

残響環境下で前処理として音声明瞭度を改善する 定常部抑圧処理の最適な抑圧率の検討*

◎程島奈緒, 荒井隆行 (上智大・理工)

1 はじめに

残響の影響により音声の聴取が難しい場合がある。事実聴覚障害者、高齢者、非母語話者は、健聴の母語話者よりも残響による音声明瞭度への影響を受けやすいという報告が数多くなされている[1-3]。公共空間における音声明瞭度の確保は、講演や構内放送などにおける明瞭な音声の提供に加え、世界規模で高齢化が進行していることや音声警報という安全面やからも今後ますます重要な課題となるであろう。

残響環境下で音声明瞭度を改善する手法として、音声に残響が付加される前に信号処理を行なう前処理[4-13]が挙げられる。Arai et al.[7, 8]は前処理として音声の定常部を抑圧する定常部抑圧処理を提案し、先行する音声区間に付加された残響の尾が後続の音声区間に影響を与える overlap-masking の影響[14]を軽減することを試みた。さらに Hodoshima et al.は健聴者に対して単音節明瞭度試験を行ない、残響時間(RT)=0.7-1.2 s の模擬残響環境下[9, 10]・RT 約 1.3 s の講堂[11]において定常部抑圧処理による明瞭度の有意な改善を得た。Yasu et al.は高齢者に対して単音節明瞭度試験を行ない、RT=1.0 s, 1.3 s の模擬残響環境下において定常部抑圧処理による明瞭度の有意な改善を得た[13]。

本研究の目的は、任意の室に対して明瞭度の低下を抑える処理を提供することである。そのためには処理の効果を最大に得ることができるように、処理のパラメータを調整する必要がある。本報告では、定常部抑圧処理のパラメータのうち overlap-masking 量に直接的に対応していると考えられる定常部を抑圧する度合い(抑圧率)と明瞭度の関係について調べた。そのため抑圧率を変化させた定常部抑圧処理を用いて、健聴者に対して残響模擬環

境下における単音節明瞭度試験を行った。

2 聴取実験

残響条件として、東京都東大和市大ホールで測定(反射板なし)されたインパルス応答(RT=1.1 s)と、そのインパルス応答に時間包絡をかけ残響時間を 1.3 s にしたものを用いた。以下では前者を IR1, 後者を IR2 とする。本研究で用いた RT は、中心周波数 500Hz, 1kHz, 2kHz のオクターブバンドで帯域分けされたインパルス応答の EDT (Early Decay Time)の平均とした。

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基本的に同じである[7, 8]。本処理は音声のスペクトル遷移を表す D [15]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。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を用いた。定常部とみなした箇所は元の波形の振幅を 0, 10, 20, 30, 40, 50, 60, 100% に抑圧した。0%とは定常部の振幅を完全に抑圧すること、100%とは原音声に相当する。

被験者は健聴な日本語母語話者 21 名(男性 10 名, 女性 11 名, 年齢 18-25 才)であった。

原音声は、ATR 研究用日本語音声データベースから単音節 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ものを用いた。V として/a/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。

実験は被験者ごとに防音室内で行なった。刺激音の提示はヘッドフォン(STAX SR-303)を用い、音圧レベルはあらかじめ被験者ごとに適したレベルに調整した。各試行では刺激音を一度提示し、提示終了後 PC の画面上に選択肢として実験で使用した 14 種類の CV をかなで表示した。被験者には、画面上の選択

* Investigating an optimum suppression rate of steady-state portions of speech that improves intelligibility as a pre-processing approach in reverberant environments by Nao Hodoshima and Takayuki Arai (Sophia University).

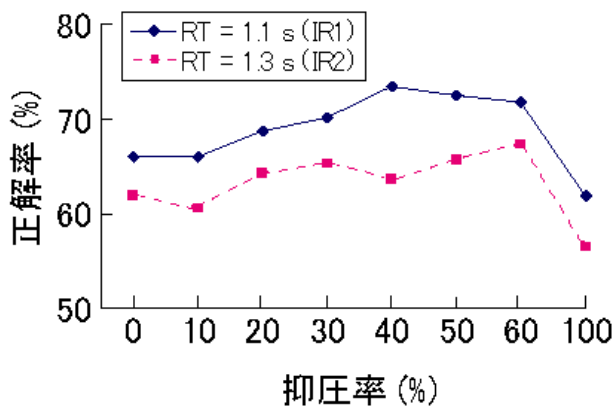


Fig. 1 抑圧率(0, 10, 20, 30, 40, 50, 60, 100%)と残響条件(IR1, IR2)における単音節14種類の正解率の平均。抑圧率100%は原音声に対応する。

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224刺激(残響2種類×14原音声×処理8種類)をランダムに提示した。

3 結果

Fig. 1に各残響・処理条件における正解率の平均を示す。両残響条件において、全処理音声の正解率の方が未処理音声の正解率よりも高く、抑圧率が同じであってもRTによって処理音声の正解率は異なることが示された。

残響・処理条件に対する繰り返しのある8×2の分散分析を行ったところ、残響による主効果($p < 0.01$)が有意であった。よってIR1の正解率の方がIR2の正解率よりも高いことが示された。

処理による主効果($p < 0.01$)は有意であった。下位検定としてSidakの多重比較を行ったところ、IR1では抑圧率100%(未処理音声)の正解率($M=61.9\%$)は抑圧率40%($M=73.5\%$, $p=0.02$)・50%($M=72.5\%$, $p=0.04$)の正解率よりも有意に低い結果が得られた。またIR2では、抑圧率100%(未処理音声)の正解率($M=56.5\%$)は、抑圧率60%の正解率($M=67.3\%$, $p=0.04$)よりも有意に低い結果が得られた。

4 おわりに

定常部抑圧処理の抑圧率を0%から60%まで変化させた結果、RT=1.1sでは抑圧率40%・50%で、RT=1.3sでは抑圧率60%で未処理音声より正解率が有意に高い結果が得られた。今回使用した処理条件のうち、RT=1.3sでは抑圧率が最も高い60%において最大の改善が

得られた。処理音声と未処理音声の正解率の差は、ある抑圧率で最大となった後に減少し、抑圧率が100%ではほぼその差はなくなると予想される。今後はさらに高い抑圧率を用い、最大の改善が得られる抑圧率を調べたい。また今回使用したRTとは異なる条件を用い、RTと抑圧率の関係を詳しく検討したい。さらに抑圧率以外のパラメータによる明瞭度の変化を調べたい。

謝辞

本研究は文部科学省の科研費(A-2, 16203041)、日本学術振興会の特別研究員奨励費(176911)から助成を得た。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の橋秀樹先生、上野佳奈子さん、横山栄さんに感謝申し上げます。

参考文献

- [1] Nábělek and Pickett, J. Speech Hear. Res., 17, 724-739, 1974.
- [2] Nábělek and Donahue, J. Acoust. Soc. Am., 75(2), 632-634, 1984.
- [3] Crandell and Smaldino, J. Lang. Speech Hear. Services Schools, 31 (4), 362-370, 2000.
- [4] Langhans and Strube, Proc. IEEE ICASSP, 7, 156-159, 1982.
- [5] Kitamura et al., Proc. ICSLP, 3, 586-589, 2000.
- [6] Kusumoto et al., Proc. IEEE ICASSP, 2, 853-856, 2000.
- [7] Arai et al., Proc. Autumn Meet. Acoust. Soc. Jpn., 1, 449-450, 2001.
- [8] Arai et al., Acoust. Sci. Tech., 23 (4), 229-232, 2002.
- [9] Hodoshima et al., Proc. Eurospeech, 1365-1368, 2003.
- [10] Hodoshima et al., Acoust. Sci. Tech., 25 (1), 58-60, 2004.
- [11] Hodoshima et al., Acoust. Sci. Tech., 26 (2), 212-214, 2005.
- [12] Kusumoto et al., Speech Comm, 45 (2), 101-113, 2005.
- [13] Yasu et al., Technical Report of IEICE Japan, 35 (2), 31-36, 2005.
- [14] Nábělek et al., J. Acoust. Soc. Am., 86 (4), 1259-1265, 1989.
- [15] Furui, J. Acoust. Soc. Am., 80 (4), 1016-1025, 1986.