

小・中規模ホール環境を想定した定常部抑圧による拡声音声の明瞭度改善

©程島奈緒, 荒井隆行, 井上豪, 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

広い室内においてしばしば講演の内容を理解することが難しい場合があり、その一つの原因として残響による影響があげられる。残響は音楽にとっては深みを与えるため必要だが、一方で音声明瞭度を減少させる。残響によって音声明瞭度が減少する原因は、主に先行する音素に付加される残響の尾が後続の音素に影響を与える overlap-masking の影響によると考えられている[1][2]。

残響環境下において音声の明瞭度を改善する方法には、主に microphone-array, post-processing, pre-processing の 3 種類の手法がある。microphone-array では空間情報を利用して目的音の到来方向を推定し、目的音を強調する。post-processing では音声が入射され、残響が付加された後で処理を行う。pre-processing は音声に残響が付加される前に音声処理を行うことにより残響の影響を軽減させる手法である。pre-processing では処理をマイクロフォンとスピーカ間で行うことから、PA (Public Address) の一部に組み込むことで実用化できる。

荒井らは、pre-processing 処理として残響による overlap-masking の影響を軽減するため、音声の定常部を抑圧する処理を行い音声明瞭度の改善を得た[3][4]。overlap-masking の影響を軽減させるためには、先行音のエネルギーをあらかじめ抑えることにより、後続する音素に付加される残響成分のエネルギー量を減らせばよい。一方、音声の遷移部は音声知覚に関して非常に重要な役割を果たしているが、母音の定常部は音節の認識において必要ではないという報告がある[5]。そこで荒井らは、エネルギーは比較的大きいが音声認識にはそれほど重要ではないとされる音声の定常部をあらかじめ抑圧することで、残響による overlap-masking の影響を軽減することに実験的に初めて成功したのである[3][4]。

本論文の目的は、定常部抑圧処理の効果を様々な残響条件下において調べることである。本研究の最終目的は、任意のホールに対して音声明瞭度の低下を抑える最適な pre-processing 処理を提供することである。そのためには残響時間と処理の効果の関連を知る必要があるが、残響時間、処理の 2 つのパラメータを同時に変化させて処理の効果を調べることは難しい。そこで我々はまず残響時間のみを変化させ、単一処理の効

果が残響時間によってどのように変化するか注目した。我々は以前、残響時間 (RT)=0.9~1.3s の場合における定常部抑圧処理の効果を調べた。その結果、特定の残響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、残響時間が比較的短い場合においても処理の効果が認められた[6][7]。本論文ではさらに短い残響時間における処理の効果を確認するため、先行研究よりも残響時間が短い RT=0.4~1.0s を用いて聴取実験を行った。

2. 聴取実験

我々は様々な残響条件を実現するため、式(1)に示すようにインパルス応答 h_n を人工的に作成した[8]。

$$h_n(t) = e^{-\frac{t}{\tau}} h_o(t) \quad (1)$$

ここで τ は時定数、 h_o は東京都東大和市大ホール(反射板なし)で測定されたインパルス応答であり、RT=1.1s であった。我々は式(1)から、 h_o の残響時間を基準として τ を変化させることにより RT=0.4, 0.6, 0.8, 0.9, 1.0s の計 5 種類のインパルス応答を作成した。

本論文で用いた定常部抑圧処理は、先行研究で用いた処理と同じである[3][4]。本処理は音声のスペクトル遷移を表す D [5]を計算し、 D が一定の閾値より小さい箇所を定常部とした。 D は Furui によって提案されたパラメータ[5]と基本的に同じであるが、本論文で使用した D は帯域分割された信号の時間包絡の対数に対する回帰係数を複数帯域に渡って二乗平均したものを用いた。そして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。

原音声は、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に挿入した。V として /a/, /i/, /e/, /o/, /u/, /ɛ/, /ɔ/, /ɪ/, /ʊ/, /ɔ̃/, /ɛ̃/, /ɔ̃/, /ɪ̃/, /ʊ̃/ の 14 種類を用いた。各刺激は ATR 研究用日本語音声データベースを用いた。刺激音として原音声(Org)、原音声に対して定常部抑圧処理を行った刺激(Proc)、Org に残響を畳み込んだ刺激 (Org_rev)、Proc に残響を畳み込んだ刺激(Proc_rev)の計 4 種類を使用した。被験者は日本語を母語とする健聴者 22 名であった。実験の手順は、各試行において刺激音を提示後、PC の画面上に選択肢として実験で使用した 24 種類の CV をかなで表示した。被験者には、画面上の選択肢を強制的に一つマウスでクリックさせ回答させた。各被験者に対して計 288 刺激 (残響 5 種類 × 24 CV × 処理 2

* Improving intelligibility of speech through PA systems by steady-state suppression in small to medium sized halls.

By Nao Hodoshima, Takayuki Arai, Tsuyoshi Inoue, Keisuke Kinoshita (Sophia Univ., Tokyo, Japan) and Akiko Kusumoto (Portland VA Medical Center, OR 97207, USA)

表 1 各条件における正解率

RT (s)	0.4	0.6	0.8	0.9	1.0
Org_rev (%)	90.7	84.3	73.9	70.1	69.5
Proc_rev (%)	92.8	86.6	82.4	79.2	73.9

Org (%)	97.0
Proc (%)	97.2

種類+24CV×処理2種類)をランダムに並べて提示した。

3. 結果、考察

表 1 に各残響条件、処理条件における正解率の平均値を示す。Org_rev, Proc_rev 全体に対して繰り返しのある 2×5 の分散分析を行ったところ、処理による主効果 [$p < 0.001$]、残響による主効果 [$p < 0.001$]、処理条件と残響条件の交互作用 [$p = 0.007$] が有意であった。処理による主効果が有意であり、Proc_rev の正解率(85.3%) が Org_rev の正解率(80.9%) よりも高いことから、定常部処理を行うことによって明瞭度が改善されたことが示された。さらに t 検定の結果から RT=0.8, 0.9, 1.0s において Proc_rev の正解率が Org_rev の正解率よりも有意に高いことが示された [RT=0.8s: $p < 0.001$; RT=0.9s: $p < 0.001$; RT=1.0s: $p = 0.05$]。残響による主効果は有意であり、RT=0.4s の正解率(91.8%) が RT=1.0s の正解率(71.7%) よりも高かったため、正解率は残響時間が長くなるにつれ減少することが示された。処理条件と残響条件の交互作用が有意であることから、処理の効果は残響時間によって異なることが示された。RT=0s である Org と Proc の正解率はほぼ等しいことから、音声の定常部を抑圧してもクリーン環境下における明瞭度は原音声と変わらないことが示された。

表 1 から残響時間が短くなるにつれ Proc_rev と Org_rev の正解率の差が減少している。これは残響時間が短くなるにつれて残響による overlap-masking の影響が減少するために、Proc_rev, Org_rev 共に正解率が上昇し、定常部抑圧処理の効果が現れなくなったからである。実験結果から、音声の定常部を 40% に抑圧した処理の効果が得られる残響時間の下限は RT=0.8s であった。

4. おわりに

本論文では、荒井らによって提案された手法[3][4]に基づき、残響環境下において overlap-masking の影響を軽減することにより音声明瞭度の低下を抑える定常部抑圧処理が有効である範囲を検討した。定常部抑圧処理を様々な残響条件において調べるため、残響時間を人工的に細かく変化させた複数のインパルス応答を用いて聴取実験を行った。結果から、RT=0.8~1.0s において処理による改善が得られた。先行研究では、RT=0.9~1.3s の範囲で本論文と同じ定常部抑圧処理を

用いて聴取実験を行った結果、RT=0.9~1.2s において処理による改善が得られた[6][7]。先行研究[6][7]と本研究の結果から、定常部を 40% に抑圧した処理による明瞭度の改善が得られる範囲は RT=0.8~1.2s であることが示された。定常部に対する抑圧の度合いを変化させることで明瞭度の改善が得られる残響時間の範囲は異なると予測されることから、今後は明瞭度が問題であるとされる多目的ホール規模の比較的長い残響時間において有効である処理の度合いを検討したい。また本論文では残響時間を人工的に作成したが、今後は実在するホールにおいて処理の効果を検討したい。

謝辞

インパルス応答のデータを提供して頂いた東京大学生産技術研究所橘秀樹先生、上野佳奈子さん、横山栄さん、ご指導頂いた上智大学認知心理学研究室道又爾先生に心から感謝申し上げます。

文 献

- [1] R. H. Bolt and A. D. MacDonald, "Theory of speech masking by reverberation," J. Acoust. Soc. Am., 21, pp. 577-580, 1949.
- [2] A. K. Nabelek and L. Robinette, "Influence of precedence effect on word identification by normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am., 63, pp. 187-194, 1978.
- [3] 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.
- [4] 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.
- [5] S. Furui, "On the role of spectral transition for speech perception," J. Acoust. Soc. Am., 80(4), pp. 1016-1025, 1986.
- [6] 井上豪, 程島奈緒, 荒井隆行, 木下慶介, 楠本亜希子, "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," 日本音響学会秋季研究発表会講演論文集, 1, pp. 377-378, 2002.
- [7] N. Hodoshima, T. Inoue, T. Arai and A. Kusumoto, "Suppressing steady-state portions of speech for improving intelligibility in various reverberant environments," Proc. China-Japan Joint Conference on Acoustics, pp. 199-202, 2002.
- [8] N. Hodoshima, T. Arai and A. Kusumoto, "Enhancing temporal dynamics of speech to improve intelligibility in reverberant environments," Proc. Forum Acusticum Sevilla, 2002.