Vol.2013-NL-211 No.6
Vol.2013-SLP-96 No.6
2013/5/23

聴覚特性に基づく重み付け反復スペクトル減算法による音質改善の検討

福森 隆寛[1,a)]　堀井 圭祐[2]　中山 雅人[3]　西浦 敬信[3]　山下 洋一[3]

概要：実環境下での音収録において，周囲の雑音が目的信号に混入し音質が大きく劣化するという問題がある．そのため，収録した音を受聴する場合，混入雑音を抑圧し目的音のみを強調することが重要である．単一マイクロホンでの音収録における雑音抑圧手法としては，SS (Spectral Subtraction) が一般的に利用されている．SS は低演算コストで雑音を抑圧できるが，ミュージカルノイズと呼ばれる聴感上不快な雑音が発生する．そこで，SS を用いて雑音抑圧後の信号を受聴する場合，ミュージカルノイズを発生させずに混入雑音を抑圧する必要がある．これまで，ミュージカルノイズ低減のために SS を反復する手法が提案されており，その有効性が確認されている．しかし，これらの手法では全周波数で一様に雑音を抑圧しており，周波数毎に雑音抑圧量を制御することで更なるミュージカルノイズの低減が期待される．そこで，本研究ではミュージカルノイズが発生しない雑音抑圧手法の構築を目指して，聴覚特性に基づく反復 SS を提案する．提案法の有効性を確認するために，客観・主観評価実験を実施した．各評価実験の結果，提案法は従来法と比較して高い雑音抑圧性能を達成しつつ，主観的にミュージカルノイズを低減できた．

1. はじめに

近年，小型マイクロホンなどの収録機器の発達により誰でも気軽に音声を収録可能であるが，雑音下音声受音においてはエアコンや PC ファンなどの背景雑音が混入するため音質が大きく劣化する．対話型ロボットやスマートフォン上の音声認識サービスなどで受聴音声を利用する場合，高精度な音声認識性能が求められており，これまでに雑音混入音声から雑音のみを抑圧する手法の研究が盛んに行われてきた．雑音抑圧手法を用いることで，雑音環境下においても高い音声認識性能を達成可能であるが，テレビ電話やボイスレコーダによる議事録のように人間が雑音抑圧後の音声を受聴する場合，聴感上不快が少なく雑音を抑圧することが重要である．

これまでに雑音抑圧手法として，マイクロホンアレー [1]，独立成分分析 [2]，ウィナーフィルタ [3] を用いた手法，SS (Spectral Subtraction)[4] 等が提案されており，これらの手法を用いることで効果的に雑音を抑圧することが可能である．マイクロホンアレーや独立成分分析は，2 つ以上のマイクロホンを用いることで雑音を抑圧する手法であるが，高精度に雑音を抑圧するためには，多数のマイクロホンが必要である．一方，ウィナーフィルタを用いた手法は単一マイクロホンで雑音を抑圧可能であるが，源信号のパワースペクトルが必要である．源信号のパワースペクトルを推定する手法は検討されているが，計算コストの増大が問題視されている．

SS は，ウィナーフィルタを用いた手法と同様に単一マイクロホンで雑音を抑圧でき，特に観測信号の無音声部分から雑音を推定するため，雑音混入音声のみを用いて低演算コストで雑音を抑圧できる．しかし，SS では雑音抑圧後の音声にミュージカルノイズ [5] と呼ばれる聴感上不快な雑音が発生する問題がある．ミュージカルノイズは受聴者にとって不快な雑音であるため，SS を用いて雑音抑圧後の音声を受聴する場合，ミュージカルノイズを発生させずに混入雑音を抑圧する必要がある．そこで本研究では，ミュージカルノイズを低減するために，聴覚特性に基づく重み付き係数を用いた反復 SS を提案する．

これまでにミュージカルノイズ低減のためには，雑音抑圧量を抑えて減算処理を繰り返し反復する手法の有効性が確認されている．しかし，従来のミュージカルノイズ低減手法は全周波数で一様な処理係数を用いて雑音を抑圧しており，ミュージカルノイズを低減するためには周波数毎に雑音抑圧量を制御することが望まれる．周波数毎に雑音抑

[1] 立命館大学 大学院情報理工学研究科
Ritsumeikan University, Kusatsu, Shiga 525–8577, Japan
[2] 立命館大学 大学院理工学研究科
Ritsumeikan University, Kusatsu, Shiga 525–8577, Japan
[3] 立命館大学 情報理工学部
Ritsumeikan University, Kusatsu, Shiga 525–8577, Japan
[a)] cm013061@ed.ritsumei.ac.jp

Vol.2013-NL-211 No.6
Vol.2013-SLP-96 No.6
2013/5/23

圧量を制御するためには，処理係数を重み付けする必要がある．また，人間は周波数毎に聴こえる音の大きさが異なるという特徴を有する．そこで，本研究では聴覚特性に基づいた重み付き係数を用いて減算処理を反復することで，ミュージカルノイズの発生を低減可能な雑音抑圧手法を実現する．なお，今回は客観的に雑音抑圧性能を，そして主観的にミュージカルノイズ残存量を評価することで提案法の有効性を確認する．また，提案法は減算処理を反復することを前提としているため最適な反復回数についても検討し，実用化に向けた雑音抑圧指標を策定する．

2. 従来法

2.1 SS (Spectral Subtraction)

SS[4] は，目的信号と雑音が無相関であると仮定して，観測信号から雑音を推定し周波数領域で減算することにより雑音を抑圧する手法である．観測信号のパワースペクトルを $|Y(\omega)|^2$，目的信号のパワースペクトルの推定量を $|\hat{X}(\omega)|^2$，観測信号から推定した雑音のパワースペクトルを $|\hat{N}(\omega)|^2$ とすると，SS は式 (1) のように表される．

$$|\hat{X}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - \alpha|\hat{N}(\omega)|^2, & \text{if (P)}, \\ \beta|Y(\omega)|^2, & \text{if (O)}, \end{cases} \quad (1)$$
$$\text{(P)} = (|Y(\omega)|^2 - \alpha|\hat{N}(\omega)|^2 > \beta|Y(\omega)|^2),$$
$$\text{(O)} = (\text{otherwise}),$$

ここで，α は減算係数，β はフロアリング係数を表し，一般的な SS では，$\alpha > 1.0$，$0 < \beta \ll 1$ の範囲を採用している．また，式 (1) における otherwise の場合は雑音推定誤差により，減算処理後に目的音声のパワーが負値になることを防ぐためにフロアリング処理が行われる．そして，推定された目的信号のパワースペクトルと観測信号の位相を用いて逆フーリエ変換することで，雑音が抑圧された信号を算出することが可能である．しかし，SS では雑音抑圧後にミュージカルノイズ [5] と呼ばれる聴感上不快な雑音が発生することが問題視されている．

2.2 ミュージカルノイズ低減手法

2.2.1 I-SS (Iterative-Spectral Subtraction)

これまでにミュージカルノイズを低減することを目指して数多くの手法 [6], [7] が提案されているが，その中でも減算処理を反復する反復 SS (I-SS:Iterative-SS)[8] が広く利用されている．反復回数を i とすると，I-SS は式 (2) のように表される．

$$|\hat{X}_i(\omega)|^2 = \begin{cases} |\hat{X}_{i-1}(\omega)|^2 - \alpha|\hat{N}_i(\omega)|^2, & \text{if (P)}, \\ \beta|\hat{X}_{i-1}(\omega)|^2, & \text{if (O)}, \end{cases} \quad (2)$$
$$\text{(P)} = (|\hat{X}_{i-1}(\omega)|^2 - \alpha|\hat{N}_i(\omega)|^2 > \beta|\hat{X}_{i-1}(\omega)|^2),$$
$$\text{(O)} = (\text{otherwise}),$$
$$i = 1, 2, 3, \ldots, n, \quad |\hat{X}_0(\omega)| = |Y(\omega)|,$$

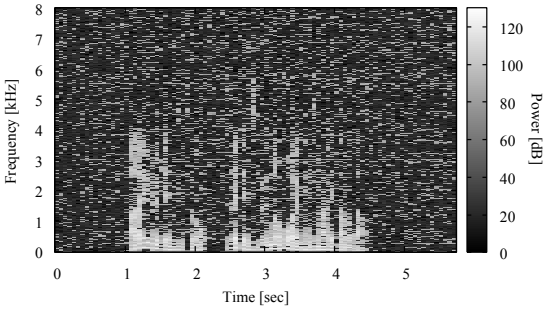


図 1 SS による雑音抑圧後のスペクトログラム
Fig. 1 Speech spectrogram after noise reduction by SS.

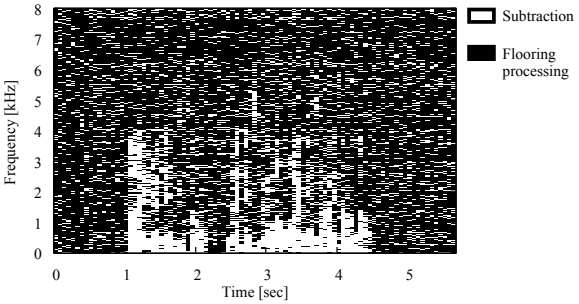


図 2 SS における雑音抑圧処理の分布
Fig. 2 Distribution of noise reduction processing in SS.

I-SS は，SS における雑音推定と減算処理を反復する手法であり，反復の度に推定雑音を更新するため，ミュージカルノイズを含めて抑圧可能である．また，反復回数が多いほど雑音抑圧量は増大するものの，音声成分も抑圧されるため音声ひずみ量も多くなる傾向がある．I-SS はミュージカルノイズが発生した後に雑音を低減する手法であり，ミュージカルノイズの発生自体を防ぐことはできない．

2.2.2 F-SS (Flooring processing-improved Spectral Subtraction)

ミュージカルノイズの発生を低減するために，従来の SS におけるフロアリング処理部を改良したフロアリング処理改良型 SS (F-SS: Flooring processing-improved SS)[9] が提案されている．一般的にフロアリング係数は，$0 < \beta \ll 1$ の非常に小さい値を用いられているが，F-SS は従来とは異なるフロアリング係数 ($0 \ll \beta < 1$) を用いて雑音を抑圧する．フロアリング係数を大きく設定することで，ミュージカルノイズの発生を低減可能であるが，1 度の減算処理では高い雑音抑圧性能は達成できない．そこで，F-SS は I-SS 同様に反復処理を行うことで，ミュージカルノイズの発生を低減しつつ高い雑音抑圧性能を達成できる．本研究では，従来手法を改良し更に効率よくミュージカルノイズの発生を低減できる SS の提案を目指す．

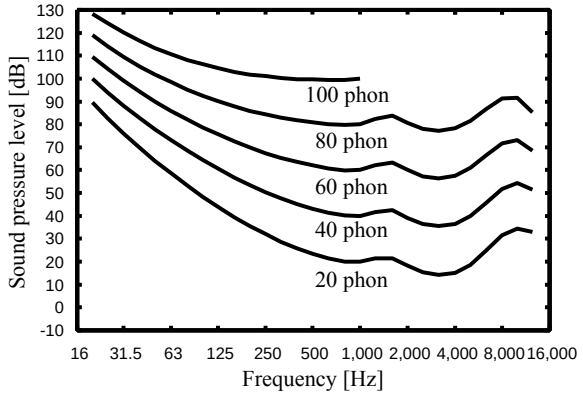


図 **3** 等ラウドネス曲線
Fig. 3 Equal loudness contour.

3. ミュージカルノイズ低減に向けた聴覚特性に基づく提案法

3.1 ミュージカルノイズの発生原因と低減の方針

これまでに，減算処理とフロアリング処理の間に生じるパワー差が，雑音抑圧後のスペクトログラム上に局所的なピークとして出現し，このピークがミュージカルノイズとして知覚されると考えられてきた [5]．特に，減算処理とフロアリング処理が混在する場合にはパワー差が生じやすく，ミュージカルノイズも発生しやすい．そのため，ミュージカルノイズの発生を低減するためには両処理間に生じるパワー差を極力低減する必要がある．

図 **1** に SS による雑音抑圧後のスペクトログラム，図 **2** に雑音抑圧処理の分布図を示す．図 1, 図 2 より，雑音部分において減算処理とフロアリング処理が頻繁に切り替わる場合，スペクトログラム上に局所的なピークが発生することを確認できる．2 つの処理間のパワー差が，スペクトログラム上に局所的なピークとして発生する原因として，フロアリング処理後のパワーが非常に小さい値となることが考えられる．フロアリング処理後のパワーが小さくなる一方，減算処理では一度の処理で大幅にパワーが低下することは少なく，減算処理部における雑音が残存してしまうため局所的なピークとして現れる．そのため，減算処理部のパワーが著しく大きくなり残存することを防ぐために，減算処理とフロアリング処理の間に大きなパワー差が発生しにくいフロアリング係数を採用することで，ミュージカルノイズの発生を低減できると考えられる．

3.2 L-SS (Loudness contour-weighted SS)

F-SS のように，1 に近いフロアリング係数を用いてフロアリング処理を行うことで，大幅なミュージカルノイズの低減を期待できるが，本研究では更に効率良く低減するために，減算処理とフロアリング処理間のパワー差を動的に制御する手法を提案する．前章で述べた従来のミュージカルノイズ低減手法では，全周波数で一様な処理係数 (α, β) を用いて雑音を抑圧しているが，周波数毎に処理係数を重み付けすることで，パワー差をより綿密に制御できミュージカルノイズの発生も低減できると考えられる．

人間の耳は周波数毎に聴こえる音の大きさが異なるため，処理係数の重み付け指標として聴覚特性に着目した．ここで，図 **3** に等ラウドネス曲線 [10] を示す．等ラウドネス曲線は，周波数毎に等しい大きさの音に聴こえる音圧レベルを結んで得られる曲線であり，低域の音が聴き取り難く 3 ∼ 4 kHz の音が聴き取り易いことを示す．本研究では，等ラウドネス曲線に基づいて処理係数を重み付けすることによって，周波数毎の雑音抑圧量を制御する．

等ラウドネス曲線は，1 kHz における音圧レベルを基準としているため，提案法においても 1 kHz における係数の値を基準値として設定する．他の周波数では等ラウドネス曲線に基づいて，受聴が困難な低域の雑音抑圧量を抑え，反対に受聴が容易な 3 ∼ 4 kHz の雑音抑圧量が多くなるように式 (3), (4) に基づいて各係数を重み付けする．

$$\alpha(\omega) = \alpha_{\mathrm{bsc}} - \alpha_{\mathrm{wt}}(L_c(\omega) - \mathrm{phon}), \tag{3}$$
$$\beta(\omega) = \beta_{\mathrm{bsc}} + \beta_{\mathrm{wt}}(L_c(\omega) - \mathrm{phon}), \tag{4}$$

ここで，α_{bsc}, β_{bsc} は各係数の基準値，$\alpha_{\mathrm{wt}}, \beta_{\mathrm{wt}}$ は各係数の重み，$L_c(\omega)$ は等ラウドネス曲線における音圧レベル，phon は 1 kHz における音圧レベル（音の大きさ）を示す．式 (3), (4) では係数の基準値や重みに依存して，α が負値になることや β が 1 以上になる可能性があるため，α は 1.0 を下限値とし β は 0.9 を上限値とする．各係数の基準値について，減算係数は従来の SS と同様に $\alpha > 1.0$ の値を用いるが，フロアリング係数は F-SS と同様に $0 \ll \beta < 1$ に設定する．フロアリング係数の基準値を 1 に近い値に設定することで，ミュージカルノイズを発生させないパワー差の制御を目指す．また，提案法は等ラウドネス曲線に基づいて算出した重み付き係数（$\alpha(\omega), \beta(\omega)$）を用いて，I-SS や F-SS と同様に減算処理を反復する手法である．提案法では，式 (5) を用いて目的信号のパワースペクトルを推定する．

$$|\hat{X}_i(\omega)|^2 = \begin{cases} |\hat{X}_{i-1}(\omega)|^2 - \alpha(\omega)|\hat{N}_i(\omega)|^2, & \mathrm{if(P)}, \\ \beta(\omega)|\hat{X}_{i-1}(\omega)|^2, & \mathrm{if(O)}, \end{cases} \tag{5}$$
$$(\mathrm{P}) = (|\hat{X}_{i-1}(\omega)|^2 - \alpha(\omega)|\hat{N}_i(\omega)|^2 > \beta(\omega)|\hat{X}_{i-1}(\omega)|^2),$$
$$(\mathrm{O}) = (\mathrm{otherwise}),$$
$$i = 1, 2, 3, \ldots, n, \quad |\hat{X}_0(\omega)| = |Y(\omega)|, \quad 0 \ll \beta < 1,$$

本研究では聴覚特性に基づく重み付き減算係数を用いた反復 SS を，ラウドネス曲線重み付け SS (L-SS: Loudness contour-weighted SS) と定義する．

表 1　実験条件
Table 1　Experimental conditions.

Speeches	50 sentences of ATR phoneme balanced sentences[11]
Speakers	Five females and five males
Sampling	16 kHz, 16 bits
Frame length	64 ms. (1024 samples)
Shift length	32 ms. (512 samples)
FFT length	64 ms. (1024 samples)
Window function	Hanning window
Noise estimation	Average of seven frames
Coef. of I-SS	α : 2.0, β : 0.01
Coef. of F-SS	α : 2.0, β : 0.7
Coef. of L-SS (Proposed SS)	α_{bsc} : 2.0, 3.0, 4.0, β_{bsc} : 0.7, 0.8, 0.9, α_{wt} : 0.05, β_{wt} : 0.005
Kind of noise	Server noise, factory noise
SNR	0, 5, 10 dB

4. 評価実験

4.1 客観評価実験

4.1.1 雑音抑圧性能の評価指標

客観評価実験では，提案法が従来法と比較して高い雑音抑圧性能を達成可能か検証するために雑音抑圧量と音声ひずみ量を評価した．雑音抑圧量は NRR (Noise Reduction Rate)，音声ひずみ量は SDR (Signal-to-Distortion Ratio) を用いて，各 SS（I-SS, F-SS, L-SS）を評価した．

NRR は，雑音抑圧前後のエネルギー比を表し，NRR が高ければ雑音抑圧量が多いことを表す．本実験において，NRR は音声が混入していない雑音のみの信号を用いて算出する．なお NRR は次式から算出される．

$$NRR = 10\log_{10}\left(\sum_{n=0}^{N-1} x^2(n) \;/\; \sum_{n=0}^{N-1} y^2(n)\right), \qquad (6)$$

ここで $x(n)$ は雑音抑圧前の信号，$y(n)$ は雑音抑圧後の信号，n は時間，そして N は信号長を表す．

SDR は源信号と雑音抑圧後信号のエネルギー比を表し，SDR が高ければ音声ひずみ量が少ないことを表す．源信号を $x(n)$，評価信号を $y(n)$，n を時間，N を信号長とすると，SDR は次式を用いて算出される．

$$SDR = 10\log_{10}\left(\frac{\sum_{n=0}^{N-1} x^2(n)}{\sum_{n=0}^{N-1}(x(n)-\gamma y(n))^2}\right), \qquad (7)$$

$$\gamma = \sum_{n=0}^{N-1}|x(n)| \;/\; \sum_{n=0}^{N-1}|y(n)|, \qquad (8)$$

4.1.2 実験条件

ここで表 **1** に実験条件を示す．本実験では，ATR 音素バランス文コーパス [11] よりランダムに選んだ 50 文に各雑音を SNR=0, 5, 10 dB で加算した信号を用いた．また，反復回数 50 回までの NRR と SDR を算出し，両指標の関係を評価する．評価雑音としては，電子協騒音データベース [12] に含まれる 2 種類の雑音（Server noise, Factory noise）に対する雑音抑圧性能を評価した．SS に用いる推定雑音は，観測信号の先頭部分を無音声区間と仮定して先頭 7 フレームの平均を利用した．また，L-SS は各係数の基準値と重みを設定する必要がある．α_{bsc} は一般的に採用されている 2.0 以上の値を用いて雑音を抑圧し，β_{bsc} はミュージカルノイズを低減するために 0.7 以上の値を採用した．なお，等ラウドネス曲線は音の大きさにより，複数存在するものの各曲線の形状に大きな差異は無いため，図 3 に示す 60 phon の曲線を用いて実験を実施した．

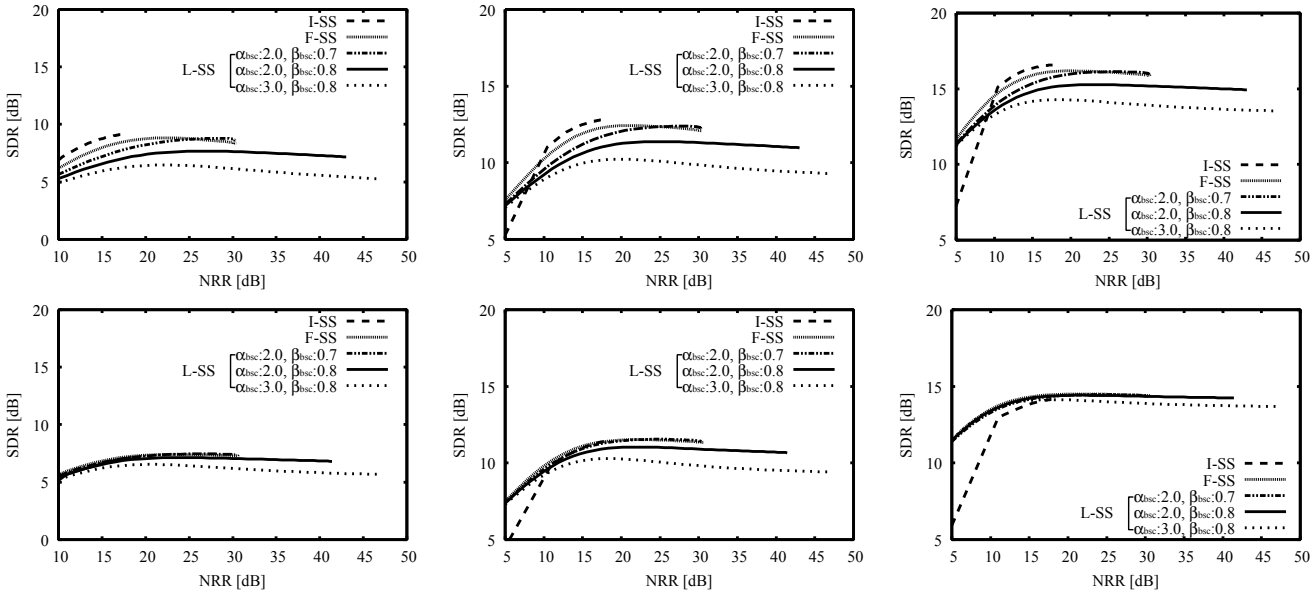
4.1.3 実験結果

図 **4** に各 SS による NRR と SDR の実験結果を示す．図 4 より，L-SS の $\alpha_{\mathrm{bsc}}, \beta_{\mathrm{bsc}}$ の値が大きいほど NRR が高くなる一方，$\alpha_{\mathrm{bsc}}, \beta_{\mathrm{bsc}}$ の値が小さいほど SDR が高くなる傾向であった．ただし，Factory noise に対してはパラメータ毎の SDR に大きな差はみられなかった．また全雑音に対して反復処理を重ねることで NRR=20 dB 程度までは，NRR・SDR 共に増加するが，更に高い NRR を達成するためには SDR が徐々に減少することを確認できた．そして信号対雑音比 (SNR) に着目すると，SNR の増加に伴い SDR も増加し，高 SNR であるほど少ない雑音抑圧量で最高の SDR となったことを確認した．

4.2 主観評価実験

4.2.1 実験条件

主観評価実験では，提案法が従来法と同程度の音声明瞭度を保ちつつミュージカルノイズを低減可能か MOS (Mean Opinion Score) を用いて検証した．まず，雑音抑圧後の音声明瞭度を雑音抑圧前と比較して 5 段階（1. 非常に聴き取り難い，2. 聴き取り難い，3. 変わらない，4. 聴き取り易い，5. 非常に聴き取り易い）で評価した．また，ミュージカルノイズが低減されていることを確認するために，雑音抑圧後のミュージカルノイズ残存量を 5 段階（1. 非常に気になる，2. だいぶ気になる，3. それほど気にならない，4. あまり気にならない，5. 全く気にならない）で評価した．評価音源としては，客観評価実験と同様の信号を利用した．また，各手法における反復回数は SDR が十分に収束していると考えられる I-SS（処理回数: 10 回）と同程度の NRR を達成可能な回数とした．表 **2** に各手法の反復回数を示す．両実験共に 7 名（女性: 2 名，男性: 5 名）の被験者に対して防音室（暗騒音レベル: 19 dBA）にて行い，66 パターン（施行回数: 2 回）で合計 132 音源をランダムに提示し，再生デバイスとしてはヘッドホン（SONY, MDR-CD900ST）を利用した．なお，その他の実験条件については客観評価



(a) SNR=0 dB　(b) SNR=5 dB　(c) SNR=10 dB

図 4　NRR と SDR の結果（上段：Server noise，下段：Factory noise）
Fig. 4　Results of NRR and SDR (Upper: Server noise, Lower: Factory noise).

表 2　主観評価実験における各手法の反復回数
Table 2　Number of times for iteration in each SS method.

			Server noise	Factory noise
I-SS			10 times	10 times
F-SS			9 times	10 times
L-SS	α_{bsc} : 2.0	β_{bsc} : 0.7	10 times	10 times
		β_{bsc} : 0.8	15 times	16 times
		β_{bsc} : 0.9	30 times	32 times
	α_{bsc} : 3.0	β_{bsc} : 0.7	10 times	10 times
		β_{bsc} : 0.8	15 times	16 times
		β_{bsc} : 0.9	30 times	32 times
	α_{bsc} : 4.0	β_{bsc} : 0.7	10 times	10 times
		β_{bsc} : 0.8	15 times	16 times
		β_{bsc} : 0.9	30 times	32 times

実験と同様のものとした．

4.2.2　実験結果

図 **5** に音声明瞭度の結果を示し，図 **6** にミュージカルノイズ残存量の結果を示す．図 5 の音声明瞭度の結果より，Server noise に対しては，全てのパラメータで従来法と同程度の MOS であるため，L-SS により音声明瞭度が劣化しないことが確認できた．特に，$\alpha_{\mathrm{bsc}} = 2.0, \beta_{\mathrm{bsc}} = 0.7$ のときに最も聴き取り易いという結果であった．また Factory noise に対しては，$\alpha_{\mathrm{bsc}}, \beta_{\mathrm{bsc}}$ の値が小さいほど高い MOS であり，$\alpha_{\mathrm{bsc}} = 3.0, \beta_{\mathrm{bsc}} = 0.8$ 以下であれば F-SS と同程度の音声明瞭度であった．

図 6 のミュージカルノイズ残存量の結果より，Server noise に対しては，β_{bsc} の値が大きいほどミュージカルノイズを低減可能であった．また Factory noise に対しては，全てのパラメータで従来法よりもミュージカルノイズを低減可能であり，$\alpha_{\mathrm{bsc}} = 4.0, \beta_{\mathrm{bsc}} = 0.9$ の場合に最も高い MOS であった．図 6 より，全ての雑音に対してパラメータ毎に大きな差異はなく，$\beta_{\mathrm{bsc}} = 0.8$ 以上であれば，従来法と比較してミュージカルノイズを低減可能であった．

4.3　反復回数最適化

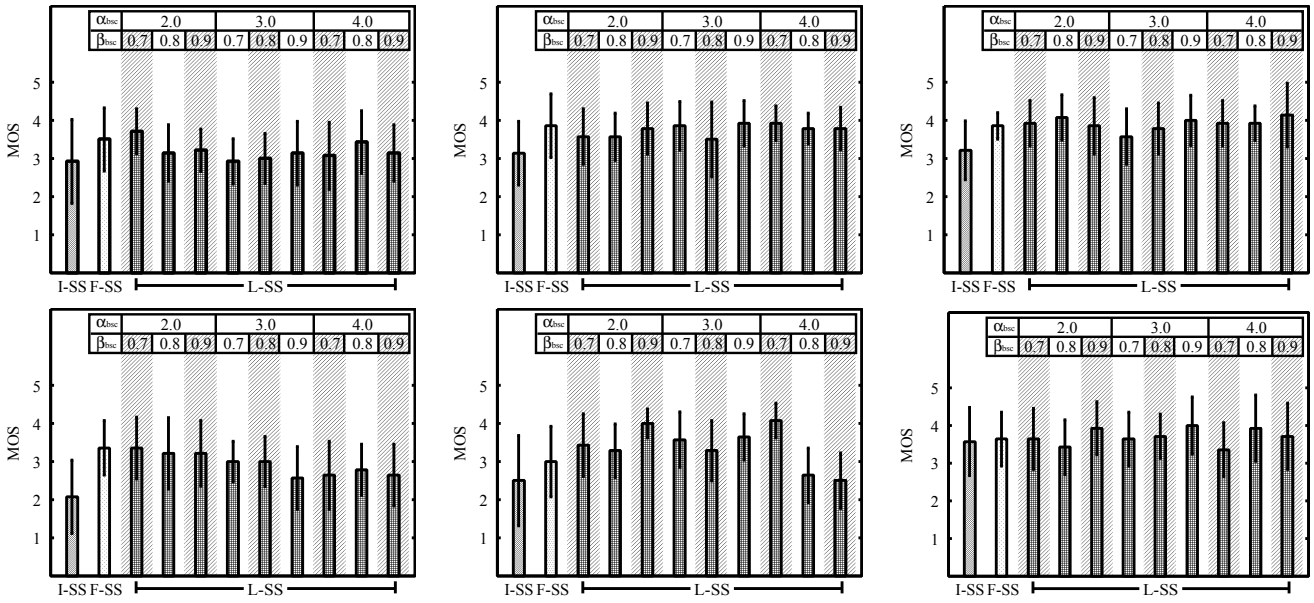
4.3.1　反復回数最適化の方針

客観・主観評価実験より提案法を用いることで，ミュージカルノイズを低減しつつ高い雑音抑圧性能を達成できることを確認した．しかし，提案法は反復処理を行うことを前提としており，反復回数の増加に伴い雑音抑圧性能も向上する反面，ミュージカルノイズが発生し音質も劣化する傾向がある．そこで，提案法の実用化に向けて最適な反復回数について検討する．

本研究では主観・客観評価実験に基づいてミュージカルノイズの発生を抑えつつ高音質に雑音を抑圧可能な反復回数を提案法における最適値として定義する．主観評価実験ではミュージカルノイズ残存量を MOS (Mean Opinion Score) により評価し，ミュージカルノイズを発生させずに雑音を抑圧できる反復回数を調査する．また，客観評価実験では雑音抑圧後の音質を PESQ (Perceptual Evaluation of Speech Quality)[13] により評価する．

本研究における最適な反復回数は，以下のアルゴリズムで算出する．

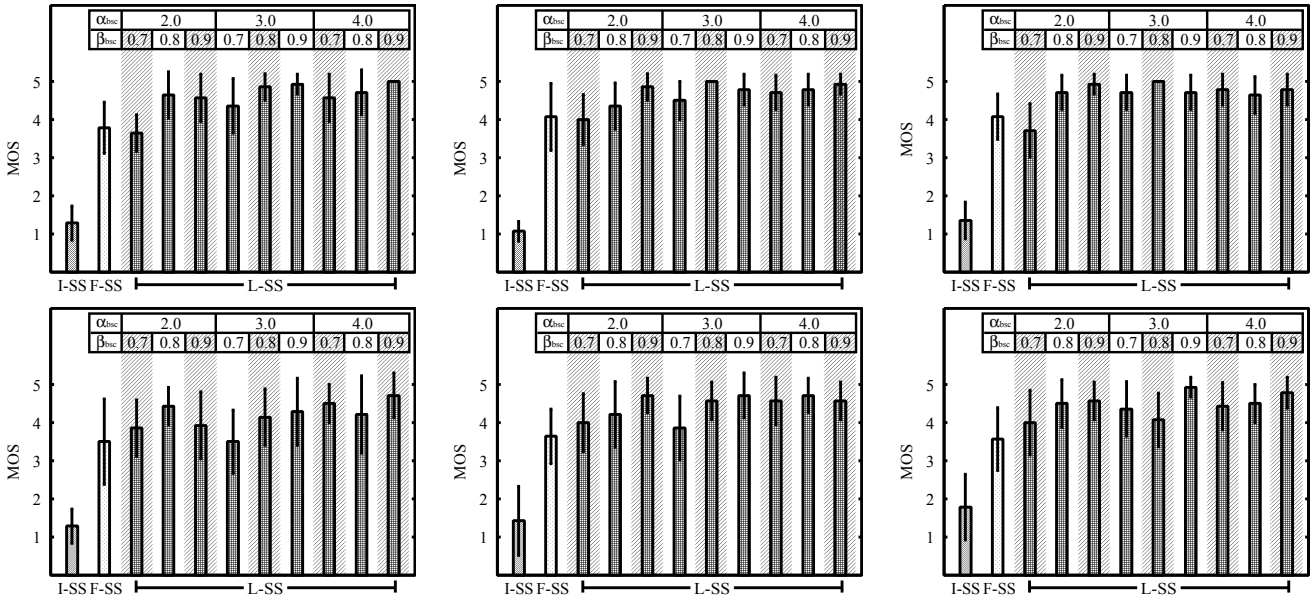
- **Step.1** 主観評価実験
  反復回数毎のミュージカルノイズ残存量を評価し，MOS が 4.0（ほとんど気にならない）以上の反復回数を算出する．



(a) SNR=0 dB (b) SNR=5 dB (c) SNR=10 dB

図 **5** 音声明瞭度の結果（上段：Server noise，下段：Factory noise）

Fig. 5 Results for the speech articulation (Upper: Server noise, Lower: Factory noise).



(a) SNR=0 dB (b) SNR=5 dB (c) SNR=10 dB

図 **6** ミュージカルノイズ残存量の結果（上段：Server noise，下段：Factory noise）

Fig. 6 Results for the amount of the musical tone (Upper: Server noise, Lower: Factory noise).

- **Step.2** 客観評価実験
  反復回数毎の PESQ を評価し，Step.1 で算出した反復回数以下で最も PESQ が高い回数を最適値とする．

4.3.2 主観評価実験

主観評価実験は，7 名の被験者（女性：2 名，男性：5 名）を対象に防音室（暗騒音レベル: 21 dBA）にて行い，ランダムに提示された音源のミュージカルノイズ残存量を 5 段階（1. 非常に気になる，2. だいぶ気になる，3. それほど気にならない，4. あまり気にならない，5. 全く気にならない）で評価した．反復回数は 5 種類の条件（10 回から 30 回まで 5 回間隔ずつ）で 2 種類の雑音（Server noise, Factory noise）を抑圧した．提案法における処理係数の基準値は，ミュージカルノイズを低減しつつ高い雑音抑圧性能を達成可能である $\alpha_{\mathrm{bsc}} = 3.0$, $\beta_{\mathrm{bsc}} = 0.8$ を採用した．

主観評価実験結果を図 **7** に示す．また，MOS が 4.0 以上である反復回数を表 **3** に示す．図 7 より，反復処理を重ねることで MOS が低下していくことから，ミュージカルノイズが徐々に発生していることを確認した．表 3 より，

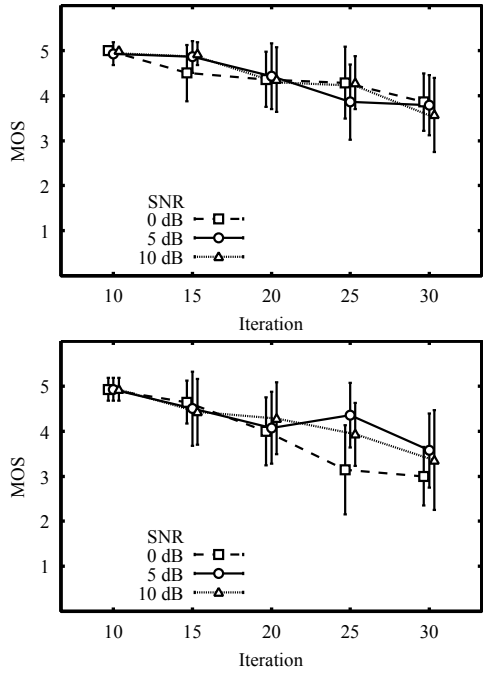


図 7　ミュージカルノイズ残存量の結果（上段：Server noise，下段：Factory noise）

Fig. 7　Results for the amount of the musical tone (Upper: Server noise, Lower: Factory noise).

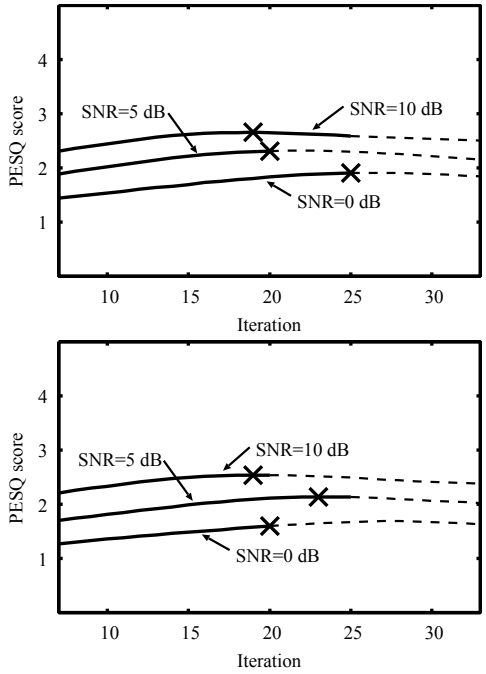


図 8　PESQ の結果（上段：Server noise，下段：Factory noise）

Fig. 8　Results for PESQ (Upper: Server noise, Lower: Factory noise).

表 3　MOS ≥ 4.0 である反復回数

Table 3　Number of times for iteration that MOS are higher than 4.0.

Noise \ SNR	0 dB	5 dB	10 dB
Server noise	25 times	20 times	25 times
Factory noise	25 times	25 times	20 times

SNR や雑音の種類による大きな差異は見られず，20 ～ 25 回程度まではミュージカルノイズの発生を低減可能できることがわかった．

4.3.3　客観評価実験

客観評価実験では，表 3 に示す主観評価実験により算出された反復回数までの PESQ を評価する．PESQ[13] は，ITU-T 勧告で定められている客観的な音声品質指標であり，源音声と雑音抑圧後の音声を用いて算出可能である．PESQ は 0.5 ～ 4.5 の範囲で算出され，値が高いほど品質が高いことを示す．また，PESQ は聴覚心理尺度を考慮しているため，主観的な評価と高い相関があることが確認されている．

客観評価実験の結果を図 **8** に示す．図中の実線は，MOS が 4.0 以上の反復回数における評価結果，点線はその後雑音抑圧処理を続けた場合の PESQ の推移を示す．また，× 印は各条件下で最も PESQ が高い反復回数であり，表 **4** に具体的な回数を示す．実験結果より，PESQ は反復を重ねることで，ある程度まで値が増加した後，徐々に減少する傾向であった．特に，SNR が高いほど少ない反復回数で最高値を達成した．なお，表 4 より雑音環境に関わらず最適な反復回数は 20 回程度であった．

表 4　雑音環境毎の最適な反復回数

Table 4　Optimum number of times for iteration.

Noise \ SNR	0 dB	5 dB	10 dB
Server noise	25 times	20 times	19 times
Factory noise	20 times	23 times	19 times

4.3.4　最適な反復回数の定式化

主観・客観評価実験結果より，$\alpha_{\mathrm{bsc}} = 3.0$, $\beta_{\mathrm{bsc}} = 0.8$ の条件において，雑音環境に依存せず最適な反復回数は 20 回程度であることを確認した．しかし，処理係数の基準値（α_{bsc}, β_{bsc}）に依存して提案法の性能は大きく変化するため，最適な反復回数も処理係数の基準値に依存すると考えられる．そこで，処理係数の基準値から最適な反復回数を一意に決定することができれば，様々な条件で提案法を利用できると考えた．

まず，処理係数の基準値と最適な反復回数の関係を明確にするために，パラメータを変更して最適な反復回数を算出した．これまでの実験より，ミュージカルノイズを低減するには $\beta_{\mathrm{bsc}} = 0.7$ 以上の値を採用する必要があるため，本実験では，9 種類のパラメータ（$\alpha_{\mathrm{bsc}} = 2.0, 3.0, 4.0$, $\beta_{\mathrm{bsc}} = 0.7, 0.8, 0.9$）を用いて実験を行った．また，前節の実験より最適な反復回数は，雑音環境に依存しないことを確認したため 1 条件（加算雑音: Server noise, SNR=5 dB）で実験を行った．

主観・客観評価実験により算出した各パラメータ毎の最適な反復回数を図 **9** に示す．評価結果より，最適な反復回数

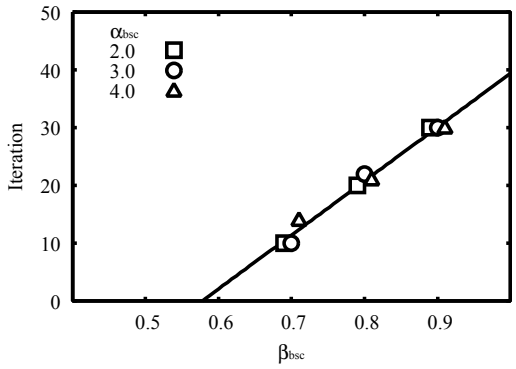


図 9 処理係数の基準値（$\alpha_{\rm bsc}, \beta_{\rm bsc}$）と最適な反復回数の関係（実線：回帰直線）

Fig. 9 Relationship among $\alpha_{\rm bsc}, \beta_{\rm bsc}$ and oputimum number of times for iteration (A solid line indicates a regression line).

は $\beta_{\rm bsc}$ の値に大きく依存し，$\beta_{\rm bsc}$ の値が大きいほど最適な反復回数も多くなることを確認した．具体的に，$\beta_{\rm bsc} = 0.7$ では 10 回程度，$\beta_{\rm bsc} = 0.8$ では 20 回程度，$\beta_{\rm bsc} = 0.9$ では 30 回程度が最適な反復回数であった．また，図 9 の 1 次直線は $\beta_{\rm bsc}$ と最適な反復回数に基づいて算出した近似直線である．y を最適な反復回数，x を $\beta_{\rm bsc}$ の値とすると近似直線の定義式は式 (9) のように表される．

$$y = ax + b \ (x \geq 0.7), \tag{9}$$

ここで，近似直線の $a = 93.3$, $b = -53.9$ の場合に相関係数が 0.98 であり，この近似直線に基づいて $\beta_{\rm bsc}$ の値から最適な反復回数を一意に算出することができた．

5. さいごに

SS は，低演算コストで雑音を抑圧可能であることから，一般的に広く利用されているがミュージカルノイズと呼ばれる聴感上不快な雑音の発生が問題視されていた．そこで本研究では，聴覚特性に基づく重み付き係数を用いた反復 SS を提案した．提案法の有効性を確認するための客観・主観評価実験を実施した結果，提案法は従来法と比較して高い雑音抑圧性能を達成しつつ，主観的にミュージカルノイズを低減できた．

今後の課題としては，様々な雑音環境下で提案法の実用化を目指して，雑音環境毎に最適なパラメータを算出するための指標を策定する必要がある．また，非定常な雑音に対して高精度な抑圧を実現することで，提案法の利便性が大きく向上すると考えられる．

謝辞　本研究の一部は，科研費の研究助成を受けた．

参考文献

[1] J.L. Flanagan, J.D. Johnston, R.Zahn and G.W. Elko, “Computer-steered microphone arrays for sound transduction in large rooms, ” Journal of the Acoustical Society of America, vol. 78, no. 5, pp. 1508-1518, 1985.

[2] Y. Takahashi, T. Takatani, H. Saruwatari and K. Shikano, “Blind spatial subtraction array with independent component analysis for hands-free speech recognition, ” Proc. International Workshop for Acoustic Echo and Noise Control 2006, CD-ROM, 2006.

[3] J. S. Lim and A. V. Oppenheim, “All-pole modeling of degraded speech, ” IEEE Transactions on Acoustic, Speech and Signal Processing, vol. ASSP-26, no. 3, pp. 197-210, 1978.

[4] S.F. Boll, “Suppression of acoustic noise in speech using spectral subtraction, ” IEEE Transactions on Acoustic, Speech and Signal Processing, vol. ASSP-27, no. 2, pp. 113-120, 1979.

[5] S.V. Vaseghi, “Advanced digital signal processing and noise reduction, ” John Wiley & Sons Ltd, 1995.

[6] H. Nakashima, Y. Chisaki, T. Usagawa and M. Ebata, “Spectral subtraction based on statistical criteria of the spectral distribution, ” IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. E85-A, no. 10, pp. 2283-2292, 2002.

[7] Z. Goh, K.C. Tan and B.T.G. Tan, “Postprocessing method for suppressing musical noise generated by spectral subtraction, ” IEEE Transactions on Speech Audio Processing, vol. 6, no. 3, pp. 287-292, 1998.

[8] K. Yamashita, S. Ogata and T. Shimamura, “Spectral subtraction iterated with weighting factors, ” Proc. IEEE Workshop on Speech Coding, pp. 138-140, 2002.

[9] T. Fukumori, M. Morise, T. Nishiura, Y. Yamashita and H. Nanjo, “The estimation of optimum subtraction parameters for iterative spectral subtraction towards musical tone reduction, ” Proc. Internoise2011, PaperID:Mon-P-21, 2011.

[10] ISO 226:2003, “Acoustics-normal equal loudness level contours, ” 2003.

[11] Y. Sagisaka, K. Takeda, M. Abe, S. Katagiri, T. Umeda and H. Kuwabara, “A large-scale Japanese speech database, ” Proc. Int. Conf. Spoken Language Processing 1990, pp. 1089-1092, 1990.

[12] 社団法人日本電子工業振興協会 電子協 騒音データベース, http://www.sunrisemusic.co.jp/database/fl/noisedata01_fl.html

[13] International Telecommunication Union, “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, ” p. 862, 2001.