

雑音・残響環境下における音声明瞭度改善のための 定常部抑圧処理の評価*

◎吉田航, 程島奈緒, 荒井隆行 (上智大)

1. はじめに

残響は音声の聴き取りを妨害することがある。この原因のひとつに直前の音素に付加された残響の尾が、後続の音素に重なる **overlap-masking** の影響が挙げられる[1]。この影響を軽減する前処理として、荒井らは定常部抑圧処理を提案した[2,3]。この処理は、音声知覚に比較的重要ではない定常部[4]のエネルギーを抑圧することで **overlap-masking** の影響を軽減し、音声明瞭度の改善を図っている。先行研究[5-8]では、残響時間 0.7 ~ 1.3 s の条件下において、定常部抑圧処理を施すことで音声明瞭度を改善することができるという結果を得ている。

ところで、空港や駅などの公共空間においては、雑音だけでなく周囲の雑音もまた音声の聴き取りの妨げとなることがある。しかし、このような雑音や残響が付加された空間においても、音声による情報の取得が求められることがあり、音声明瞭度を保持しなければならない。しかしながら、これまで定常部抑圧処理の効果は、雑音と残響の両方が存在する音場では確かめられていない。

本研究は、公共空間での音声明瞭度改善を念頭に置いて、雑音と残響の両方が存在する環境下で定常部抑圧処理の効果を検査するために、これらを模擬した環境下で単音節明瞭度試験を行った。また、雑音と残響それぞれの環境下での、定常部抑圧処理の効果の度合いを比較するため、雑音のみ付加された環境下と残響のみ付加された環境下においても同様の試験を行った。

2. 実験

2.1 聴取実験

2.1.1 刺激

原音声は、先行研究[5-8]と同様、ATR 研究用日本語音声データベースより日本語の単音節 CV (子音-母音) をターゲットとし、キャ

リアセンテンス「題目としては__といます」に挿入して作成したものを用いた。ターゲットの V として /a/, C として /p, t, k, b, d, g, s, j, h, dz, dʒ, tʃ, m, n/ の 14 種類を用いた。

刺激作成の手順を Fig. 1 に示す。残響を付加しない音声の作成手順は①-③のとおりである。

- ① : 先行音効果[9]を考慮するため、原音声と、原音声に定常部抑圧処理を施した音声に、本実験で使用したインパルス応答の直接音から 50 ms の成分 (h_{50}) のみを畳み込む。
- ② : ①の音声を A 特性で重み付けをする (A-filter)。
- ③ : 実効値 (RMS) を算出し、①の音声 28 種類毎に異なる値を掛けて、音圧レベル (SPL) を揃える。ここで、雑音と残響が付加されてなく、かつ処理を施していない音声を S とし、処理を施している音声を S_pro とする。

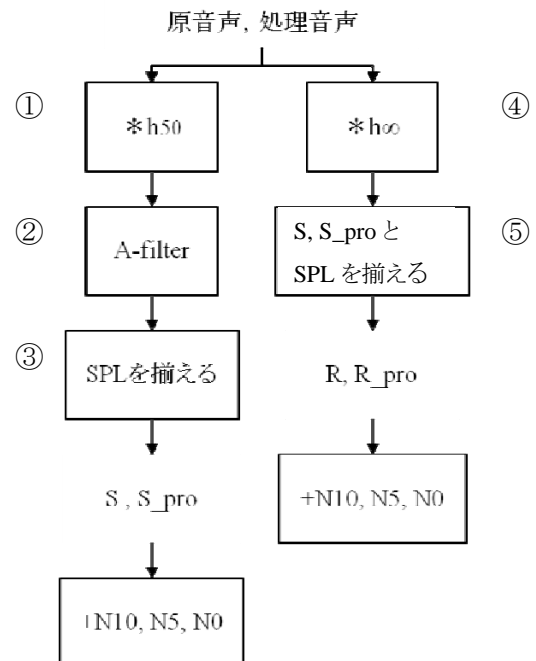


Fig. 1 刺激作成の手順

* Evaluation of steady-state suppression for improving speech intelligibility in noisy and reverberant environments, Wataru Yoshida, Nao Hodoshima and Takayuki Arai (Sophia Univ.).

残響付加した音声の作成手順は④-⑤のとおりである。

- ④ :原音声と処理を施した音声に、本実験で使用したインパルス応答の後期反射を含んだ全成分 (h_{∞}) を畳み込む。
- ⑤ :A 特性を考慮した S, S_pro と SPL を揃えるため、③で SPL を揃える際に用いた 28 種類毎に異なる値を、④の同じ種類の音声に掛ける。ここで、残響が付加されて、かつ処理を施していない音声を R とし、処理を施している音声を R_pro とする。

本実験で使用したインパルス応答は、定常部抑圧処理の効果が確認された先行実験[7]と同じ、XEBEC ホールで測定されたもの(残響時間は 0.9 s)を使用した。

雑音は、公共空間での人の話し声や足音を想定して NOISEX-92 からバブルノイズを用いた。雑音も音声と同様にまず A 特性で重み付けし、RMS を算出した。その後、S との SNR が 10 dB, 5 dB, 0 dB となるように雑音の RMS を調整して作成した (N10, N5, N0)。ここで S を基準信号として SNR をとったため、R, R_pro との SNR は 10 dB, 5 dB, 0 dB ではない。

本実験では、N10, N5, N0 を S, S_pro に各々足し合わせることで、雑音のみ付加された音声を作成した。また、N10, N5, N0 を R, R_pro に各々足し合わせることで、雑音と残響が付加された音声を作成した。なお、本論文では雑音を付加していない音声の SNR は ∞ dB とする。

以上の条件より、総刺激数は 224 (ターゲット 14 種 \times 処理 2 種 \times 残響 2 種 \times 雑音 4 種) となる。

2.1.2 実験手順

実験は防音室で行い、刺激音の提示はヘッドホン (STAX SR-303) による diotic 受聴とした。本番前に練習を 6 回行い、刺激音の提示レベルは全参加者で 65.0 dBA とした。実験では、各試行において刺激音を一度だけ提示した後に、コンピュータの画面上に選択肢を平仮名で提示し、参加者はその中から聴き取った刺激音一つを回答した。選択肢は使用したターゲット 14 種類 と、実験では使用していない/a/と、「その他」を加えた 16 種類とした。「その他」を選択した場合は、聴き取っ

た CV を別途紙に書き取ることとした。なお、参加者には選択肢の中の/a/と「その他」の 2 種類は、実験では使用していないことを知らせなかった。刺激は始めに S, S_pro の 28 刺激を除く 196 刺激を参加者ごとにランダムに提示し、最後に残った S, S_pro の 28 刺激を同様の方法でランダムに提示した。

2.1.3 参加者

実験参加者は、日本語を母語とする 19~23 歳の健聴者 21 名 (男性 11 名, 女性 10 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。

2.2 結果・考察

2.2.1 実験結果

全雑音条件におけるターゲットの正解率の平均を Fig. 2 に示す。ここで、雑音環境下で処理を施していない音声を Sig, 処理を施している音声を Sig_pro, 雑音・残響環境下で処理を施していない音声を Rev, 処理を施している音声を Rev_pro とする。

SNR が ∞ dB の音声を除いた処理 2 条件、残響 2 条件、雑音 3 条件に関して分散分析を行った。その結果、音声に定常部抑圧処理を施すことで音声明瞭度は改善した ($p < 0.01$)。また、雑音が付加されることで明瞭度は低下することが示された ($p < 0.01$)。また、残響と雑音の交互作用も有意 ($p < 0.01$) であったことから、雑音の大きさや残響時間の長さによって、明瞭度低下の度合いが異なることが示された。

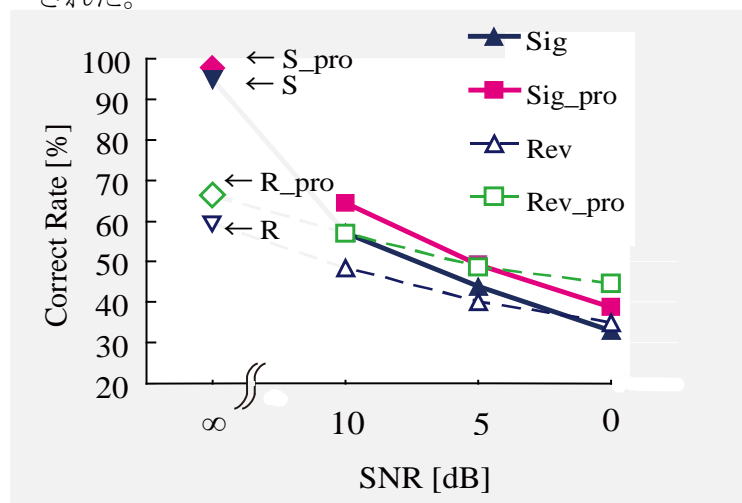


Fig. 2 SNR 毎のターゲットの正解率

各雑音条件下における処理の効果を調べるために、下位検定として Sidak の多重比較を行った。その結果、雑音・残響環境下において、SNR が 10 dB, 5 dB, 0 dB いずれの場合も、処理なしの音声より処理ありの音声の方が正解率は有意に高かった (SNR=10 dB : $p<0.05$, SNR=5, 0 dB : $p<0.01$)。

残響のみ付加された環境下での処理の効果を調べるため、 t 検定を行った。その結果、残響のみの環境下で処理の効果は有意であった ($p<0.01$)。これより、先行研究同様、残響のみ付加された環境下において、定常部抑圧処理を施すことで音声明瞭度を改善できることが示された。

2.2.2 残響のみの環境下での評価

残響が付加されていない S, S_pro を聴き取った場合、それらの正解率は各々 94.9 %, 97.6 % と聴き取りにほぼ支障がないのに対し、残響を付加した音声 R, R_pro の正解率は各々 59.2 %, 66.3 % と減少している。R と R_pro の正解率の間に有意差があったことから、処理の効果が確認でき、先行研究[5-8]の結果が裏づけられた。

2.2.3 雑音のみの環境下での評価

Sig, Sig_pro において SNR が小さくなるにつれて、正解率が減少した。これは、SNR が小さくなるほど、雑音によってターゲットがよりマスクされ、子音を識別するのがより困難となり、正解率が下がったと考えられる。

処理なしと処理ありの音声の正解率に有意差はなかった。この要因としては、定常部抑圧処理は残響による overlap-masking の影響を軽減する処理であるためと考えられる。しかしながら、処理なしの音声の正解率より処理ありの音声の正解率の方が高く、改善の傾向が確認された。これは定常部抑圧処理を施すことで、母音の定常部を抑圧し、相対的に子音部や母音への遷移部を強調することで子音が聴き取りやすくなったためと考えられる。

本実験の参加者は健聴者であったのに対し、小林らは、雑音も残響も付加されていない環境下において高齢者を対象に定常部抑圧処理の効果を調査した[10]。その結果、処理を施すことで明瞭度が改善された。高齢者や聴覚障害者は、雑音の影響によって健聴者よりも

聴き取りづらくなる[11]ので、雑音が存在する環境下においても、定常部抑圧処理の効果があるか調査する必要がある。

2.2.4 雑音・残響環境下での評価

雑音・残響環境下において、SNR が小さくなるにつれて、Rev, Rev_pro の正解率が減少した。また、SNR が 10 dB では処理の有無に関わらず、雑音のみ付加された音声の方が、雑音と残響の両方が付加された音声より正解率は有意に高かった (Sig と Rev, Sig_pro と Rev_pro : $p<0.05$)。しかし、SNR が 5 dB より小さくなると、有意ではなかったものの、雑音・残響環境下の音声の方が正解率は高くなった。

この要因について、佐藤らは、騒音レベルが高い音場では残響による音声の補強効果があると述べている[12]。本研究では、SNR が 0 dB の雑音環境下において、残響が付加された音声の正解率の方が、付加されていない音声より高く、さらに処理なしに比べ、処理ありの方が正解率の差は顕著に見られた。これより、雑音は先行するキャリアセンテンスからの overlap-masking の影響を軽減すると同時に、ターゲットにかかる残響によって音声が増強されたため、Rev_pro の正解率が最も高かったと考えられる。

雑音・残響環境下の SNR 毎の正解率を Table 1 に示す。Table 1 より、処理を施した音声と施していない音声の正解率の差を比較すると、SNR が小さくなるにつれて、改善の割合が小さくなることはなかった。このことより、残響環境下に雑音が存在する場合、雑音の影響で音声明瞭度が低下するものの、残響のみ付加された環境下と比較して改善率は低下しない。つまり、残響環境下で雑音が付加されても、定常部抑圧処理の効果は健在すると考えられる。しかしながら、ターゲットが完全に雑音にマスクされるほど SNR が小さい場

Table 1 雑音・残響環境下の SNR 毎の正解率[%]

SNR	雑音・残響環境下		
	Rev	Rev_pro	改善率
∞dB	59.2	66.3	7.1
10dB	48.3	56.8	8.5
5dB	40.1	48.6	8.5
0dB	35.0	44.6	9.5

合は、処理による改善が得られなくなるものと予想される。

3. おわりに

本報告の目的は、雑音・残響環境下(1種類の残響と3種類の雑音)において、荒井らが提案した定常部抑圧処理[2,3]を音声に施すことの効果とどの程度あるかを調査するものであった。残響を模擬した環境下で聴取実験を行った結果、SNRが10~0 dBの雑音と残響時間が0.9 sの残響が付加された環境下において、処理を施すことで音声明瞭度は有意に改善された。また、SNRが0 dBの雑音のみ付加された環境下より、同じレベルの雑音と、残響の両方が付加された環境下の方が、有意ではなかったものの、処理を施した音声の明瞭度は高かった。これらのことより、雑音と残響が存在する環境下でも音声明瞭度が改善される可能性が示された。

本実験は、残響、雑音共に使用した条件が少なかったことから、今後は雑音と残響の条件を増やし、定常部抑圧処理の効果と残響時間及びSNRの範囲や、雑音と残響の相互関係などをさらに詳細に調査していきたい。さらに、公共空間などで実際に案内放送や講演に定常部抑圧処理を実用することを想定し、聴取実験をdichotic受聴で行う、あるいは単語理解度試験や文章理解度試験を行うなどによって処理の効果を追ってみたい。

謝辞

本研究を行うにあたって、TOA株式会社の栗栖清浩さんにご助言を頂いた。

また、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] R. H. Bolt and A. D. MacDonald, J. Acoust. Soc. Am., 21(6), 577-580, 1949.
- [2] 荒井隆行 他, 音講論(秋), 1, 449-450, 2001.
- [3] T. Arai *et al*, Acoust. Sci. Tech., 23(4), 229-232, 2002.
- [4] S. Furui, J. Acoust. Soc. Am., 80(4), 1016-1025, 1986.

- [5] N. Hodoshima *et al*, J. Acoust. Soc. Am., 119(6), 4055-4064, 2006.
- [6] N. Hodoshima *et al*, Acoust. Sci. Tech., 26(2), 2122-214, 2005.
- [7] 中田有貴 他, 音講論(秋), 693-694, 2005.
- [8] N. Hayashi *et al*, Proc. Interspeech, 1741-1744, 2005.
- [9] A. J. Watkins, J. Acoust. Soc. Am., 106(5), 2933-2944, 1999.
- [10] 小林敬 他, 日本音響学会誌, 64(5), 278-289, 2008.
- [11] J. S. Bradley *et al*, J. Acoust. Soc. Am., 106(4), 1820-1828, 1999.
- [12] 佐藤洋 他, 日本建築学会計画系論文集, 484, 1-8, 1996.