

## 残響環境下における前処理を用いた音声明瞭度の改善 —単語の聞き取りに効果的な処理の検討—\*

◎千葉亜矢子, 荒井隆行 (上智大・理工), 栗栖清浩 (TOA),  
程島奈緒, 吉田航(上智大・理工)

### 1 はじめに

残響の影響によって音声が聞き取りにくくなることがあるが, 残響環境下で音声明瞭度を改善する前処理の一つとして, 荒井らは定常部抑圧処理を提案した[1,2]。この処理は音声知覚において比較的重要ではない定常部[3]を抑圧することにより *overlap-masking* の影響を軽減し, 音声明瞭度を改善するものである。

先行研究[4]において, 単音節明瞭度試験では定常部抑圧処理の効果が得られているが, 単語了解度試験や文章了解度試験では処理の十分な効果が得られていない。しかし, 日常生活でこの処理を活用する場合, 人が単音節のみを聞き取る場面はほぼなく, 単語や文章などを聞き取り易くしなければならない。ゆえに, 本研究ではターゲットを単語に絞り, 音声明瞭度が改善されるような定常部抑圧処理を調査するために, 2つの聴取実験を行った。

本論文では, 今まで提案されてきた定常部抑圧処理[1,2,4]に加え, 新しい定常部抑圧処理として母音定常部抑圧処理(Vsss)を提案した。従来の処理では, 子音・母音に関わらず全定常部を抑圧したり(sss1[1,2]), 一部のエネルギーの弱い子音を抑圧せずに保持しながらも多くの子音の定常部を抑圧したり(sss3[4])していた。しかし, 定常部であっても子音は音声知覚に重要な部分であると考えられる。ゆえに, 子音を抑圧せず, 母音の定常部のみ抑圧する処理を提案した。この処理を施すことで子音は抑圧されないため, 従来の処理以上の明瞭度改善が期待される。

実験 I では, 単語了解度試験で正解率が高くなる処理を検討するために, 従来の処理 sss1, sss3 と提案処理の効果の比較を行った。実験 II では, 原音声として合成音声を用いる

実験を行った。合成音声を用いることで, 子音・母音の判別がしやすい音声波形が得られ, 実験 I よりも精度良く, 母音の定常部のみ抑圧することができ, 明瞭度の改善が期待される。

### 2 実験 I

#### 2.1 聴取実験

##### 2.1.1 原音声

NTT-AT 親密度別単語了解度試験用音声データベース(FW03)の女性話者(hi)より, ターゲットとなる 4 モーラ語を選択した。原音声はキャリアフレーズ「モクテキ」を用いて, 「モクテキ○○○○」(○○○○:ターゲット)とした。そして, キャリアフレーズとターゲットの音圧レベル比を 1:0.7 とした。ターゲットとして単語親密度 4.0~2.5 の単語を用いた。

##### 2.1.2 処理条件

Table 1 に, 本実験に用いた処理を示す。本実験では, Vsss を施すにあたり, まず sss3 で抑圧箇所を自動判定した後, 音声波形やスペクトログラムの変化を基に, 実験者の判断で子音, 母音の遷移部, 撥音が抑圧されていた場合, 抑圧しないようにするなど, 手動で修正した。Fig.1 に sss1, sss3 と Vsss の, 抑圧の仕方の違いを示す。sss1, sss3 の処理は, Fig.1 のように遷移部と定常部の間で抑圧の度合いを徐々に変化させ抑圧した。Vsss の手動修正を行った箇所のみ, 遷移部と定常部の間で抑圧の度合いを急激に変化させた。

##### 2.1.3 刺激

原音声に Table 1 に示した 6 種類の処理を施しインパルス応答を畳み込んだ音声と, 処理を施さずにインパルス応答を畳み込んだ音声(unproc)を刺激とした。また, 話者の発話の明瞭性を確認するために, インパルス応答が畳み込まれていない原音声(org)も用意した。ゆ

\* Improving speech intelligibility by preprocessing approach in reverberant environments: Investigation of effective processing for word perception by CHIBA, Ayako, ARAI, Takayuki (Sophia Univ.) KURISU, Kiyohiro (TOA), HODOSHIMA, Nao and YOSHIDA, Wataru(Sophia Univ.).

Table 1 実験 I に用いた処理の種類

	抑圧率 40%	抑圧率 70%
sss1	sss1_40	sss1_70
sss3	sss3_40	sss3_70
Vsss	Vsss_40	Vsss_70

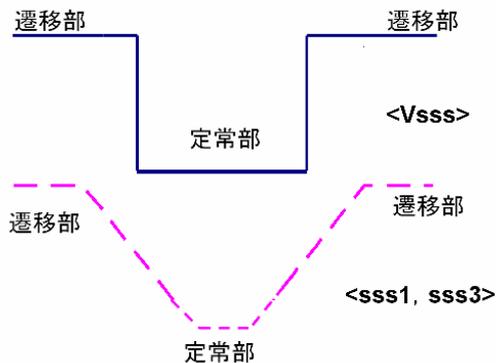


Fig. 1 実験 I の提案法(一部のみ)と従来法の立ち上がり立ち下りの違い

えに計 8 条件の音声刺激を作成した。

#### 2.1.4 残響条件

先行研究[4]で使用された残響時間 RT=2.0s, D 値=67%のインパルス応答を, 残響条件として用いた。このインパルス応答は, 日本建築学会音声伝送品質研究ワーキンググループデータベースに含まれる, 講演ホールで測定されたインパルス応答を基にして, 時間包絡を変えて RT を変化させたものである。

#### 2.1.5 参加者

実験参加者は, 日本語を母語とする 19~34 歳の健聴者 24 名(男性 9 名, 女性 15 名)であった。参加者全員は 250Hz ~ 4kHz の純音が 20dB HL で聞き取れており, 全ての参加者の聴力について問題はないと判断した。

#### 2.1.6 実験手順

実験は防音室内で行った。刺激音はヘッドホン(STAX SR-303)から提示した。実験中は刺激音を一度だけ提示した後に, 刺激を提示しているコンピュータとは別のコンピュータ上のスプレッド・シートに, 聞こえたターゲットをひらがなで入力させた。参加者がターゲットの一部を聞き取って, 既知の単語を推測して回答することを避けるために, 「聞き取る単語には, 意味がないこともある」と教示した。全 8 条件に対し各条件 10 刺激ずつ割り当て, 始めにインパルス応答を畳み込んだ音声 70 刺激, その後に原音声 10 刺激, 計 80

刺激を提示した。単語毎に処理の効果や残響環境下での聞き取りの難易度が異なると考えられるため, 全刺激が全条件で参加者に提示されるようにカウンタバランスを取った。

## 2.2 結果

原音声(org)の正解率が 95%であったことから, 話者の発話は十分明瞭であったと言える。処理毎の正解率を Fig. 2 に示す。正解は, ターゲットの 4 モーラ全てが正解することを条件とした。その結果, 残響環境下で処理を施さない音声(unproc)よりも, 抑圧率 70%の sss1(sss1\_70)と抑圧率 70%の Vsss(Vsss\_70)を施した音声の方が, 正解率が高かった。処理(7 条件)間で *t* 検定を行った結果, 各処理条件間に, 有意差が見られなかった。

## 3 実験 II

### 3.1 聴取実験

#### 3.1.1 原音声

合成音声としてアルカディアの「SpeeCAN SFT 5」を用いて作成したものを使用した。実験 I と同様, 女性話者による音声を用いた。また, ターゲットとなる 4 モーラ語も実験 I と同じ単語を用いた。原音声はキャリアフレーズ「題目としては」を用いて, 「題目としては○○○○」(○○○○:ターゲット)とした。

#### 3.1.2 処理条件

Vsss の抑圧率 40%と 70%を用いた。実験 II では, 抑圧箇所の持続時間を測定し, それに基づき抑圧するという, 実験 I とは異なる Vsss の施し方を行った。そして, 全ての定常部と遷移部間の立ち下がり, 立ち上がり時間を 10ms とし, 定常部と遷移部間では Fig. 1 下図の sss1, sss3 のような処理を施した。

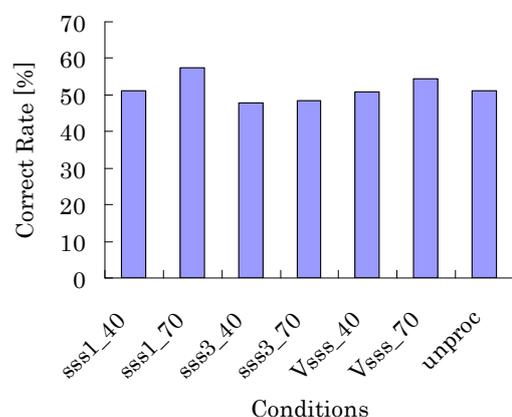


Fig. 2 実験 I の処理毎の正解率

### 3.1.3 刺激

原音声に2種類(Vsssの40%, 70%)の処理を施しインパルス応答を畳み込んだ音声と、処理を施さず、原音声にインパルス応答を畳み込んだ音声(unproc)を刺激とした。また、合成音声の発話の明瞭性を確認するために、インパルス応答が畳み込まれていない原音声(org)も用意した。計4条件の音声刺激を作成した。

### 3.1.4 残響条件

実験Iと同じインパルス応答を用いた。

### 3.1.5 参加者

実験参加者は実験Iに参加していない日本語を母語とする18~24歳の健聴者20名(男性8名, 女性12名)であった。但し、健聴者かどうかの判断は参加者の自己申告によるものとした。

### 3.1.6 実験手順

実験Iと同じ手順で行った。全4条件に対し各条件10刺激ずつ割り当て、始めにインパルス応答を畳み込んだ音声30刺激、その後原音声10刺激、計40刺激を提示した。今回もカウンタバランスをとって実験を行った。

## 3.2 結果

原音声(org)の正解率が97%であったことから、合成音声の発話は十分明瞭であったと言える。処理毎の正解率をFig. 3に示す。実験Iと同様、正解はターゲットの4モーラ全てが正解することを条件とした。その結果、抑圧率40%(Vsss\_40)の処理を施した音声、抑圧率70%(Vsss\_70)の処理を施した音声、施していない音声(unproc)の順で正解率が高かった。しかし、処理(3条件)間でt検定を行った結果、各処理条件間に、有意差が見られなかった。

## 4 考察

### 4.1 処理に関して

#### 4.1.1 従来処理と提案処理の比較

実験Iより、単語了解度試験で処理なしよりも、有意に正解率が高くなる処理を見つけ出すことはできなかった。結果を詳しく見てみると、有意差はないものの、Fig. 2より、sss1を施した音声の正解率が、処理の効果が一番期待されていたVsssの正解率よりも高かった。ゆえに、定常と判定された子音部分の抑圧は、単語了解度を低下させない可能性が考えられる。なお、sss3はVsssよりも音声明瞭

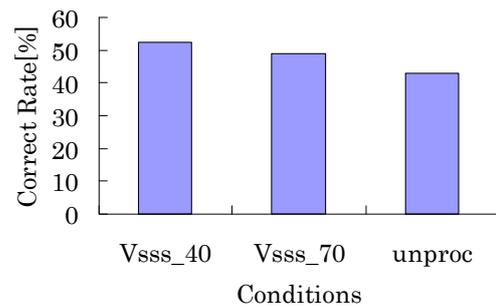


Fig. 3 実験IIの処理毎の正解率

度が低かったが、音声波形を確認すると、sss3によって子音保持がされずに、1モーラ目の子音部を抑圧している事が多かった。ゆえに、1モーラ目の遷移部の抑圧が音声明瞭度に大きく影響していると考えられるため、本実験ではVsssとの比較は行わなかった。

#### 4.1.2 抑圧率・抑圧箇所に関して

実験I・IIどちらも抑圧率条件間で有意差は見られなかった。ゆえに抑圧率の違いが単語の正解率に大きく影響しないと考えられる。

有意差は見られなかったが、実験Iで抑圧率が70%の方が、抑圧率が40%よりも正解率が高かったのに対し、実験IIでは抑圧率が40%の方が、抑圧率が70%よりも正解率が高かった。ゆえに、実験IIで合成音声を用いて精度良く母音の定常部を抑圧した結果、実験Iよりも聞き取りに重要な部分が保持され、処理による明瞭度の改善が見られた可能性が考えられる。また、実験IIでは遷移部と定常部の間の、立ち下がり、立ち上がり時間を10msとしたため、一部の定常部と遷移部の間の抑圧を急激に変化させていた実験Iよりも、聞き取りに不自然さがなくなり、処理によって明瞭度が改善した可能性も考えられる。

#### 4.2 残響に関して

実験I・II共に、残響環境下では処理の有無に関わらず、正解率に有意差がなかった。この原因として、先行研究[5]で得られた、処理による明瞭度の改善が見られる残響時間0.9~1.2sに比べ、使用したインパルス応答の残響時間が長かったことが考えられる。

残響時間を長くすると、overlap-masking量が増加する。よって本実験では、抑圧処理を施してoverlap-masking量を減らしてもoverlap-masking量が多過ぎることにより、明瞭度が改善するほどの変化が見られなくなっ

てしまったことが考えられる。ゆえに、単音節明瞭度試験で処理の効果が最も得られた残響時間を用い、overlap-masking 量を同じにして単語了解度試験を行う必要がある。

しかし、その残響条件では処理の有無に関わらず、処理の効果を見るには正解率が高すぎてしまい、処理の効果が見られない可能性があると考えられる。そのようになることを避けるために、白色雑音を付加して明瞭度を下げる、もしくは、原音声として劣化音声を用いて実験を行うなどの工夫をして聴取実験を行うことが考えられる。

#### 4.3 単音節明瞭度試験との比較

ターゲットが単音節の場合、先行研究[4]では処理の効果が得られていた。しかし、本実験でターゲットを単語にすると、処理の効果は得られなかった。そこで、キャリアフレーズの長さの違いとターゲットの音素数の違いに注目する。

先行研究[1,2,4]で行われた単音節明瞭度試験では、キャリアフレーズの長さが 1.17s だった。一方、実験 I ではキャリアフレーズの長さは 0.85s だったため、先行研究[4]で行われた単音節明瞭度試験の方が実験 I で行われた単語了解度試験より、ターゲットにかかる overlap-masking 量が多いと考えられる。そして、キャリアフレーズに処理をかけると、キャリアフレーズが長い方が抑圧される部分も多いため、ターゲットにかかる overlap-masking 量はより減少する。よって、単音節明瞭度試験で音声明瞭度の改善が見られたと考えられる。

それを踏まえ、実験 II ではキャリアフレーズの長さは 1.26s を用いた。しかし、明瞭度は改善されなかった。これより、キャリアフレーズの増加によって overlap-masking がより減少しても、明瞭度が改善しないことがわかった。原因として、先に述べたように残響時間が長かったことにより明瞭度が改善するほどの変化が見られなかった可能性に加え、本実験では、単音節明瞭度試験で効果が得られた話者とは異なる話者による音声を用いたことにより、処理の効果が見られなかった可能性が考えられる。

## 5 おわりに

本研究は、残響環境下での音声明瞭度改善

を最終目的としている。よって、単語をターゲットにした際に効果が得られる定常部抑圧処理を調査する目的で 2 つの聴取実験を行った。

実験 I では、計 6 種類の処理を用いて単語了解度試験を行った。そして実験 II では、原音声として合成音声を用いて、実験 I よりも精度の良く母音の定常部を判定する処理を音声に施して、実験 I と同様の単語了解度試験を行った。

2 つの実験により、精度良く母音の定常部を抑圧することによって明瞭度が改善する可能性、使用したインパルス応答の残響時間が長過ぎるために処理の効果が得ることができなかった可能性、処理の効果が得られている先行研究と本研究とでは実験で用いた音声の話者が異なることによって効果が得られなかった可能性が見られた。

処理を施しても、明瞭度が低下することはなかった。理由として、overlap-masking が抑圧部分に付加されたため、抑圧による聴感上の不自然さが解消されたと考えられる。しかし、本実験では生じた overlap-masking 量が多過ぎたため、明瞭度改善にはつながらなかった。ゆえに、単音節よりも単語を聞き取る時に、処理を施す際の抑圧率と残響時間のバランスが重要であることが考えられる。

今後、残響環境下での単語の聞き取りに有効な処理を検討していくことが必要である。そのために、残響時間の短いインパルス応答を使用し、原音声に劣化音声を用いる実験や、合成音声を使った単音節明瞭度試験などに取り組みたい。

#### 謝辞

本研究にあたり、技術協力をいただきました株式会社アルカディアに感謝申し上げます。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「人間情報科学プロジェクト」の支援を受けて行われた。

#### 参考文献

- [1] 荒井隆行 他, 音論 (秋), 1, 449-450, 2001.
- [2] T. Arai *et al.*, *Acoust. Sci. Tech.*, 23(4), 229-232, 2002.
- [3] S. Furui *et al.*, *J. Acoust. Soc. Am.*, 80(4), 1016-1025, 1986.
- [4] 中田有貴, 上智大学大学院理工学研究科電気・電子工学専攻 2007 年度修士論文.
- [5] 程島奈緒 他, 電子情報通言学会技術報告, SP2002-65, 47-51, 2002.