

定常部抑圧処理による残響環境下の音声強調

高橋 慶[†] 安 啓一[†] 程島 奈緒[†] 荒井 隆行[†] 栗栖 清浩[‡]

[†] 上智大学 理工学部 電気・電子工学科 〒102-8554 東京都千代田区紀尾井町 7-1

[‡] TOA 株式会社 〒665-0043 兵庫県宝塚市高松町 2-1

E-mail: [†] kei-taka@sophia.ac.jp

あらまし 荒井ら[音講論 2001; Acoust. Sci. Tech., 2002]は, 残響に対して音声明瞭度を改善する定常部抑圧処理を提案し, その有効性を示している. 一方, 安武ら[電子情報通信学会ヒューマン情報処理研究会, 2005]は, 雑音環境下での補聴と実時間処理を念頭に置いた子音強調処理を提案している. 我々の最終目的は, 定常部抑圧処理を実時間で表現することである. 本論文では, 母音の抑圧は結果的に子音を強調している点に着目し, 2つの先行研究が統一的に扱えるかを考察した. その結果, 安武らによる子音強調処理と同じ枠組みで定常部抑圧処理を実現できることを確認した.

キーワード 音声強調, 母音抑圧, 残響, 実時間処理

Enhancing Speech in Reverberation by Steady-State Suppression

Kei TAKAHASHI[†] Keiichi YASU[†] Nao HODOSHIMA[†] Takayuki ARAI[†] and
Kiyohiro KURISU[‡]

[†] Department of Electrical and Electronic Engineering, Sophia University 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554
Japan

[‡] TOA Corporation 2-1 Takamatsu-cho, Takarazuka, Hyogo, 665-0043 Japan

E-mail: [†] kei-taka@sophia.ac.jp

Abstract In order to improve speech intelligibility in reverberation, steady-state suppression was proposed by Arai *et al.* [Proc. Autumn Meet. Acoust. Soc. Jpn., 2001; Acoust. Sci. Tech., 2002]. On the other hand, a consonant enhancement technique for real-time processing in noise was proposed by Yasutake *et al.* [Tech. Rep. of IEICE Japan, 2005]. Our ultimate goal is to implement steady-state suppression in real-time. We discussed whether these two techniques are treated within a single unified framework because suppressing energy in vowels essentially enhancing consonants. As a result, we confirmed that the consonant detection method used in Yasutake *et al.* was included in steady-state suppression for real-time processing.

Keyword speech enhancement, vowel suppression, reverberation, real-time processing

1. はじめに

講義や講演において拡声が行われる際, 音声の明瞭性が劣化することがある. これは, 先行する音声に付加された残響の尾が後続の音素に影響を与える *overlap-masking* が原因であると考えられている[1, 2]. 例えば, 図 1 の左図は, 'october' という発話を音素ごとにセグメンテーションしたものであり, 右図は, その発話に残響時間 1.1 秒のインパルス応答を畳み込んだものである. 図 1 から, /k/, /t/, /b/ のようなエネルギーの弱い子音は, そ

の前に存在する /o/ のようなエネルギーの強い音素に付加された残響によって覆われてしまうことがわかる.

overlap-masking に対して, 荒井ら[3, 4]は, 音声明瞭度を改善するために定常部抑圧処理を提案しており, Hodoshima *et al.* [5]や後藤ら[6]により特定の残響環境下において, その有効性が実証されている. 一方, 安武ら[7]は, 雑音環境下での補聴を目的として, 実時間処理を念頭に置いた子音強調処理を提案している. 安武ら[7]の子音強調処理に用いられる子音判定法は, 時間軸上の連続するフレームのエネルギー変化をみるとい

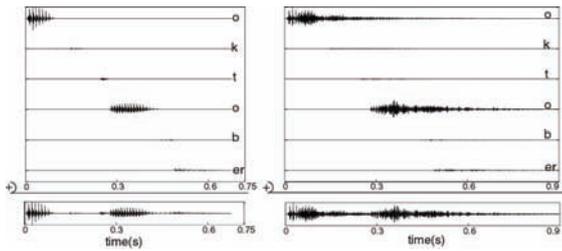


図 1：発話 ‘october’を用いた overlap-masking の例 [5]
(左図:原音声，右図:原音声に残響をふかした音声)

う単純な手法のため，従来の子音判定に用いられるような複雑な処理より，処理を軽くすることができる．

我々の最終目標は，定常部抑圧処理を PA (public address) に組み込んで実時間で実現することである．それに向けて本論文では，定常部抑圧処理が母音を抑圧することで結果的に子音を強調しているという点に着目し，2つの先行研究[3,4,7]のアルゴリズムを統一的に扱えることを考察した．そして，安武らによる子音強調処理と同じ枠組みで，子音判定法だけを変えることで定常部抑圧処理が実現できることを確認した．

2. 音声強調処理

安武らが提案した子音強調処理[7]では，時間軸上の隣接した2つのフレーム(w_1 , w_2)を見て，その2つのフレーム内のレベルを比較することで，子音部の判定を行っている(図 2)． $w_1 > w_2$ なら子音とし，またその逆なら母音としている．一方，荒井らによって提案された定常部抑圧処理[3,4]でも前後フレームのエネルギー変化を見て，その変化の傾きが閾値よりも大きい場合は遷移部，小さい場合は定常部として判定を行っている．子音は定常性が低い一方，母音の中央部は定常性が高いので，結果的に母音の中央部が定常部として判定されることが多い．

2つの処理[3,4,7]は前後フレームのエネルギー変化を見てフレームごとにレベル処理をするかどうかを判定しているという点で共通している．一方，相違点は主に以下の4つの点にまとめられる：

- 1) 子音強調処理では前後2フレームだけがエネルギー変化を見る範囲であるのに対し，定常部抑圧処理では前後2フレームずつ，合わせて5フレームが対象．
- 2) 子音強調処理では全帯域のエネルギー変化を見てのに対して，定常部抑圧処理では聴覚特性に近い1/3-oct帯域ごとのエネルギー変化を平均的に見ている．
- 3) 子音強調処理では補充現象の補償を目的とした強調処理が行われているのに対し，定常部抑圧処理

ではマスキング量を減らす目的で定常部を抑圧する処理が施されている．

- 4) 子音強調処理では子音から母音への入りわり部のみを処理の対象としているのに対して，定常部抑圧処理では出わり部に対しても処理の対象としている．

本研究では安武らによる子音強調処理の枠組みの中で定常部抑圧処理が実現できるかどうか調べることが目的である．上記のうち，1)と2)に関しては特に実時間処理を考えた際に問題になる点である．1)はアルゴリズムに内在する遅延を決定することになる．前後2フレームのみを用いた場合，フレーム長が子音判定に及ぼす影響については，2.1節で検討する．2)は処理の計算量に関する．処理が軽いことも DSP (digital signal processor) 上での実装のし易さに関係してくることから，本研究では全帯域を一括に扱うこととする．3)と4)については，安武らが用いている補充現象を補償する関数の形状に関することであり，これについては2.2節で検討する．

2.1. フレーム長の検討

図3にフレーム長を10msから60msまで変化させ，そのときの子音判定の精度を示す．図3よりフレーム長が10ms，と20msでは，子音と母音の誤判定が多く，また30ms以上では判定結果がほぼ変化しなかった．PAに実装したときのことを考えると，子音部の誤判定をある程度防ぎながら処理の時間遅延を最小限に抑えるため30msを用いることとした．なお，各フレーム内におけるレベルとしては，実効値を用いた．

2.2. 補充現象の補償を目的とした関数の検討

安武ら[7]が用いていた補充現象の補償を目的とした関数上にあるニーポイント(増幅特性の切り替わり点)を，本論文では，定常部と遷移部の境目とみなし，ニーポイントを設定した．

ニーポイントの設定方法としては，代表的な14種の単音節 (/p/, /t/, /k/, /b/, /d/, /g/, /s/, /h/, /ə/, /ɔ̃/, /ɔ̃ŋ/, /dz/, /m/, /n/)を用いて，各々の子音部を抑圧しないように目視で確認しながら調整した．また，この際のニーポイントは立ち上がり部分と立ち下がり部分の2点求めた．

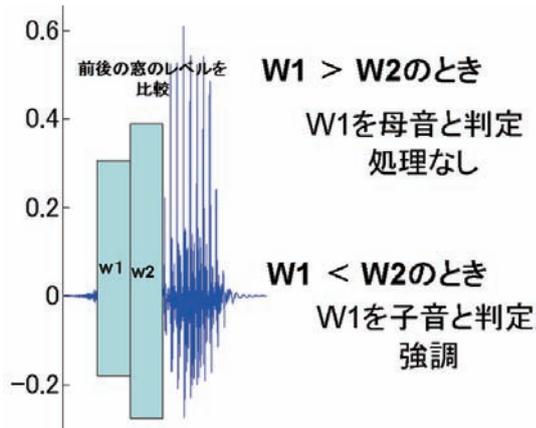


図 2 : 安武ら[3]の子音判定アルゴリズム

結果として、補充現象を補償する関数上に表した定常部抑圧処理を図 4 に示した。ニーポイントの設定方法から立ち上がりの点を -0.72dB と置き、また、立ち下りの点を 6dB と置いた。そして、ニーポイントの間を定常部として考え、抑圧することとした。また、抑圧率は、定常部抑圧処理の抑圧率[7]で効果が得られている抑圧率(40%)と対応させるため、定常部の箇所ではレベルを 7.95dB 下げることとした。ここに提案する定常部抑圧処理を施した音声を図 5 に示した。

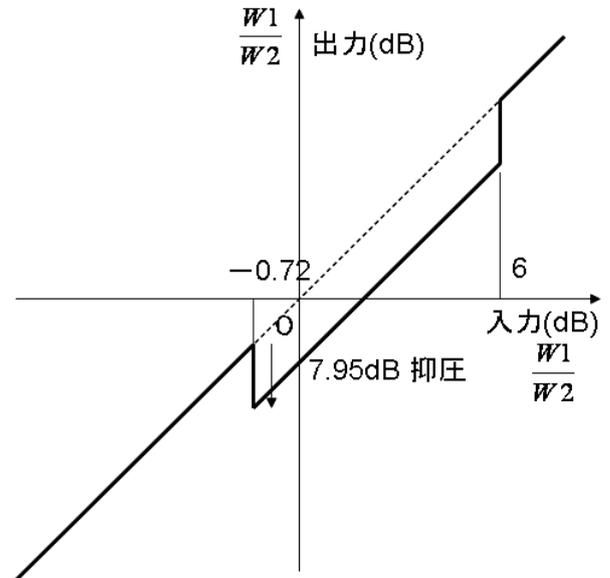


図 4 安武ら[7]が用いた補充現象を補償することを目的とした関数を定常部抑圧処理用に修正したもの

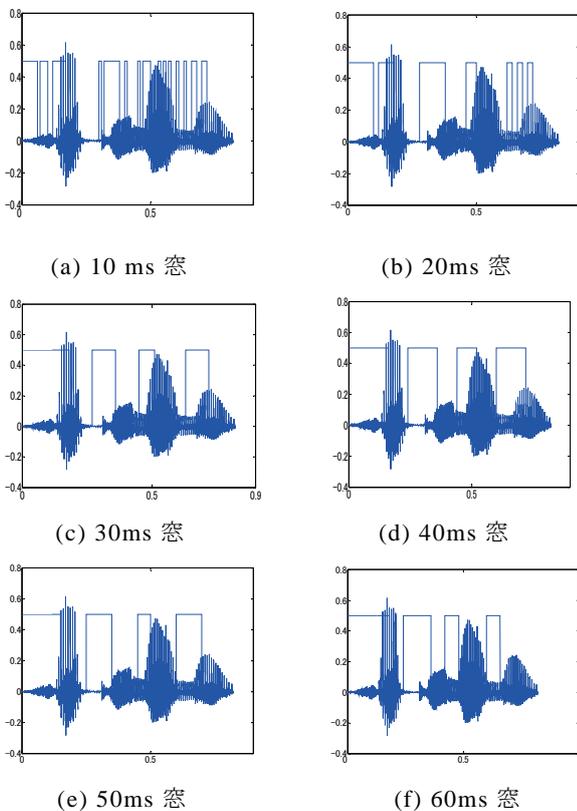
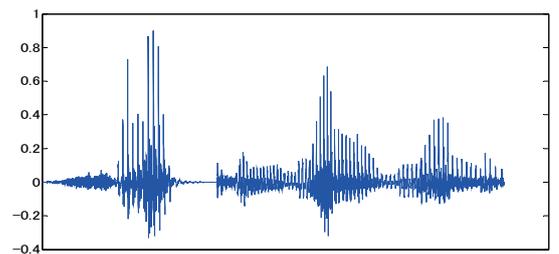
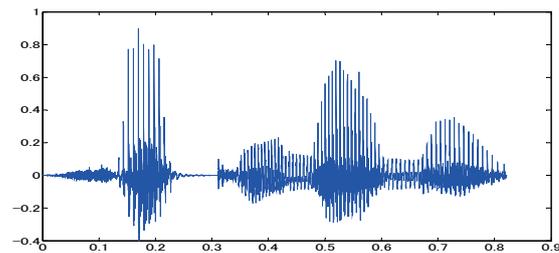


図 3 発話「しゃくなげ」の子音部検出結果



(a) 定常部抑圧処理(提案法)



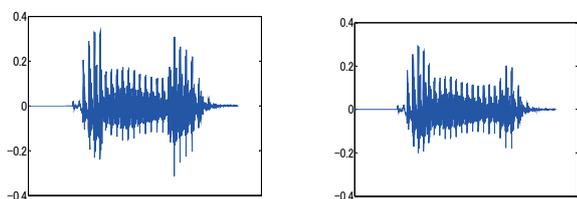
(b) 原音声(提案法)

図 5 定常部抑圧処理(提案法)を施した音声と原音声

3. 結果・考察

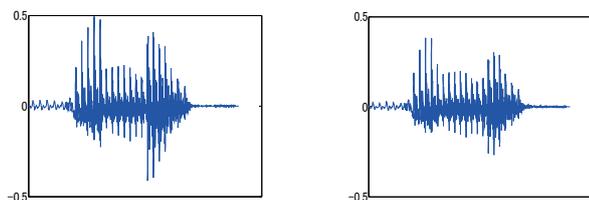
3.1. 従来法と提案法との比較

従来の定常部抑圧処理[3,4]を基に、本論文で示す提案法によって検出された定常部判定法にどの程度違いがあるかを調べた。その際、従来法の抑圧箇所を基準とし、提案法の抑圧箇所との一致率(%)を求めた。比較の際に用いた単音節は、ニーポイントの設定の際に用いた14種類の単音節のうちの無声閉鎖音/p/,有声閉鎖音/d/,無声摩擦音/ə/,無声破擦音/ʈə/,有声破擦音/ʂɨ/,鼻音/m/の6種類とした。また、この際のターゲットは、Hodoshima *et al*[5]の実験に用いられている刺激文からターゲットの単音節の部分を抜き出したものである。



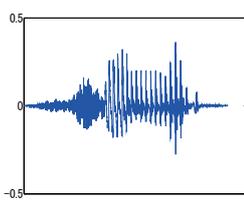
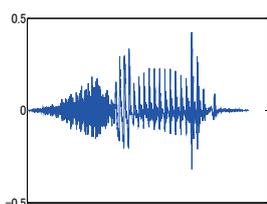
(a) 提案法 (b) 従来法

図6 無声閉鎖音/p/の場合



(a) 提案法 (b) 従来法

図7 有声閉鎖音/d/の場合



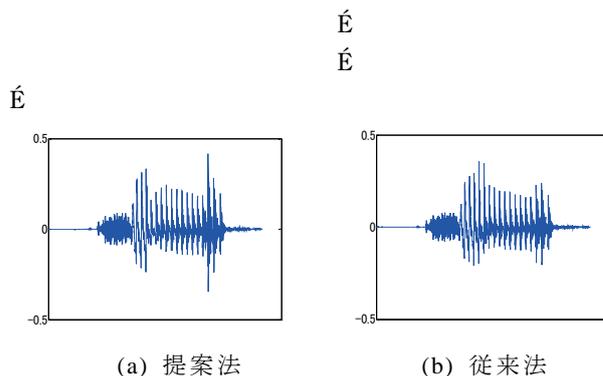
(a) 提案法 (b) 従来法

図8 無声摩擦音/əの場合

É

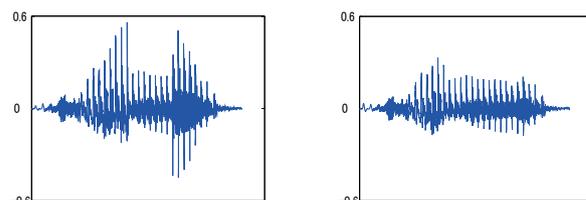
表1 提案法と従来法の定常部判定の一致度 [%]

単音節	pa	da	əaÉ	ʈəaÉ	ʂɨaÉ	ma
一致率	79.8	82.8	97.2	82.4	76.3	79.5



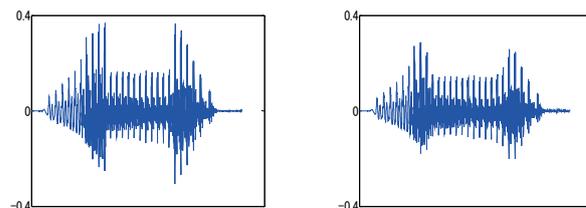
(a) 提案法 (b) 従来法

図9 無声破擦音/ʈə/の場合



(a) 提案法 (b) 従来法

図10 有声破擦音/ʂɨ/の場合

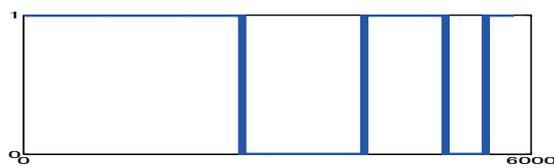


(a) 提案法 (b) 従来法

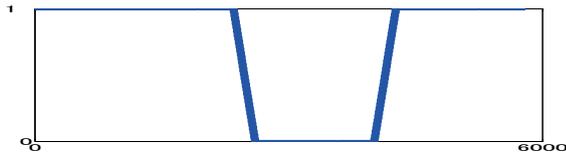
図11 鼻音/m/の場合

図6-図11には、提案法と従来法の時間波形を示し、表1には、6種類の単音節の提案法と従来法の一致率を示した。また、鼻音/m/+母音/aについては、定常部判定がどのようになっているかを調べた。遷移部なら1とし、定常部なら0とした。また、従来法は、抑圧の際、急激に振幅が変化しないように台形補正を行っているが、この部分に関しては、定常部内であるので、0とした。なお、ここでは母音定常部の検出精度を比較したいので、従来法[3,4]に加え子音定常部を検出しないための追加処理を行った[9,10].

ÉÉÉÉÉ



(a) 提案法



(b) 従来法

図 12 ‘ma’における提案法と従来法の定常部判定
(子音部を 1 とし、母音部を 0)

判定を図 12 に示す。提案法の定常部判定は、立ち上がり部では、従来法より 14.9ms 遅れてから定常部と判定し、立ち下がり部でも、従来法より 14.9ms 早く遷移部として判定している。

表 1 より母音抑圧処理と従来法の抑圧箇所は、約 80% の一致であるということがわかった。しかし、この一致率は、フレームの移動により大幅に変動する可能性がある。そのことについては、3.2 節で述べる。

3.2. 窓長と窓のタイミングの再検討

子音強調処理では前後 2 フレームだけがエネルギー変化を見る範囲であるのに対し、定常部抑圧処理では前後 2 フレームずつ、合わせて 5 フレームが対象であった。このエネルギー変化を見るための時間幅は、実時間処理をする際にこれらのアルゴリズムが生み出す遅延を決定することになる。提案法では前後 2 フレームのみを用いているが、フレーム長を 30ms にしたため分析対象範囲が入力されるのを待つためには最大で 30ms の遅延が生じる (2 フレームの入力が完了した時点で最新フレームに対して処理を施すので 1 フレーム分の遅延)。一方、前後 5 フレームを用いる従来法の場合、フレーム長を 10ms と短く設定すれば、遅延は最大で 30ms となる (5 フレームの入力が完了した時点で中央フレームに対して処理を施すので 3 フレーム分の遅延)。ただし、フレーム長を短くすることが問題になるかは今後調べる必要がある。この場合、フレーム長は 20-30ms 程度のしながらフレームシフト幅を 10ms に設定するなどフレームをオーバーラップさせる工夫で、遅延を抑えながら検出率を高めることが期待される。

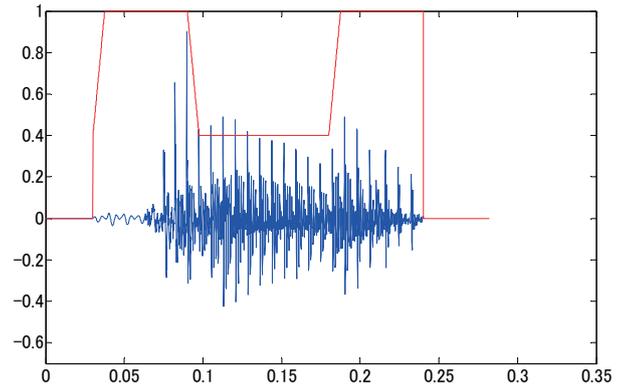


図 13-1 フレームの開始タイミングを 10ms 遅らせた場合の定常部判定結果

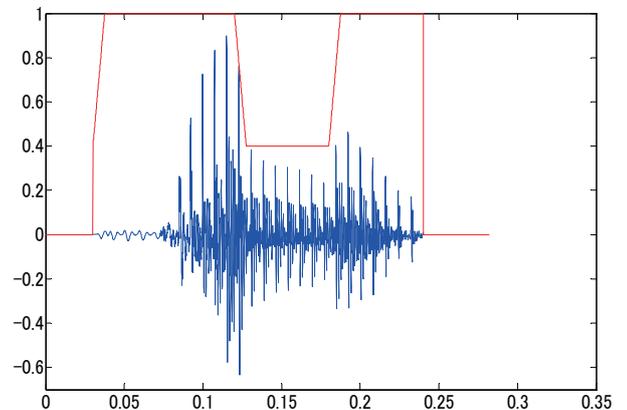


図 13-2 フレームの開始タイミングを 20ms 遅らせた場合の定常部判定結果

3.1 節では約 8 割もの定常部と判定する箇所が一致したが、図 13-1、図 13-2 のように時間を遅らせるとフレームのタイミングがずれ、それによって定常部と判定される区間が変わってくる。上記で述べたフレーム長を短くすることは、この検出区間がずれる問題の解決にもつながると期待される。

3.3. 補充現象と定常部抑圧処理の関係

聴覚障害者は健聴者に比べて継時マスクング量が増加することが報告されている[11]。継時マスクングの補償を目的とした補聴処理としては、Automatic Gain Control (AGC) [12]などの振幅圧縮手法が代表的であるが、一方で処理による包絡線の平坦化などの問題点も指摘されている[13]。黒澤ら[14]は振幅圧縮に加えて包絡線強調処理を施すことによって、模擬難聴環境下での単語了解度の改善を示した[14]。荒井ら[15]は聴覚障害者の継時マスクング量の増加に着目し、定常部抑

圧処理[3, 4]の補聴器応用を提案した。そして小林ら[14, 15]がその提案を受けて、高齢者を対象とした実験を行った。その結果、処理による音声明瞭度の改善が確認された[9, 10]。本報告の子音検出の枠組みとなった安武ら[7]の子音強調処理は補充現象の補償を目的としているが、従来の定常部抑圧処理[3,4]および本報告の提案法も安武らの子音強調処理と同じく相対的に子音を強調させる処理である。よって、定常部抑圧処理の従来法および提案法は補充現象・継時マスキングの両方を補償する処理であると考えられる。

4. まとめ

安武ら[7]が提案する子音判定アルゴリズムを応用して、補充現象を補償することを目的とする関数上で、定常部抑圧処理を擬似的に再現した。また、今後は、主観評価として聴取実験を行う予定である。さらに、ターゲット音節にかかるマスキング量を処理前後で比較するなどの客観評価を検討することも考えられる。

5. 謝辞

本研究は、科学研究費補助金（A-2, 16203041）の助成を受けて行った。

文 献

- [1] R. H. Bolt and A. D. MacDonald, "Theory of Speech masking by reverberation," J. Acoust. Soc. Am., 21(6), 577-580, 1949.
- [2] 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.
- [3] 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", 日本音響学会秋季研究発表会講演論文集, Vol.1, pp.449-450, 2001.
- [4] 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.
- [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(6), 4055-4064, 2006.
- [6] 後藤崇公, 井上豪, 大畑典子, 程島奈緒, 荒井隆行, "実環境における音声明瞭度改善への前処理の効果—上智大学大講堂の場合—," 日本音響学会秋季研究発表会講演論文集, 613-614, 2003.a
- [7] 安武達朗, 中島祥好, "準実時間子音強調システム", 電子情報通信学会技術研究報告, HIP2005-94, pp 79-84, 2005.
- [8] N. Hodoshima, T. Goto, N. Ohata, T. Inoue and T. Arai, "The effect of pre-processing approach for improving speech intelligibility in a hall: comparison between diotic and dichotic listening conditions," Acoust. Sci., & Tech. 26, 212-214, 2002.
- [9] 小林敬, 八田ゆかり, 安啓一, 程島奈緒, 荒井隆行, 進藤美津子, "高齢者を対象にした定常部抑圧処理による単音節強調処理の検討," 電子情報通信学会技術研究報告, 104(695), 7-12, 2005.
- [10] K. Kobayashi, Y. Hatta, K. Yasu, S. Minamihata, N. Hodoshima, T. Arai and M. Shindo, "Improving speech intelligibility for elderly listeners by steady-state suppression," 電子情報通信学会技術報告, SP2005-168, 31-36, 2006.
- [11] S. E. Gehr and M. S. Sommers, "Age differences in backward masking," J. Acoust. Soc. Am., 106(5), pp.2793-2799, 1999.
- [12] B. C. J. Moore and B. R. Glasberg, "A compression of four methods implementing automatic gain control (AGC) in hearing aids", Brit. J. Audiol., 22, 93-104, 1988.
- [13] R. Plomp, "The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function," J. Acoust. Soc. Am., 83, 2322-2327, 1988.
- [14] 黒澤智幸, 西村竜一, 鈴木陽一, "振幅圧縮型補聴処理における包絡線強調の有効性に関する検討", 日本音響学会春季研究発表会講演論文集, 459-460, 2003.
- [15] T. Arai, K. Yasu and N. Hodoshima, "Effective speech processing for various impaired listeners," In Proc. of the International Congress on Acoustics, II, 1389-1392, Kyoto, 2004.