

Title	マイクロホン対を用いたスペクトルサブトラクションによる雑音除去法
Author(s)	水町, 光徳; 赤木, 正人
Citation	電子情報通信学会論文誌 A, J82-A(4): 503-512
Issue Date	1999-04-20
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/4719
Rights	Copyright (C)1999 IEICE. 水町光徳, 赤木正人, 電子情報通信学会論文誌 A, J82-A(4), 1999, 503-512. http://www.ieice.org/jpn/trans_online/
Description	

マイクロホン対を用いたスペクトルサブトラクション による雑音除去法

水町 光徳[†] 赤木 正人[†]

Noise Reduction by Paired-Microphones Using Spectral Subtraction

Mitsunori MIZUMACHI[†] and Masato AKAGI[†]

あらまし 本論文では、マイクロホン対で受信した雑音が含まれる信号から雑音のスペクトルを解析的に推定し、スペクトルサブトラクション(SS)を用いて受信信号のスペクトルから雑音スペクトルのみを減算する雑音除去法を提案する。マイクロホンアレーを用いた雑音除去法としては、適応処理を用いた手法が主流であるが、本手法は雑音の推定に関して解析的に構成する減算形アレーを使用する。このため、本手法は演算量が少なく、しかも雑音の経時変化に対する雑音除去能力低下が生じ難いという性質をもつ。そして、従来のSSは定常雑音の除去を目的としているが、本手法では時々刻々雑音スペクトルの推定を行うため、突発性雑音の除去も可能である。本雑音除去法の性能を評価するため、雑音除去の計算機シミュレーション、並びに残響の少ない実環境での雑音除去実験を行い、雑音によりスペクトル包絡に生じたひずみが低減できることを確認した。

キーワード 雑音除去, マイクロホン対, スペクトルサブトラクション, 音声認識

1. ま え が き

音声認識システムは、既に一部実用化されており、各分野への普及には大きな期待が寄せられている。しかし、実使用環境には雑音が存在し、雑音によりひずみが生じた音声が入力されると認識性能は大幅に低下する[1]。したがって、早急にこの問題を解決する必要がある。

耐雑音性能を向上させるためにさまざまな研究が行われているが、その問題を解決するためのアプローチは大きく二つに分類できる。

まず一つは、音声認識の照合部における耐雑音処理である。例えば、HMMベースの認識アルゴリズムで、クリーンな音声に対する状態と雑音の状態とを組み合わせることにより、変動性雑音まで考慮した音声認識アルゴリズムも発表されているが、まだ満足できるレベルには達していない[1]。

音声認識システムの耐雑音性能を向上させるためのもう一つのアプローチとして、音声認識システムのプ

リプロセッサとしての雑音除去が考えられる。この手法の特徴は、雑音除去アルゴリズムと認識アルゴリズムとが完全に独立しているため、いかなる認識アルゴリズムにも対応できることである。

そこで、本論文では、後者の音声認識システムの雑音除去プリプロセッサとしての雑音除去法を提案し、その有用性について検討する。

本雑音除去法は、短時間フレームごとに、減算形マイクロホンアレーにより雑音の振幅スペクトルを解析的に推定する。そして、スペクトルサブトラクション(SS)[2]を用いて、受信信号の振幅スペクトルから推定した雑音の振幅スペクトルを引き去る。

本手法の雑音除去性能を評価するため、計算機シミュレーション、並びに残響の少ない実環境での雑音除去実験を行った。その結果、LPC対数スペクトル包絡ひずみを低減できることを確認した。

2. 雑音除去アルゴリズム

本雑音除去法では、図1に示すように、3本の無指向性マイクロホンを等間隔に直線配列したマイクロホンアレーを使用する。信号はすべて平面波で到来する、すなわち音源とアレーは十分離れていると仮定する。

音声を $s(t)$ とし、これがアレーの両端のマイクロ

[†] 北陸先端科学技術大学院大学, 石川県
School of Information Science, Japan Advanced Institute of
Science and Technology (JAIST), Ishikawa-ken, 923-1292
Japan

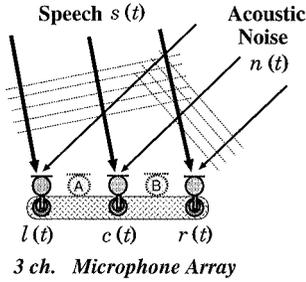


図1 マイクロホンアレーと音声、雑音の関係
Fig. 1 Relationship among a microphone array, speech and an acoustic noise.

ホンから成るマイクロホン対（主対）へ到来する際の時間差を 2ζ とする。また、除去対象となる雑音 $n(t)$ の主対への到来時間差を 2δ とする。

アレーに到来する信号には、加法性の関係が成り立つと仮定すると、それぞれのマイクロホンでの受信信号 $l(t)$, $c(t)$, $r(t)$ と音声 $s(t)$, 雑音 $n(t)$ との間には

$$l(t) = s(t - \zeta) + n(t - \delta) \quad (1)$$

$$c(t) = s(t) + n(t) \quad (2)$$

$$r(t) = s(t + \zeta) + n(t + \delta) \quad (3)$$

の関係が成り立つ。

$l(t)$, $c(t)$, $r(t)$, 及び $s(t)$, $n(t)$ の Fourier 変換をそれぞれ $L(\omega)$, $C(\omega)$, $R(\omega)$, $S(\omega)$, $N(\omega)$ とすると、式 (1), (2), (3) はそれぞれ

$$L(\omega) = S(\omega)e^{-j\omega\zeta} + N(\omega)e^{-j\omega\delta} \quad (4)$$

$$C(\omega) = S(\omega) + N(\omega) \quad (5)$$

$$R(\omega) = S(\omega)e^{j\omega\zeta} + N(\omega)e^{j\omega\delta} \quad (6)$$

となる。

本手法では、以下に示すステップで雑音除去を行う。

1. 3本のマイクロホンの受信信号を用いて、音声（目的信号）及び雑音（除去対象信号）の到来方向を推定する。
2. 3本のうちの2本のマイクロホン（マイクロホン対）で受信した信号を用いて、中央のマイクロホン位置における雑音のスペクトルを推定する。
3. 中央のマイクロホンでの受信信号から、先に推定した雑音成分を引き去ることにより雑音除去を行う。

以下、音声及び雑音の到来方向推定、雑音スペクトルの推定、雑音除去の順に説明する。

2.1 音声及び雑音の到来方向推定セクション [3]

信号が平面波としてマイクロホンへ到来する場合、信号の到来方向の推定は、信号が二つのマイクロホンへ到来する際の時間差を推定することに相当する。よって、本論文では、信号の到来方向として信号の二つのマイクロホンへの到来時間差の推定を行う。

2.1.1 時間差推定の基礎アルゴリズム

平面波として到来する信号の到来時間差の推定に関しては、二つのマイクロホンで受信した信号の相互相関関数を計算し、その最大値を与えるタイムラグを信号の到来時間差とする手法がある [4]。この時間差推定法は、受信信号の周波数帯域が広いほど、また受信信号のデータ長が長いほど推定精度が高くなる。そこで、本論文では二つのマイクロホンで受信した信号 $x_1(t)$, $x_2(t)$ の Fourier 変換 $X_1(\omega)$, $X_2(\omega)$ を用いて、式 (7) により相互相関関数 $\phi_{x_1x_2}(t)$ を計算し、式 (8) により信号の到来時間差 d を求める。

$$\phi_{x_1x_2}(t) = \text{IFFT} \left[\frac{X_1(\omega)X_2^*(\omega)}{|X_1(\omega)||X_2(\omega)|} \right] \quad (7)$$

$$d = \underset{t}{\text{argmax}} [\phi_{x_1x_2}(t)] \quad (8)$$

式 (7) のように、クロススペクトルを受音信号のパワーで正規化することにより、信号の振幅スペクトル形状に依存せず、より高精度の時間差推定が可能になる [5]。以下、これを白色化相関法と呼ぶ。

2.1.2 音声及び最大雑音の到来時間差推定

本時間差推定法では、短時間フレームごとに目的信号及び最大雑音の到来時間差を推定する。

STEP 1: [音声及び雑音の存在判定]

フレームごとに受信信号のパワーをチェックする。それがしきい値を越えた場合に限り、そのフレームにおいて信号の到来時間差推定、雑音除去を行う。

STEP 2: [一方の信号の到来時間差推定]

受信信号に含まれる信号のうち、最もパワーの大きな信号の到来時間差を推定する。

まず、マイクロホンアレーの左端、中央、右端のマイクロホンで受信した信号 $l(t)$, $c(t)$, $r(t)$ のそれぞれの Fourier 変換 $L(\omega)$, $C(\omega)$, $R(\omega)$ を計算する。マイクロホン間隔が大きいほど高精度の到来時間差推定が可能であるため、 $L(\omega)$ と $R(\omega)$ を用いて、白色化相関法により最もパワーの大きな信号の到来時間差を推定する。ただし、式 (7) において使用する信号を 100 Hz から 6 kHz までの音声帯域に限定して相互相関関数 $\phi_{lr}(t)$ を計算し、

$$2d_1 = \operatorname{argmax}_{t \in C_1} [\phi_{lr}(t)] \quad (9)$$

により隣接マイクロホン間への到来時間差 d_1 を推定する．ここで，アレー両端のマイクロホンの間隔は，隣接マイクロホン間隔の 2 倍となっている．また，探索範囲 C_1 は前フレームで推定した二つの信号の到来時間差の近傍，具体的には到来時間差の変化を 2pt (1pt はサンプリング間隔) 以内とする．これは，音源は高速では移動しないという仮定に基づく制約条件である．なお，前フレームにおいて時間差推定を行わなかった場合は，本フレームにおいて新たに音源が生じたと考え，式 (9) において t の探索範囲 C_1 に制約を設けない．

STEP 3 : [他方の信号の到来時間差推定]

受信信号に含まれる信号のうち，2 番目にパワーの大きな信号の到来時間差を推定する．

ここでは，STEP 2 において推定した隣接マイクロホンに時間差 d_1 で到来する信号を完全に抑圧し，それ以外の成分の到来時間差を推定する [6]．まず，受信信号 $l(t)$ と $c(t)$ を用いて，到来時間差 d_1 方向にアレーの指向性の死角 (零点) を形成するようなビームフォーマ $g_{lc}(t)$:

$$g_{lc}(t) = \frac{1}{4} \left[\{l(t+d_1+\tau) - l(t+d_1-\tau)\} - \{c(t+\tau) - c(t-\tau)\} \right] \quad (10)$$

を考える．ここで， τ は 0 以外の任意の定数とする．STEP 2 で推定した到来時間差 d_1 は，音声の到来時間差 ζ か雑音の到来時間差 δ のいずれかである． d_1 が音声の到来時間差である場合，すなわち $d_1 = \zeta$ の場合，式 (10) の Fourier 変換 $G_{lc}(\omega)$ は，式 (1)，(2)，(4)，(5) より

$$G_{lc}(\omega) = N(\omega) e^{j\omega \frac{\zeta-\delta}{2}} \sin \omega \tau \sin \omega \frac{\delta-\zeta}{2} \quad (11)$$

となり，音声 $S(\omega)$ に関する項は存在しない．また， $d_1 = \delta$ の場合，式 (10) の Fourier 変換 $G_{lc}(\omega)$ は

$$G_{lc}(\omega) = S(\omega) e^{-j\omega \frac{\zeta-\delta}{2}} \sin \omega \tau \sin \omega \frac{\zeta-\delta}{2} \quad (12)$$

となり，雑音 $N(\omega)$ に関する項は存在しない．つまり，式 (10) は，図 1 の仮想マイクロホン A の受信信号のうち，時間差 d_1 で到来する信号を完全に抑圧した信号に相当する．同様に， $c(t)$ と $r(t)$ を用いて，時間差 d_1 で到来する信号を完全に抑圧した仮想マイクロホン B での受信信号 $g_{cr}(t)$ を求める．

次に， $g_{lc}(t)$ と $g_{cr}(t)$ の音声帯域の成分のみを使用して相互相関関数 $\phi_{g_{lc}g_{cr}}(t)$ を計算し，

$$d_2 = \operatorname{argmax}_{t \in C_2} [\phi_{g_{lc}g_{cr}}(t)] \quad (13)$$

により，受信信号の中で 2 番目にパワーの大きな信号の到来時間差 d_2 を推定する．探索範囲 C_2 は最もパワーの大きな信号の到来時間差 d_1 の近傍，具体的には 2pt 以内を除くと定めた．

STEP 4 : [音声及び雑音の到来時間差判定]

ここで， d_1 と d_2 のどちらが音声の到来時間差 ζ で，どちらが雑音の到来時間差 δ かを判定する．

d_1, d_2 はそれぞれ，過去に推定した音声及び雑音の到来時間差の推定結果に近い方を選択するものとした．これにより， ζ, δ が定まる．ただし，本手法では雑音除去処理の開始時点において，音声あるいは雑音の少なくとも一方の到来時間差は既知であるという仮定を設けている．本手法では，無音区間が長時間続いた場合は時間差推定に誤りが生じる可能性がある．この問題を解決するためには，音声あるいは雑音に関する何らかの知識を保持しておき，適宜，音声/雑音の判定を行う機構が必要となるであろう．

2.2 雑音推定セクション [6], [7]

式 (1)，式 (3) にそれぞれ $\zeta \pm \tau_1, -\zeta \pm \tau_1$ の遅延を与えた信号を加減算することにより，目的信号である音声 $s(t)$ を完全に抑圧するようなビームフォーマ $g_{lr}(t)$:

$$g_{lr}(t) = \frac{1}{4} \left[\{l(t+\zeta+\tau_1) - l(t+\zeta-\tau_1)\} - \{r(t-\zeta+\tau_1) - r(t-\zeta-\tau_1)\} \right] \quad (14)$$

を定める． $g_{lr}(t)$ の Fourier 変換 $G_{lr}(\omega)$ は

$$G_{lr}(\omega) = N(\omega) \sin \omega(\delta - \zeta) \sin \omega \tau_1 \quad (15)$$

となる． $\tau_1 = \delta - \zeta$ と定めると，式 (15) は

$$G(\omega) = N(\omega) \sin^2 \omega(\delta - \zeta) \quad (16)$$

となり，雑音のスペクトル $N(\omega)$ は

$$\hat{N}(\omega) = G(\omega) / \sin^2 \omega(\delta - \zeta) \quad (17)$$

と推定できる．ただし，これでは雑音スペクトルのうち

$$\omega = \frac{n\pi}{\delta - \zeta} \quad n: \text{整数} \quad (18)$$

となる周波数成分は計算できず、その近傍では推定誤差が生じる。例えば、隣接マイクロホン間隔を 10 cm とし、音声为正中面方向 ($\zeta = 0$)、雑音が正中面より 30° 方向 ($\delta = 146 \mu\text{s}$) から到来する場合、約 $3.4 \cdot n \text{ kHz}$ (n : 整数) 付近が計算できない帯域に相当する。

そこで、式 (14) で定義した主対ビームフォーマで雑音を推定できない周波数成分を、アレーの中央と右端のマイクロホンから成るマイクロホン対 (副対) での受信信号を利用して推定することを考える [3]。副対ビームフォーマ $g_{cr}(t)$ は、式 (10) と同様に

$$g_{cr}(t) = \frac{1}{4} \left[\{c(t + \tau_2) - c(t - \tau_2)\} - \{r(t - \zeta + \tau_2) - r(t - \zeta - \tau_2)\} \right] \quad (19)$$

と定める。ここで、 $\tau_2 = (\delta - \zeta)/2$ とすると、 $g_{cr}(t)$ の Fourier 変換 $G_{cr}(\omega)$ は

$$G_{cr}(\omega) = N(\omega) e^{j\omega \frac{\delta - \zeta}{2}} \sin^2 \omega \frac{\delta - \zeta}{2} \quad (20)$$

となる。副対ビームフォーマ $g_{cr}(t)$ で雑音を正確に推定できない周波数帯域は

$$\omega \approx \frac{2m\pi}{\delta - \zeta} \quad m: \text{整数} \quad (21)$$

となる。したがって、副対ビームフォーマ $g_{cr}(t)$ を使用することにより、主対ビームフォーマ $g_{tr}(t)$ で雑音を正確に推定できない周波数帯域のうち、 n が奇数の場合は推定できるようになる。なお、 n が偶数の場合は、副対ビームフォーマ $g_{cr}(t)$ を使用しても雑音スペクトルを正確には推定できない。

ここで、簡単のため $\zeta = 0$ 、つまり音声は正中面方向から到来すると仮定する。このとき、雑音スペクトルの推定値 $\hat{N}(\omega)$ は、二つのしきい値 $\varepsilon_1, \varepsilon_2$ を設け、以下により求める。

$$\hat{N}(\omega) = \begin{cases} G_{tr}(\omega)/\sin^2 \omega \delta, & \sin^2 \omega \delta > \varepsilon_1 & \text{(a)} \\ G_{cr}(\omega) e^{-j\omega \frac{\delta}{2}} / \sin^2 \omega \frac{\delta}{2}, & & \\ \sin^2 \omega \delta \leq \varepsilon_1 \text{ and } \sin^2 \omega \frac{\delta}{2} > \varepsilon_2 & \text{(b)} \\ G_{tr}(\omega)/\varepsilon_2^2, & \sin^2 \omega \frac{\delta}{2} \leq \varepsilon_2 & \text{(c)} \end{cases} \quad (22)$$

つまり、 $\sin \omega \delta$ が十分大きい周波数帯域 (a) では、マイクロホン間隔が大きい主対ビームフォーマを用いて雑音スペクトルを推定する。 $\sin \omega \delta$ が小さくなるに従い計算誤差が大きくなるので [7]、その周波数帯域 (b) においては副対を用いて雑音スペクトルを推定する。副対を使用しても雑音スペクトルを推定できない周波数帯域 (c) に限って、雑音スペクトルを近似的に定める。

2.3 雑音除去セクション

本手法は、音声認識システムの雑音除去プリプロセッサを目指しているため、雑音により位相スペクトルに生じたひずみについては考慮しない。これは、現在の認識アルゴリズムの多くが、振幅スペクトルに基づいた距離尺度を利用しているため、雑音により振幅スペクトルに生じたひずみさえ回復すればよいからである。そこで、周波数領域における減算である SS [2] を利用し、受信信号を Fourier 変換した $C(\omega)$ の振幅スペクトル $|C(\omega)|$ から、先に推定した雑音の振幅スペクトル $|\hat{N}(\omega)|$ を引き去ることを考える。

このとき、音声の振幅スペクトルの推定値 $|\hat{S}(\omega)|$ は、Boll の SS [2] を改良した非線形 SS:

$$|\hat{S}(\omega)| = \begin{cases} |C(\omega)| - \alpha \cdot |\hat{N}(\omega)|, & |C(\omega)| \geq \alpha \cdot |\hat{N}(\omega)| \\ \beta \cdot |C(\omega)|, & \text{otherwise} \end{cases} \quad (23)$$

により推定する。ここで、 α, β は定数とする。

2.4 本雑音除去アルゴリズムの特徴

本手法の雑音推定セクションでは、解析的に減算形アレーを構成している。

減算形アレーとしては、Griffiths-Jim 形アレー [8] や AMNOR 方式 [9] がよく知られている。最近では、永田らが音源追尾機構を備えた雑音抑圧法 [10], [11] を提案している。しかし、これらは適応フィルタを使用しているため、雑音の変化への追従が難しく、音声と雑音との相関が高い場合など、雑音の種類によっては性能が大幅に低下する [4]。一方、本手法は、信号の到来時間差推定を行い解析的に雑音を推定するため、雑音の経時変化に対する雑音除去能力低下が生じにくく、しかも演算量が少ないという利点をもつ。

次に、本手法の雑音除去セクションでは、非定常雑音の除去も可能である。

Boll が提案した SS は、定常雑音の除去を目的とし

ているため、非音声区間で受音した信号の振幅スペクトルを利用する [2]。時間の経過とともに、非音声区間で受音した信号の振幅スペクトルを変更する手法 [12] も提案されているが、ここでは音響特性が短時間で大幅に変化する雑音、例えば突発性雑音などには対応できない。また、2チャンネル SS として、本手法と同様に目的信号を除去し、雑音の振幅スペクトルを推定する手法 [13] が提案されているが、雑音を正確に推定できないという問題点がある。一方、本手法は、雑音の振幅スペクトルを短時間フレームごとに正確に推定できるため、時々刻々変化する雑音の除去が可能となり、以上の SS では不可能であった突発性雑音も除去できる。

3. 雑音除去実験

本雑音除去アルゴリズムの性能を評価するため、客観的評価尺度を設定し、雑音除去の計算機シミュレーション、並びに残響が少ない実環境における雑音除去実験を行った。

3.1 客観的評価尺度の設定

客観的評価尺度としては、一般に信号対雑音比 (SNR) やスペクトルひずみ (SD) が用いられている。しかし、本手法は位相を全く考慮していないため、時間波形の SNR を雑音除去性能の評価尺度として用いるのは不適切である。

そこで、本論文では、SD と同様に周波数領域で定義される評価尺度を設定する。まず、音声認識アルゴリズムの多くが LPC 分析に基づくスペクトル距離尺度を利用しているという現状を考慮し、LPC 分析を行い対数スペクトル包絡を計算する。そして、短時間フレームごとに、雑音によりその対数スペクトル包絡に生じたひずみを SD と同様に計算する。以下、これを LPC 対数スペクトル包絡ひずみ (LPC-SED) と呼ぶ。LPC-SED の計算にはユークリッド距離を用いるが、これはユークリッド距離が距離尺度の対称性、正値性を満たすことができ、音声認識の距離尺度として利用価値が高いためである [16]。

実際の計算方法については、評価の基準となる信号 (クリーンな音声) と評価対象信号 (雑音除去前、後の信号) のそれぞれに対し、各フレームごとに LPC 対数スペクトル包絡を表 1 に示す条件で求め、それらを $S_c(\omega)$, $S_x(\omega)$ とする。なお、本論文では LPC-SED を計算するために、雑音除去後の音声の振幅スペクトルと受音信号の位相スペクトルを用いて、逆 Fourier

表 1 LPC 分析条件
Table 1 Conditions for LPC analysis.

分析パラメータ	設定値
サンプリング周波数	12 kHz
フレーム長	21.3 ms
フレーム周期	5.3 ms
窓関数	hamming 窓
LPC 次数	16
プリエンファシス	0.98

変換により時間領域へ変換した信号を作成し、これを LPC 分析して求めたスペクトル包絡を $S_x(\omega)$ としている。これは、音声認識で通常用いられている分析条件を考慮し、雑音除去処理の分析条件 (表 2) とは異なる分析条件 (表 1) で評価を行うためである。

このとき、LPC-SED を次式で定める。

$$\text{LPC-SED} = \sqrt{\frac{1}{N} \sum_{i=1}^N \{S_x(i) - S_c(i)\}^2} \quad [\text{dB}] \quad (24)$$

N : 最高周波数 (6 kHz) に対応するサンプル番号

3.2 雑音除去の計算機シミュレーション

3.2.1 シミュレーションデータ

計算機シミュレーションのためのデータは、計算機上でクリーンな音声に雑音を付加して作成した。これは、暗騒音や残響が全く存在しない理想的な環境 (自由音場) に相当する。

音声は、ATR 音声データベース [17] に収録されているタスクコード SY 単音節データセットから、男性話者 mht 発話の単母音 /a/, /i/, /u/, /e/, /o/ の五つを使用し、それらの到来時間差を 0 (正中間方向から到来) と設定した。一方、雑音は継続時間が 50 ms で帯域幅が 200 Hz、中心周波数が 1.5 kHz と 2.5 kHz の 2 種類の狭帯域雑音が、音声区間で突発的に発生するよう作成した。これらの雑音は、前者が到来時間差 8 pt (右 35° に相当)、後者が到来時間差 10 pt (右 45° に相当) と設定した。

音声と雑音との時間関係については、音声データ /a/ を使用した場合、音声は約 10 ms から約 150 ms まで存在し、中心周波数 1.5 kHz の雑音が 30 ms から 80 ms まで、中心周波数 2.5 kHz の雑音が 90 ms から 140 ms まで存在する。

また、雑音除去アルゴリズムの各種パラメータは、表 2 に示す値に設定した。フレーム長を非常に短く設定した理由は、突発性雑音などの時変性雑音にも対

表 2 雑音除去アルゴリズムのパラメータ設定値
Table 2 Parameter settings for noise reduction.

パラメータ	設定値
隣接マイクロホン間隔	10 cm
サンプリング周波数	48 kHz
フレーム長	5.3 ms
フレーム周期	2.7 ms
窓関数	hamming 窓
しきい値 ε_1	0.5
しきい値 ε_2	0.2
サブトラクション係数 α	1
フロアリング係数 β	0.001

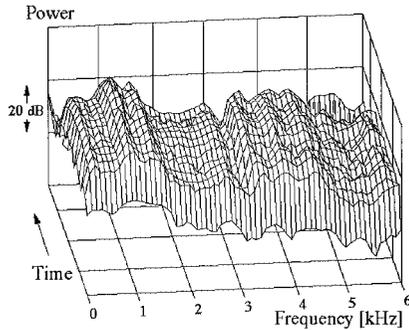


図 2 クリーンな音声の対数スペクトル包絡
Fig. 2 Log spectral envelope of clean speech.

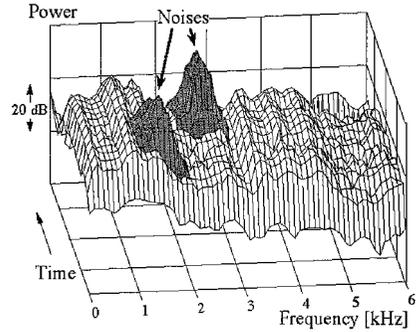


図 3 雑音を付加した音声の対数スペクトル包絡
Fig. 3 Log spectral envelope of noise-added speech.

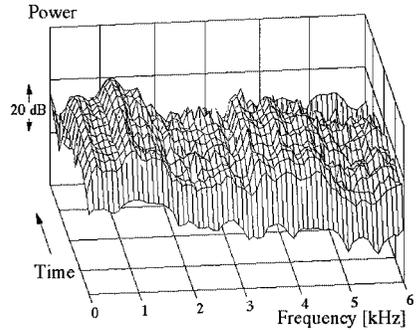


図 4 雑音除去後音声の対数スペクトル包絡
Fig. 4 Log spectral envelope of noise-reduced speech.

応できるようにするためである．本論文では，信号の到来時間差の推定精度を向上させるため，到来時間差推定と雑音推定・除去処理とは異なる分析条件を採用した．表 2 に示した分析条件のもと，第 $(4n + 1)$ フレーム ($n = 0, 1, 2, \dots$) において，第 $(4n + 4)$ フレームまでの 4 フレーム分の信号をいったん統合して到来時間差推定を行い，第 $(4n + 2)$ ，第 $(4n + 3)$ ，第 $(4n + 4)$ フレームでは第 $(4n + 1)$ フレームで推定した信号到来時間差を利用して雑音推定・除去処理を行った．

3.2.2 雑音除去の計算機シミュレーション結果

クリーンな音声/a/の対数スペクトル包絡を図 2 に，それに SNR が -10 dB となるように突発性雑音を付加した音声を図 3 に，雑音除去を行った結果を図 4 に示す．図 2~図 4 の対数スペクトル包絡は，それらにより細かな差異を見るため，LPC 分析ではなく，FFT ケプストラムより求めたものである．

ここで，LPC-SED のフレームごとの計算結果を図 5 に示す．図 5 は，上段から順に，SNR が 0 dB である雑音除去前音声の LPC-SED，雑音除去後音声の LPC-SED，雑音除去による LPC-SED の改善量で

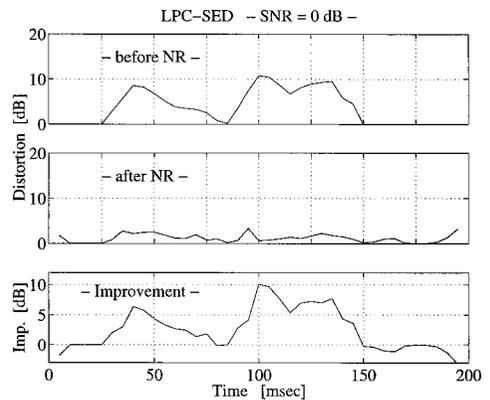


図 5 LPC-SED の変化 (上段：雑音除去前の LPC-SED，中段：雑音除去後の LPC-SED，下段：雑音除去による LPC-SED の改善量)

Fig. 5 Changes of LPC-SED (Upper panel: LPC-SED before noise reduction, Mid panel: LPC-SED after noise reduction, Lower panel: Improvement of LPC-SED by noise reduction).

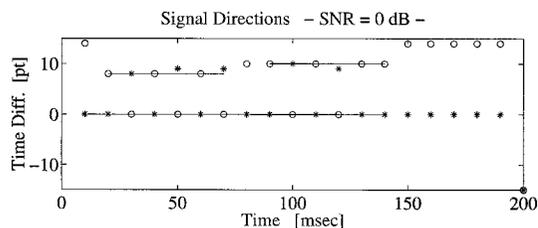


図6 音声及び雑音の到来時間差の設定値とその推定結果
Fig. 6 Set time differences of speech and noises, and results of their estimations.

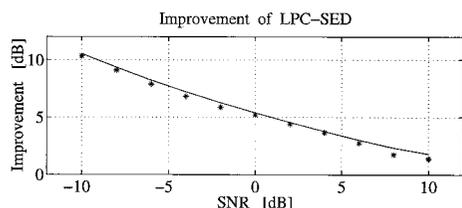


図7 各 SNR に対する LPC-SED の平均改善量
Fig. 7 Mean improvement of LPC-SED for each SNR.

ある．音声と雑音が共存する区間における LPC-SED の平均改善量は約 4.85 dB であった．この場合の音声及び雑音の到来時間差推定結果を図 6 に示している．図 6 の縦軸は、隣接マイクロホンへの到来時間差を表している．図中の実線は、音声及び雑音の到来時間差の設定値、‘*’、‘○’ はそれぞれ最初に推定された信号の到来時間差、2 番目に推定された信号の到来時間差である．

また、雑音を付加した音声の SNR が異なる場合の雑音除去性能については、音声と雑音が共存する区間における LPC-SED の平均改善量を図 7 にまとめて示している．図中の ‘*’ は音声及び雑音の到来時間差を自動推定した場合、実線がそれらを既知とした場合の LPC-SED の平均改善量である．

/a/以外の音声データに対する雑音除去結果は、音声と雑音が共存する区間における LPC-SED の平均改善量として表 3 に示す．

LPC-SED と音声認識率との関係については、LPC-SED による評価結果と、ほぼ良好に認識結果を反映している評価尺度 NDSR [18] による評価結果との比較により検討した．ここで使用した音声は、ATR 音声データベース [17] に収録されている男性話者 mht 発話の音韻バランス単語 216 語とし、到来時間差 0 (正

表 3 日本語 5 母音に対する LPC-SED の平均改善量
Table 3 Mean Improvement of LPC-SED for Japanese 5 vowels.

母音	LPC-SED 平均改善量 [dB]	
	計算機シミュレーション	実環境 (Mean [SD])
/a/	4.85	3.14 [0.10]
/i/	7.79	5.13 [0.16]
/u/	10.14	6.61 [0.43]
/e/	4.15	3.33 [0.14]
/o/	8.08	4.03 [0.18]

表 4 LPC-SED 及び NDSR による雑音付加音声、雑音除去後音声の評価結果

Table 4 Results of evaluating noise-added and noise-reduced speech on LPC-SED or NDSR.

	LPC-SED [dB]	NDSR [dB]
before NR	11.15 (1.54)	-6.57 (0.63)
after NR (DS)	8.22 (1.27)	-3.33 (0.53)
after NR (proposed)	5.15 (0.68)	-2.30 (0.26)

中面方向から到来)と設定した．雑音は、125 Hz から 6 kHz までのランダム帯域雑音とし、到来時間差 8 pt (右 35°に相当)と設定した．これらを SNR が +10 dB となるように計算機上で加算し、表 2 に示す条件で雑音除去を行い、LPC-SED、NDSR による評価を行った．参考のため、マイクロホン配置が同じ場合の遅延和アレーによる雑音除去結果に対しても両評価尺度で評価を行った．雑音除去前 (before NR) と、遅延和アレー (after NR (DS)), 提案法 (after NR (proposed)) で雑音除去を行った信号の各々に対して計算した LPC-SED、NDSR の平均値 (標準偏差) を表 4 に示す．LPC-SED、NDSR は、両者ともそれらの値が 0 dB に近いほど、評価対象音声のひずみが小さいことを意味する．表 4 より、LPC-SED による評価結果は、NDSR による評価結果とほぼ同様の傾向を示すことを確認できる．

遅延和アレーと提案法との性能をより詳細に比較するため、遅延和アレーのマイクロホン数と NDSR との関係性を調査した．ATR 音声データベース [17] に収録されている音声に、125 Hz から 6 kHz までのランダム帯域雑音を SNR が +10 dB となるように計算機上で加算し、マイクロホン間隔を 6 cm と設定した遅延和アレーと提案法で雑音除去を行った結果を図 8 に示す．図 8 より、遅延和アレーで提案法と同等の雑音除去能力を得るためには、6 本のマイクロホンが必要であることがわかる．

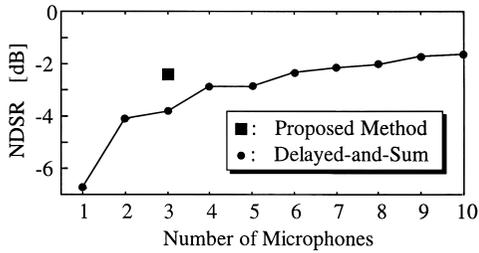


図 8 遅延和アレーと提案法との性能比較
Fig. 8 Performance comparison between delayed-and-sum array and proposed method.

3.3 残響の少ない実環境における雑音除去実験

3.3.1 実験データ

残響の少ない実環境における雑音除去実験には、計算機シミュレーションで使用した音声信号及び雑音信号のそれぞれをスピーカを用いて実音場に放射し、マイクロホンアレーで収録した信号を使用した。本実験においては、各母音に対し、同一環境において 10 回の収録を行い、それらに対して雑音除去を行った。雑音除去アルゴリズムのパラメータは、計算機シミュレーションの場合と同様、表 2 に示す値に設定した。ただし、実験機材の都合上、二つの突発性雑音の到来時間差は共に 8 pt (右 35°に相当)とした。実験に使用した音場は、4 m × 4.5 m × 2.4 m の防音室であるが、反射音の影響を抑えるために室側面に吸音材を設置した。残響時間は、500 Hz 以上の帯域においては約 50 ms、低周波数帯域に関しては 125 Hz で約 320 ms、250 Hz で約 100 ms であった。

3.3.2 実験結果

母音/a/に対し、雑音除去前後の LPC-SED と、雑音除去によるその改善量を図 9 に示す。受信信号の SNR は約 -0.6 dB で、音声と雑音が共存する区間における LPC-SED の平均改善量は、同一環境で収録した 10 個の信号に対して雑音除去を行った結果、平均が 3.14 dB であった。

次に、/a/以外の日本語母音に対する雑音除去結果を表 3 に示す。表 3 に示した実験結果は、10 個の収録信号に対して雑音除去を行った場合の LPC-SED の平均改善量の平均値 (Mean) と標準偏差 (SD) である。

3.4 考 察

従来の SS は、雑音を正確に推定できないこと、雑音除去後に新たなひずみが生じることが問題であった。雑音除去前後の音声のスペクトル包絡を示した図 3、

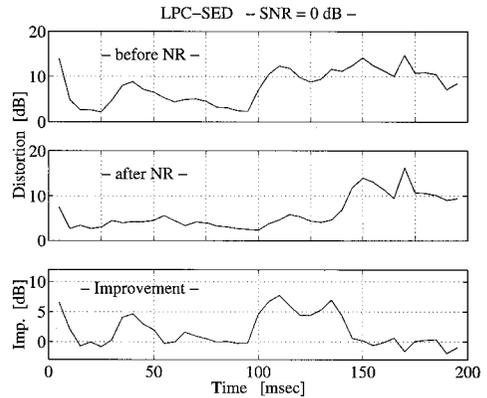


図 9 実環境における雑音除去結果 (上段: 受信信号の LPC-SED, 中段: 雑音除去後の LPC-SED, 下段: 雑音除去による LPC-SED の改善量)

Fig. 9 Results of noise reduction experiment in a real environment (Upper panel: LPC-SED before noise reduction, Mid panel: LPC-SED after noise reduction, Lower panel: Improvement of LPC-SED by noise reduction).

図 4 を比較すると、本手法は突発性雑音をほぼ除去できており、従来の SS の問題点の前者は解決できたといえる。なお、式 (22) (c) に相当する帯域では、雑音スペクトルを正確に推定できないが、これは音声帯域のうちたかだか数パーセントにすぎない。ところが、雑音除去後の音声のスペクトル包絡には、雑音が存在していた帯域を中心に若干のひずみが生じている。これは、SS の本質的な問題点であり、今後の更なる検討が必要である。

本手法により雑音除去を行った結果、計算機シミュレーション (図 5, 図 7, 表 3)、残響の少ない実環境における実験 (図 9, 表 3) 共に LPC-SED が改善されている。図 9 最上段の雑音除去前音声に関しては、雑音を付加した区間以外でも LPC-SED が大きくなっている。この原因は、音声の立上り部以前は暗騒音の影響で、音声の立下り部付近は暗騒音と残響の影響であると考えられる。図 9 の雑音除去後音声の LPC-SED (中段)、雑音除去による LPC-SED の改善量 (最下段) を見ると暗騒音は除去できているが、残響は除去できていないことがわかる。残響除去は今後の課題である。表 3 における母音による雑音除去性能の差は、音声のホルマント周波数と二つの雑音の中心周波数の関係に起因するものと考えられる。また、SNR が 0 dB の場合の信号到来時間差推定結果 (図 6)

を見ると、雑音の到来時間差に若干の推定誤りが生じている。しかし、信号の到来時間差が既知、それを自動推定した場合の雑音除去性能を比較した図 7 より、今回の実験条件においては、信号の到来時間差推定誤差が小さい場合にはそれが雑音除去結果へ及ぼす影響は非常に小さいことがわかった。

4. む す び

本論文では、減算形マイクロホンアレーにより雑音を推定し、SS により受信信号より雑音を除去する雑音除去法を提案した。

本手法は、雑音の推定に関して解析的手法を用いることにより、演算量が少なく、適応形アレーと比較して雑音の経時変化に対する雑音除去能力低下が生じにくいという特徴をもつ。更に、時々刻々雑音スペクトルの推定が可能であるため、雑音に関する過去の情報を一切必要としない。これにより、従来の SS では不可能であった突発性雑音の除去も可能になった。本論文では、LPC-SED の平均改善量により評価を行った。雑音除去実験の結果、突発性雑音に対する本雑音除去法の有効性を確認できた。

今後は、本手法の問題点である SS により生じるひずみへの対策について検討する予定である。また、より実環境に近い環境を想定し、雑音源が単一ではなく複数存在する場合、残響のある環境下での雑音・残響除去法について検討していきたい。

謝辞 本研究の一部は、科学技術振興財団による戦略的基礎研究推進事業 CREST、並びに文部省科学研究費補助金 (No.10680374) による援助を受けて行われた。

文 献

- [1] J.C. Junqua and J.P. Haton, "Robustness in automatic speech recognition," Kluwer Academic Publishers, Boston, 1996.
- [2] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. ASSP, vol.27, no.2, pp.113-120, April 1979.
- [3] 水町光徳, 赤木正人, "雑音除去を前提とした音声及び雑音の到来方向推定法," 音響学会講論集, 2-5-16, March 1998.
- [4] 大賀寿郎, 山崎芳男, 金田 豊, "音響システムとデジタル処理," 電子情報通信学会, 1995.
- [5] C.H. Knapp and G.C. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. ASSP, vol.24, no.4, pp.320-327, Aug. 1976.
- [6] M. Akagi and M. Mizumachi, "Noise reduction by paired microphones," Proc. 5th European Conf.

on Speech Comm. and Tech., Rhodes, Greece, no.T1D.3, pp.335-338, Sept. 1997.

- [7] 水町光徳, 赤木正人, "複数マイクロホンを用いたスペクトルサブトラクションによる雑音除去法," 音響学会講論集, 2-4-7, Sept. 1997.
- [8] L.J. Griffiths and C.W. Jim, "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. AP, vol.30, no.1, pp.27-34, Jan. 1982.
- [9] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," IEEE Trans. ASSP, vol.34, no.6, pp.1391-1400, Dec. 1986.
- [10] 永田仁史, 坪井宏之, "2ch ビームフォーマによる雑音抑圧処理に関する検討," 音響学会講論集, 2-P-19, March 1997.
- [11] Y. Nagata and H. Tsuboi, "A two-channel adaptive microphone array with target tracking," Proc. 5th European Conf. on Speech Comm. and Tech., Rhodes, Greece, no.T1D.5, pp.343-346, Sept. 1997.
- [12] 谷口賢一, 津村尚志, 福留公利, "スペクトルサブトラクション法における雑音推定方式," 音響学会講論集, 1-Q-17, Oct. 1994.
- [13] 金 学胤, 浅野 太, 鈴木陽一, 曾根敏夫, "短時間振幅スペクトル推定を用いた 2 チャンネル音声強調法における振幅スペクトル推定について," 音響学会講論集, 1-6-24, Sept. 1995.
- [14] 水町光徳, 赤木正人, "複数マイクロホンを用いたスペクトルサブトラクションによる雑音除去法," 信学技報, SP 97-35, July 1997.
- [15] M. Mizumachi and M. Akagi, "Noise reduction by paired-microphones using spectral subtraction," Proc. 23th Intl. Conf. on Acoustics, Speech, and Signal Processing, Seattle, USA, no.SP30.3, May 1998.
- [16] 古井貞熙, "デジタル音声処理," 東海大学出版会, pp.154-158, 1985.
- [17] 武田一哉, 勾坂芳典, 片桐 滋, 阿部匡伸, 桑原尚夫, "研究用日本語音声データベース利用解説書," ATR Tech. Rep. TR-I-0028, 1988.
- [18] 井上雅晶, 山田武志, 中村 哲, 鹿野清宏, "種々のマイクロホンアレーによる音声認識の検討," 信学技報, SP 96-89, Jan. 1997.

(平成 10 年 5 月 11 日受付, 10 月 9 日再受付)



水町 光徳 (学生員)

平 7 九州芸工大・芸術工・音響設計卒。平 9 北陸先端科学技術大学院大学情報科学研究科博士前期課程了。現在、同大学博士後期課程在学中。小規模マイクロホンアレーを用いた雑音・残響除去法に関する研究に従事。日本音響学会会員。



赤木 正人 (正員)

昭54名工大・工・電子卒。昭59東工大大学院博士後期課程情報工学専攻了。工博。同年電電公社(現NTT)研究所入社。以来、ATR視聴覚機構研究所、NTT基礎研究所を経て、現在、北陸先端科学技術大学院大学情報科学研究科助教授。この間、昭63米国MIT客員研究員、平5オランダアムステルダム大学客員研究員。音声信号処理、聴覚機構のモデル化の研究に従事。昭62年度本会論文賞受賞。平9年度日本音響学会論文賞受賞。IEEE, ASA, ESCA, 日本音響学会各会員。