことにする。また，従来法である定常部抑圧処理（sss）も原音声に施し，CE と共に聴取実験に使用した。この時，単音節を対象とする先行研究において効果が確認された抑圧率 40%を用いることとする [10]。なお，ここで抑圧率とは，元の音声の振幅を 100%としたときの抑圧後の音声の振幅が 40%になることを意味する。

以上，処理条件は原音声（non），CE の 5 条件，sss の計 7 条件である。

3.3 残 響 条 件

残響については，鎌倉芸術館内にある大ホールと小ホールで測定されたインパルス応答を用い，それぞれの残響時間は 1.43, 2.57 s であった。なお，これ以降，

それぞれの残響時間に対応して，これらの残響条件をRT1（1.43 s）と RT2（2.57 s）と呼ぶことにする。これらの残響を原音声・処理音声にそれぞれ畳み込み，正規化したものを刺激とした。

3.4 刺　　激

提示条件は，処理条件と残響時間を組み合わせて 14 条件となる。これらすべての条件ごとに 42 語が存在するので，全刺激は 588 語になる。ただし，ある単語が 1 人の参加者に一度だけしか割り当てられないようにしたため，1 人の参加者が聞く刺激は 42 語であり，1 条件につき 3 語ずつが割り当てられた。各参加者で単語と条件の組み合わせが異なるようにカウンタバランスを取り，全体で 14 の刺激セットを用意した。1 人の参加者は，そのうちの一つの刺激セットを聴取した。

3.5 参　加　者

参加者は 18～25 歳（平均 20.6 歳）の 28 名の日本語母語話者で，その内訳は男性 16 名，女性 12 名であった。なお，全員が 1, 2, 4, 8 kHz での聴力レベルが 20 dB 以下であったため，聴力に問題はないと判断した。

3.6 実 験 方 法

聴取実験は上智大学言語聴覚研究センターの遮音室で行った。刺激をヘッドホン（STAX SRM-323A）より提示した。参加者は実験を開始する前に，練習として本番と同じ試行を 3 回行った。その際，提示レベルを参加者の聴き易いレベルに設定した。

参加者には刺激を 1 度だけ聴かせ，聴こえた音を Matlab の GUI 上の入力欄に仮名で入力させた。

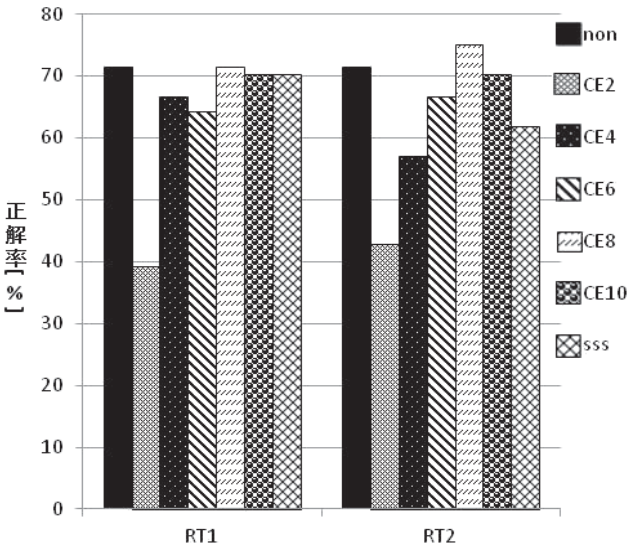
4. 結　　果

4.1 単語了解度試験の結果

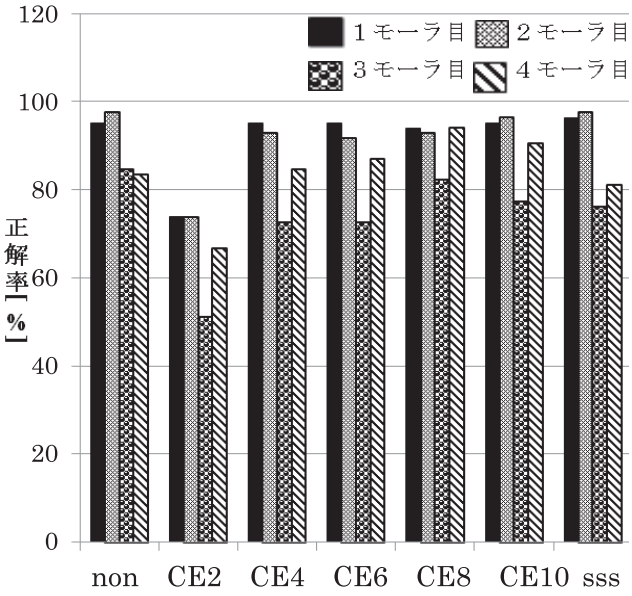
単語了解度試験によって得られた結果を図–2 に示す。この結果は，4 モーラすべてが正解したときを正解率 100%，それ以外を 0%とし，全参加者の結果を平均したものである。

その結果，原音声よりも高くなった条件は，残響時間が RT2 のときの CE8 のみとなった。また，統計ソフト SPSS により，残響時間と処理条件の 2 要因で分散分析（残響時間は 2 水準，処理条件は 7 水準）を行った結果，残響時間には主効果が見られなかった。そのため，残響時間が長くなると聴き取りが低下した先行研究 [2, 3, 5] と同様の結果は得られなかった。一方，処理条件には主効果が見られた（$p < 0.01$）。残響条件と処理条件の間には交互作用は見られなかった。

また，Sidak による多重比較を行ったところ，CE2 は原音声よりも有意に正解率が低かった（$p < 0.01$）。その他 CE4, CE6, CE8, CE10, sss と原音声との間には有意な差は見られなかった。よって，この結果より残響環境下での CE の効果は確認されなかった。また，原音声と sss の間に有意差は見られなかった。



図–2　単語了解度試験の結果

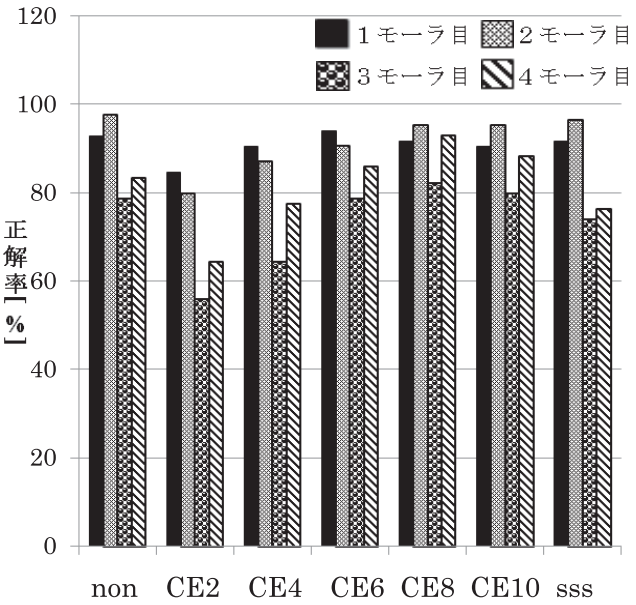


図–3　モーラごとの正解率（RT1）

4.2 モーラごとの正解率

モーラごとに正答であったかを判定した結果を図–3, 4 に示す。この結果に対し，残響条件，処理条件，モーラの 3 要因で分散分析を行った。なお，残響条件は 2 水準，処理条件は 7 水準，モーラは 4 水準であった。その結果，処理条件とモーラには主効果が見られた（$p < 0.01$）。一方，残響時間には主効果が見られなかった。交互作用は処理条件とモーラ間でのみ有意であった（$p < 0.05$）。

更に，Sidak による多重比較を行ったところ，1 モーラ目と 2 モーラ目は，3 モーラ目と 4 モーラ目よりも有意に正解率が高かった（$p < 0.01$）。また，3 モーラ



図–4 モーラごとの正解率（RT2）

目と 4 モーラ目を比較すると，4 モーラ目の方が有意に正解率が高かった（$p < 0.01$）。

5. 考　　察

5.1 単語了解度の考察

今回の実験では，明確な CE の効果を確認することはできなかった。まず CE2 に関しては，いずれの残響条件においても，原音声よりも有意に低い了解度となった（$p < 0.01$）。CE2 とは最大振幅が C:V=1:0.2 となっているものを指す。これは，子音と母音の振幅を増幅・抑圧しすぎると，その間の遷移部の振幅を急激に変化させることとなるため，これが音声の不自然さ・不明瞭さを生み，了解度の低下の原因となった可能性が考えられる。

通常の音声は，子音ごと，母音ごとに振幅が異なるが，CE の手法では，例えば CV 比が 1.0 の場合，子音と母音の最大振幅をともに一定値に揃えることにより，ダイナミックレンジを有効に使うことが可能となる。今回の実験では CV 比が 0.8 のときに最も高い了解度となった。子音と母音の最大振幅が等しいときよりも，子音の最大振幅が母音よりある程度大きいときの方が，明瞭な音声になる可能性を有しているが，このことは今後検証していく必要がある。

5.2 モーラごとの比較による考察

モーラごとの正解率を分析すると，1 及び 2 モーラ目は，3 及び 4 モーラ目よりも有意に正解率が高く（$p < 0.01$），4 モーラ目は 3 モーラ目よりも有意に正解率が高かった（$p < 0.01$）。まず，1 及び 2 モーラ目の正解率が高かった理由としては，単語（ターゲット）の直前にキャリアが存在しなかったことが挙げられる。キャリアが存在しないと，残響が畳み込まれても，1 及び 2 モーラ目には overlap-masking による影響が少ないため，明瞭性はあまり低下しない。しかし，3 及び 4 モーラ目には，1 及び 2 モーラ目の残響の尾がオーバラップするため，明瞭度が低下し易くなると考えられる。

次に，3 モーラ目と 4 モーラ目を比較すると，原音声や sss では 3 モーラ目と 4 モーラ目の正解率はほぼ同程度となったのに対し，CE では 4 モーラ目は 3 モーラ目よりも有意に 10%程度改善している（$p < 0.05$）。また，CE8 の 4 モーラ目と原音声の 4 モーラ目の正解率を比較すると，有意ではなかったが 10%程度 CE8 の方が高かった。この理由として，単語の発話がすべて終了した後に着目したとき，刺激の最後には主に 4 モーラ目の残響の尾が存在していることが考えられる。1 から 3 モーラ目は，その後に次の音素が続くため，それぞれの残響の尾を単独で聴くことはできない。しかし 4 モーラ目は，次に続く音素がないため，その残響の尾を単独で聴くことができる。このことから，CE では，母音定常部が抑圧され，かつ 4 モーラ目の子音が強調されたことによって，overlap-masking 量が減り，4 モーラ目の正解率が高くなったものと推測される。

これに関連するものとして，荒井 [11] が提案した定常部零挿入処理がある。これは定常部の中央に零系列を挿入することによって残響環境下での overlap-masking の影響を低減し，音声明瞭度を高めるものである。更に，これを応用して，音節間や文節間に零系列を挿入する手法も考えられる。この音節間に零系列を挿入する手法と，CE とを組み合わせることにより，残響環境下での聴き取りの改善が期待される。

6. ま　と　め

本研究では，残響環境下での単語・文章の聴き取りの改善を目的とした，子音強調及び母音抑圧法について述べた。処理音声に残響を畳み込んだ刺激を用いて単語了解度試験を行った結果，各子音部の最大振幅を 1.0 とし，各母音部の最大振幅を 0.4〜1.0 とした場合に関しては，原音声と比較して有意差はなかった。一方，各母音部の最大振幅を 0.2 とした場合に関しては有意に了解度が低下した。このように健聴者に対して，母音を抑圧し過ぎると了解度が低下するという知見を得たが，健聴者に対する結果に有意差が得られない場合であっても高齢者にとって了解度が改善した例が先行研究で報告されていることから [12, 13]，対象を高齢者にまで広げて処理の効果を検討していきたい。

謝　　辞

本研究は，文部科学省私立大学学術研究高度化推進

事業上智大学オープン・リサーチ・センター「人間情報科学研究プロジェクト」の支援を受けて行われた。

聴取実験に当たり，インパルス応答を提供して下さった東京大学生産技術研究所（当時）の橘秀樹先生，上野佳奈子先生，横山栄先生に感謝申し上げます。また，技術的なアドバイス，助言を下さった TOA 株式会社の栗栖清浩氏と，東海大学の程島奈緒先生に深く御礼申し上げます。

文　献

[1] A.K. Nabelek, T.R. Letowski and F.M. Tucker, "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Am.*, **86**, 1259–1265 (1989).

[2] 荒井隆行, 木下慶介, 程島奈緒, 楠本亜希子, 喜田村朋子, "音声の定常部抑圧処理の残響に対する効果," 音講論集, Vol. 1, pp. 449–450 (2001.10).

[3] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto and T. Kitamura, "Effects on suppressing steady-state portions of speech on intelligibility in reverberant environments," *Acoust. Sci. & Tech.*, **23**, 229–232 (2002).

[4] S. Furui, "On the role of spectral transition for speech perception," *J. Acoust. Soc. Am.*, **80**, 1016–1025 (1986).

[5] N. Hodoshima, T. Arai, A. Kusumoto and K. Kinoshita, "Improving syllable identification by a preprocessing method reducing overlap-masking in reverberant environments," *J. Acoust. Soc. Am.*, **119**, 4055–4064 (2006).

[6] 千葉亜矢子, "残響環境下における処理音声の「聴き取りにくさ」による評価—定常部抑圧処理の効果の検討—," 上智大学大学院理工学研究科理工学専攻 2010 年度修士論文 (2010).

[7] M. Morimoto, H. Sato and M. Kobayashi, "Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces," *J. Acoust. Soc. Am.*, **116**, 1607–1613 (2004).

[8] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," Version 5.2.23, retrieved from http://www.praat.org/ (2011).

[9] 天野成昭, 近藤公久, 坂本修一, 鈴木陽一, "親密度別単語了解度試験用音声データセット（FW03）," NII 音声資源コンソーシアム (2006).

[10] 村上善昭, 程島奈緒, 中田有貴, 林奈帆子, 宮内裕介, 荒井隆行, 栗栖清浩, "残響環境下における音声明瞭度改善のための前処理—定常部抑圧処理の抑圧率と残響の関係—," 音講論集, pp. 649–650 (2006.3).

[11] T. Arai, "Padding zero into steady-state portions of speech as a preprocess for improving intelligibility in reverberant environments," *Acoust. Sci. & Tech.*, **26**, 459–461 (2005).

[12] Y. Miyauchi, N. Hodoshima, K. Yasu, N. Hayashi, T. Arai and M. Shindo, "A preprocessing technique for improving speech intelligibility in reverberant environments: The effect of steady-state suppression on elderly people," *Proc. Interspeech*, pp. 2769–2772 (2005).

[13] T. Arai, N. Hodoshima and K. Yasu, "Using steady-state suppression to improve speech intelligibility in reverberant environments for elderly listeners," *IEEE Trans. Audio Speech Lang. Process.*, **18**, 1775–1780 (2010).