

高齢者を対象とした残響環境下における 音声定常部抑圧処理の評価実験

安 啓一[†] 宮内 裕介[†] 程島 奈緒[†] 林 奈帆子[†] 井上 豪[†]
荒井 隆行[†] 進藤 美津子[‡]

† 上智大学理工学部電気電子工学科 ‡ 上智大学言語障害研究センター
〒102-8554 東京都千代田区紀尾井町 7-1
E-mail: † k-yasu@sophia.ac.jp

あらまし 残響環境下においては音声の明瞭度が低下し聞き取りが困難になる。これは先行する音声に付加された残響の尾が後続する音声にかかる overlap-masking により、音声の明瞭度が劣化したためである。講演やスピーチを聞き取る際、残響は健聴者だけでなく、高齢者や聴覚障害者にとっては更なる問題となる (Fitzgibbons and Gordon-Salant, 1999)。残響による音声の明瞭度の劣化を防ぐ方法は主に、前処理 (pre-processing) と後処理 (post-processing) の二つに分類される。荒井ら (2002, 2001) は overlap-masking の影響を軽減するための前処理として、音声の知覚にはそれほど重要でないといわれる音声定常部のエネルギーを抑圧する定常部抑圧を前処理として使用することにより、残響環境下における音声の明瞭度の改善を示唆している。また、先行研究 (程島ら 2003, 2004) では防音室内において、原音声にインパルス応答を畳み込んだ音声を提示することで模擬残響環境を実現し、定常部抑圧処理の効果を調査した。健聴者を対象に行った聴取実験の結果、残響時間 0.7s から 1.2s において音声の明瞭度が有意に改善された。本研究では高齢者を対象に聴取実験を行い、定常部抑圧により残響時間 1.0s と 1.3s において音声の明瞭度が有意に改善された。また被験者のうち、老人性難聴群においては処理による音声の明瞭度の改善は健聴者群に比べて大きいものとなった。

キーワード 音声強調, 高齢者, 残響, 定常部抑圧, 音声明瞭度

Evaluation of steady-state suppression of speech for elderly people in reverberant environments

Keiichi YASU[†] Yusuke MIYAUCHI[†] Nao Hodoshima[†] Nahoko HAYASHI[†]
Tsuyoshi INOUE[†] Takayuki ARAI[†] Mitsuko SHINDO[‡]

† Dept. of Electrical and Electronics Eng., Sophia University ‡ Research Center for Communication Disorders,
Sophia University, 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan
E-mail: † k-yasu@sophia.ac.jp

Abstract In reverberant environments, speech intelligibility is reduced and it is difficult for people to perceive speech. This is due to overlap-masking by reverberation components of previous segments (Bolt et al., 1949). For elderly people and hearing-impaired people, it is more critical issue to perceive and to understand speech in reverberant environments (Fitzgibbons and Gordon-Salant, 1999). There are two general approaches for improving speech intelligibility in reverberant environments: pre-processing and post-processing. To obviate the deterioration of speech intelligibility, Arai et al. suppressed steady-state portions of speech (steady-state suppression) that contained more energy compared with transitions but which were less crucial for speech perception, and confirmed a promising result for improving speech intelligibility (Arai et al., 2002). In previous studies, the effect of steady-state suppression was evaluated by measuring speech intelligibility of original and steady-state suppressed speech in simulated reverberation environments by convoluting speech with an impulse response in an auditorium at a sound proof room (Hodoshima et al., 2003, 2004). The results of the experiments with young normal hearing people showed that there were significant improvements in speech intelligibility using steady-state suppression in reverberant times between 0.7 s and 1.2 s. In this research, we conducted an experiment for evaluating steady-state suppression using reverberation times of 1.0 s and 1.3 s with fifty elderly people. We found that there were significant improvements in both reverberant conditions. Also, steady-state suppression yielded more improvement in intelligibility for a presbycusis group than a normal hearing group.

Keyword speech enhancement, elderly people, reverberation, steady-state suppression, speech intelligibility

1. はじめに

講演ホールや教会など残響時間の比較的長い広い室内においては音声の明瞭度が低下し、スピーチの聞き取りや内容の理解が困難になることがある。その原因の1つとして残響による影響が挙げられる。先行する音声区間の残響の尾が後続する音声区間をマスクする overlap-masking により、後続する音素のエネルギーが先行音素のエネルギーより小さい場合 overlap-masking の影響は大きくなる[1]。ホール等はその用途によって残響時間はさまざまで、音楽を目的とした比較的残響時間の長いホール等でスピーチや講演を行う場合には何らかの処置が必要となることが考えられる。

残響環境下において音声明瞭度の劣化を防ぐ対策として電気音響的なアプローチとして、前処理 (pre-processing) と後処理 (post-processing) がある。後処理として代表的なものには逆フィルタを用いることで残響除去を目的とする処理などがある[2]。前処理はマイクロフォンとスピーカの間で行われることから PA (public address) の一部に本処理を取り込むことで実用化が可能である。荒井らは、音声の遷移部に比べて音声知覚にはそれほど重要でないと言われる音声定常部のエネルギーをあらかじめ抑圧し、残響の尾による音声明瞭度の劣化を防ぐための前処理として“定常部抑圧処理”を考案した[3]。若い健聴者を対象に聴取実験を行ったところ、残響時間 0.7 s から 1.2 s において処理による主効果が有意であることが示された[4-6]。

残響は特に聴覚障害者や高齢者にとって深刻な問題である。残響の影響に加え、加齢に伴う聴力の低下と時間分解能の低下は音声明瞭度にさらなる影響を及ぼすと報告されている[7]。日本社会では高齢化が急速に進行し、平成 16 年度内閣府高齢者白書によると 2040 年には人口の 1/3 が高齢者になると推計されているが

[8]、高齢者を考慮した聞き取りやすい音声環境を提供する準備がまだ整っていない現状がある。アンケート調査[9]の結果、80 歳以上の高齢者のうち、全体の半分以上が講演会や集会でよく聞きとれないと回答した。またアナウンスがよく聞きとれないと答えた者は全体の 80% 以上に達した。

以上の理由から高齢者や聴覚障害者にとって聞き取りやすい音声を提供できる環境を提供することは今後の社会にとって必須事項である。よって本研究では定常部抑圧を前処理として利用し、処理音声の高齢者への効果を確認するために、高齢者 50 名を対象に音声明瞭度実験を行った。

2. 定常部抑圧処理

定常部抑圧処理は、音声信号のうち、エネルギーは大きいが生知覚には重要ではないとされる定常部のみを抑圧する。そうすることで、本処理は残響による overlap-masking の影響を最小限に抑えることを目的としている。

図 1 に定常部抑圧処理のブロックダイアグラムを示す。はじめに、標準化周波数 16kHz でサンプリングされた音声信号を、バンドパスフィルタを用いて 1/3 オクターブごとに帯域分割する。各帯域において時間包絡をヒルベルト変換によって抽出する。次に時間包絡にカットオフ周波数 30Hz のローパスフィルタをかける。続いて、この時間包絡を 100Hz にダウンサンプリングしてから、包絡成分の対数をとる。その対数軌跡から前後 2 点、計 5 点の回帰係数 (デルタ) をサンプル毎に計算し、全帯域に渡り回帰係数の 2 乗平均を求める。この 2 乗平均を D とし、この D を古井[10]にならない音声のスペクトル遷移を表すパラメータとする。

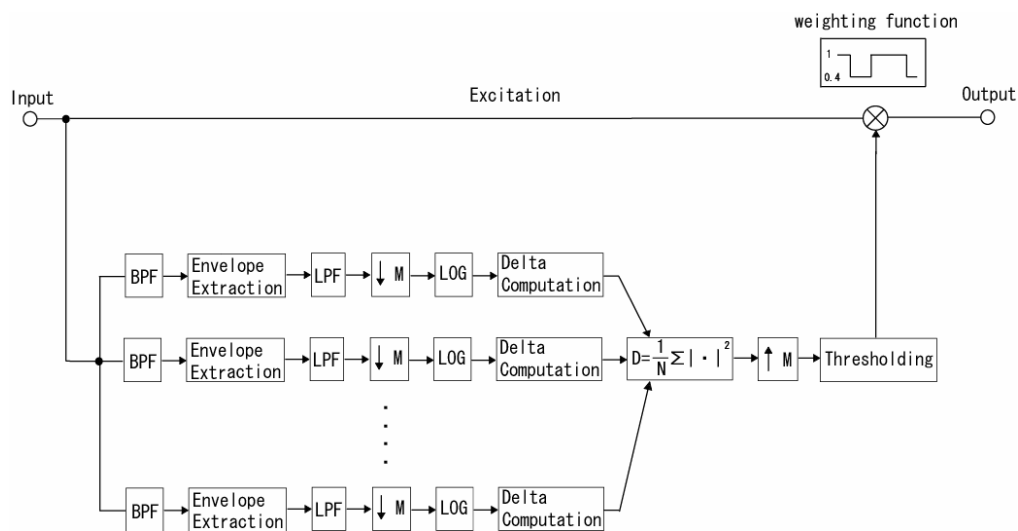


図 1 定常部抑圧処理のブロックダイアグラム

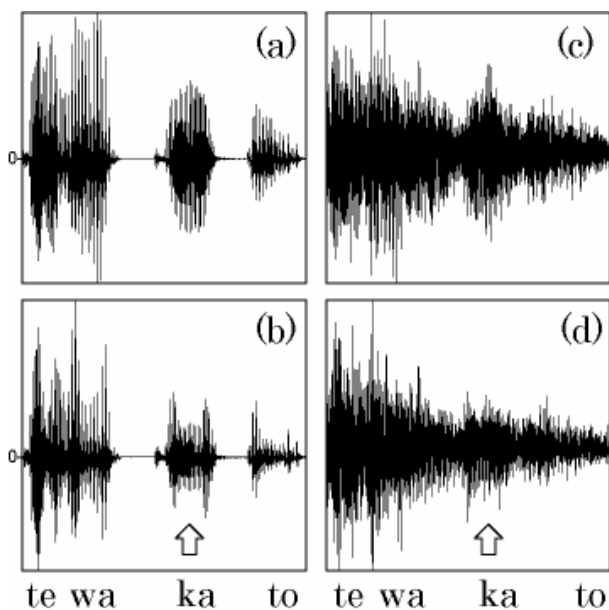


図 2 (a): 実験で使用した刺激の一部, (b): (a)に定常部抑圧処理を行った波形, (c): (a)に残響を畳み込んだ波形, (d): (b)に残響を畳み込んだ波形。矢印はターゲットを示す。

元の標本化周波数に戻した後に、 D が一定の閾値よりも小さい箇所を音声定常部とする。定常部とみなした箇所では元の時間波形振幅を抑圧する。振幅の抑圧率は先行研究と同様 40%とした。振幅を抑圧する際、急激に振幅が変化しないように台形補正をしている。図 2 に実験で使用した刺激の一部とそれに処理・残響を付加した音声波形。先行する音声区間に付加された残響の尾が後続する音素をマスクしている様子が見られる。

3. 実験

本研究の目的は高齢者を対象に聴取実験を行い、定常部抑圧処理の効果を調べることである。先行研究で健聴者を対象に実験を行ったところ、残響時間 0.7s から 1.2s にかけて定常部抑圧処理の効果が有意であるという結果が得られた[4-6]。本報告では、残響条件 2 条件を用いて処理の効果を調査した。

3.1. 被験者

被験者は日本語を母語とする年齢 56 歳 ~ 90 歳 (平均年齢 67.9 歳) の計 50 名 (男性 21 名、女性 29 名) であった。被験者は痴呆等の症状がなく、実験を遂行できるだけの運動レベル・知的レベルが保たれていた。実験を行う前にオーディオメータ(RION, AA-77)を用いた聴力検査を行い、各被験者のオーディオグラムを作成した。実験に際して被験者にアンケートをとり、耳鼻

咽喉関係の病気の既往歴を調査した。いずれの被験者も補聴器は未装着であった。

3.2. 実験に用いた刺激

原音声は、程島ら[4-6]と同様に、ATR 研究用日本語音声データベース (話者: MAU、40 歳男性) 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__といいます」に挿入して作成した。ここで V として /a/ を、C として /p/, /t/, /k/, /b/, /d/, /g/, /s/, /ʃ/, /h/, /dz/, /dʒ/, /tʃ/, /m/, /n/ の 14 種類を用いた。すべての音声サンプルで、キャリアセンテンスからターゲットに対するマスキング量の統制をとるため、ターゲットの母音の開始位置はキャリアセンテンス終了後 150ms に統一した。また、すべての刺激においてターゲットのエネルギーの実効値とキャリアセンテンスのエネルギーの実効値の比を一定にした。

実験に用いたインパルス応答は上智大学 10 号館講堂で測定したものの (残響時間 1.3s) と東京都東大和市大ホールのインパルス応答の残響時間を人工的に 1.0s にしたものである。

原音声に上記のインパルス応答を畳み込むことによって擬似的な残響環境を実現し、刺激音を作成した。実験には原音声に残響を畳み込んだ刺激セット (処理なし) と原音声に対して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 (処理あり) の 2 種類を使用した。表 1 に本実験で使用した 14 種類の CV を示す。

表 1 14 種類の単音節 (CV)

| | Voiceless C + Vowel | Voiced C + Vowel |
|---------------------|---------------------|------------------|
| Stop C + Vowel | /pa/ /ta/ /ka/ | /ba/ /da/ /ga/ |
| Fricative C + Vowel | /sa/ /ʃa/ /ha/ | |
| Affricate C + Vowel | /tʃa/ | /dza/ /dʒa/ |
| Nasal C + Vowel | | /ma/ /na/ |

3.3. 実験手法

実験は防音室内のコンピュータを用いて行った。刺激音はヘッドフォン (STAX SR-303) を用いて提示し、刺激音は各被験者の聞き取りやすいレベルに調整し、実験中は一定であった。実験形式に慣れるため、被験者は本番の前に本番と同様の形式をとった練習を 7 回行った。

実験中は使用した 14 種類の CV を選択肢としてあら

はじめ示し、刺激音を一度だけ提示した後に被験者に選択肢の中から一つ選択させ書き取らせた。被験者が回答を選択し、マウスをクリックすると次の刺激が提示された。以上の試行を計 56 刺激(14 単音節×残響 2 種類×処理あり/なし 2 条件)に対して行った。なお、刺激は被験者ごとに異なるランダム系列で提示された。また、実験は被験者のペースで進められた。

4. 実験結果と考察

4.1. 残響条件による考察

表 2 に本実験における被験者 50 名の各残響条件(東大和ホール: 残響時間 1.0s、上智大学 10 号館講堂: 残響時間 1.3s)における処理あり、処理なしの正解率の平均値を示す。各残響条件において t 検定を行ったところ、両残響条件とも定常部抑圧処理による効果が有意であることが示された ($p<0.01$)。

表 2 より両残響条件ともに処理なしの正解率よりも処理ありの正解率のほうが高いことから定常部抑圧処理が高齢者に対しても音声明瞭度を改善する効果が示された。

表 2 各残響条件における正解率の平均値

| 残響条件 | 東大和ホール | | 上智大学 10 号館講堂 | |
|---------|--------|------|-----------------|------|
| | 処理なし | 処理あり | 処理なし | 処理あり |
| 正解率 [%] | 34.4 | 45.8 | 37.3 | 44.9 |

両残響条件において定常部抑圧処理により明瞭度は改善されたが、その改善の度合いに差が見られた。処理を行うことで東大和ホールの正解率が上智大学 10 号館講堂の正解率を上回った。このことから定常部抑圧処理の効果は空間により異なることが示唆される。

東大和ホールと上智大学 10 号館講堂の残響条件に着目すると、処理なしの時には上智大学 10 号館講堂のほうが正解率が高いことが分かる。インパルス応答より計算した東大和ホールの残響時間は 1.0s、上智大学 10 号館講堂の残響時間は 1.3s であった。一般的に残響時間が長くなると音声明瞭度は低下するが、本実験ではそれとは逆の傾向が示された。しかし本実験では異なる空間のインパルス応答を用いている。よって本実験の残響条件による正解率の違いは残響時間以外にも要因があると思われる。

各残響条件における D (Deutlichkeit) 値(全体のエネルギーに対する初期のエネルギー比)を計算すると東大和ホールは 64.2%、上智大学 10 号館講堂は 65.5%であった(表 3)。本実験における各残響条件において処理なしの正解率と D 値の間には正の相関の傾向がみられた。

表 3 各残響条件における D 値

| 残響条件 | 東大和ホール | 上智大学 10 号館講堂 |
|-----------|--------|-----------------|
| D 値 [%] | 64.2 | 65.5 |

4.2. 聴力形による考察

オージオグラムを基に聴力型によるグループを求めた結果、高音域の閾値が上昇している者は 16 名、全体的に閾値の上昇している者は 1 名、健聴と判断された者は 22 名、その他は 11 名であった。聴力型と明瞭度、定常部抑圧処理の効果に対する比較を行うため、全体的に閾値の上昇していた 1 名とその他のグループに属する者に関しては比較の対象からはずし、健聴である高齢者群を健聴群、高音域の閾値が上昇している高齢者群を高音漸傾群として、実験結果に関して比較・検討を行った。被験者の平均年齢、平均聴力は表 4 に示した。また、図 3 と図 4 に各群のオージオグラムを示す。

表 4 高齢者グループ

| 被験者群 | 人数 [人] | 平均年齢 [歳] | 平均聴力 [dBHL] |
|-------|-----------|-------------|----------------|
| 健聴群 | 22 | 65.4 | 17.3 |
| 高音漸傾群 | 16 | 70.4 | 25.2 |

各群・残響条件における処理あり、処理なしの正解率の平均値を表 5 に示す。健聴群、高音漸傾群ともに処理なしよりも処理ありのほうが正解率が高い。 t 検定の結果、両群とも東大和ホールの残響条件では $p<0.01$ 、上智大学 10 号館講堂の残響条件では $p<0.05$ で処理間の有意差が認められた。

表 5 各残響条件における正解率の平均値 [%]

| 被験者群 | 各残響条件における正解率 [%] | | | |
|-------|------------------|------|--------------|------|
| | 東大和ホール | | 上智大学 10 号館講堂 | |
| | 処理なし | 処理あり | 処理なし | 処理あり |
| 健聴群 | 41.5 | 55.7 | 47.1 | 53.6 |
| 高音漸傾群 | 27.7 | 36.2 | 28.1 | 36.6 |

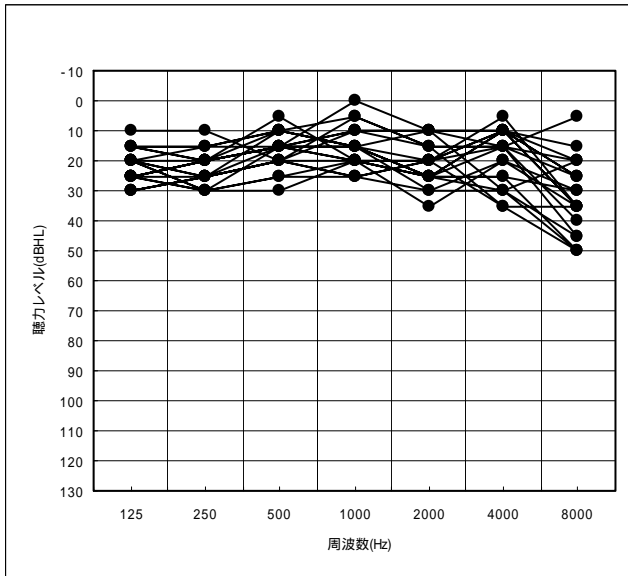


図3 健聴高齢者群のオーディオグラム

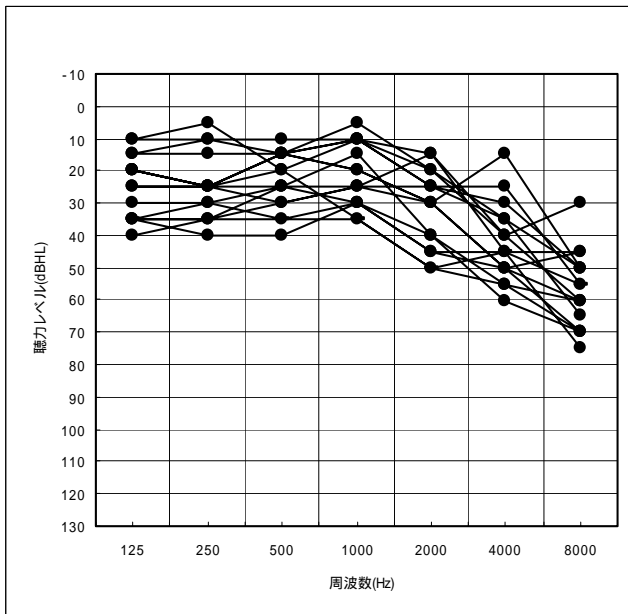


図4 高音域閾値上昇群のオーディオグラム

表5より、高音域閾値上昇群と健聴高齢者群を比較すると処理なし、処理ありともに健聴高齢者群の正解率のほうが高くなる結果となった。高音域閾値上昇群には残響条件の間での処理による明瞭度の改善度合いには差は見られなかったが、健聴高齢者群では残響条件の間で処理による明瞭度の改善度合いに大きな差が見られた。このことから定常部抑圧処理による効果は残響条件だけでなく被験者によっても異なると思われる。

5. おわりに

本研究により、定常部抑圧処理によって高齢者の残響環境下における明瞭度が改善することが示された。また、残響条件や被験者により定常部抑圧処理の効果が異なることが示唆された。本実験の受聴環境は原音声にインパルス応答を畳み込むことで擬似的な残響環境を実現したものであったが、実環境では明瞭度が上昇すると報告されており、先行研究[11]においても同様の傾向が見られる結果となった。以上のことを踏まえ、今後は高齢者による実環境での聴取実験を行い、処理の効果の差を検討したい。

6. 謝辞

本研究は、科学研究費補助金（A-2, 16203041）の助成を受けて行った。実験にご協力頂いた千代田区シルバー人材センター及び被験者の皆様に感謝いたします。

文 献

- [1] A. K. Nabelek and J. M. Pickett, "Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired," *Journal of Speech and Hearing Research*, 17, pp. 724-739, 1974.
- [2] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, 66(1), pp. 165-169, 1979.
- [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] N. Hodoshima, T. Inoue, T. Arai, A. Kusumoto, K. Kinoshita, "Suppressing steady-state portions of speech for improving intelligibility in various reverberant environments," *Acoust. Sci. Tech.*, 25(1), pp. 58-60, 2004.
- [5] N. Hodoshima, T. Arai, T. Inoue, K. Kinoshita and A. Kusumoto, "Improving speech intelligibility by steady-state suppression as pre-processing in small to medium sized halls," *Proc. Eurospeech*, pp. 1365-1368, 2003.
- [6] 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*, Vol. 65, pp. 47-51, 2002.
- [7] P. J. Fitzgibbons and S. Gordon-Salant, "Profile of auditory temporal processing in older listeners," *J. Speech, Language, and Hearing Reserch*, 42, pp. 300-311, 1999.
- [8] 内閣府, "平成15年度版高齢者政府白書," 2003.
- [9] 進藤美津子(研究代表者), "老人性難聴者の自覚意識の評価法とリハビリテーションに関する研究," 平成12年度-平成13年度科学研究費補助金(基盤研究(C)(2))研究成果報告書, 2002.
- [10] S. Furui, "On the role of spectral transition for speech perception," *J. Acoust. Soc. Am.*, 80(4), pp.

1016-1025, 1986.

- [11]N. Hodoshima, T. Goto, N. Ohata, T. Inoue and T. Arai, "The effect of pre-processing approach for improving speech intelligibility in a hall: Comparison between diotic and dichotic listening conditions," *Acoustical Science and Technology*, 26(2), 2005.