

## 反復処理を利用した改良スペクトル引き算

山下 浩平<sup>†a)</sup> 緒方 伸哉<sup>†</sup> 島村 徹也<sup>†</sup>

Improved Spectral Subtraction Utilizing Iterative Processing

Kohei YAMASHITA<sup>†a)</sup>, Shin'ya OGATA<sup>†</sup>, and Tetsuya SHIMAMURA<sup>†</sup>

あらまし 本論文では、雑音付加音声の雑音低減の手法であるスペクトル引き算に、反復処理とそれに適したパラメータ設定を施した、新しい雑音抑制技術を提案する。反復処理とは、一度雑音低減処理を施した推定音声再度入力信号とみなし、音声強調処理を施す手段であり、残留雑音の低減が見込まれる。反復ごとにパラメータを調整することで、音声の劣化を抑えた更なる残留雑音低減が可能となる。また、提案法を実行する際に、スペクトル引き算のもつリアルタイム性を保持する手法も同時に提案する。2種類の提案法の特性を、白色雑音、自動車雑音、人混み雑音を付加した実音声を用い、従来のスペクトル引き算及びその改良法と比較する。主観評価及び客観評価により、各提案法はすべての雑音環境に対して優れた結果を示すことが確認された。

キーワード 音声強調、雑音抑制、スペクトル引き算、ミュージカルノイズ、反復処理

## 1. ま え が き

近年、携帯電話の普及や音声認識技術の発達に伴い、デジタル音声信号の雑音低減技術が必要とされている。音声を送受信する際には様々な要因により雑音が付加される。しかし、本来所望とされるのは音声信号のみであるため、雑音の付加により、通話時の不快感や認識率の低下をもたらすこととなる。したがって、雑音付加音声に対する雑音除去は、音声通信における重要な研究テーマとなっている。

音声通信システムにおける雑音除去（音声の品質改善）の技術の一つにスペクトル引き算がある [1], [2]。スペクトル引き算はその計算アルゴリズムが簡潔であるため、現在幅広く用いられている方法である [3]。雑音付加音声は、多くの場合、もとの音声と雑音が足し合わされた信号と考えることができる。そこで、雑音付加音声を音声と雑音に分離するために、雑音のみをスペクトル領域で引き去る方法がスペクトル引き算法である。この方法では、まず雑音スペクトルの推定を行うことが必要とされるが、その際に完全な推定結

果を得ることは一般に困難である。なぜなら、雑音は統計量としてのみ扱うことが可能であり、与えられる雑音付加音声信号に対して、十分な統計処理を行うことは本来不可能であるためである。よって、多くの場合、推定誤差に対応する雑音が残留することになる。このような残留雑音には、ミュージカルノイズと呼ばれる人工的な雑音を含む場合が多い。

スペクトル引き算は、この残留雑音を低減するために、今日までに様々な改良が試みられている。Bohらは様々な改良パラメータを統括して、スペクトル引き算の一般形を導き出した [4]。しかし、パラメータ設定を行うことは、トレードオフの関係にある残留雑音と音声ひずみに関して、どちらに重みを置いて考えるかに基づいており、音声の品質を保持しつつ残留雑音を低減させる手法とはいえない。

また、雑音を低減するために、音声に前処理や後処理を加える方法も提案されているが、いまだ決定的な方法は導き出されておらず、様々な環境、状況に応じて有効な方法を選択、模索している段階といえる [5] ~ [11]。

また、スペクトル引き算法では入力信号対雑音比 (SNR) が低くなる場合、雑音低減の精度が著しく低下することが知られている。3GPP (the Third Generation Partnership Project) [12] では、自動車雑音では 6dB 以下、白色雑音では 9dB 以下を高雑音環境

<sup>†</sup> 埼玉大学大学院理工学研究科, さいたま市  
Graduate School of Science and Engineering, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama-shi, 338-8570 Japan

a) E-mail: k-yama@sie.ics.saitama-u.ac.jp

と指定している．これは，SNR が 0 dB 程度の超高雑音環境に有用なスペクトル引き算が提案されていないことを意味しよう．しかし，通信技術が多様化する今日では，電車のプラットフォームや工事現場，高速走行中の車内等，高雑音環境下における雑音抑制技術も望まれている．

そこで本論文では，高雑音環境下において従来のスペクトル引き算法が与える残留雑音の低減と，ミュージカルノイズの抑制を目的とし，反復処理を用いたスペクトル引き算法 [13] を使い，反復に併せて効果的に性能を發揮するパラメータ設定法 [14] を提案する．反復スペクトル引き算法では，基本的には残留雑音を含む出力音声信号を再度入力信号として用いる単純な操作を行い，これを反復的に繰り返す．しかし，提案法では反復に応じてパラメータを変化させることにより，音声を大きくひずませることなく，ミュージカルノイズを大幅に低減する．また，反復処理に適合した副処理法を提案し，特に高雑音環境下に有用なリアルタイムスペクトル引き算を目指す．本論文での各提案法は，簡単な処理により残留雑音成分をより多く除去し，更にミュージカルノイズを低減させることを目的とする．

## 2. スペクトル引き算法

本章では，スペクトル引き算法の原理と，その具体的な処理方法を述べる．

### 2.1 スペクトル引き算法

雑音付加音声信号を  $y(n)$ ，もとの音声信号を  $x(n)$ ，付加雑音を  $w(n)$  で表すとすると，このとき

$$y(n) = x(n) + w(n) \quad (1)$$

となる信号が与えられると仮定する．このような雑音の付加は，周波数成分でも同様に考えられるので，周波数成分において，雑音成分のみを引き去れば，音声信号が残ることになる．

スペクトル引き算法は，基本的にはフレーム処理で実行される．具体的には，

$$|\hat{X}_w(f)| = |Y_w(f)| - |\hat{W}_w(f)| \quad (2)$$

$$\hat{X}_w(f) = |\hat{X}_w(f)| \frac{Y_w(f)}{|Y_w(f)|} \quad (3)$$

である．ここで， $Y_w(f)$ ， $X_w(f)$ ， $W_w(f)$  は，それぞれ窓かけされた信号  $y_w(n)$ ， $x_w(n)$ ， $w_w(n)$  のフーリエ変換を表している． $\hat{X}_w(n)$ ， $\hat{W}_w(f)$  は  $X_w(f)$ ， $W_w(f)$  の推定値である． $w$  はフレーム番号に対応す

る．式 (2) にて算出された  $|\hat{X}_w(n)|$  に，雑音付加音声信号の位相を用いて  $\hat{X}_w(n)$  を算出し，これを逆フーリエ変換し，時間波形に変換した後，オーバラップ処理することにより音声信号の推定値  $\hat{x}(n)$  が得られる．これは，人間の聴覚が位相情報に対し鈍感であることを利用している．よって，通常，スペクトル引き算では位相情報の考慮はしない．以下に説明するすべてのスペクトル引き算の改良法では，同様に雑音付加音声信号の位相情報を用いることとし， $|\hat{X}_w(f)|$  の推定法のみ言及する．また，式 (2) での  $|\hat{W}_w(f)|$  は，通常いくつかの無音区間から推定され，用いられる．

### 2.2 整流処理

式 (2) のスペクトルの引き算を実際に行う場合には，推定誤差により，ある周波数においては雑音成分が音声成分を上回ることがあり得る．具体的には， $|Y_w(f)| - |\hat{W}_w(f)| < 0$  となる．本来，付加された雑音成分がもとの音声成分を上回るとは不合理であるため，通常は整流処理によりスペクトル引き算処理の補正を行う．このような整流処理の方法には次の 2 通りがある [15]．

#### (i) 半波整流処理

$$|\hat{X}_w(f)| = 0 \quad (4)$$

#### (ii) 全波整流処理

$$|\hat{X}_w(f)| = |\hat{W}_w(f)| - |Y_w(f)| \quad (5)$$

半波整流処理は，一度に多くの雑音量を除去することができるため，雑音除去の精度は高くなる．一方，全波整流処理は，雑音除去精度は半波整流処理に劣るものの，入力音声信号の SNR において，半波整流処理より広範囲の SNR で改善が得られることで知られている [15]．ミュージカルノイズに関しては，半波整流処理においてはその発生が多く，全波整流処理は半波整流処理に比べて少なくなる．

スペクトル引き算法では，その雑音除去能力の優位性より，一般に半波整流処理を用いていることが多い．しかし，反復スペクトル引き算法において，これは必ずしも成り立たない．反復回数における整流処理の違いについては，3. において述べる．

### 2.3 非線形スペクトル引き算法

スペクトル引き算の改良法の一例として，非線形スペクトル引き算法を紹介する．スペクトル引き算は，前述したとおり今日までに様々な改良が施されている．その方法の一つが非線形スペクトル引き算法であ

る [9]．非線形スペクトル引き算法は、しばしばスペクトル引き算の発展的方法の代表例として比較対象に挙げられる．その基本処理は典型的なスペクトル引き算法を用いる [1], [5], [16]．具体的には、

$$|\hat{X}_w(f)| = |\dot{Y}_w(f)| - |\dot{W}_w(f)| \quad (6)$$

である．ここで、 $|\dot{Y}_w(f)|$ 、 $|\dot{W}_w(f)|$  は以下の式で表されるスムージングされた振幅スペクトルである．

$$|\dot{Y}_w(f)| = \lambda_Y |\ddot{Y}_{w-1}(f)| - (1 - \lambda_Y) |Y_w(f)| \quad (7)$$

$$|\dot{W}_w(f)| = \lambda_W |\ddot{W}_{w-1}(f)| - (1 - \lambda_W) |W_w(f)| \quad (8)$$

$\lambda_Y$ 、 $\lambda_W$  はスムージング係数であり、通常、 $0.1 \leq \lambda_Y \leq 0.5$ 、 $0.5 \leq \lambda_W \leq 0.9$  の範囲で用いられる．

非線形スペクトル引き算法では、雑音成分の過剰推定を行う．具体的には、過去の 40 フレームを保持し、それぞれのフレームで保持された周波数成分の最大値をとり、次のように係数  $\alpha_w(f)$  を得る．

$$\alpha_w(f) = \max_{w-40 \leq \tau \leq w} (|W_\tau(f)|) \quad (9)$$

この  $\alpha_w(f)$  に重み係数を掛け合わせ、推定雑音スペクトルとする．重み係数は

$$\rho_w(f) = \frac{|\dot{Y}_w(f)|}{|\dot{W}_w(f)|} \quad (10)$$

を利用し、 $\frac{1}{1+\rho_w(f)}$  とする．このとき雑音推定スペクトルは

$$\Phi_w(f) = \frac{\alpha_w(f)}{1 + \rho_w(f)} \quad (11)$$

と表せる．ただし  $\Phi_w(f)$  には以下のようなしきい値を用いる．

$$|\dot{W}_w(f)| \leq \Phi_w(f) \leq 3|\dot{W}_w(f)| \quad (12)$$

また、整流処理として、基本的には半波整流を用い、フロア係数  $\beta$  を導入し、最終的には

$$\begin{aligned} |\hat{X}_w(f)| &= \begin{cases} |\dot{Y}_w(f)| - \Phi_w(f), & |\dot{X}_w(f)| > \beta |\dot{Y}_w(f)| \\ \beta |\dot{Y}_w(f)|, & \text{else} \end{cases} \end{aligned} \quad (13)$$

とする．この処理はそれぞれの周波数において前後フレームの周波数成分に依存しない値をとる．これが非

線形と呼ばれるゆえんである．非線形スペクトル引き算法は、主に自動車内の音声認識に用いられる手法で、特に自動車内のエンジン雑音に対して効果を発揮する報告がなされている [9]．その他、様々な手法でスペクトル引き算の特性改善を図ろうとする試みがあるが、どの方法も雑音のスペクトルを変化させる基本原理に帰着される．その延長上には、おのずと特性改善の限界があり、またパラメータ設定の過程に置いて、膨大な計算量の増加を避けられない場合も多い．

### 3. 反復スペクトル引き算法

本章では、反復スペクトル引き算法の概要について述べる．

#### 3.1 反復スペクトル引き算法

本章では、反復スペクトル引き算法 [13] について述べる．原理は非常に単純で、スペクトル引き算法を用いて算出された音声信号  $\hat{x}(n)$  を再度入力信号として扱うことである．反復回数を  $k$  とすると、 $k$  回目の音声信号は  $y(k, n) = x(k, n) + w(k, n)$  で与えられると仮定する． $k$  回目のスペクトル引き算は

$$|\hat{X}_w(k, f)| = |Y_w(k, f)| - |\hat{W}_w(k, f)| \quad (14)$$

$$\hat{X}_w(k, f) = |\hat{X}_w(k, f)| \frac{Y_w(k, f)}{|Y_w(k, f)|} \quad (15)$$

で求められる．ここで、 $|\hat{X}_w(k, f)|$ 、 $|Y_w(k, f)|$ 、 $|\hat{W}_w(k, f)|$  は、それぞれ反復回数  $k$  回目の音声信号  $\hat{x}_w(k, n)$ 、雑音付加入力信号  $y_w(k, n)$ 、付加雑音  $\hat{w}_w(k, n)$  のフーリエ変換を表す． $k+1$  回目の反復処理では、 $k$  回目の反復処理で得られる出力信号  $\hat{x}(k, n)$  が、

$$y(k+1, n) = \hat{x}(k, n) \quad (16)$$

として、再度入力信号とされる．このとき、 $y(k+1, n)$  の雑音成分は、 $k$  回目のスペクトル引き算で抑制することのできなかつた残留雑音成分となる．反復スペクトル引き算法では、 $y(k+1, n)$  から各フレームで用いる推定雑音スペクトル  $|\hat{W}_w(k+1, n)|$  を求めるので、残留雑音を更に抑制する方法となる．

#### 3.2 反復処理における整流処理の挙動

反復処理における整流処理の違いについて検討する．先述したように、半波整流処理は、一度に多くの雑音を除去することができる．ゆえに、全波整流処理は半波整流処理に比べ、この点では劣っている．しかし、反復処理を加えることにより、全波整流処理の欠点で

ある、雑音除去の少なさを補うことができる。図 1 は、白色雑音、自動車雑音、人混み雑音を付加させた実音声信号 (SNR=0dB) に、スペクトル引き算を単純に反復させた際のセグメンタル SNR 改善度の平均値を示している。セグメンタル SNR は、一般に全区間 SNR に比べ、客観評価により近い値を示すことが知られている [15]。改善度とは、入力音声の SNR と出力音声の SNR の差をとったものである。実験に使用した音声は、NTT アドバンステクノロジー (株)「20ヶ国語音声データベース」からの、日本人男女各 2 名ずつの約 11 秒の連続音声で、サンプリング周波数は 10 kHz であり、3.4 kHz の帯域制限がかけられている。以下、実音声を用いた実験には、すべてこれらの音声を用いることにする。図 1 の反復 1 回は、従来のスペ

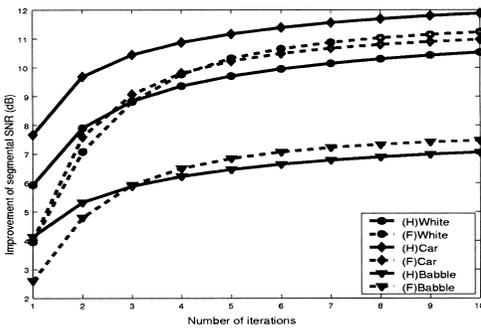


図 1 セグメンタル SNR 改善度 (H: 半波整流, F: 全波整流)

Fig. 1 Improved segmental SNR (H: half-wave rectification, F: full-wave rectification).

クトル引き算を表す。図 1 から、半波整流処理は 1, 2 回目までの反復において、おおよそ全波整流処理に勝っているが、反復回数を 5 回程度まで増やすと、全波整流処理の雑音除去量も半波整流処理と同等程度またはそれ以上に向上しているのが見てとれる。よって、反復処理を行う場合には、一般的に認識されているような半波整流処理の雑音除去能力に関する優位性は、確立されないものと考えられる。また、半波整流処理にて繰返し処理を行うと、全波整流処理に比べ、各周波数に部分的に現れる残留雑音のスペクトルの推定誤差を強調する結果となり、ミュージカルノイズを強調することとなる。図 2 に、男声の白色雑音付加音声信号 (SNR=0dB) に、半波整流処理と全波整流処理を用いてスペクトル引き算を反復 5 回で処理を施した出力信号のスペクトログラムを示す。この図より、半波整流に多く斑点模様が見られているのが分かる。ミュージカルノイズは、スペクトルのピークがランダムに現れるため、斑点のように観測されることが知られている [7]。これらの結果より、反復スペクトル引き算には全波整流処理が、半波整流処理よりも適していると判断し、以下、各提案法では全波整流処理を用いることとする。

#### 4. 提案法

本章では、反復スペクトル引き算の改良法を提案する。

##### 4.1 重み係数付き反復スペクトル引き算

まず、重み係数を用いた反復スペクトル引き算

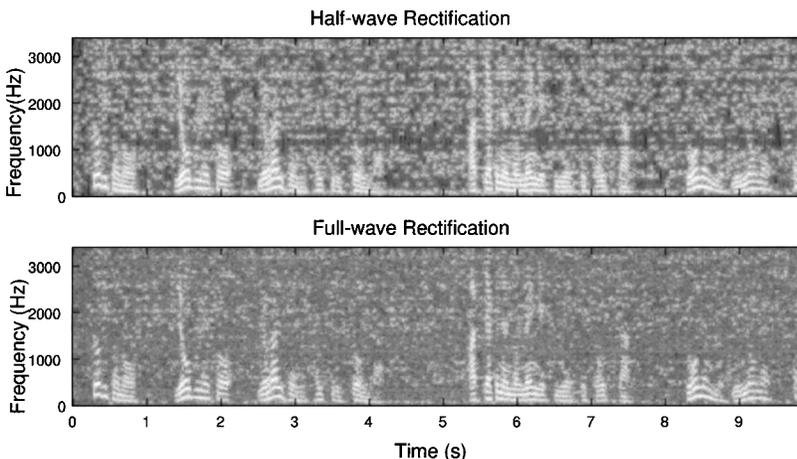


図 2 半波整流と全波整流によるスペクトログラムの比較

Fig. 2 Spectrogram comparison by half-wave and full-wave rectifications.

(以後、提案法 1 と記述する) を提案する。提案法 1 のブロック図が図 3 に示される。本法の具体的な処理手順は以下のように記述される。

- 1: 反復回数  $k$  を  $k = 1$  と初期化する。
- 2:  $k = 1$  の場合は、 $y(n)$  の無音区間から雑音スペクトル  $|\hat{W}_w(k, f)|$  を推定する。一方  $k \neq 1$  の場合は、反復処理により得られた  $\hat{x}(k-1, n)$  の無音区間から  $|\hat{W}_w(k, f)|$  を推定する。
- 3: 推定雑音スペクトル  $|\hat{W}_w(k, f)|$  に重み係数  $\lambda_k$  を掛け合わせ、その減算量を調整する。
- 4: 3 で得られた重み係数付き推定雑音スペクトル  $\lambda_k |\hat{W}_w(k, f)|$  を用い、 $k = 1$  の場合は  $y(n)$  に対し、 $k \neq 1$  の場合は  $\hat{x}(k-1, n)$  に対し、スペクトル引き算処理を各フレームごとに実行する。そして、反復回数  $k$  回目における出力信号  $\hat{x}(k, n)$  を求める。
- 5:  $k = k + 1$  とし、2~4 の処理を繰り返す。

$K$  を最終的に繰り返される反復回数とすると、本法では結果として  $\hat{x}(K, n)$  が算出された推定音声信号となる。

本法においては、各反復回数ごとに用いられる雑音スペクトルが、前段階の反復処理で残留した雑音成分から推定される。このとき、1 回のスペクトル引き算処理において、含有される雑音成分量は低減されるため、原理的には、本法での反復回数を増大させることにより、雑音量の大幅な低減が図られる。

重み係数  $\lambda_k$  は、単独で用いられる場合、 $\lambda_k \geq 1$  のように設定されることが多い [5]。これは、雑音スペクトルを過剰推定することにより、ミュージカルノイズを抑圧できるためである。しかし、雑音スペクトルを過剰推定する場合には、もとの音声にひずみを与えることにもなる。そこで本法では、少ない反復回数では  $\lambda_k$  の値を小さくし、反復回数が多くなるにつれて  $\lambda_k$  の値を増大させることにする。すなわち

$$1 \leq \lambda_k < \lambda_{k+1} \tag{17}$$

とする。反復的に行われるスペクトル引き算により、雑音成分が小さくなったときに重み係数の値を増大させていくことで、音声信号に対しての重み係数による品質の劣化の影響を小さくするためである。このようにすることにより、音声を大きくひずませることなくミュージカルノイズを抑圧することができると思う。

#### 4.2 リアルタイム性を考慮した重み係数付き反復スペクトル引き算法

前述の重み係数付き反復スペクトル引き算法では、入力信号としてのすべての音声信号を得た後に反復処理を行うことを基本とする。しかし、実際に音声信号を処理する際には、連続的に入力される音声を各フレームごとに随時処理を行う必要性もある。そこで、反復回数に応じた雑音スペクトルを先に推定し保持することにより、リアルタイム的な処理を考える。この手法を提案法 2 と呼ぶこととする。具体的には次のようになる。

- 1: 無音区間より推定雑音スペクトル  $|\hat{W}_w(1, f)|$  を算出する。
- 2: 雑音推定の際に用いられた無音区間のみで提案法 1 のスペクトル引き算を行い、各反復回数ごとに  $\lambda_k |\hat{W}_w(k, f)|$  ( $k = 1, 2, 3, \dots, K$ ) を推定し、保持する。
- 3: 各フレームごとに、2 で算出された推定雑音信号を用い、反復的にスペクトル引き算を行う。その際算出されるスペクトルは、

$$|\hat{X}_w(k, f)| = \begin{cases} \left( |Y_w(k, f)| - \lambda_k |\hat{W}_w(k, f)| \right), & \text{---(P)} \\ \left( \lambda_k |\hat{W}_w(k, f)| - |Y_w(k, f)| \right), & \text{---(N)} \end{cases} \tag{18}$$

$$(P) \dots |Y_w(k, f)| > \lambda_k |\hat{W}_w(k, f)|$$

$$(N) \dots \text{otherwise}$$

となる。

4:  $Y_w(f)$  の位相情報を用いて、算出された  $|\hat{X}_w(K, f)|$  を逆フーリエ変換し、 $\hat{x}_w(n)$  を求める。

5: 各  $\hat{x}_w(n)$  をオーバーラップさせ、 $\hat{x}(n)$  を推定音声信号として算出する。

提案法 2 のブロック図が図 4 に示される。この方法は、フレーム処理を基本とした方法であり、スペクトルの引き算処理は一つの入力フレームに対して反復的に行われる。反復ごとの音声信号のフレームのオーバ

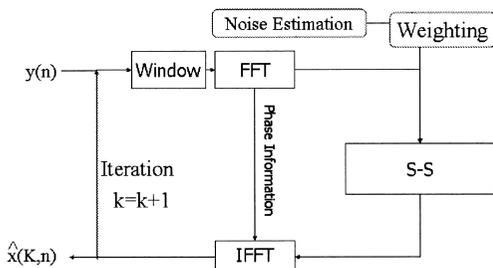


図 3 提案法 1 のブロック図  
Fig. 3 Block diagram of proposed method 1.

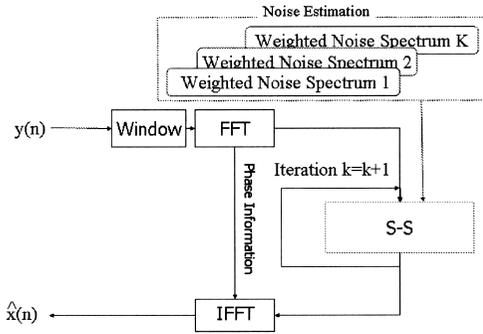


図 4 提案法 2 のブロック図  
Fig. 4 Block diagram of proposed method 2.

ラップ処理が入らないため、前述した提案法 1 と同じ出力音声は得られないことになる。

## 5. 実験

本章では、前章で述べた提案する二つの方法の特性を従来の方法と比較し、評価する。実験に用いたデータは、先述したものと同一である。使用した音声は、NTT アドバンステクノロジー(株)「20ヶ国語音声データベース」の、日本人男女各 2 名ずつ、計 4 名の約 11 秒の連続音声で、サンプリング周波数は 10 kHz、3.4 kHz の帯域制限がかけられている。発話内容は「人々の屏風絵と如来像に対する興味は、800 年の年月によって生じた、表面の微妙な色彩変化にある」、「彼は以前から、科学技術の進歩と人間の勇気が、遙かな宇宙への旅を、可能にしたのだと考えていました」の 2 種類である。雑音は定常雑音として白色雑音及び自動車のエンジン雑音(以下、自動車雑音)、非定常の雑音としてある展示会の会場の雑音(以下、人混み雑音)を用いる。ただし、これらの各雑音も 10 kHz でサンプリングされ、3.4 kHz の帯域制限が施されている。

比較対照とする手法は、典型的なスペクトル引き算法、非線形スペクトル引き算法、重み係数を使用しない反復スペクトル引き算法である。非線形スペクトル引き算法は、前述したとおり、特に自動車雑音に強じんな手法で、定常雑音の抑制には定評がある [9]。ゆえに、特に定常雑音に関しての優劣を評価するために、ここで用いた。また、重み係数を用いない反復スペクトル引き算法は、各提案法と直接比較することにより、重み係数の有用性を調べるために用いた。また、本研究では、残留雑音、特にミュージカルノイズの低

減を目的としているので、典型的なスペクトル引き算法と重み係数を用いない反復スペクトル引き算法においても、ミュージカルノイズの発生が少ない全波整流処理を用いることとする。計算機で実行したスペクトル引き算法では、共通して 512 ポイントのハニング窓をかけ、それぞれのフレームをフレーム長の半分ずつオーバーラップして処理させてある。また、各フレームでのフーリエ変換には 512 ポイントの FFT が用いられている。雑音推定は、共通して先頭 5 フレームを無音区間とし、各フレームにより得られたスペクトルの平均値を雑音スペクトルとした。また、非線形スペクトル引き算のパラメータとしては、2.3 における  $\lambda_Y = 0.9$ ,  $\lambda_W = 0.1$ ,  $\beta = 0.1$  を用いている。これは、非線形スペクトル引き算法において、最も典型的であるとされる値である [9]。

まず、提案法 1 を基本とし、反復回数と反復に対応する重み係数の値を決定するための予備実験を行った。ここでは、式 (17) に従い、重み係数  $\lambda_k$  の値を以下の式のように設定した。

$$\lambda_k = 1 + \alpha \left( \frac{k-1}{K-1} \right)^\gamma \quad (19)$$

ただし、 $\alpha > 0$  とする。ここで、 $\alpha$  の値の変化による雑音抑制精度を調べることにする。予備実験では、SNR=0dB の白色雑音、自動車雑音、人混み雑音付加音声に対し評価を行った。これは、本論文が特に高雑音環境下の音声品質向上を目指しているためである。反復回数  $K$  は 10 回まで変化させた。また  $\lambda_k$  の値が  $k = 10$  で 3, 6, 10 となるように、 $\gamma$  の値を 0.2~1 まで 0.2 刻みで、また 1 から 5 まで 1 刻みで調整し、反復的なスペクトル引き算の挙動を調べた。これは、通常のスぺクトル引き算では、重み係数の値は 6 前後が限界であることが知られていることに基づいている [5]。しかし、この重み係数 6 が反復処理に必ずしも適合しているとはいえないので、上下に幅をもたせた。図 5 に、それぞれの反復回数に応じた重み係数の値を示す。このような  $\lambda$  の推移に対するセグメンタル SNR 改善度を測った。図 6~図 8 に男女 4 人の平均の結果を示す。図より、反復回数が少ない段階で  $\lambda$  の値を大きくすると、音声を大きくひずませることが分かる。一方、重み係数を小さくすると、音声を大きくひずませることなく雑音を抑制できるが、反復回数をむやみに増大させることは、不要な計算を重ねることとなり、合理的ではない。ゆえに、早い段階ですべての雑音環境下で十分に高い特性を得られる  $\lambda$  を選ぶこ

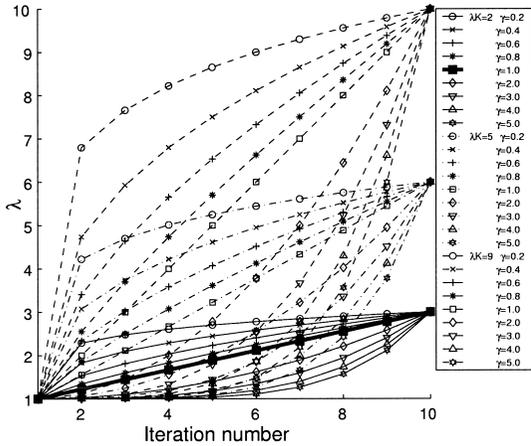


図 5 反復回数と重み係数  
Fig. 5 Iteration number and weighting factors.

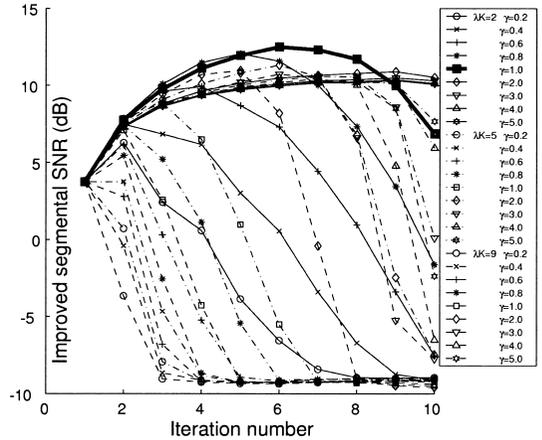


図 7 自動車雑音下での反復回数と重み係数によるセグメンタル SNR 改善度の依存性  
Fig. 7 Dependency of SNR improvement on iteration number and weighting factors in car noise.

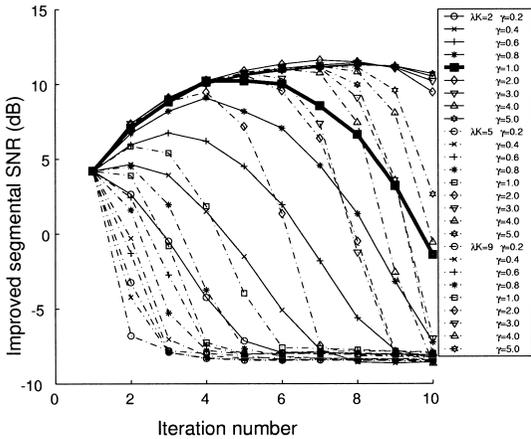


図 6 白色雑音下での反復回数と重み係数によるセグメンタル SNR 改善度の依存性  
Fig. 6 Dependency of SNR improvement on iteration number and weighting factors in white noise.

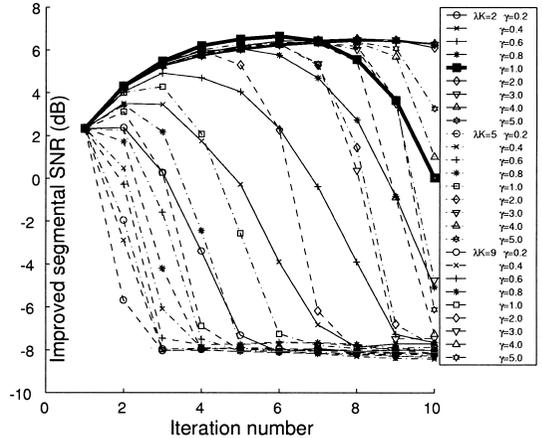


図 8 人混み雑音下での反復回数と重み係数によるセグメンタル SNR 改善度の依存性  
Fig. 8 Dependency of SNR improvement on iteration number and weighting factors in babble noise.

とを考える．よって，反復回数をむやみに大きくせず，最も効率良く雑音抑制効果を向上させる重み係数の調整を，実験では

$$\lambda_k = 1 + 0.2(k - 1) \quad (20)$$

とし，反復回数を  $K = 5$  とすることにした．図 5 ~ 図 8 での太線のプロットがそれにあたる．これに伴い，重み係数を用いないスペクトル引き算法の反復回数も，同様に 5 回と設定した．また，雑音抑制と音声の聞きやすさの両方の観点より，反復回数と重み係数の妥当性を評価するため，試聴実験を行った．評価法には DMOS (Differential Mean Opinion Score) [17]

を用いた．DMOS は，原音声との差を，音声のひずみと雑音の量により判定する評価法となる．ゆえに，DMOS は雑音の低減と音声のひずみを考慮した相対的な評価法である．試聴実験では，まず無雑音の音声を再生し，次いで未処理の雑音付加音声を再生する．無雑音の音声の評価を“5”に固定し，これを最高評価とする．そして，再生された音声に対し，5 段階の評価を被験者にしてもらった．評価値は順に，5：非常に良い，4：良い，3：普通，2：悪い，1：非常に悪い，となる．未処理の雑音付加音声を評価の基準値として実験を行うので，仮に雑音付加音声を音声強調を

施し、品質が劣化したと判断された場合には、未処理の音声で最低評価とならない場合があることに注意されたい。ゆえに音声の品質の改善度は、それぞれの値と未処理の評価値との比較により得られる。一度の試験実験で、各手法にて出力された音声信号が 2 回ずつランダム再生され、連続的に評価をしてもらった。試験実験に用いた音声は、男女各 1 名で、被験者数は各音声につき 10 名である。ゆえに、各数値は累計 40 回分の平均値となる。今回比較に用いられた反復回数は 1, 3, 5, 8 回で、重み係数を用いない場合と重み係数を用いた場合の両方を比較する。結果が表 1 に示される。表より、反復回数が  $K = 5$  で十分な雑音抑制効果をもち、更に反復回数を増大させると、音質を劣化させることが分かる。また、提案法 1 の場合、演算量は反復回数に比例して増大するため、重み係数を利用することにより、少ない反復回数で効率的な音声の品質改善をもたらすことができることも分かる。

これらの設定に基づき、セグメンタル SNR 改善度により実行精度の比較をした。表 2 にその結果を示す。ここでは、各入力信号の SNR に対する男声と女声の平均の結果がそれぞれ示してある。太字は、各 SNR における最良の値である。この表 2 より、それぞれの環境下において各提案法は、雑音抑制の観点から、優れた特性改善を与えることが分かる。また、非線形スペクトル引き算法は、典型的スペクトル引き算法と比較した際に、白色雑音、人混み雑音に比べ、自動車雑音環境下にて最も高いセグメンタル SNR 改善特性を与えていることが分かる。提案法 2 では、提案法 1 と比較すると、多少改善精度が劣ることが分かる。これは先述したとおり、反復ごとのオーバーラップ処理が入らないことに起因するものと考えられる。また、入力 SNR が高くなるほど特性改善が低くなるのは、もとの雑音成分が小さいため、反復処理による残留雑音も

表 1 反復回数と重み係数による DMOS の依存性  
Table 1 Dependency of DMOS on iteration number and weighting factors.

付加雑音	白色雑音	自動車雑音	人混み雑音
未処理	1.52	2.65	1.70
反復 1 回	2.45	3.87	2.50
反復 3 回 重みなし	4.67	4.50	3.27
反復 5 回 重みなし	3.67	4.82	3.17
反復 8 回 重みなし	3.62	4.62	3.15
反復 3 回 重みあり	3.57	4.57	<b>3.35</b>
反復 5 回 重みあり	<b>3.95</b>	<b>4.90</b>	3.30
反復 8 回 重みあり	3.67	4.32	3.12

小さく、特性改善の余地が少ないためと考えられる。

次に、音声のひずみを測り、各提案法が音声信号をひずませることなく、雑音抑制処理できているかを調べる。ここでの実験では、LAR 距離 (Log Area Ratio Measure) を用いた。LAR とは、声道を断面積の異なる複数の無損失音響管をつないだものとモデル化し、その接続点の断面積比 (反射係数) を利用したパラメータである。もとの音声信号である無雑音信号と処理された出力信号の差を比べることにより、音声のひずみを測定することができる。LAR 距離は各フレームごとに算出されるので、今回の実験では各々のフレームの平均値をとり、以下の式で与えた。

$$\eta_x(i) = \log_{10} \left( \frac{1 + k_x(i)}{1 - k_x(i)} \right) \quad (21)$$

$$\eta_{\hat{x}}(i) = \log_{10} \left( \frac{1 + k_{\hat{x}}(i)}{1 - k_{\hat{x}}(i)} \right) \quad (22)$$

$$\text{LAR} = \frac{1}{M} \sum_{m=1}^M \sqrt{\frac{1}{I} \sum_{i=1}^I |20\{\eta_{\hat{x}}(i) - \eta_x(i)\}|^2} \quad (23)$$

ここで、 $M$  はフレームの個数、 $x$ 、 $\hat{x}$  はそれぞれ原音声、出力音声を表す。 $k_x(i)$ 、 $k_{\hat{x}}(i)$  はそれぞれ原音声、出力音声の  $i$  次の反射係数である。反射係数の次数は  $I$  となる。LAR は実際の音声信号と導出された音声信号との差を表す値なので、値が小さいほど優良な結果を示す。図 9 ~ 図 11 に結果が示される。図中の SS はスペクトル引き算法、NSS は非線形スペクトル引き算法、ISS は反復スペクトル引き算法、W-ISS は提案法 1、R-W-ISS は提案法 2 を表す。白色雑音、自動車

表 2 セグメンタル SNR 改善度の比較  
Table 2 Comparison of improved segmental SNR.

雑音の種類	入力音声の SNR (dB)	0	5	10
白色雑音	スペクトル引き算法	3.94	4.01	3.69
	非線形 SS	6.67	6.33	5.51
	反復 SS	10.32	9.53	<b>8.30</b>
	提案法 1	<b>10.64</b>	<b>9.57</b>	8.23
	提案法 2	9.92	8.99	7.64
自動車雑音	スペクトル引き算法	3.52	4.19	4.00
	非線形 SS	9.32	8.89	8.22
	反復 SS	10.22	9.77	9.04
	提案法 1	<b>12.33</b>	<b>11.58</b>	<b>10.57</b>
	提案法 2	11.24	10.60	9.70
人混み雑音	スペクトル引き算法	2.61	2.66	6.09
	非線形 SS	5.21	5.04	4.61
	反復 SS	6.85	6.52	5.88
	提案法 1	<b>7.34</b>	<b>6.93</b>	<b>6.13</b>
	提案法 2	6.42	6.09	5.43

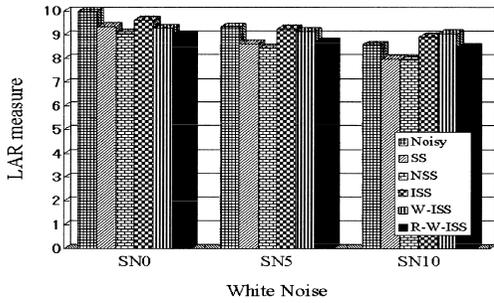


図 9 白色雑音下での LAR  
Fig.9 LAR in white noise.

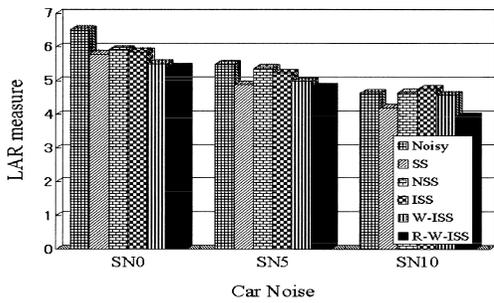


図 10 自動車雑音下での LAR  
Fig.10 LAR in car noise.

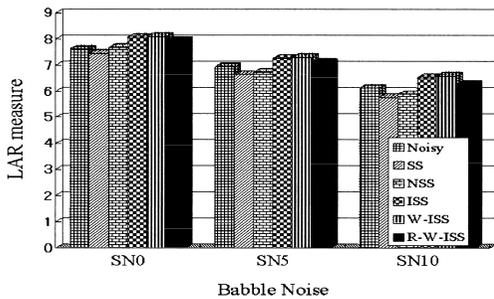


図 11 人混み雑音下での LAR  
Fig.11 LAR in babble noise.

雑音においては、SNR が低い環境下において、各提案法が優良な結果を示している。人混み雑音においては、各提案法は、音声のひずみが増大する結果を示した。これは、人混み雑音のような非正常雑音環境下においては、スペクトル引き算自体がその性能を十分に発揮できないことに起因していると考えられる。また、単純な反復処理を行う場合と、提案する重み係数を用いた場合とを比較すると、すべての環境下にて提案法は従来の反復スペクトル引き算法と同等か、優れた結果を示している。これにより、重み係数は音声のひずみを増大させる要因とはならないことが分かる。

また、ミュージカルノイズの低減を量るために、図 12 ~ 図 15 に各手法にて処理した音声信号のスペクトログラムを示す。図より、反復処理を施すことによって、大幅な雑音の低減がなされていることが分かる。更に、反復スペクトル引き算法と、提案法 1 を比較すると、重み係数を用いた方が、斑点模様が減っていることが見てとれる。これは、重み係数がミュージカルノイズの除去に有効であることを示しているといえる。

しかしながら、様々な環境下においては、反復処理による雑音抑制を行うことにより、多少の音質の劣化が生じることは避けられないことが分かる。ただし、音声を実際に聞いた際に、このひずみがどの程度まで人間に煩わしさを与えるかはこの結果のみでは言及しかなる。なぜなら、音声の品質は雑音量とひずみの両観点から考慮する必要があるからである。

そこで、次に試聴実験による評価を行った。評価法は DMOS を用いる。用意した音声信号は、それぞれ重み係数を用いない反復スペクトル引き算法、非線形スペクトル引き算法、そして提案法 1、提案法 2 によって得られたものである。試聴実験に用いた音声は、予備実験と同じく男女各 1 名で、被験者数は各音声につき 10 名である。ゆえに、各数値も同様に累計 40 回分

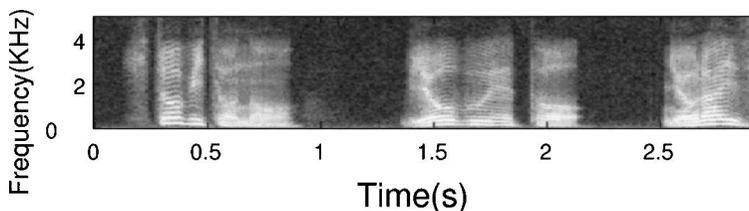


図 12 無雑音下における女声のスペクトログラム  
Fig.12 Spectrogram for clean female speech.

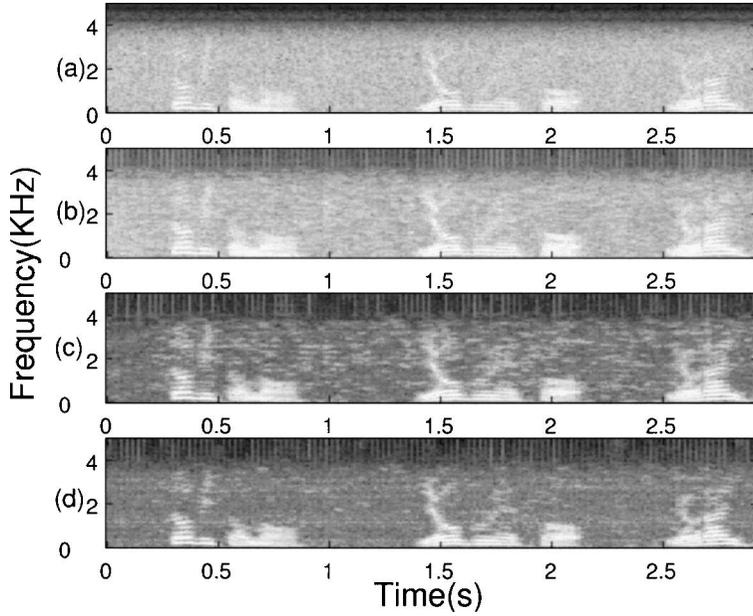


図 13 白色雑音下における女声のスペクトログラムの比較 (a) 未処理, (b) 非線形 SS, (c) 反復 SS, (d) 提案法 1

Fig. 13 Comparison of spectrograms for female speech in white noise. (a) Noisy, (b) NSS, (c) ISS, (d) W-ISS.

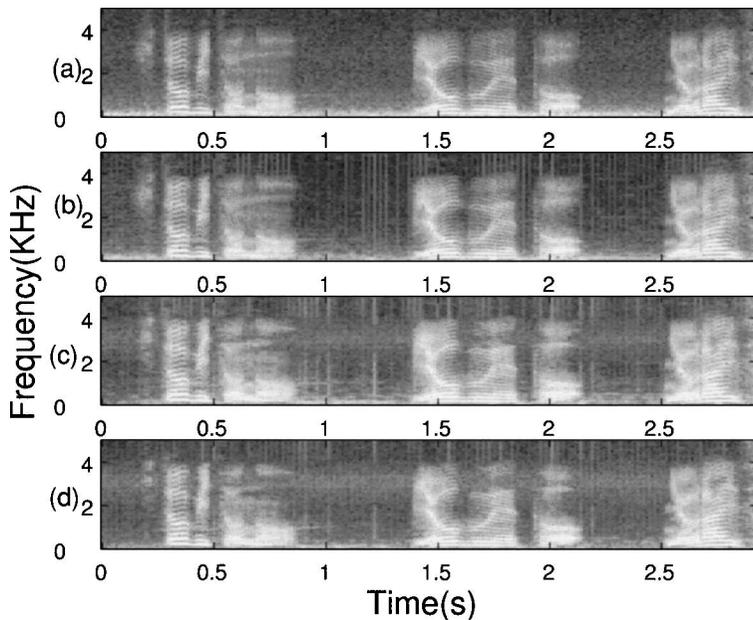


図 14 自動車雑音下における女声のスペクトログラムの比較 (a) 未処理, (b) 非線形 SS, (c) 反復 SS, (d) 提案法 1

Fig. 14 Comparison of spectrograms for female speech in car noise. (a) Noisy, (b) NSS, (c) ISS, (d) W-ISS.

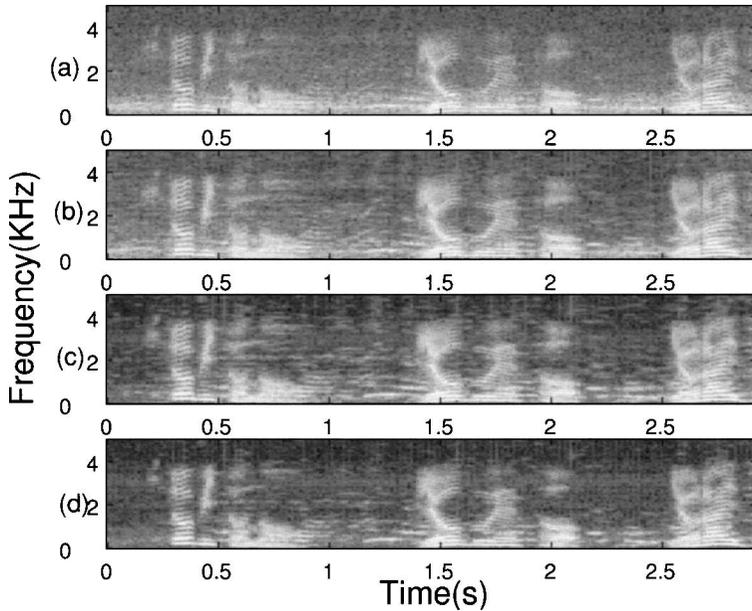


図 15 人混み雑音下における女声のスペクトログラムの比較 (a) 未処理, (b) 非線形 SS, (c) 反復 SS, (d) 提案法 1

Fig. 15 Comparison of spectrograms for female speech in babble noise. (a) Noisy, (b) NSS, (c) ISS, (d) W-ISS.

表 3 DMOS による比較  
Table 3 Comparison of DMOS.

雑音の種類	入力音声の SNR	0 dB	5 dB	10 dB
白色雑音	未処理	1.20	1.10	1.55
	スペクトル引き算法	2.00	1.85	2.05
	非線形 SS	2.93	2.55	2.48
	反復 SS	3.35	3.78	3.38
	提案法 1	<b>3.45</b>	<b>4.05</b>	<b>3.83</b>
	提案法 2	3.43	3.55	3.78
自動車雑音	未処理	1.98	1.73	2.45
	スペクトル引き算法	2.45	2.23	2.90
	非線形 SS	3.20	3.10	3.98
	反復 SS	3.75	3.75	4.35
	提案法 1	<b>4.05</b>	<b>4.03</b>	<b>4.48</b>
	提案法 2	4.00	3.93	4.43
人混み雑音	未処理	1.53	1.65	1.83
	スペクトル引き算法	2.10	2.25	2.43
	非線形 SS	2.63	2.75	2.98
	反復 SS	3.20	3.03	3.20
	提案法 1	<b>3.25</b>	<b>3.45</b>	<b>3.68</b>
	提案法 2	3.23	2.93	3.23

の平均値となる．結果が表 3 に示される．反復のみのスペクトル引き算法と提案法 1 の結果を比較すると，提案法 1 が良好な結果を示している．これは重み係数を用いることにより，雑音抑制の特性が向上したことを示している．また表 3 から，各提案法は各環境にお

いて高い評価を得ていることが分かる．提案法 2 の場合，提案法 1 と比較すると，多少の劣化が見られる．ゆえに，リアルタイム性を考慮せずともよい場合には，提案法 1 を用いることが最良といえる．しかしながら，リアルタイム性を考慮した場合においても，その結果は非線形スペクトル引き算法及び，従来の反復スペクトル引き算法よりも良好で，各提案法の有効性がうかがえ，各提案法の実用的な利用が期待できる．

## 6. む す び

反復処理を用いたスペクトル引き算法に，改良を加えた 2 種類のスペクトル引き算法を提案した．各提案法は，反復回数に応じて重み係数を変化させる方法であり，過剰な重み係数により大きく音声をひずませることなく雑音を効果的に除去し，ミュージカルノイズを抑制する方法である．また，リアルタイム性を考慮した，フレームベース型の重み係数付き反復スペクトル引き算法を同時に提案した．各提案法は，スペクトル引き算法の計算の簡潔さを保持しつつ，従来のスペクトル引き算法に更なる雑音抑制の特性改善を与える．実験により，各提案法は従来のスペクトル引き算及びその改良法に比べ，より優れた雑音抑制効果を与える

ことが確かめられた。

## 文 献

- [1] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-27, no.2, pp.113-120, April 1979.
- [2] J.S. Lim, "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-26, no.5, pp.471-472, Oct. 1978.
- [3] S.V. Vaseghi, Advanced Digital Signal Processing and Noise Reduction, Second ed., Wiley, 2000.
- [4] L.S. Boh, C.T. Yit, J.S. Chang, and T.T. Chin, "A parametric formulation of the generalized spectral subtraction method," IEEE Trans. Speech Audio Process., vol.6, no.4, pp.328-337, July 1998.
- [5] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," Proc. IEEE Conference on Acoustics, Speech, and Signal Processing, pp.208-211, April 1979.
- [6] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," IEEE Trans. Speech Audio Process., vol.7, no.2, pp.126-137, March 1999.
- [7] Z. Goh, K.C. Tan, and B.T.G. Tan, "Postprocessing method for suppressing musical noise generated by spectral subtraction," IEEE Trans. Speech Audio Process., vol.6, no.3, pp.287-292, May 1998.
- [8] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Speech Audio Process., vol.9, no.5, pp.504-512, July 2001.
- [9] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and the projection, for robust speech recognition in cars," Speech Commun., vol.11, pp.215-228, Nov. 1992.
- [10] K. Onoe, H. Seki, T. Kobayakawa, S. Sato, S. Homma, T. Imai, and A. Ando, "Filter bank subtraction for robust speech recognition," IEICE Trans. Inf. & Syst., vol.E86-D, no.3, pp.483-488, March 2003.
- [11] H. Nakashima, Y. Chisaki, T. Usagawa, and M. Ebata, "Spectral subtraction based on statistical criteria of the spectral distribution," IEICE Trans. Fundamentals, vol.E85-A, no.10, pp.2283-2292, Oct. 2002.
- [12] 杉山昭彦, 芹沢昌宏, 加藤正徳, "携帯電話用雑音抑制技術の標準化," 信学技報, SP2002-1, April 2002.
- [13] 緒方伸哉, 島村徹也, "反復スペクトル引き算によるミュージカルノイズの低減," 音響春季講論集, pp.387-388, March 2001.
- [14] K. Yamashita, S. Ogata, and T. Shimamura, "Spectral subtraction iterated with weighting factors," Proc. IEEE Speech Coding Workshop, pp.138-140, Oct. 2002.
- [15] J.R. Deller, Jr., J.G. Proakis, and J.H.L. Hansen, Discrete-Time Processing of Speech Signals, Macmillan, 1993.
- [16] D.B. Paul, "The spectral envelope estimation vocoder," IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-29, no.4, pp.786-794, Aug. 1980.
- [17] 守谷健弘, 音声符号化, コロナ社, 1998.  
(平成 17 年 2 月 21 日受付, 6 月 8 日再受付, 7 月 21 日最終原稿受付)



山下 浩平 (学生員)

平 14 埼玉大・工・情報システム卒。現在、同大学院博士後期課程に在学。デジタル音声処理に関する研究に従事。日本音響学会会員。



緒方 伸哉 (正員)

平 12 埼玉大・工・情報システム卒。平 17 同大学院博士後期課程了。同年メディアドライブ(株)入社。工博。在学中はデジタル音声処理に関する研究に従事。日本音響学会, 信号処理学会各会員。



島村 徹也 (正員)

昭 61 慶大・理工・電気卒。平 3 同大学院博士課程了。工博。同年埼玉大・工・助手。平 10 同助教授, 現在に至る。この間, 平 7 ラフバラ大学, 平 8 ベルファーストクイーンズ大学(ともに連合王国)客員研究員。デジタル信号処理とその音声, 通信システムへの応用に関する研究に従事。IEEE, EURASIP, アメリカ音響学会, 日本音響学会各会員。