簡易式定常部抑圧処理による残響環境下の音声明瞭性*

☆武田昌也，荒井隆行（上智大・理工），
栗栖清浩，網谷智博（TOA），安啓一（上智大・理工）

1　はじめに

駅のホームや広い講堂などで放送を行うと，音が響いて発話が聴き取りづらくなる．このような残響下における音声明瞭度の低下は，直前の音声に付与された残響の尾が，後続の音声に覆い被さる，overlap-masking が原因の 1 つとして指摘されている[1].

この現象を軽減させる手法として，音声が室内に放射される前に処理を行う前処理[2, 3]と，室内に放射され，残響が音声に付与されてから処理を行う後処理[4]の 2 種類がある．荒井ら[2, 3]は，前処理の一つの方法として，定常部抑圧処理を提案した．この処理は，発話の聴き取りにおいてさほど重要でない音声の定常部[5]（主に母音）の振幅を抑圧し，残響のマスキング量を減らすものである．これにより，特定の残響環境下において overlap-masking の影響が減少し，単音節明瞭度が向上することが報告されている[2-3, 6].

本研究の最終的な目標は，実際の環境において実時間で前処理を行い，残響環境下の音声明瞭度を向上させることである．高橋ら[7]は，安武ら[8]の子音判定アルゴリズムを応用した簡易式定常部抑圧処理（Simple Algorithm for steady-state suppression，以後 SA 法）を提案した．この処理は従来の定常部抑圧処理と異なり，音声の隣接する時間フレームのエネルギ比のみに注目するため，処理が短時間で済み，実時間処理に適している．高橋ら[7]は SA 法において，母音の定常部のみを抑圧するパラメータを検証した.

本報告では，SA 法を施した音声の残響環境下の明瞭性を，聴取実験を通して検証する．先行研究で明瞭度改善の効果が示されている処理[2-3,6]（Filter Bank 法，以後 FB 法）と SA 法の 2 種類の処理方法を用いて聴取実験を行い，その結果を基に SA 法の有用性を検討する.

2　簡易式定常部抑圧処理（SA法）

高橋ら[7]の提案した SA 法は，音声信号をある一定の長さの時間長でフレーム分けする．そして，時間軸上で前フレームと後フレームのエネルギ比を算出し，そのエネルギ値が閾値の範囲内ならば定常部と判断して抑圧する．その際，フレーム長（以後 Frame Size）は子音部の誤判定を抑えつつ，DSP に実装した時の処理時間を短くするため，Frame Size = 30 ms とした．また，入力信号における定常部の範囲の，入りわたりのニーポイント（以後 Lower Limit）を－0.72 dB, 出わたりのニーポイント（以後 Upper Limit）を 6.0 dB とした．このニーポイントは, 代表的な 14 種の単音節 (/k/, /g/, /s/, /ʃ/, /dz/, /dʒ/, /t/, /tʃ/, /d/, /n/, /h/, /b/, /p/, /m/)で，母音の定常部を抑圧し，かつ子音の定常部を抑圧しないように目視で調節されている．定常部における抑圧率（以後 Gain Steady）に関しては, FB 法と同様に 40%とし，それに相当する 7.95 dB 抑圧した．また，抑圧箇所の立ち下がりにかかる時間（以後 Fall Time）と, 立ち上がりにかかる時間（以後 Rise Time）を共に 10 ms とした．以上をまとめると，連続した 2 つの 30 ms フレーム W_1, W_2 のエネルギ E_1, E_2 を算出し，

-0.72 dB $< 10\log_{10} E_1/E_2 <$ 6.0 dB の時,

W_1 の振幅を 7.95 dB 抑圧した.

Fig. 1 に実験で用いた単音節/ka/の原音声と，SA 法による処理後の音声，Fig. 2 に SA 法で用いた隣接するフレーム間のエネルギ比に対する入出力関数を示す.

* Intelligibility of speech processed by simple algorithm for steady-state suppression in reverberant environments, by TAKEDA, Masaya, ARAI, Takayuki (Sophia Univ.), AMITANI, Tomohiro, KURISU, Kiyohiro (TOA) and YASU, Keiichi (Sophia Univ.)

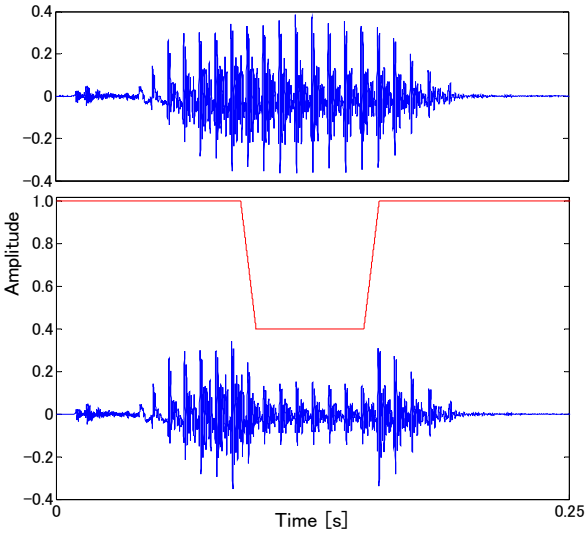

Fig. 1　単音節/ka/
（上：原音声の時間波形，下：SA 法による処理後の時間波形と抑圧関数）

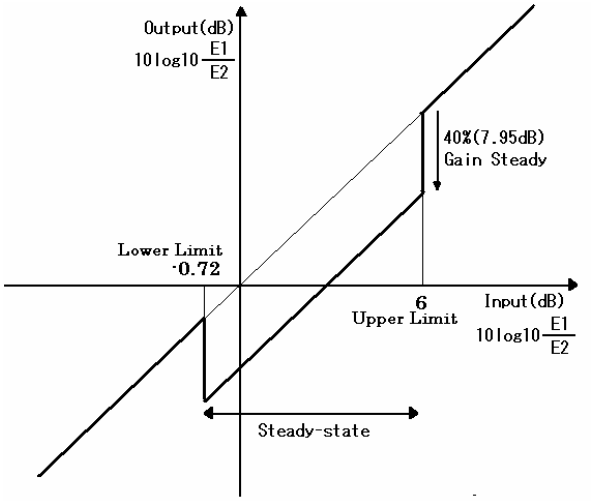

Fig. 2　SA 法で用いた隣接するフレーム間のエネルギ比に対する入出力関数
（[7]を参考に改変）

3　実験

3.1 目的

単音節明瞭度試験を行うことで，SA 法を施した音声の明瞭性を評価することを目的とした．

3.2 刺激

聴取実験の刺激作成に用いた原音声は，[6]と同じものを用いた．この原音声は，ATR 研究用日本語音声データベース（話者：MAU，40 歳男性）を用いて作成されている．日本語の単音節 CV をターゲット音節とし，キャリアセンテンス「題目としては＿といいます」に挿入されている．

ターゲット音節の子音は代表的な 14 種の単音節(/k/, /g/, /s/, /ʃ/, /dz/, /dʒ/, /t/, /tʃ/, /d/, /n/, /h/, /b/, /p/, /m/)，母音は/a/が用いられている．キャリアセンテンスとターゲット音節の音圧レベル比は 1:0.7 である．なお，ターゲット音節へのオーバーラップのエネルギ量を単音節毎で固定するため，「題目としては」からターゲット音節の母音/a/までの時間は全単音節で 150 ms に統一されている．

この原音声に FB 法と SA 法を施し，処理の条件は未処理，FB 法処理，SA 法処理の 3 条件とした．また，インパルス応答を各刺激に畳み込むことによって，残響を付与した．使用したインパルス応答は東大和市大ホール（残響時間：RT = 1.3 s）の RT を変化させたものである．RT の変更は，インパルス応答における時間包絡の時定数を変化させることで行い[9]，実験では RT が 0.8 s, 1.0 s, 1.2 s の，3 種類のインパルス応答を作成した．なお，使用したキャリアセンテンスとターゲット音節，インパルス応答の標本化周波数は 16 kHz，量子化ビット数は 16 bit であった．

以上より，単音節 14 種類，処理方法 3 条件，残響時間 3 条件の計 126 刺激（14×3×3）を実験に用いた．

3.3 実験参加者

実験参加者は 18～31 歳（平均 23 歳）の健聴者 17 名（男性 9 名・女性 8 名）であり，いずれも日本語母語話者であった．健聴者か否かは，参加者の自己申告で確認した．

3.4 実験手順

実験は防音室で行われた．実験機材については，パソコンと USB オーディオインターフェース (Roland UA-25EX)を接続し，ヘッドフォン（STAX SRM-313）を USB オーディオインターフェースの出力端子に接続した．

参加者はヘッドフォンを着用し，聴こえた単音節をパソコンのスクリーン上で回答した．スクリーン上には単音節と同じ 14 個のボタンがあり，そのボタンをマウスでクリックして回答した．なお，刺激の順番はランダムにし，各刺激は一度だけ提示されるものとした．

3.5 仮説

実験を行うにあたり，以下の仮説を立てた．

仮説①　未処理の刺激と比べ，SA 法を施した刺激は，FB 法の刺激と同様に音声明瞭度が向上する．

仮説② SA 法の音声明瞭度は，各残響条件で FB 法と同程度である．

3.6 結果

実験結果を Fig. 3 に示す．横軸は条件，縦軸はターゲット音節の正解率を表す．実験結果に対し，処理 3 条件×残響時間 3 条件において，統計ソフト SPSS を用いて分散分析を行った. その結果, 処理条件の主効果($p < 0.01$), 残響条件の主効果($p < 0.01$)が見られ，処理条件と残響条件の交互作用は確認されなかった.

また，すべての残響条件において未処理と SA 法の正解率に有意差が見られ($p < 0.01$), 未処理と FB 法の正解率にも有意差が見られた($p < 0.01$)．しかし，SA 法と FB 法の正解率には有意差が見られなかった.

以上の結果を基に，仮説の検討を行う．①に対しては，未処理と SA 法，未処理と FB 法の，それぞれの正解率の差は有意であることから，仮説は支持された．②に対しては，SA 法と FB 法の正解率の差は有意ではないことから，仮説は支持された.

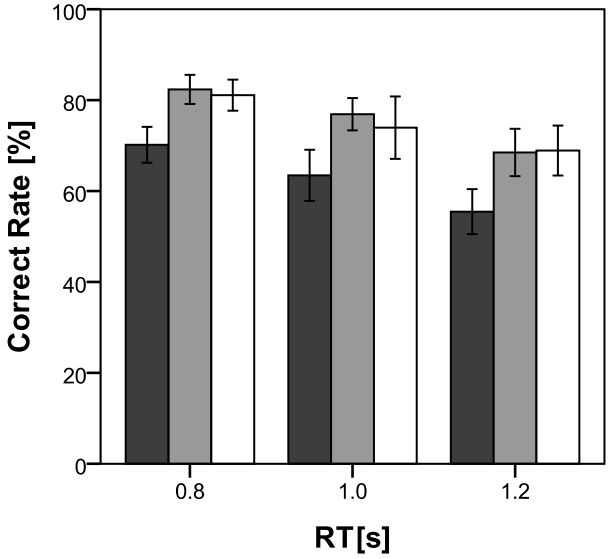

Fig. 3 未処理（黒），SA 法（灰），FB 法（白）のターゲット音節の正解率
（図中のバーは 95%信頼区間を示す）

4 考察

4.1 子音ごとの検討

処理条件に対する, 子音ごとの正解率を Fig. 4 に示す．Fig. 3 と同様に，棒グラフの色で処理条件を示している.

SA 法における正解率が FB 法より高かった子音は/k/（無声軟口蓋破裂音），/s/（無声歯茎摩擦音），/ʃ/（無声歯茎硬口蓋摩擦音），/t/（無声歯茎破裂音）, /d/（有声歯茎破裂音）, /h/（無声声門摩擦音）, /b/（有声両唇破裂音）, /p/（無声両唇破裂音）であり，FB 法における正解率が SA 法より高かった子音は/g/（有声軟口蓋破裂音）, /tʃ/（無声歯茎硬口蓋破擦音）, /n/（有声歯茎鼻音）であった.

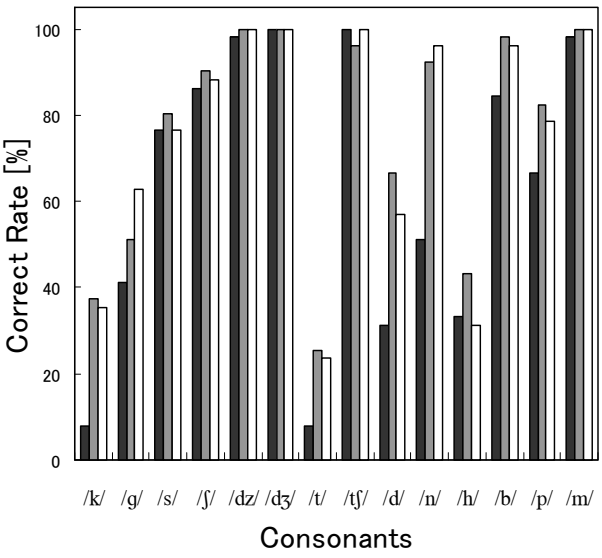

Fig. 4 子音ごとの正解率
（黒：未処理 灰：SA 法 白：FB 法）

4.2 リアルタイム化への試み

高橋ら[10]は，SA 法の処理を TI 社の TMS320C6713DSK に実装し, リアルタイム化を目指した．実装後，MATLAB 上で動作させた際の結果と，DSP の開発環境である Code Conmposer Studio ver. 2.21 の Simulator 上（以後 CCS Simulator）での結果を比較することにより，実装の正確性を検証した．単音節の出力波形において，フレーム間レベル差が MATLAB 上と CCS Simulator 上で一致する（誤差 0.001%未満）ことにより，正確に実装されていることを確認した．しかし，DSK 実機において単語「しゃくなげ」における出力波形を MATLAB 上のものと比較したところ, 両者の抑圧箇所は一致していなかった．このように, DSK 実機で定常部と判定されている箇所が MATLAB 上とは異なるため，両者で処理した音声の明瞭度に差が出る可能性がある.

この問題を考慮し，実環境を想定したハードウェアを目指すため, TMS320C6720DSP を搭載した試作機 SPeeCH システムが開発されている．SPeeCH システム本体の外観を Fig. 6

に示す.

SPeeCH システムにおける動作の正確性を確認するため，単語「しゃくなげ」を SPeeCH システムと MATLAB 上に入力し，出力音声波形を比較した．また，SPeeCH システムと MATLAB 上でエネルギ比の計算を行うフレームの位置を調節した．なお，抑圧箇所を比較しやすくするため，Gain Steady = 0%, Rise Time = 0 ms, Fall Time = 0 ms に設定した．

結果は，MATLAB，SPeeCH システムの両者において，抑圧箇所がフレーム単位で一致していることがわかった．Fig. 7 (a)に「しゃくなげ」の原音声，(b)に MATALB 上の出力波形，(c)に SPeeCH システムの出力波形を示す．これより，SPeeCH システムにおける動作の正確性を確認した．

Fig. 6　SPeeCH システム（試作機）本体

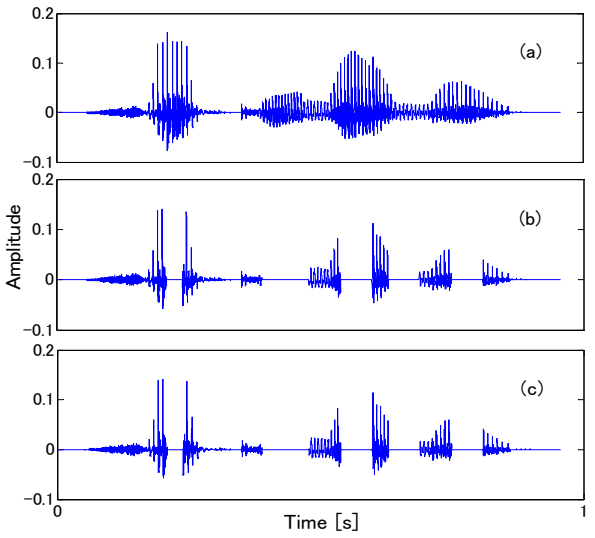

Fig. 7　単語「しゃくなげ」
(a)原音声　(b) MATLAB の出力波形
(c) SPeeCH システムの出力波形

5　まとめ

SA 法を施した音声の，残響下における音声明瞭性を評価するため，単音節明瞭度試験を行った．単音節は 14 種，処理は未処理，FB 法, SA 法の 3 条件, 残響時間は 0.8 s, 1.0 s, 1.2 s の 3 条件で, 総計 14×3×3 = 126 刺激を用いて実験を行った．実験の結果，未処理の音声と比較して，FB 法，SA 法の音声明瞭度は同程度改善された．また，SA 法が実装された SPeeCH システムにおいて，動作の正確性を確認した．

本報告の実験では SA 法を MATLAB 上で動作させている．そこで今後は，SPeeCH システム実機を用いて単音節や単語，文章の明瞭度試験を行い，さらに実環境に近い評価をしていきたい．

謝辞

本研究の一部は，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「人間情報学研究センター」の支援を受けて行われた．

参考文献

[1] Nábêlek *et al.*, *Acoust. Soc. Am.*, 86(4), 1259-1265, 1989.
[2] 荒井他，音講論(春), 449-450, 2001.
[3] Arai *et al.*, *Acoust. Sci. Tech.*, 23(4), 229-232, 2002.
[4] Allen *et al.*, *J. Acoust. Soc. Am.*, 62(4), 912-915, 1977.
[5] Furui, *J. Acoust. Soc. Tech.*, 80(4), 1016-1025, 1986.
[6] Hodoshima *et al.*, *J. Acoust. Soc. Am.*, 119(6), 4055-4064, 2006.
[7] 高橋他，電子情報通信学会技術研究報告, 107(26), 11-16, 2007.
[8] 安武他，電子情報通信学会技術研究報告, 105(479), 79-84, 2005.
[9] Hodoshima *et al.*, *Proc. Forum Acusticum Sevilla*, 2002.
[10] 高橋他，音講論(春)，865-866, 2008.