

定常部抑圧による異なる残響環境下での音声明瞭度の改善

◎井上 豪, 程島奈緒, 荒井隆行, 木下慶介(上智大・理工)

楠本亜希子(Dept. of Veterans Affairs, Portland VA Medical Center, OR 97207, USA)

1. はじめに

ホールなどで音楽演奏を聞くと、ほどよい残響がかかり、音が豊かに聞こえることがあるが、かかり過ぎる残響は特に講演演説などに対し聞こえを悪くさせる。このことは特に高齢者や聴覚障害者には健聴者よりも深刻な問題である。

このように残響が聞こえを劣化させる理由の1つとして、残響が時間的に後方に尾を引くため、直前の音声信号に対する残響が現在の音声信号をマスクすることから、音声の明瞭度が低下すると考えられている[1][2]。

残響環境下における音声明瞭度を改善させる手法は pre-processing と post-processing の2つに、大きく分けられる。pre-processing は残響が掛かる前の音声信号に処理を施す、つまりマイクロフォンとスピーカ間で音声処理を施す方法[1][2][4][5][6]、であるのに対して、post-processing はすでに残響が付加された音声に対して処理を行い残響成分を抑圧する方法である(例えば[5])。

荒井らは、pre-processing として音声の遷移部はそのまま保存し、音声の定常部のみ抑圧する処理を行い、残響によるマスクの影響を減らす結果を得ている[1][2]。また楠本らは pre-processing として音声知覚に重要とされる変調周波数帯4 Hz を強調する処理を行い、音声明瞭度の改善する結果を得ている[4]。

我々は残響時間を変化させた数種類のインパルス応答と、Kusumoto et al.[4]の変調フィルタ処理を用いて聴取実験を以前行なっている[6]。実験結果から変調フィルタリングの効果は残響時間によって異なり、特定の条件下では明瞭度の低下を抑えることができると確認された。Hodoshima et al.[6]では変調フィルタ処理に関して実験を行っているのに対し、本論文では、定常部抑圧処理を様々な残響条件において調べることで残響時間と処理の効果の関連を得ることを目的とした。

そこで、残響時間を変化させた数種類のインパルス応答と Arai et al.[1][2]で用いた定常部抑圧処理を用いて聴取実験を行った。

2. 定常部抑圧処理

定常部抑圧処理は、Arai et al.[1][2]と同様の処理を用いた。まず、音声信号を 1/3-oct に帯域分割し、各帯域において包絡を抽出した。次にダウンサンプリングされた包絡成分の対数を取り、前後計5点の回帰係数を計算し、帯域に渡って回帰係数の2乗平均(以下では D とする)を求めた。ここで、 D は Furui にならない音声のスペクトル遷移を表すパラメータを表す[3]。元の標準化周波数に戻した後、 D が一定の閾値より小さい箇所を定常部とし、定常部とみなした箇所は元の波形の振幅を40%に抑圧した。

3. 聴取実験

我々は様々な残響条件を実現するため、Hodoshima et al.[6][7]にならない数種類のインパルス応答を人工的に作成した。各々の残響時間を表1に示す。

刺激は、日本語の単音節 CV (子音-母音) をターゲットとし、日本語のキャリアセンテンス「題目としては__といます」に挿入した。V として /a/, /i/, /e/, /o/ を、C として /p/, /t/, /k/, /b/, /d/, /g/, /s/, /ʃ/, /h/, /tʃ/, /dz/, /dʒ/, /m/, /n/ の14種類を用いた。実験では単音節24種類をターゲットに使用した。各刺激は ATR 研究用日本語音声データベース(話者: MAU, 40才男性)を用いた。ターゲットは単音節データを使用したのに対し、キャリアセンテンスは文章データの中から2文を選択し、その一部ずつを使用した(前半部と後半部で異なる文を使用した)。全ての音声サンプルで、直前の音節からのターゲットに対するマスク量の統制をとるため、ターゲットの母音の開始位置はその直前のキャリアセンテンスの終了部から 150ms に統一した。

*Improvement of speech intelligibility under various reverberant environments by the steady-state suppression

By Tsuyoshi Inoue, Nao Hodoshima, Takayuki Arai, Keisuke Kinoshita (Sophia Univ., Tokyo, Japan) and Akiko Kusumoto (Dept. of Veterans Affairs, Portland VA Medical Center, OR97207, USA)

また、ターゲットの実効値は同じ母音の CV セットごとに正規化した。

刺激音はオリジナルの音声信号に残響を畳み込んだ刺激セット(org_rev)と、定常部抑圧処理を行った後に残響を畳み込んだ刺激セット(proc_rev)の2種類を用意した。

被験者は日本語を話す健聴者24人(男性14人、女性10人、年齢18-26才)であった。

表1 本論文で使った残響条件.

インパルス応答	h1	h2	h3	h4	h5
残響時間(s)	0.9	1.0	1.1	1.2	1.3

4. 実験結果

各残響条件、処理条件における子音の正解率の平均値を図1ならびに表2に示す。ただし母音の正解率は、いずれの条件においても99%以上であった。以下では外れ値である2名を除き、22人を分析の対象とした。繰り返しのある2×5の分散分析を行なったところ、処理による主効果($p < .001$)、残響による主効果($p < .001$)が有意であった。処理条件間でのt検定では、h1~h4条件において有意差が得られた (h1: $p = .034$; h2: $p = .026$; h3: $p = .003$; h4: $p < .001$)。

表2 各条件における子音の正解率.

	h1	h2	h3	h4	h5
org_rev (%)	68.3	63.5	61.4	55.1	58.1
proc_rev (%)	73.1	68.3	67.4	64.2	58.5

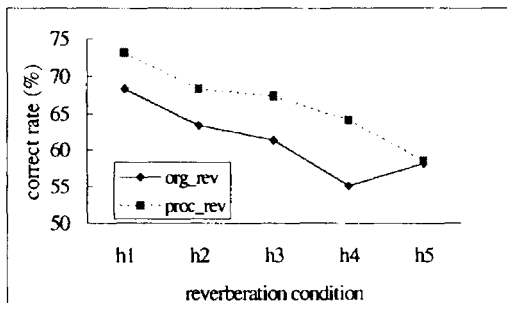


図1 各条件における子音の正解率.

5. 考察

残響条件における主効果は有意であり、さらに主効果の成分を多項式展開すると一次成分で有意であったため、正解率は処理によらず残響時間が長くなるにつれて単調減少することが確認された。実験の結果から、全ての残響条件においてproc_revの方が正解率は高

く、さらに残響時間が0.9~1.2秒ではproc_revの方がorg_revよりも正解率が有意に高いことが示された。残響時間が1.2秒において最も処理による改善が得られたことから、この残響条件においては直前の残響成分によるターゲットへのマスクング量を最も抑えることができたと考えられる。またproc_revとorg_revの正解率の間には、残響時間が短くなるにつれて差がみられなくなると予想されるが、本実験で用いた条件の中で残響時間が最も短い0.9秒においても処理による改善が得られていることから、残響時間がh1より短い場合においても処理による効果のある条件が存在するものと考えられる。以上より、定常部抑圧は前処理として音声明瞭度の改善にとって有効であり、残響時間によって処理の効果が異なることが確認された。

- [1] 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, 1, pp. 449-450, 2001.
- [2] 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, 23, pp. 229-232, 2002.
- [3] S. Furui, “On the role of spectral transition for speech perception,” J. Acoust. Soc. Am., 80(4), pp. 1016-1025, 1986.
- [4] A. Kusumoto, T. Arai, M. Takahashi and Y. Murahara, “Modulation enhancement of speech as a preprocessing for reverberant chambers with the hearing-impaired,” Proc. IEEE ICASSP, pp. 933-936, 2000.
- [5] T. Langhans and H. W. Strube, “Speech enhancement by nonlinear multiband envelope filtering,” Proc. IEEE ICASSP, pp. 156-159, 1982.
- [6] N. Hodoshima, T. Arai and A. Kusumoto, “Enhancing temporal dynamics of speech to improve intelligibility in reverberant environments,” Proc. Forum Acusticum Sevilla, 2002.
- [7] 程島奈緒, 井上豪, 荒井隆行, 木下慶介, 楠本亜希子, “定常部抑圧による音声明瞭度改善のための前処理 -異なる残響環境下について-,” 電子情報通信学会技術報告, SP2002-65, pp. 47-51, 2002.