論文

変形DFT対に基づく逐次スペクトル検出を用いた 音声重畳雑音除去

中西 功¹, 湊 勇気¹, 李 仕剛¹, 藤井 健作²

¹ 鳥取大学大学院 工学研究科, 鳥取県鳥取市湖山町南 4-101

² 兵庫県立大学大学院 工学研究科, 兵庫県姫路市書写 2167

Noise Reduction on Speech Using Sequential Spectrum Detection Based on Modified DFT Pair

Isao Nakanishi¹, Yuuki Minato¹, Shigang Li¹ and Kensaku Fujii²

¹Graduate School of Engineering, Tottori University, 4-101 Koyama-minami, Tottori 680-8552, Japan

²Graduate School of Engineering, University of Hyogo, 2167 Shosha, Himeji, Hyogo 671-2201, Japan

E-mail: nakanishi,li@ele.tottori-u.ac.jp, fujiken@eng.u-hyogo.ac.jp

Abstract A single-channel method for noise reduction on speech is proposed in this paper. A noise added speech is decomposed into frequency signals using the modified DFT. From each frequency signal, signal and noise elements are sequentially detected and then noise subtraction is performed by subtracting the noise element from the signal one. Detectors for the signal and noise elements are implemented by simple averaging circuits. However, the noise level detector needs a coefficient to compensate differences between the detected levels and proper ones. In this paper, we introduce automatic setting of the coefficient into the proposed method. The effectiveness of the proposed noise reduction method is evaluated in computer simulations of speech noise reduction and subjective evaluations using processed signals.

Keywords: noise reduction, single-channel system, modified discrete Fourier transform pair, sequential spectrum detection, noise subtraction coefficient

1. はじめに

雑音環境下における通話品質を改善するため、音声に重畳する雑音除去に関する研究が盛んに行われており、大別すると単一の入力信号だけを用いる方法と複数の入力信号を用いる方法になる。以下では便宜上それぞれ単一マイクロフォン方式、複数マイクロフォン方式と呼ぶ。

複数マイクロフォン方式は、適応ノイズキャンセラやマイクロフォンアレイといった構成に代表され [1],

複数の信号（情報）が同時に得られるため、高い雑音除去性能が期待できる。一方、単一マイクロフォン方式は、得られる情報は一つしかないために雑音除去性能は劣るが、機器の小型化、低コスト化の面で有利であるため、小型・携帯型機器での応用に向いている。

単一マイクロフォン方式として有名なものにスペクトル減算 (Spectral Subtraction; SS) 法 [2] がある。これは、事前の無音声区間（ブロック）において高速フーリエ変換 (FFT) により雑音スペクトルを検出しておき、その後の雑音重畳区間において同様に求めた

雑音重畳音声スペクトルから差し引くことで音声スペクトルを得るものである。ここで、SS法は一つのマイクロフォンを、雑音用と雑音重畳音声用に時分割で使用しているとも考えられ、そういう面では完全な単一マイクロフォン方式ではない。そのため、雑音検出後にその特性が変化した場合、雑音除去性能が劣化するという問題があり、改良法がいくつか提案されている [3, 4, 5]。

そこで、本論文では、一つのマイクロフォンを切り替えて使用する擬似的な方式ではない完全な単一マイクロフォン方式による音声重畳雑音除去を提案し、その実現可能性について検証する。まず、変形離散フーリエ変換 (Modified Discrete Fourier Transform ; MDFT) 対 [6] を用いて雑音重畳音声を周波数毎の信号に分解し、周波数毎に設けた信号ならびに雑音レベル検出回路によりそれぞれのスペクトル量を検出することで逐次的にスペクトル減算を実現する [7]。MDFT 対は単純な FIR フィルタバンクで実現できるため、FPGA 等のハードウェア実現に適する [8]。ただし、検出される雑音スペクトル量は近似的なものであるため、補正するための係数が必要になる [9]。本論文では、その自動設定に関して一つの知見が得られたので示す。また、提案法の雑音除去性能について、計算機シミュレーション実験や主観評価実験を通して評価を行ったので報告する。

以下、2章では提案する MDFT 対に基づく逐次スペクトル検出による音声重畳雑音除去の構成について述べ、3章ではその有効性を音声重畳雑音除去に関する計算機シミュレーションにより検証する。さらに、4章ではシミュレーションにより得られた音声を用いて複数被験者による主観評価実験により提案構成の有効性を検証する。最後に5章でまとめる。

2. 逐次スペクトル検出による音声重畳雑音除去

2.1 システム構成

提案するシステムの全体構成を図1に示す。入力信号 x_i は MDFT により周波数毎の信号 $X_{k,i} (k = 0, 1, \dots, N/2 - 1)$ に分解される。次に、各周波数信号 $X_{k,i}$ から信号レベル検出器 (Signal Level Detector)、雑音レベル検出器 (Noise Level Detector) により、信号スペクトル量 $DS_{k,i}$ 、雑音スペクトル量 $DN_{k,i}$ がそれぞれ逐次的に検出される。そして、これらの検出された量からスペクトル制御回路 (Controller) により雑音除去が行われる。なお、レベル検出器では MDFT 出

図1 提案システム構成

Fig. 1 Proposed system diagram

力の逐次変動を抑えるために平均化処理が行われており、そのために雑音除去後の各周波数信号は平均数 M に相当する標本数だけ遅延される。最後に、雑音除去が行われた各周波数信号 $X'_{k,i}$ は逆変換により音声信号 y_i へと戻される。以下では、これらの処理を詳しく説明する。

2.2 MDFT 対

MDFT 対は、離散フーリエ変換 (DFT) とその逆変換 (IDFT) の演算を簡単化したものであり、変換 (MDFT) と逆変換 (MIDFT) はそれぞれ式 (1)、式 (2) で定義される [6]。

$$X_{k,i} = \sum_{n=0}^{N-1} x_{i-n} \cos(2\pi nk/N) \quad (1)$$

$$x_i = \frac{X_{0,i}}{N} + \frac{2}{N} \sum_{k=1}^{N/2-1} X_{k,i} \quad (2)$$

ここで、 n, i は時間に関する、 k は周波数に関する添え字である。 N は DFT 演算で用いる標本 (解析) 数を表し、ここでは偶数であると仮定する。

これらの式から明らかなようにすべての演算は実数演算だけで実現される。さらに、MDFT は FIR フィルタの構成式と一致し、 N を偶数とすれば、 $1/N$ 、 $2/N$ の乗算はシフト演算で実現できるため、MIDFT は加算だけで実現される。したがって、周波数成分毎に FIR フィルタを設け、それらの出力を加え合わせるだけで MDFT 対演算が実現される。言い換えれば、MDFT は入力信号に含まれる各周波数成分を、位相関係を伴い

図2 信号レベル検出器
Fig. 2 Signal level detector

ながら分離するものであり、よって、MDFT ではそれらを再合成するだけで元の信号が再現できるという仕組みになっている。

逆に MDFT 対は複素数演算を用いないことから、位相制御を伴うような応用には向かない。よって、出力信号の位相は入力信号のそれで代用することになる。このような代用は SS 法でも用いられている。

さらに、DFT の解析区間と入力信号の基本周期が一致しない場合にスペクトル漏れが生じることは良く知られた事実である。これは、MDFT においても同様である。そのため、MDFT への窓関数の導入も提案されている [10, 11]。\$w(n)\$ を窓関数とすると MDFT の定義式は次式となる。

$$X_{k,i} = \sum_{n=0}^{N-1} x_{i-n} \cos \{2\pi(n - N/2)k/N\} w(n) \quad (3)$$

式 (1) の \$\cos(2\pi nk/N)\$ が \$N/2\$ シフトされたものになる。そのため、MDFT 出力はすべて \$N/2\$ だけ遅れることになる。詳しくは先の文献を参照していただきたい。

2.3 レベル検出器

次に MDFT により分解された周波数毎の信号から信号ならびに雑音スペクトル量を検出するのだが、それぞれの特性の違いを利用するため、以下のような仮定を行う。すなわち、雑音スペクトルは広い周波数帯域に分散し、特性の時間的変化は比較的ゆるやかなものとする。一方、音声は固有のスペクトルを持ち、特性の時間的変化は急峻であるとする。

2.3.1 信号レベル検出器

周波数信号から信号レベルを検出するための回路を図 2 に示す [12]。雑音重畳音声を入力とした場合、MDFT により分解された各周波数の信号 (MDFT 出力) は、音声と雑音の両方のスペクトル成分を含む。よって、信号レベルは音声と雑音が混ざったものになる。

まず、MDFT 出力は基本的に正弦波であり、その振幅は振幅スペクトルに相当する [6]。そこで、全波整流するため、二乗して電力スペクトルとする。なお、得られる電力スペクトルは細かな変動を含むため、移動平均することでそれを抑圧する。平均電力スペクトル \$PS_{k,i}\$ は以下で求められる。

$$PS_{k,i} = \frac{1}{M} \sum_{n=0}^{M-1} X_{k,i-n}^2 \quad (4)$$

ここで、\$M\$ は移動窓の長さ (平均数) である。

次に、これを 1 標本数過去の信号レベル検出器の出力 \$DS_{k,i-1}\$ と比較し、現時点での平均電力スペクトル量 \$PS_{k,i}\$ の方が大きい場合には、図中のスイッチ (sw) を a 側に倒す。これより回路の出力 \$DS_{k,i}\$ は現時点での平均スペクトル量になる。

$$DS_{k,i} = PS_{k,i} \quad (5)$$

一方、逆の (小さい) 場合はスイッチを b 側に倒す。この場合、回路出力 \$DS_{k,i}\$ は以下で定義される大きな時定数 \$\alpha\$ を有する平滑化回路の出力となる。

$$DS_{k,i} = \alpha DS_{k,i-1} + (1 - \alpha) PS_{k,i} \quad (6)$$

これらの処理を具体的に述べると以下ようになる。入力に音声加わると音声に含まれる周波数成分の平均電力スペクトル量は急激に増加する。信号レベル検出器はそのような平均スペクトル量の増加に追従し、スペクトルの変化を直接出力する。ただし、音声が続く間であっても音声 (MDFT 出力) は周期信号であるため、スペクトル量は周期的に変動する。しかし、この細かな変動は信号レベルの変動として捉えたくない。そこで、スペクトル量が減少する場合には、大きな時定数の平滑化回路を通すことにより、その変化が出力側に伝わり難い構成になっている。ただし、音声の持続が終了し、信号レベル全体が減少を始めた場合にはその変化に緩やかに追従する。

2.3.2 雑音レベル検出器

次に各周波数信号から雑音レベルを検出するための回路を図 3 に示す [12]。まず、信号レベル検出器と同様に平均電力スペクトル \$PS_{k,i}\$ が算出される。その後、以下の式で示す平滑化回路で処理される。

$$DN_{k,i} = \beta DN_{k,i-1} + (1 - \beta) PS_{k,i} \quad (7)$$

ここで、時定数 \$\beta\$ は 2 つの異なる値 \$\beta_1\$ と \$\beta_2\$ (\$\beta_1 > \beta_2\$) を持ち、以下のような基準で切り替えられる。

図3 雑音レベル検出器
Fig. 3 Noise level detector

現時点での平均電力スペクトル $PS_{k,i}$ と1 標本数過去の回路出力 $DN_{k,i-1}$ を比較し、平均電力スペクトル量が大きい場合には大きい方の時定数 β_1 が選択され、そうでない場合には小さい方の時定数 β_2 が選択される。ここで、小さい方の時定数 β_2 は、信号レベル検出器の立ち下がりと同様にするためのものであり、 α に近く設定する。

具体的に説明する。音声に加わりスペクトルが増加し始めても大きな時定数を有した平滑化回路により検出器の出力は早急には変化せず、非常に緩やかに追従しようとする。一方、スペクトルが減少する場合は、信号レベル検出器と同様に振る舞う。MDFT 出力の細かな周期変化には追従せず、音声の終了時には緩やかに追従する。一般に一つの音素が持続する時間は短く、音声はそのような音素の繰り返しである。したがって、緩やかに増加しては緩やかに減るという動作を繰り返すことで音声の有無に関わらず定常的に存在する雑音量が検出される。

ただし、このような仕組みにより検出される雑音レベルは、厳密な雑音量を検出するものではない。この問題については後述する。

2.4 スペクトル制御回路

スペクトル制御回路では、信号スペクトル量から雑音スペクトル量を差し引くことにより雑音除去を行う。雑音スペクトルを差し引くという考え方はSS法と同じである。ただし、提案法は雑音減算処理を逐次的（標本毎）に行なうため、事前に雑音量を推定する必要はない。

具体的な雑音減算は以下の式で行う。なお、減算結果が負値となる場合は0とする。

$$SS_{k,i} = DS'_{k,i} - \delta DN'_{k,i} \quad (8)$$

ここで、減算処理では各レベル検出器の出力の平方根をとった振幅スペクトル $DS'_{k,i}$, $DN'_{k,i}$ を用いる。し

たがって、 $SS_{k,i}$ は雑音が除去された振幅スペクトルに対応する。

加えて、音声の存在しない雑音だけの区間において、雑音レベル検出器と信号レベル検出器の出力が同じであるのが理想であるが、提案法では両者の特性が異なるために一致しない。そこで、雑音レベル検出器の出力値に雑音減算係数 δ を乗じて補正を行っている。

次に、雑音除去を行った振幅スペクトル $SS_{k,i}$ は、以下の式により周波数信号 $X'_{k,i}$ に戻される。

$$X'_{k,i} = \frac{SS_{k,i}}{DS'_{k,i}} X_{k,i} \quad (9)$$

前述の通り出力信号の位相は、入力信号（すなわち、 $X_{k,i}$ ）のものをを用いる。

2.5 逆変換

逆変換（MIDFT）は雑音除去された各周波数信号を加起来合わせるだけで実現され、雑音除去された信号 y_i が得られる。

$$y_i = \frac{X'_{0,i}}{N} + \frac{2}{N} \sum_{k=1}^{N/2-1} X'_{k,i} \quad (10)$$

2.6 雑音減算係数の自動設定

2.4 節でも示したように提案法は雑音減算係数を必要とし、従来は事前実験により求めた一定値を用いていた [9]。

ここで、例えば、雑音だけが存在する無音声区間では、雑音の引き残しが無いように雑音減算係数は大きめに設定することが望ましい。一方、音声区間では、雑音の引き過ぎにより生じる原音声の劣化を防止するために雑音減算係数は小さめに設定するのが良い。よって、雑音減算係数は音声の有無に応じて調整する必要がある。

そこで、本論文では、信号ならびに雑音レベル検出器の出力を用いて以下のように自動設定することを提案する。

$$\delta_{k,i} = \chi \frac{\sum_{n=0}^{m \times k} DN'_{k,i-n}}{\sum_{n=0}^{m \times k} DS'_{k,i-n}} \quad (11)$$

ここで、 χ は大きさを調整するための係数である。また、それぞれのレベル検出器の出力は、過去のものを

積算したものとなっており、積算に用いる個数 ($m \times k$) は周波数に比例させている。これは、周波数が高くなるほど検出器出力の変動が大きくなるためである。

具体的には、音声区間において式 (11) の分母は大きくなり、雑音減算係数は小さくなる。一方、無音声区間では分母は小さくなるため、雑音減算係数は大きくなり、先に述べた仕様を満たす。

なお、提案する設定法は、自動化が目的であり、最適な設定を保証するものではない。一方で新たな定数 (χ と m) が導入され、その設定に経験的知識が必要となるが、音声や雑音の種類に大きく依存するものではないことは後のシミュレーション実験で確認する。

2.7 実信号によるレベル検出の概観

レベル検出器の働きを理解してもらうため、実際の信号を用いて処理の流れを概観する。まず、レベル検出器で必要となる設定値 (時定数) を表 1 に示す。こ

表 1 レベル検出器の設定値
Table 1 Setting values of level detectors

	α	β_1	β_2
$k < L$	0.991	0.9999999	0.998
$k \geq L$	0.998	0.99999	0.998

こで、一般に音声成分は低い周波数帯域に偏在するため、高周波数帯では雑音が主となる。このような違いに対応するため、 L を境として帯域を区分し、それぞれの帯域に合わせて設定を行っている。今回は $L = 8$ とする。なお、これらの値は実験において見いだされたものである [9]。

16 ビット、8 kHz で標本化された音声信号に平均 0 の白色信号を重畳したものを入力信号とし、DFT の解析数 N を 62、平均スペクトルを算出する際の平均数 M を 10 として検出器の動作を検証した。窓関数はハミング窓を用いた。

低周波数帯域での例として周波数 $k = 1$ におけるレベル検出結果を図 4 に、高周波数帯域での例として周波数 $k = 20$ における結果を図 5 に示す。(a) は平均電力スペクトル $PS_{k,i}$ 、(b) は信号レベル検出器出力 $DS_{k,i}$ 、(c) は雑音レベル検出器出力 $DN_{k,i}$ の時間変化を表している。なお、2.4 節で示した雑音除去の過程と一致させるため、いずれも振幅スペクトルで表している。

音声は標本数が 4000 付近から始まり、19000 付近

図 4 レベル検出器の出力例 ($k = 1$ の場合)

Fig. 4 Outputs of level detectors ($k = 1$)

図 5 レベル検出器の出力例 ($k = 20$ の場合)

Fig. 5 Outputs of level detectors ($k = 20$)

で終了している。図 4 の観察より、音声の主となる周波数では、信号レベル検出器の出力は音声の出現 (立ち上がり) に素早く追従し、かつ、音声が続く過程での細かな変動には大きく影響されず、音声全体の終了 (立ち下がり) に追従して雑音レベルまで下がっていることが見て取れる。一方、雑音レベル検出器の出力は音声の出現や消滅に大きく影響されず、一定量が得られることが分かる。

また、図 5 より、雑音が主である周波数では、信号レベル検出器の出力は雑音の細かな変動には影響されず、雑音レベル検出器も一定量を出力していることが分かる。ただし、雑音が主であれば、信号レベル検出器の出力は雑音レベル検出器のそれと同じになるのが理想であるが、回路特性が異なるために一致していない。これは図 4 の無音声区間でも観測されることである。雑音減算係数 δ はこの差を埋めるためのもので

図6 波形例：男性発話 (A)，白色雑音，SNR=5dB
Fig. 6 Waveforms： Male speech (A), White Noise, SNR=5dB

ある。

(c) の雑音スペクトル量に雑音減算係数 δ を掛けて補正し、それを (b) の信号スペクトル量から差し引くことで (b) と (c) で挟まれた部分が音声スペクトルとして取り出される。これが提案システムにおける雑音除去の仕組みである。

3. 雑音除去シミュレーション実験

逐次スペクトル検出に基づく音声重畳雑音除去システムの実現性、有効性を検証するため、計算機シミュレーションを行った。使用した音声は NTT アドバンステクノロジ (株) の「Multilingual Speech Database 2002」にある日本人男女各 1 名による 2 種類の発話内容、(A)：「省エネルギーが叫ばれています」と (B)：「あなたは物事を楽天的に考えすぎる」である。標本化周波数は 8 kHz，量子化は 16 ビットである。

雑音は、NOISEX-92 [13] の白色雑音、ピンク雑音、話し声 (Babble Noise)，工場騒音 (Factory Noise) の四つを用いた。前者の二つは定常であり、後者の二つは非定常なものである。話し声雑音は、多くの人が同時に会話しているものを背景雑音とするものであり、緩やかな非定常性を示す。工場騒音は、複数の機械の動作音から構成されており、突発的な音が周期的に発生するため、非定常性は比較的強い。

図7 波形例：女性発話 (B)，ピンク雑音，SNR=5dB
Fig. 7 Waveforms： Female speech (B), Pink noise, SNR=5dB

DFT 解析数 N は 128，レベル検出器における平均数 M は 10 とした。式 (11) の雑音減算係数の設定で必要となる定数 χ ， m はそれぞれ 100，7 とした。なお、雑音検出器は大きな時定数を有する平滑化回路であるため、出力の初期値を 0 とすると出力が安定するまでに時間（標本数）を要する。そこで、最初の 100 標本までは平滑化回路を用いるのではなく、それまでの出力の平均をとったもので代替した。

まず、処理波形の例を図 6，7 に示す。図 6 は男性発話 (A) に白色雑音を、図 7 は女性発話 (B) にピンク雑音を重畳させた場合である。それぞれ、(a) が原音波形、(b) が雑音を重畳させた波形 (SNR=5.0dB)，(c) は従来法 [9] による処理後の波形、(d) が提案法によるものである。なお、従来法では雑音減算係数は固定であり、1kHz 未満の周波数では 2.5，1kHz 以上では 4.0 とした。これらは事前実験により見いだしたものである。また、提案法と同様、雑音検出結果は、最初の 100 標本までは出力平均で代替した。

波形の面から提案法が音声に重畳する雑音を除去できていることが確認できる。出力 SNR は、図 6 の場合で 12.3dB，図 7 の場合で 10.1dB であり、定量的にも雑音除去の効果が認められる。

また、従来法の場合、出力 SNR はそれぞれ 11.2dB，

図 8 入出力 SNR (白色雑音の場合)
Fig. 8 Input SNR vs output SNR (White Noise)

図 9 入出力 SNR (ピンク雑音の場合)
Fig. 9 Input SNR vs output SNR (Pink Noise)

9.2dB であったことから従来法より 1dB 程度改善されている。これは、音声の有無に応じた雑音減算係数の設定が有効に働いたものと考えられる。

なお、発話前の残余雑音が発話後のそれと比較して少ないことに関しては、雑音検出器での出力平均代替が影響している。図 4, 5 を詳しくみると処理開始後 5000 標本あたりまでは雑音検出レベルはわずかながら大きくなっている。式 (11) で定めるように雑音区間では引き残しが無いように雑音減算係数が大きくなるようにしているが、それ以上に大きくなり、結果、定めた以上の雑音除去が行われたことが原因である。なお、代替処理は従来法でも行っており、比較する上では問題はないと考える。

次に、入力 SNR を -10dB から 10dB まで変化させたときの出力 SNR の変化を図 8-11 に示す。それぞれ、白色雑音、ピンク雑音、話し声、工場騒音の場合であり、(a) は男性、(b) は女性の音声の場合である。発話内容は (A) である。また、比較として改良型 SS 法 [4] による結果を点線で示す。この改良型 SS 法は、雑音の事前推定を必要としない。言い換えれば SS 法に基づく完全単一マイクロフォン方式を目指したものと

図 10 入出力 SNR (話し声の場合)
Fig. 10 Input SNR vs output SNR (Babble Noise)

図 11 入出力 SNR (工場騒音の場合)
Fig. 11 Input SNR vs output SNR (Factory Noise)

言える。その点で提案する完全単一マイクロフォン方式の実現性を評価する場合の比較対象として適していると考えられる。

これらの結果より、提案法は、発話者や雑音の種類、SNR によらず安定した雑音除去性能を示すことが分かる。また、雑音減算係数の自動設定も提案した方法により可能であることが確認できた。ただし、用いた定数 χ と m が最適であるかどうかの保証はない。それについては、今後より多くの実験を通じて検証する必要がある。

一方、非定常性の雑音の場合、改良 SS 法の方が概ねより良い出力 SNR を示している。2.3 で示したように提案法は、雑音の特性変化が比較的ゆるやかなものであることを仮定しており、音声や突発的な騒音への対処機能は有していないことから予期した結果である。ただし、改良型 SS 法では高い入力 SNR 状態における出力 SNR の劣化が顕著である。一般の雑音除去システムは、雑音成分を除去するだけでなく、本来残すべき音声成分も除去することが多く、そのため、高 SNR 状態での雑音除去性能が劣化する傾向がある。それに対し、提案法では、高 SNR 状態でも雑音除去性

能の劣化は生じておらず、音声の有無や周波数に応じて雑音減算係数を調整したことが有効に働いたためと考える。

4. 主観評価実験

次に、二つの主観評価実験を行った。一つは雑音除去の効果を絶対的に評価するもの、もう一つは改良 SS 法との比較による相対評価である。

4.1 評価法

使用した音声は先のシミュレーション実験で用いたものと同じである。雑音は、ピンク雑音、話し声、工場騒音の三つを用い、入力 SNR は -5dB と 5dB の 2 種類とした。被験者は 11 名の男子大学生で、各種の情報機器が並ぶ大学の研究室においてコンピュータで処理した音声（モノラル）をヘッドフォンにより試聴してもらった。

手順は、最初に原音声を聞かせ、その後に異なる二つの音（A, B）を交互に 2 回聞かせ、どちらが良いかを評価基準に従って評価してもらった。評価基準は表 2 のとおりである。なお、被験者にはそれぞれの音

表 2 評価基準

Table 2 Evaluation standards	
+2	A に比べて B は良い
+1	A に比べて B はどちらかといえば良い
0	A と B は同等
-1	A に比べて B はどちらかといえば悪い
-2	A に比べて B は悪い

がどのような処理を行ったものなのかは教えず、さらに、順番効果を防ぐために二つの音の提示順序はその都度ランダムに並び替えた。

4.2 雑音除去性能の絶対評価

まず初めに、雑音重畳音声と雑音除去後の音声の二つを比較対象として評価実験を行った。結果を図 12 に示す。ここで、図中の記号は、最初が雑音の種類、次が発話者の種類、その次は SNR 値、最後の数値は平均評価値である。例えば、PM-5,-0.18 は、ピンク雑音 (P) における男性 (M) 発話で、SNR が -5dB の場合に平均評価値が -0.18 であったことを示す。その他、B は話

図 12 雑音除去性能の主観評価実験

Fig. 12 Subjective evaluation of noise reduction performance

し声、F は工場騒音、F は女性の場合である。グラフの棒が右に伸びるほど良いと評価されたことになる。

結果として、定常なピンク雑音や、緩やかな非定常の話し声の場合は、評価値が概ね 0 以上であり、提案法の雑音除去結果が確認できる。一方、突発的な音を含む工場騒音の場合は、雑音除去の効果が評価されていない。図 11 に示した SNR での評価では雑音除去の効果が確認されており、著者らが聞く限りでも雑音が削減されていることは明白であった。しかしながら、処理後の音声には残余雑音などによる人工的な音が生じており、このことが被験者に悪く評価されたものと考えられる。同様なことは SNR が 5dB の場合に評価値が悪いという結果にも言える。雑音が少ない状態では、人は自身の聴覚機能により発話内容を聞き取ることができるため、逆に発生した人工的な音を悪く評価したものと考えられる。

4.3 雑音除去性能の相対評価

次に、改良 SS 法による処理結果との比較による主観評価実験を行った。結果を図 13 に示す。図中の表記は先と同じである。小さいながらも大部分において正值の評価が得られたことから、提案法の処理後の音声は改良 SS 法と比べて良いといえる。

ここで、SS 法においても残余雑音などにより人工的な音（ミュージカルノイズ）が発生することは良く知られている。著者らが試聴する限りでは、改良 SS 法による雑音除去後の音声は、原音声成分への影響がうねりやかすれとなり、提案法のそれと比べて若干目立つものであった。

改良 SS 法は、雑音の事前推定は行わず、過去の L フレームの中でパワースペクトルの分散が最小となる場

図 13 改良 SS 法との比較による主観評価実験

Fig. 13 Subjective evaluation comparing with the modified SS method

合のそれを雑音量と仮定して処理を行っている。よって、真の雑音量と推定されたものとの間に時間的なずれが生じたり、雑音が非定常な場合には実際の量と推定されたものとの間に差が生じたりする。提案法においても雑音はある程度の定常性を仮定しており、その部分では改良 SS 法と同じである。しかし、提案法は逐次処理であり、時間的なずれの問題は、基本的にはないと言える。そのあたりの違いが結果として評価値に現れたのかも知れない。

5. まとめ

入力信号を変形離散フーリエ変換 (MDFT) により周波数毎の信号に分解し、レベル検出回路により信号ならびに雑音スペクトル量を検出することで、逐次的にスペクトル減算を実現する新しい音声重畳雑音除去法を提案した。SS 法のように雑音の事前推定、すなわち、マイクロフォンを時分割使用しないため、完全な単一マイクロフォン形式による雑音除去が実現できる。ただし、検出される雑音スペクトル量は近似的なものであり、補正するための係数が必要になるため、その自動設定法を併せて提案した。そして、雑音除去のシミュレーション実験ならびに主観的評価の実験を通して提案構成の実現性、有効性を確認した。

今後の課題は、被験者実験で問題となった処理後の音声に生じる人工的な音の軽減である。SS 法を基本とした手法においてもミュージカルノイズの軽減が課題となっている [14, 15]。今回は提案構成の実現性を確認するのが目的であったため、それへの対応はまったく行っていない。例えば、周波数毎に独立した処理を行うだけでなく、周波数間での協調動作のような処理を加え、残余雑音の突発的な発生や、雑音の引き過

ぎによる無信号状態の発生を抑えることで人工的な音声の発生を抑え、処理後の音声の評価値を向上させることが可能になるのではと考える。その場合、より人間の聴覚特性に近い定量評価や詳細な評価観点による主観評価実験も必要となろう。

一方、最近では DFT (FFT) 以外の変換を導入したり [16, 17, 18]、適応フィルタやカルマンフィルタを用いたりする方法 [19, 20] も提案されている。入力信号に対するより詳細な解析結果とより柔軟な追従性能は、雑音除去能力を向上させる反面、多くの演算量を必要とする。今後の改良に関しては、演算量の増加を抑えつつ、雑音除去性能を向上させることが重要である。加えて、これらの方法との比較も課題である。

参考文献

- [1] 辻井重男, 久保田一, 古川利博, 趙晋輝: 適応信号処理, 昭晃堂, 東京, 1995.
- [2] S. F. Boll: Suppression of Acoustic Noise in Using Spectral Subtraction, IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-27, No. 2, pp. 113-120, Apr. 1979.
- [3] S. Yoon and C. D. Yoo: Speech Enhancement Based on Speech / Noise-Dominant, IEICE Trans. Inf. & Syst., Vol. E85-D, No. 4, pp. 744-750, Apr. 2002.
- [4] 幸田晃, 宇佐川毅, 江端正直: 雑音スペクトル推定にパワー変動を利用したスペクトルサブトラクション法, 日本音響学会誌, Vol. 58, No. 8, pp. 493-500, Aug. 2002.
- [5] T. Murakami, T. Hoya and Y. Ishida: Speech Enhancement by Spectral Subtraction Based on Subspace Decomposition, IEICE Trans. Fundamentals, Vol. E88-A, No. 3, pp. 690-701, Mar. 2005.
- [6] S. Yoneda, I. Nakanishi, I. Sasaki and A. Ogihara: Switched-Capacitor DFT and IDFT Circuit, Int. J. Electronics, Vol. 67, No. 6, pp. 839-851, Dec. 1989.
- [7] I. Nakanishi, Y. Itoh, Y. Fukui and K. Fujii: Noise Reduction System Using Modified DFT Pair, Proc. of 2001 International Symposium on Circuits and System (ISCAS2001), Sydney, Australia, Vol. II, pp. 9-12, May 2001.
- [8] 早惇一, 中西功: 変形離散フーリエ変換対のハードウェア実現, 第 10 回 DSPS 教育者会議予稿集, pp. 74-75, Sep. 2008.
- [9] Y. Nagata, I. Nakanishi, Y. Itoh and Y. Fukui: Performance Improvement of Noise Reduction System Using Modified DFT Pair, Proc. of 2006 International Technical Conference on Circuits/Systems, Computers and Communications, Chiang Mai, Thailand, Vol. II, pp. 365-368, Jul. 2006.
- [10] 中西功, 浅倉健徳, 伊藤良生, 副井裕: 周波数領域 ALE を用いた音声重畳雑音除去の研究 - 窓関数導入による特性改善 -, 信学技報, SIP2004-155, pp. 87-92, Mar. 2005.
- [11] I. Nakanishi, Y. Nagata, T. Asakura, Y. Itoh and Y. Fukui: Speech Noise Reduction System Based on Frequency Domain ALE Using Windowed Modified DFT Pair, IEICE Trans. Fundamentals, Vol. E89-A, No. 4, pp. 950-959, Apr. 2006.

- [12] 星野勉, 坂本淳一, 藤井健作, 大賀寿郎: 1 タップ固定小数点 DSP による音響エコーキャンセラを用いたハンズフリー通話系, 第 13 回デジタル信号処理シンポジウム講演論文集, pp. 195-200, Nov. 1998.
- [13] <http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html>
- [14] C. Breithaupt, T. Gerkmann and R. Martin: Cepstral Smoothing of Spectral Filter Gains for Speech Enhancement without Musical Noise, IEEE Signal Processing Letters, Vol. 14, No. 12, pp. 1036-1039, Dec. 2007.
- [15] S. Hayashi, H. Inukai and M. Sugimoto: A Subtractive-Type Speech Enhancement Using the Perceptual Frequency-Weighting Function, IEICE Trans. Fundamentals, Vol. E92-A, No. 1, pp.226-234, Jan. 2009.
- [16] J. Jensen and R. Heusdens: Improved Subspace-Based Single-Channel Speech Enhancement Using Generalized Super-Gaussian Priors, IEEE Trans. Audio Speech and Language Processing, Vol. 15, No. 3, pp. 862-872, Mar. 2007.
- [17] W. Bobillet, R. Diversi, E. Grivel, R. Guidorzi, M. Najim and U. Soverini: Speech Enhancement Combining Optimal Smoothing and Errors-in-Variables Identification of Noisy AR Processes, IEEE Trans. Signal Processing, Vol. 55, No. 12, pp. 5564-5578, Dec. 2007.
- [18] SF. Lei and YK. Tung: Wavelet-Based Speech Enhancement Using Time-Adapted Noise Estimation, IEICE Trans Fundamentals, Vol. E91-A, No. 9, pp. 2555-2563, Sep. 2008.
- [19] W. Jin, X. Liu, and MS. Scordilis: Speech Enhancement Using Harmonic Emphasis and Adaptive Comb Filtering, IEEE Trans. Audio Speech and Language Processing, Vol. 18, No. 2, pp. 356-368, Feb. 2010.
- [20] A. Ikuta and H. Oromoto: Adaptive Noise Suppression Algorithm for Speech Signal Based on Stochastic System Theory, IEICE Trans. Fundamentals, Vol. E94-A, No. 8, pp. 1618-1627, Aug. 2011.

李 仕刚 1985 年 7 月中国北京清华大学电机工程卒業。1987 年来日。1993 年大阪大学院博士課程了。同年大阪大学基礎工学部助手を経て、現在、鳥取大学大学院工学研究科教授。現在、コンピュータビジョン、ロボットビジョン、映像メディア、複合現実などの研究に従事。電子情報通信学会、電気学会、IEEE 各会員。

藤井 健作 1972 年大阪府立大学大学院修士課程了。同年富士通入社。2001 年より姫路工業大学(現兵庫県立大)教授。博士(工学)。音響エコーキャンセラ、能動騒音制御などの音響信号処理、適応信号処理の研究に従事。1993 年度電子情報通信学会論文賞受賞。1997 年広島大客員教授、1999~2000 年鳥取大客員教授。電子情報通信学会、日本音響学会各会員。

(2011 年 9 月 7 日受付; 2011 年 11 月 8 日再受付)

中西 功 1986 年大阪府立大学大学院修士課程了。同年三洋電機(株)入社。1992 年鳥取大学教育学部講師。現在、同大学院工学研究科准教授。博士(工学)。デジタル信号処理、特に音声信号処理、生体認証に関する研究に従事。2007~2008 年アメリカ合衆国ミシガン州立大学訪問研究員。電子情報通信学会、情報処理学会、IEEE 各会員。

湊 勇気 2008 年鳥取大学大学院工学研究科修士課程了。現在、京セラ株式会社勤務。在学中、音声重畳雑音除去の研究に従事。