

# 時間反転残響における音声明瞭度の低下

荒井 隆行

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

E-mail: arai@sophia.ac.jp

**あらまし** 時間反転した残響が付加された音声は、通常の（時間反転していない）残響下における音声に比べて明瞭度が下がることを聴取実験によって調べた。我々は通常、残響を意識することなく音声コミュニケーションを行っているが、残響の掛かった音声信号を時間的に逆に再生して聞いてみると、その残響音がより明確に知覚されることが古くから知られている。しかし、残響のみが時間反転された場合に明瞭度がどれだけ低下するかが定量的に示された報告はあまりない。そこで、同じ残響を時間的に反転あるいは非反転させた信号と音声信号をたたみ込むことにより刺激文を作成し、文了解度試験を行った。使用した残響の残響時間（RT）は1.0, 1.5, 2.0 sであった。24名の被験者が24文を聞いた結果、98.6%が92.2%に（RT=1.0 s）、95.4%が82.2%に（RT=1.5 s）、93.9%が63.7%に（RT=2.0 s）有意に低下した。この結果は、人間の持つ聴覚特性の時間的非対称性と、我々が発する音声波における時間的な非対称性の両方の存在を反映しているものと考えられる。

**キーワード** 時間反転残響, 音声明瞭度, 知覚的オーバーラップマスキング, 音節のオンセットとコーダ, Haas 効果

## Degradation of Speech Intelligibility in Time-Reversed Reverberation

Takayuki ARAI

Dept. of Electrical and Electronics Eng., Sophia University 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan

E-mail: arai@sophia.ac.jp

**Abstract** We investigated whether speech in time-reversed reverberation is less intelligible than that in normal (non-time-reversed) reverberation. We generally perform speech communication without noticing the existence of reverberation. Moreover, it is known that reverberation is evident when we play a speech signal recorded in a reverberant room backward in time. However, few quantitative studies have shown speech intelligibility in time-reversed reverberation. We, therefore, conducted a perceptual experiment to compare speech intelligibility between signals convolved with normal and time-reversed impulse responses of a room. The reverberation times (RTs) used were 1.0, 1.5, and 2.0 s. Twenty-four subjects participated in the perceptual experiment; each subject was instructed to listen to 24 sentences under different experimental conditions. As a result, speech intelligibility significantly decreased at all RTs: from 98.6% to 92.2% (RT=1.0 s), from 95.4% to 82.2% (RT=1.5 s), and from 93.9% to 63.7% (RT=2.0 s). This might reflect the temporal asymmetry of both the human auditory system and a speech sound itself.

**Keyword** time-reversed reverberation, speech intelligibility, perceptual overlap masking, onset and coda of syllable, Haas effect

### 1. はじめに

音声は室内で産出されると、その音声波はあらゆる方向に伝搬し、その一部は直接耳に到達すると同時に、大部分は部屋の壁や天井、床などの表面で反射してから我々の耳に到達する。つまり、我々は日常的に残響の掛かった音声信号を耳にしている。しかし多くの場合、我々は直接音と反射音を別々の音として知覚することはなく、残響音を気にせずに会話をすることが可能である[1]。

直接音から 50ms までの初期反射音は通常、独立して知覚されず、直接音と融合される[2]。Haas [3]が、このように初期反射音が音声明瞭度に寄与することを観測したことから、この効果は Haas 効果 (Haas effect) と呼ばれる。それに加え、反射音は音源の方向の判断にもあまり影響を及ぼさず、全体の音の聞こえてくる方向は、最初に耳に到達する音でほぼ決定されることが知られている[1]。このように、直接音と反射音の間における知覚や定位にまつわる現象は先行音効果 (precedence effect) と呼ばれる[4]。

通常、我々はこのように残響を意識することなく音声コミュニケーションを行っているが、残響の掛かった音声信号を時間的に逆に再生して聞いてみると、その残響音がより明確に知覚されることが古くから知られている。例えば、A. J. M. Houtsma, T. D. Rossing, and W. M. Wagenaars による “Auditory Demonstrations” [5] では、いわゆる “Harvard tapes” の “ghoulies and ghosties” というデモンストレーションの例があげられている。そこでは、過渡的な音の例としてレンガをハンマーで叩く音、そしてより持続的な音の例としてスコットランドの祈りの言葉を用い、無響室、典型的な会議室、そしてかなり残響のある部屋という3つの異なる残響環境で聞いた場合のデモンストレーションが収められている。さらにそのデモンストレーションでは、それぞれの録音を時間的に逆に再生した場合にどのように聞こえるかも示している。通常の再生では、聞き手は残響をさほど気にせずに音を聞くことができる。しかし、逆再生した録音を聞くと、残響が目立つようになることがわかる。

また Griesinger (2004) は、残響が掛かった音声の明瞭度を客観的に調べる際、音声信号の「オンセット」に着目して音節数をカウントする方法に基づく評価法を提唱している[6]。建築音響分野では、音声明瞭度のための指標として STI (speech transmission index) が広く用いられているが、この STI は室内で収録された音響信号の時間包絡の周波数特性に基づいており、その周波数分析では振幅特性のみが用いられている。そのため、時間的に反転した残響を音声信号にたたみ込んでも、STI 値は変化しない。しかし、時間反転した残響が付加された音声信号を実際に我々が聞いた場合に聞きづらくなることから、STI は人間の主観的評価に完全には一致しない。一方、音声信号の「オンセット」に着目して音節数をカウントする手法は、残響の時間方向によって結果が変わってくる。つまり、通常の残響の場合にはオンセットははっきりとカウントされるが、時間反転残響の場合はオンセットははっきりせずカウント数は極端に下がる。これは人間の主観的評価と一致する、と Griesinger は主張している[6]。

このように、残響の掛かった音声信号の知覚に関する時間方向の非対称性についていろいろと議論がされているが、時間反転された残響下における音声について明瞭度を実際に調べた研究はあまり報告されていない。そこで本研究では、(現実の世界ではあり得ないが)時間反転された残響の付加された音声の明瞭度が、通常の残響音声の明瞭度に比べ低下する様子を聴取実験によって調べた。

## 2. 実験

時間反転残響における音声の明瞭度が、通常の(時間非反転)残響における音声明瞭度に比べ低下することを調べるために、聴取実験を行った。

### 2.1. 刺激

音声サンプルとして、NTT アドバンステクノロジー(株)から提供されている日本語音素バランス 1000 文を使用した。話者 MYK の 1000 文の中から、まず発話時間が 4~5 秒程度の文を選び、さらにその中から親密度の低い固有名詞や意味の取りづらい単語を含む文を除いて、最終的に 36 文を選んだ。そのうちの 12 文を練習用、24 文を本番用とした。これらの文のモーラ数は、25~33 個(平均 28 個程度)であった。

残響に関しては、東大和市大ホール(ハミングホール)において測定されたインパルス応答を基にした。そのインパルス応答の残響時間(reverberation time, RT)は 1.1 秒であった(500, 1k, 2k Hz のオクターブバンドで求めた残響時間の平均)。このインパルス応答の時間包絡を、指数的に減少する関数で近似し、その時定数を操作することによって、3 種類の RT(1.0 s, 1.5 s, 2.0 s)を有するインパルス応答を人工的に作成した[7]。そして、時間反転残響は、それらのインパルス応答の時間軸を反転させることによって実現した。結果として 6 種類のインパルス応答が得られるが、その 6 つの残響条件を Table 1 に示す。

残響の掛かった音声刺激は、Table 1 に示された 6 種類のインパルス応答を、上記で述べた音声サンプルに計算機上で Fig. 1 のようにたたみ込むことによって作成した(練習用 72 刺激文、本番用 144 刺激文)。刺激の提示に対しては、diotic 受聴とした。

### 2.2. 被験者

被験者は上智大学に在籍している学生 24 人で、男性 16 人、女性 8 人である。全員が日本語を母語とする健聴者で、年齢は 19~24 才(平均 21.8 才)であった。

Table 1: 本研究で使った残響条件

| 時間非反転 | 時間反転 | 残響時間 (s) |
|-------|------|----------|
| h1    | h1_r | 1.0      |
| h2    | h2_r | 1.5      |
| h3    | h3_r | 2.0      |

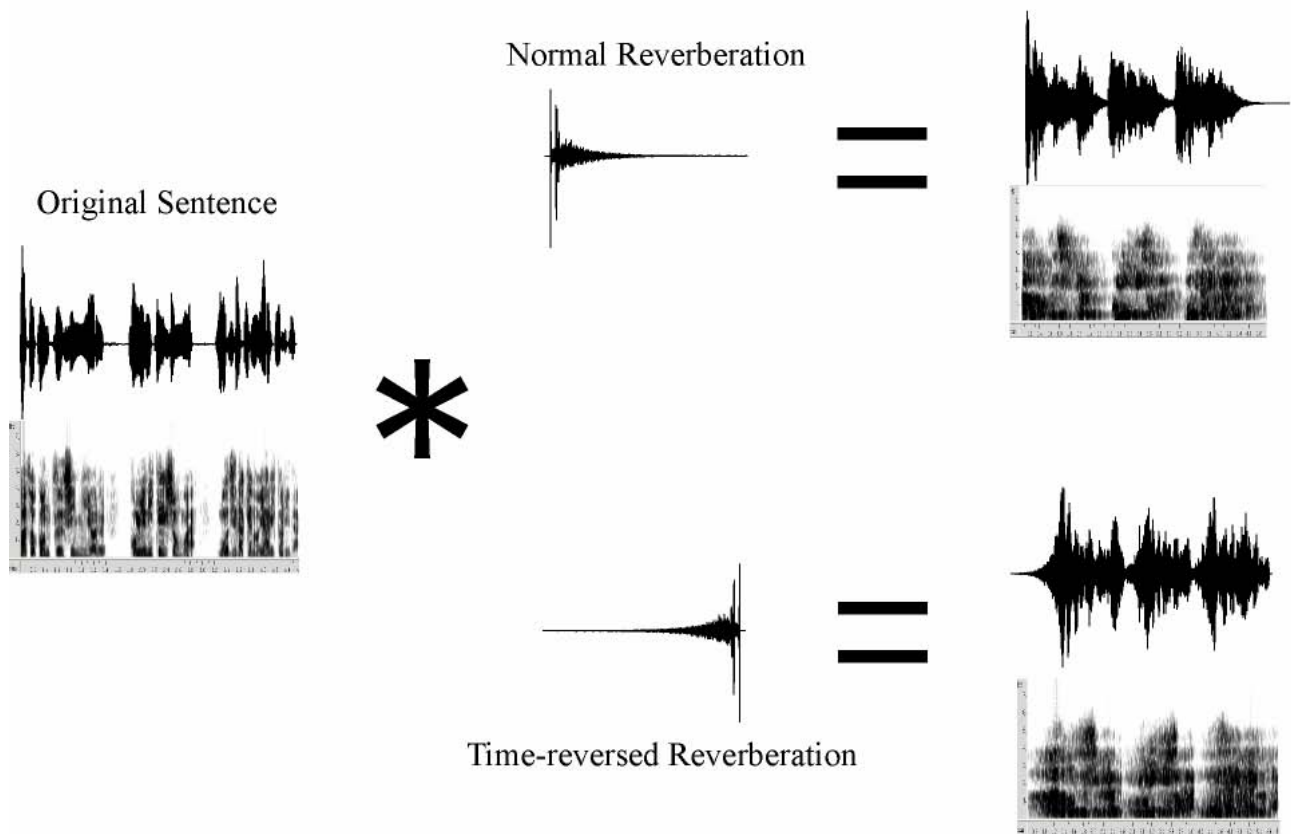


Fig. 1 : 音声信号が通常の（時間非反転）残響および時間反転残響とたたみ込まれる様子

### 2.3. 方法

実験は荒井研究室内の防音室にて行われた。被験者には防音室内のコンピュータの前に座ってもらい、実験に関する説明が書かれている画面を見てもらいながら、口頭でも実験に関して説明を行った。その後、ヘッドフォン（STAX SR-303）を装着してもらった。実験手順に慣れてもらうため、練習用に作成した刺激音を用いて実験の本番と同じ手順で12回練習を行った。また練習の間に、提示する刺激音の音量を被験者が聞き取りやすいレベルに調整した。なお、ヘッドフォンからは左右にそれぞれまったく同じ信号が提示されるようにした。練習の後に本番が始まり、本番では24回の試行が行われた。

実験手順としては、被験者にマウスをクリックしてもらい、ヘッドフォンから聞こえてきた音声を所定の回答用紙に書き取ってもらった。実験用の画面はマウスをクリックするだけで次に進むようになっており、マウスをクリックして画面が進むと同時に刺激音を提示した。同一試行内では計3度刺激音を聞いてもらい、画面上には現在何問中何問目で何回目を聞いているかが示されるようにした。同一試行内において、2度目

と3度目の音声刺激提示後は、書き取った文字列を訂正してもよいこととした。もし聞き取れなかった部分があれば、その部分を“・・・”で示してもらった。3度音声聞いた後は、再びマウスをクリックすれば次の試行に進めるようにした。なお音声の書き取りは、ひらがな又はカタカナのみで行ってもらった。

刺激文は全部で144文(本番用)用意されているが、1度過去に提示された文は記憶に残ってしまうことから、1人の被験者に同一の文を複数回(違う残響条件等で)使うことはできない。そこで、1人の被験者にはもともとの音声サンプルの数である24文(本番用)しか提示しないこととした。被験者ごとに異なる組み合わせの残響条件による24文を提示し、その残響条件の組み合わせを24人の被験者に渡ってカウンタバランスを取るものとした。その結果、1人の被験者に対して、6種類の残響条件ごとに4つの文が割り当てられ、その割り当て方は被験者ごとにシフトするようにした。練習用音声サンプル12文に関しては、6種類の残響条件ごとに2文を割り当て、カウンタバランスは取らずに全被験者に対して同じ文セットを使用した。

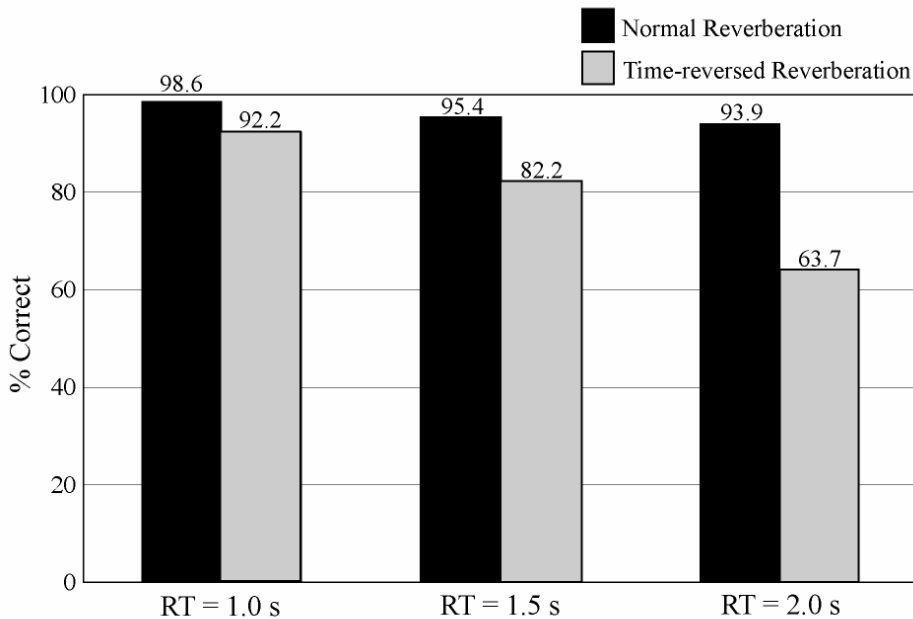


Fig. 2 : 通常の（時間非反転）残響および時間反転残響における文了解度試験の結果

## 2.4. 実験結果

被験者全員に対する文了解度の平均を残響条件ごとに求めたものを、Fig. 2 に示す。各文の了解度については、モーラごとの正答率で表した。

結果の分析は時間反転・非反転 2 種類 × 残響時間 3 種類の 2 要因 6 条件に分けて被験者ごとに正答率を出し、条件ごとに全被験者で平均をとった。本実験における 2 × 3 のデータについて、繰り返しのある分散分析を行ったところ、残響の時間反転・非反転の主効果は有意であり ( $F(1,23)=149.074$ ,  $MSe=66.378$ ,  $p<.001$ ) 正答率の平均値は時間非反転残響条件 (96.0%) よりも時間反転残響条件 (79.4%) の方が低かった。また残響時間 (1.0 s, 1.5 s, 2.0 s) の主効果は有意であり ( $F(1,23)=149.074$ ,  $MSe=74.815$ ,  $p<.001$ ) 正答率の平均値は高い順に 1.0 s, 1.5 s, 2.0 s であった (1.0 s: 95.4%, 1.5 s: 88.8%, 2.0 s: 78.8%)。また残響の種類と残響時間の種類の間の交互作用は、( $F(1,23)=60.876$ ,  $MSe=56.363$ ,  $p<.001$ ) より有意であった。この交互作用について残響の時間反転・非反転ごとに残響時間の単純主効果の検定を行った。時間非反転残響における残響時間の単純主効果は有意であり ( $F(1,23)=18.852$ ,  $MSe=13.665$ ,  $p<.001$ ) 正答率の平均値は 1.0 s > 1.5 s > 2.0 s (98.6% > 95.4% > 93.9%) であった。時間反転残響における残響時間の単純主効果は有意であり ( $F(1,23)=74.626$ ,  $MSe=130.917$ ,  $p<.001$ ) 正答率の平均値は 1.0 s > 1.5 s > 2.0 s (92.2% > 82.2% > 63.7%) であった。

以上から、通常の残響がかかった音声で残響時間が長くなるにつれてその明瞭度が減少するのと同様に、時間反転残響においても残響時間の長い音声ほど正答率が低くなることがわかった。さらに、時間反転残響のかかった音声は 3 種類の残響時間すべてにおいて、時間非反転残響のかかった音声のいずれよりも正答率は低かった。よって時間反転残響は時間非反転残響に比べ、さらに聞きづらいことが分かる。また、残響時間の長さに伴う明瞭度の変化は、時間反転残響のほうが時間非反転残響よりも急激であった。

## 3. 考察

本研究の実験結果は、予測通り、「時間反転残響における音声明瞭度が時間非反転残響の場合に比べて低下する」という仮説を支持するものとなった。残響の時間方向を反転させてもさせなくてもその長時間スペクトルは変わらないので、残響がたたみ込まれた後の音声信号の長時間スペクトルも残響の時間方向には依存しないはずである。それにも関わらず、本研究の実験結果において音声明瞭度に差が現れた。以下では、その理由について考察する。

残響が音声の明瞭度を下げる要因としては、self masking と overlap masking の 2 つがあるとされている [8-10]。一つ目の self masking はそれぞれの音素内でマスキングが起こる結果、音素自身が変形するもので、音の立ち上がりや立ち下がりといった遷移部が特になる。もう一方の overlap masking は、先行する音素に

伴う残響の尾が後続する音素をマスクするというもので、特に先行する音素のエネルギーが大きく後続する音素のエネルギーが小さい場合、その影響は大きくなると考えられる[11,12]。

Haas 効果(先行音効果)によると、直接音から 50ms までの初期反射音は通常、独立して知覚されず、直接音と融合される。融合されるためには後続する反射音のレベルが弱くなければならないが、実際の残響環境では反射音のほうが弱いのが普通である。しかし、時間反転残響の場合、後続音のほうが強くなり Haas 効果が起きにくい状況になっていたと考えられる。その結果、後続音が融合されにくく、知覚的な overlap masking のマスクング量が大きくなったと考えられる。

Haas 効果は主に両耳聴において報告されているが、少なくとも「直接音と初期反射音の融合の現象」は両耳聴のみならず monaural もしくは diotic 受聴でも議論されている[4,13,14]。本実験は diotic 受聴環境で行われておりその点においても興味深い結果となったが、上記の考察は、人間の聴覚特性の持つ時間の非対称性によるものと言える。

一方、我々が発する音声波においても、時間的な非対称性が存在する。それは音節におけるオンセットとコーダに関係している。音節は通常、母音などの聴こえの大きな言語音が音節の核をなし、その前後に聴こえの小さな子音が配置するという構造を持つ(核の子音がオンセット、核の後ろの子音がコーダ)。オンセットの子音はたいてい標準的に発音され、発音のバリエーションが少なく安定している。一方、コーダの子音は削除(deletion)や弱化(reduction)の対象となる傾向が多い[15]。例えばコーダの破裂音は必ずしも破裂の解放を伴うとは限らず、多数の異音を持つ場合もある[16]。実際、英語において CVC 構造を持つ単音節語を使って Modified Rhyme Test (MRT) を行くと、コーダの子音に対する正答率が 68%であるのに対し、オンセットの子音に対する正答率は 80%にまで達することが報告されている[17]。結局、我々が言語音を認識する際にオンセットの役割は極めて重要と言える。

Griesinger [6]が室内の音声明瞭度の指標に音節数をカウントする手法を提案した背景には、音節のオンセットのほうがコーダよりも overlap masking の影響が少ないということがありと解釈することができる。つまり、オンセットは残響があまり掛かっていない(overlap masking が少ない)上に発話も明瞭であるのに対し、コーダは残響が多く掛かっている(overlap masking が多い)上に発話が明瞭でない。その上に、知覚的な overlap masking の非対称性が合わさることで、通常の残響と時間反転残響との間で音声明瞭度の差が現れたのではないかと推測される。

#### 4. おわりに

本研究では、時間反転した残響が付加された音声は、通常の(時間反転していない)残響下における音声に比べて明瞭度が下がることを文理解度試験によって調べた。24名の被験者が24文を聞いた結果、98.6%が92.2%に(RT=1.0s)、95.4%が82.2%に(RT=1.5s)、93.9%が63.7%に(RT=2.0s)有意に低下した。この結果は、人間の持つ聴覚特性の時間的な非対称性と、我々が発する音声波における時間的な非対称性の両方の存在を反映しているものと考えられる。その両方において音節のオンセットがよりウェイトを持つような仕組みが出来上がっていることから、その非対称性と残響の持つ物理的な時間的な非対称性が合う、合わないによって、音声明瞭度に差が現れるのではないかと推測された。

#### 5. 謝辞

本研究を遂行するにあたり、きっかけを与えてくださった同志社大学の力丸裕先生、実験を主に進めてくれた上智大学荒井研究室卒業生の西館恵美さん、そして実験をサポートしてくれた同じく荒井研の奥田拓馬さん、程島奈緒さん、同大学心理学科の道又爾先生と鎌田浩史さん、インパルス応答のデータを提供して下さった東京大学の橘秀樹先生、上野佳奈子さん、横山栄さんに心から感謝いたします。本研究は科学研究費補助金(A-2, 16203041)の助成を受け行ったものである。

#### 文 献

- [1] B. C. J. Moore 著, 大串健吾監訳, 聴覚心理学概論, 誠信書房, 東京, 1994.
- [2] A. J. Watkins, "The influence of early reflections on the identification and lateralization of vowels," *J. Acoust. Soc. Am.*, vol.106, no.5, pp.2933-2944, Nov. 1999.
- [3] H. Haas, "Über den Einfluss eines Einfachechos an die Hörsamkeit von Sprache," *Acustica*, vol.1, pp.49-58, 1951.
- [4] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The precedence effect," *J. Acoust. Soc. Am.*, vol.106, no.4, pp.1633-1654, Oct. 1999.
- [5] A. J. M. Houtsma, T. D. Rossing, and W. M. Wagenaars, *Auditory Demonstrations*, Acoust. Soc. Am., Sep. 1987.
- [6] D. Griesinger, "Measurement of acoustic properties through syllabic analysis of binaural speech," *Proc. Int'l Cong. on Acoustics*, vol.1, pp.29-32, Apr. 2004.
- [7] Hodoshima, N., Arai, T., and Kusumoto, A., "Enhancing temporal dynamics of speech to improve intelligibility in reverberant environments," *Proc. Forum Acusticum*, Sevilla, 2002.
- [8] R. H. Bolt and A. D. MacDonald, "Theory of speech masking by reverberation," *J. Acoust. Soc. Am.*, vol.21, pp.577-580, 1949.

- [9] A. K. Nábělek and L. Robinette, "Influence of precedence effect on word identification by normally hearing and hearing-impaired subjects," *J. Acoust. Soc. Am.*, vol.63, pp.187-194, 1978.
- [10] A. K. Nábělek, T. R. Letowski and F. M. Tucker, "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Am.*, vol.86, pp.1259-1265, 1989.
- [11] 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, Oct. 2001.
- [12] T. Arai, K. Kinoshita, N. Hodoshima, A. Kusumoto and T. Kitamura, "Effects on suppressing steady-state portions of speech on intelligibility in reverberant environments," *Acoustical Science and Technology*, vol.23, no.4, pp.229-232, 2002.
- [13] B. Rakerd, J. Hsu, and W. M. Hartmann, "The Haas effect with and without binaural differences," *J. Acoust. Soc. Am.*, vol.101, no.5, p.3083, 1997.
- [14] R. Y. Litovsky, M. L. Hawley, and H. S. Colburn, "Measurement of precedence in monaural listeners," *Meeting of the American Speech and Hearing Association*, Boston, MA, 1997.
- [15] S. Greenberg and T. Arai, "What are the essential cues for understanding spoken language?," *IEICE Trans. Inf. & Syst.*, vol.E87-D, no.5, May 2004.
- [16] M. Davenport and S. J. Hannahs, *Introducing Phonetics and Phonology*, Arnold, 1998.
- [17] A. S. House, C. E. Williams, M. H. L. Hecker, and K. D. Kryter, "Articulation-testing methods: Consonantal differentiation with a closed-response set," *J. Acoust. Soc. Am.*, vol.37, no.1, pp.158-166, Jan. 1965.
-