

定常部抑圧処理による残響下での音声明瞭度改善 —抑圧率とモーラ位置・母音の関係—

柳橋 歩* 程島 奈緒† 荒井 隆行‡

* 東海大学大学院情報通信学研究科 〒108-8619 東京都港区高輪 2-3-23

† 東海大学情報通信学部情報メディア学科 〒108-8619 東京都港区高輪 2-3-23

‡ 上智大学理工学部情報理工学科 〒102-8554 東京都千代田区紀尾井町 7-1

E-mail: * 2ljinm001@mail.tokai-u.jp, † hodoshima@mail.tokai-u.jp, ‡ arai@sophia.ac.jp

あらまし 公共空間には残響が存在し、これによる overlap-masking よって音声の明瞭度が低下する場合があります、この影響を軽減するために定常部抑圧処理が提案されている。原音声の振幅に対する抑圧後の振幅の割合(抑圧率)を40%にすることにより、単音節の明瞭度改善効果を得たが、単語では明瞭度改善効果を得られていない。そこで、正解率の低下する先頭モーラは定常部のエネルギーを残し、後続モーラは徐々にエネルギーを削減する定常部抑圧処理を提案した。処理音声に2種類のインパルス応答(残響時間: 2.5 s, 3.5 s)を畳み込んだ刺激を用いて単語理解度試験を若年者24名に行った。モーラ位置正解率より、残響時間2.5 sでは第1・2モーラが40-80%、第3モーラは30-100%、第4モーラでは10-20%、残響条件3.5 sでは第1・2モーラが40-80%、第3モーラは40%、第4モーラでは10-40%の抑圧率が各残響条件の各モーラ位置での適正な抑圧率であると考えられる。本実験では残響下での単語理解度の改善は得られなかったが、結果より得られた各モーラ位置の適正な抑圧率を使用することにより単語理解度改善に繋がる可能性がある。

キーワード 残響, 音声明瞭度, 定常部抑圧処理

The Steady-State Suppression improving for speech intelligibility in reverberant environment

—The relation between suppression rate, syllable position and vowel—

Ayumu YANAGIBASHI*, Nao HODOSHIMA† and Takayuki ARAI‡

* † Graduate School of Information and Telecommunication Engineering, Tokai University 2-3-23 Takanawa,
Minato-ku, Tokyo, 108-8554 Japan

‡ Department of Information Media Technology, Tokai University 2-3-23 Takanawa, Minato-ku, Tokyo, 108-8554 Japan

‡ Department of Information and Communication Sciences, Sophia University 7-1 Kioityo, Cyoda-ku, 102-8554 Japan

E-mail: * 2ljinm001@mail.tokai-u.jp, † hodoshima@mail.tokai-u.jp, ‡ arai@sophia.ac.jp

Abstract Speech intelligibility in reverberant environments decreases due to overlap-masking. Steady-state suppression method has proposed as a preprocessing to reduce overlap-masking in reverberant environments. Improvements in intelligibility were obtained for monosyllables, but not for words. We revised the method by gradually suppressing steady-state portions of target words. Unprocessed and processed speech signals were convolved with two impulse response (reverberation times of 2.5 s and 3.5 s) were used in word identification tests by 24 young people. As results, appropriate suppression rates were 40-80% for the first and second morae, 30-100% and 40% for the third mora, 10-20% and 10-40% for the fourth mora at reverberation time of 2.5 s and 3.5 s respectively. The proposed method did not improve word identification scores, however, applying for appropriate suppression rates for each mora may lead to improve speech intelligibility.

Keywords Reverberation, Speech intelligibility, Steady-state suppression

1. はじめに

アナウンス音声は、駅や空港などの公共空間には欠かすことのできないシステムであり、事故や天災の際には避難誘導のための情報伝達手段として使用される。しかし、環境内に存在する残響による overlap-masking^[1]の影響により、アナウンス音声が聞き取りづらい場合がある。この現象は先行する音節の残響が後続の音節に被さることにより発生し、音声明瞭度の低下の要因となる。これを改善し音声明瞭度を向上させるための信号処理として、定常部抑圧処理^[1]が提案されている。この処理は、音声知覚に比較的重要ではないとされるが、エネルギーの大きい定常部を抑圧することにより後続の過渡部への overlap-masking を軽減する処理である。この処理で、定常部を抑圧する割合を抑圧率とし、原音声の振幅に対する抑圧後の振幅の割合を 40% にすることにより、単音節明瞭度試験では、明瞭度改善の効果が得られた^[1]が、単語了解度試験では改善効果を得られていない。

単語了解度試験での改善効果を得るために著者らは、先頭モーラの正解率が後続のモーラに比べ低下する^[6]ことに着目し、モーラ毎に抑圧率を変化させる定常部抑圧処理を提案した。これは、文章を 4~5 モーラで区切り 1 フレームとし、フレーム内の先頭モーラのエネルギーを残し、後続のモーラは徐々にエネルギーを削減する定常部抑圧処理であり、モーラ間の self-masking とフレーム間の overlap-masking を効率的に軽減することが可能と考えられる。そこで、信号処理を付加した音声にインパルス応答 2 種（残響時間：2.5 s、3.5 s）を畳み込んだ刺激音を用いて聴取実験を行った。その結果、残響時間 2.5 s の条件では、提案手法の平均モーラ・単語正解率が最も高くなったが、従来法と提案手法間に有意差は見られなかった。有意差が見られなかった原因として、ターゲット語の単語正解率の差が関係していると考えられる。平均正解率が最も高い単語は 91.7%、最も低い単語は 4.2% となり、大きな差が見られた。

本稿は、以上の実験結果よりモーラ位置正解率と平均母音正解率を算出し、モーラ位置と母音の違いによって処理効果や正解率がどのように変化するか検証した。モーラ位置正解率より、各モーラ位置に適正な抑圧率を付加することが可能になれば、単語了解度改善効果が得られると考えられる。また、平均母音正解率よりモーラ位置の正解率の変動が処理効果か母音の種類によるものか考察することで、提案手法の改善に繋がると考える。

2. 聴取実験

2.1. 原音声

刺激の原音声は、関東方言話者 1 名（年齢 24 歳、男性）にターゲット語、キャリア文を別々に防音室で発声させ収録した。収録音の振幅を正規化し、キャリア文にターゲット語を挿入して原音声を作成した。キャリア文は「このような場合には（ターゲット語）することがあります。」を使用し、ターゲット語へのマスキング量の統制を取るため、直前のキャリア文終了部から 150 ms をターゲット語の開始点とした。ターゲット語は親密度別単語了解度試験用音声データセット 2003 (FW03)^[7]より、48 種の単語（親密度 1.0~2.5 の 4 モーラ語）を選定した。選定単語は、各モーラを構成する母音が 3 種 (/a/i/u/) であり、各モーラ位置での母音の合計が 16 個ずつとなるものである。これは、モーラ位置正解率を算出する際に、母音の種類を 3 種に限定することにより、各モーラ位置で処理効果を比較しやすくするためである。

2.2. 信号処理

本実験で用いた定常部抑圧処理の各モーラ位置の抑圧率 Table 1 に示す。条件 1 は処理を付加せずに原音声を使用し、条件 2 は単音節明瞭度試験で改善効果の得られた抑圧率 40% を使用した。条件 3・4 は、単音節明瞭度試験で効果を得た抑圧率 40% が 4-5 モーラの中央である第 2、第 3 モーラに設定され、且つ先頭モーラのエネルギーが十分確保できるように条件 3 は先頭モーラから 20% ずつ抑圧率を低くし、条件 4 は先頭モーラから 10% ずつ抑圧率を低くした。これらの抑圧率を用いた理由として、条件 3・条件 4 の正解率の比較から各モーラ位置で処理効果のある適正抑圧率の範囲を絞り込むために Table 1 の値を設定した。

Table 1 : 各モーラ位置の抑圧率

条件	抑圧率 [%]				
	第 1 モーラ	第 2 モーラ	第 3 モーラ	第 4 モーラ	第 5 モーラ
1	処理なし				
2	従来法 (抑圧率: 全モーラ 40)				
3	80	60	40	20	10
4	50	40	30	20	10

2.3. 刺激

原音声に処理条件4種 (Table 1) を付加し、インパルス応答2種 (残響時間: 2.5 s、3.5 s) を畳み込んで刺激を作成した。処理条件 (4 条件) × 残響条件 (2 条件) の計 8 条件である。

2.4. 被験者

被験者は聴力に問題がなく、日本語を母語とする若年者 24 名 (平均年齢: 21.9 歳、男性: 21 名、女性: 3 名) であり、健聴者か否かの判断は自己申告である。

2.5. 実験手順

被験者には防音室内でパソコン (hash MiniITX ION Fanless Silent) 上のインタフェースを操作させ、オーディオインタフェース (TASCAM US-144MK II) を介して、ヘッドホン (STAX SR-303) から刺激を提示した。刺激は 48 刺激 (各条件 6 音声 × 8 条件) を提示し、各単語で処理効果や残響下での聞き取りの難易度が異なると考え、全単語が全条件で被験者に提示されるように統制を取り、ランダムに提示した。被験者には刺激を一度だけ提示し、聴取したターゲット語を回答用紙に記入させた。

3. 結果

統計解析ソフトウェア (IBM SPSS) を用いてそれぞれの結果に対して二元配置の分散分析を行い、条件間で Sidak による多重比較を行った。

3.1. モーラ正解率、単語正解率

各条件の全被験者の平均モーラ正解率を Figure 1、平均単語正解率を Figure 2 に示す。処理 4 条件と残響 2 条件の平均モーラ・単語正解率に対して二元配置の分散分析を行った結果、平均モーラ正解率は処理条件 ($p < 0.05$) と残響条件 ($p < 0.01$) の主効果が有意となった。処理条件間の多重比較では従来法である条件 2 が提案手法である条件 4 より有意に正解率が高かった ($p < 0.05$) が、その他の条件間では有意差は見られなかった。

また、平均単語正解率でも処理条件 ($p < 0.01$) と残響条件 ($p < 0.05$) で主効果が有意となった。しかし、両結果共に処理条件と残響条件間の交互作用は有意ではなかった。処理条件間の多重比較では平均モーラ正解率と同様に条件 2 が条件 4 より有意に正解率が高かった ($p < 0.05$) が、その他の条件間では有意差は見られなかった。以上の結果より、平均モーラ正解率と平均単語正解率で最も正解率が高かった提案手法である条件 3 に有意差はなく、条件 4 は条件 2 より有意に正解率が低くなり、これらの結果からは、提案手法による明瞭度改善は得られなかった。

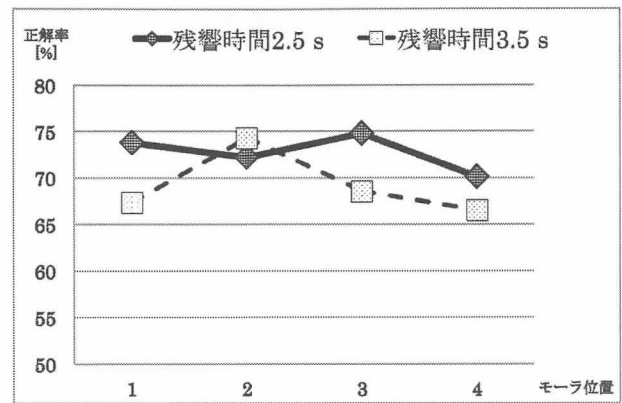


Figure 1: 平均モーラ正解率

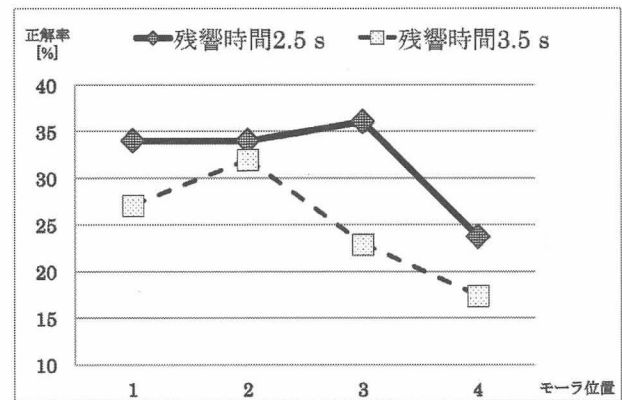


Figure 2: 平均単語正解率

3.2. モーラ位置正解率

残響条件 2.5 s のモーラ位置正解率を Figure 3、残響条件 3.5 s のモーラ位置正解率を Figure 4 に示す。3.1. 章と同様にモーラ位置 4 条件と処理条件 4 条件のモーラ位置正解率に対して二元配置の分散分析を行い、処理条件間で Sidak による多重比較を行った。また、各モーラ位置の処理条件に対して一元配置の分散分析を行った。

残響条件 2.5 s のモーラ位置正解率 (Figure 3) は、モーラ位置条件 ($p < 0.01$) の主効果は有意であったが、処理条件は有意ではなかった。さらに、モーラ位置と処理条件の交互作用も有意ではなかった。モーラ位置条件間の多重比較では、第 3 モーラと第 4 モーラ間は有意差が見られなかったが、その他のモーラ位置条件間では有意差が見られた ($p < 0.05$)。また、各モーラ位置の処理条件に対する一元配置の分散分析では、第 2 モーラのみ処理条件 ($p < 0.05$) の主効果が有意となったが、処理条件間の多重比較では有意差は見られなかった。

以上の結果より、残響条件 2.5 s では第 3・4 モーラを除く、他のモーラ位置正解率の差が有意となったことから、先行研究^[6]と同様にモーラ位置による正解率の変化が見られた。また、各モーラの処理条件間には

有意差が見られなかったため、各モーラ位置の明瞭度改善効果は得られなかった。

残響条件 3.5 s のモーラ位置正解率 (Figure 4) ではモーラ位置 ($p<0.01$) と処理条件 ($p<0.05$) の主効果が有意となったが、モーラ位置と処理条件の交互作用は有意ではなかった。モーラ位置条件に対する Sidak による多重比較では、第 1 モーラと第 3 モーラ間 ($p<0.05$)、第 2 モーラと第 3 モーラ間 ($p<0.01$) に有意差が見られた。さらに、処理条件間の多重比較では、処理条件 2 と処理条件 4 間のみに有意差 ($p<0.01$) が見られた。また、各モーラ位置の処理条件に対する一元配置の分散分析を行った結果、第 3 モーラのみ処理の主効果が有意となった。処理条件に対する Sidak による多重比較では、処理条件 2 と処理条件 3 ($p<0.05$) と処理条件 2 と処理条件 4 の間 ($p<0.01$) に有意差が見られた。

以上の結果より、残響条件 3.5 s では第 3 モーラが第 1・2 モーラより正解率が有意に低くなり、第 3 モーラ内で処理条件 2 より処理条件 3・4 の正解率が有意に低いことから、第 3 モーラは抑圧率を低く設定することにより正解率が低下する結果となった。

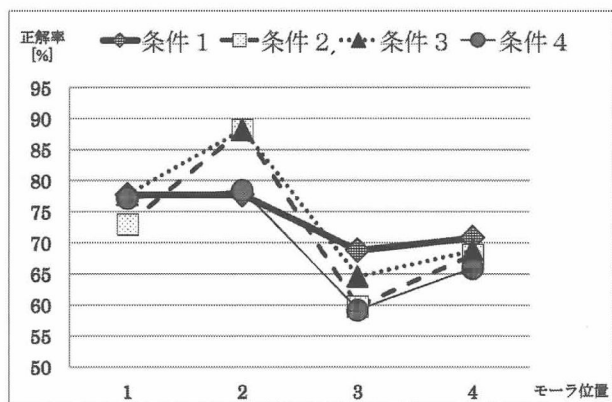


Figure 3 : モーラ位置正解率(残響時間 : 2.5 s)

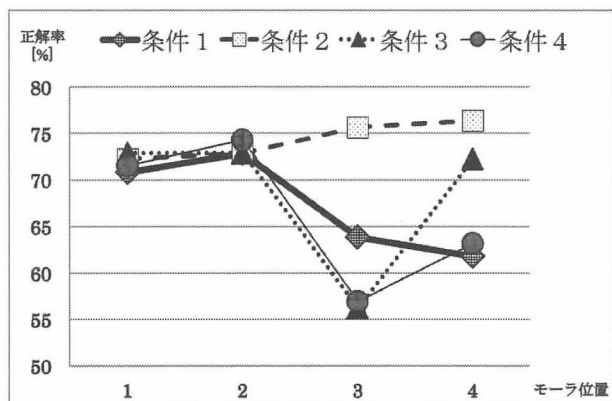


Figure 4 : モーラ位置正解率(残響時間 : 3.5 s)

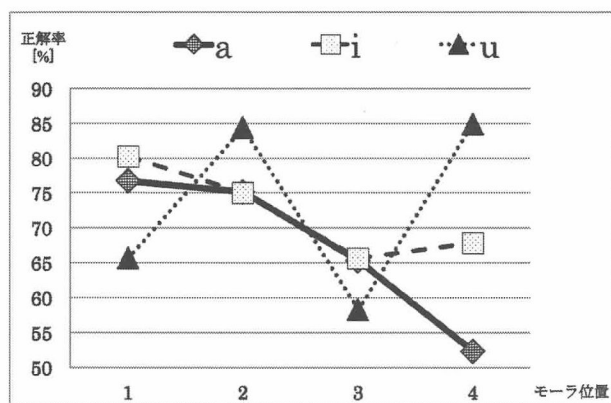


Figure 5 : 平均母音正解率

3.3. 平均母音正解率

全条件の各モーラ位置の平均母音正解率を Figure 5 に示す。母音 3 条件とモーラ位置 4 条件の平均母音正解率に対して二元配置の分散分析を行い、モーラ位置条件 ($p<0.01$) と母音条件 ($p<0.01$) の主効果が有意となった。また、モーラ位置条件と母音条件 ($p<0.01$) の交互作用が有意となった。モーラ位置条件に対する Sidak による多重比較では、全てのモーラ位置条件間に有意差が見られた ($p<0.05$)。また、母音条件に対する Sidak による多重比較では、/a/と/i/、/a/と/u/の間に有意差が見られた ($p<0.01$)。

4. 考察

平均モーラ正解率と平均単語正解率 (3.1.章) の結果より、単語了解度改善効果を得ることができなかった。平均モーラ・単語正解率で条件 4 が条件 2 より有意に正解率が低下したことにより、各モーラで抑圧率を変化させることは音声明瞭度に関係があり、極端に低い抑圧率設定は明瞭度を低下させることがわかった。また、残響条件 2.5 s で最も正解率が高い条件 3 が残響条件 3.5 s で正解率が大きく低下していることから、残響条件に対しても適した抑圧率が存在し、本手法は残響時間が長い環境には適していないと考えられる。

モーラ位置正解率 (3.2.章) より、残響条件 2.5 s (Figure 3) では各モーラ間に処理による有意な差は得られなかったが、条件 3 の処理改善のための課題を得ることができた。第 1 モーラは、抑圧率を 80% としたが条件 2 と有意差がなく、且つ第 2 モーラの正解率が条件 2 と等しいことから、第 1・2 モーラは抑圧率を 40-80% の値にすることで条件 1 と他の処理を付加した条件は等しい正解率を得られると考えられる。正解率が低下した第 3 モーラでは、各条件間に有意差が見られないことに加え、第 4 モーラの処理条件間の正解率に有意差がないことから、抑圧率は 30-100% の範囲から選択することが適しているといえる。さらに、第

4 モーラは条件 1 と条件 3・4 間の正解率に有意差がないことから、抑圧率を 10-20% に設定し、後続音節への overlap-masking 軽減を行うことが可能である。また、残響条件 3.5 s (Figure 4) では、第 1・2 モーラの正解率に有意差が見られないことから、残響条件 2.5 s (Figure 3) と同様に抑圧率が 40-80% であれば、条件 1 と他の処理を付加した条件は等しい正解率を得られる。第 3 モーラは条件 2 が条件 3・4 に対して正解率が有意に高いことから、抑圧率 40% が第 3 モーラでは適正抑圧率であるといえる。第 4 モーラは有意差がないことから 10-40% の抑圧率が適正であり、後続の音節への overlap-masking の影響より、低い抑圧率が望ましいと考える。以上により、各残響条件でのモーラ位置正解率の結果からも残響条件によって適正な抑圧率が存在することがいえる。

各結果で有意差を得られなかった共通の原因として、各単語の正解率の差が関係していると考えられる。親密度と母音の種類による統制を取ったにも関わらず、平均正解率が最も高い単語は 91.7%、最も低い単語では 4.2% となり大きな差が見られた。また、平均母音正解率 (3.3.章) の結果より、母音/u/の第 3 モーラ、母音/a/の第 4 モーラの正解率 (Figure 5) とモーラ位置正解率 (3.2.章) の第 3・4 モーラの正解率の低下には残響条件 3.5 s (Figure 4) での条件 3・4 の第 1・2 モーラの抑圧率 (Table 1) を高く選定したことによる overlap-masking による影響があった可能性がある。

5. おわりに

本稿は、残響下の単語理解度改善を目的とし、抑圧率を各モーラ位置で変化させた信号処理による改善効果と各モーラ位置での適正抑圧率について考察した。聴取実験より平均モーラ正解率と平均単語正解率では、提案手法である条件 3 が最も正解率が高くなったが、処理付加なしの条件 1 と従来法である条件 2 に対して有意差は得られなかった。また、モーラ位置正解率では、残響条件 2.5 s で条件 3 が第 1・2 モーラで最も正解率が高くなったが、第 3・4 モーラでは条件 1 の正解率を下回った。残響条件 3.5 s では、条件 2 の正解率が全てのモーラ位置で高くなった。これらの結果から、本実験では提案手法による単語理解度の改善効果は得られなかった。

以上のことから、提案手法による単語理解度改善の可能性とその課題を得た。本実験で使用した条件 3 は、有意差を得ることはできなかったが、残響条件 2.5 s で平均モーラ・単語正解率が最も高くなり単語理解度改善の可能性があると見える。その課題として、第 3・4 モーラの抑圧率を再検討し、正解率を向上させることが必要である。残響条件 3.5 s では、第 3 モーラのモ

ーラ位置正解率が条件 2 に対して条件 3・4 が有意に下回ったことから、先行研究^[5]で報告された抑圧率を過度に低く設定することによる音声明瞭度の低下があったと考えられる。提案手法の抑圧率改善の方法として、本稿のモーラ位置正解率から考察した結果、残響条件 2.5 s では、第 1・2 モーラが 40-80%、第 3 モーラは 30-100%、第 4 モーラでは 10-20%、残響条件 3.5 s では第 1・2 モーラが 40-80%、第 3 モーラは 40%、第 4 モーラでは 10-40% の抑圧率が各条件・モーラ位置での処理効果の可能性のある適正な抑圧率の範囲であると考えられる。しかし、これらの抑圧率はモーラ位置正解率の結果のみを考慮した値であるため、母音の種類によって処理効果の変動する可能性がある。各母音の適正抑圧率を設定することが可能であれば、モーラ位置正解率より得られた適正抑圧率の範囲をさらに絞り込み、提案手法による単語理解度の改善に繋がる可能性がある。今後は、母音に適した抑圧率を明らかにし、提案手法の改善を目指したい。

文 献

- [1] A. K. Nabelek *et al*, "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Am.*, 86, 1259-1265 (1989).
- [2] 荒井隆行 他, "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, 1, 449-450 (2001).
- [3] T. Arai *et al*, "Effects on suppressing steady-state portions of speech on intelligibility in reverberant environment," *Acoustical Science and Technology*, 23, 229-232 (2002).
- [4] N. Hodoshima *et al*, "Improving syllable identification by a preprocessing method reducing overlap-masking in reverberant environments," *J. Acoust. Soc. Am.*, 119, 4055-4064 (2006).
- [5] 辻美咲 他, "残響環境における音声明瞭度を目的とした子音強調・母音抑圧による前処理," 日本音響学会誌, 69(4), 179-183 (2013).
- [6] 坂本修一 他, "単語理解度試験におけるモーラ同定に対する親密度の影響," 日本音響学会誌, 60(7), 351-357 (2004).
- [7] 天野成昭 他, "新密度別単語理解度試験用音声データセット 2003 (FW03)," 音声資源コンソーシアム (2006).