

## 雑音抑圧法とモデル適応法を組み合わせた 耐雑音音声認識手法

古明地 秀治<sup>†1</sup> 荒川 隆行<sup>†1</sup> 越 仲 孝文<sup>†1</sup>

音声認識のための従来の耐雑音方式には、主に雑音抑圧法とモデル適応法の二つの方式がある。前者は、入力信号から推定雑音を抑圧することでクリーンな信号を得る方式である。しかし、抑圧の効果は雑音の推定精度に大きく依存する。一方、後者は、音響モデルの適応により雑音の分散も考慮するため、雑音の推定誤差に対して頑健である。しかし、雑音環境の変動に対する追従性は高くない。本稿では、抑圧係数と適応係数の二つの係数を導入し、従来の二つの方式の連続的な切り替えを実現する方法を提案する。Aurora2を用いた実験を通して、従来の二つの方式よりも高い認識性能を得る抑圧係数と適応係数が存在することを明らかにし、提案法の有効性を示す。

### A Hybrid method of Noise Suppression and Model Adaptation for Robust Speech Recognition

SHUJI KOMIJI,<sup>†1</sup> TAKAYUKI ARAKAWA,<sup>†1</sup>  
and TAKAFUMI KOSHINAKA<sup>†1</sup>

Noise suppression and model adaptation are two major approaches for robust speech recognition under noisy conditions. The former obtains clean speech by eliminating successively estimated noise from input speech, and the effectiveness of this approach strongly depends on the accuracy of noise estimation. The latter, on the other hand, is more robust against noise estimation error because it takes into account the variance of noise signals on an acoustic model. The calculation of variance, however, requires a large number of noise observations. This paper proposes a novel speech recognition method that continuously combines those two approaches using two weight coefficients, i.e., suppression coefficient and adaptation coefficient. A series of experiments on a speech recognition task under noisy conditions (Aurora2) shows that the proposed method is possible to achieve better performance than both conventional noise suppression and model adaptation do.

#### 1. はじめに

雑音環境下での音声認識について、これまで多くの研究がなされてきた。これらの研究では、認識対象である入力信号と、認識器、特に音響モデルとの環境のミスマッチをいかに低減するかという問題が扱われてきた。入力信号と音響モデルとのミスマッチを低減する方法は、大きく二つの方式に分類される。一つは入力信号を加工して音響モデルに近づける雑音抑圧法、もう一つは音響モデルを加工して入力信号に近づけるモデル適応法である。

雑音抑圧法として、例えば、スペクトル減算法(以下SS法)やウィナーフィルタ法(以下WF法)が知られている<sup>1)</sup>。これらの方式は、逐次推定された雑音を入力信号から除くことで雑音を抑圧し、入力信号を音響モデルに近づけることができる。この方式の利点は、雑音の抑圧に逐次推定された雑音を用いる点にある。すなわち、雑音環境に応じて推定された雑音を用いて、雑音抑圧することが可能である。しかし、短時間に激しく変化するバブル雑音などでは、雑音の推定が難しく、有効性が薄れる。

一方、モデル適応法として、例えば、HMM合成法や、ヤコビ法、Vector Taylor Series(以下VTS法<sup>2)</sup>)が知られている。これらの方法では、無音区間などで観測された雑音の平均や分散の統計量を音響モデルに反映する適応化を行い、音響モデルを入力信号の分布に近づけることができる。モデル適応法の利点は、雑音の分散を考慮しているため、雑音推定が難しい雑音(短時間の変化が激しいバブル雑音など)に対する頑健性が高いことである。一方で、雑音の平均や分散の計算には比較的長い時間の無音区間が必要であるため、雑音環境の変動に対する追従性は高くない。雑音推定が正しく動作する場合は、雑音の逐次推定に基づく雑音抑圧法の方が有利である。

以上より、雑音環境に応じて、雑音抑圧法とモデル適応法を連続的に切り替えることで、従来の二つの方式の短所が現れる状況を回避する音声認識システムが実現できると考えられる。二者択一的な切り替えではなく連続的な切り替え(重み付け)を採用することで、例えば、雑音推定の難易度が中程度の雑音環境下では、従来方式の単独実装よりも高い認識精度が期待できる。著者らは雑音抑圧法とモデル適応法にそれぞれの寄与度を規定する係数、すなわち抑圧係数と適応係数を各々導入することで、雑音抑圧法とモデル適応法の連続的な切り替えを実現する方法を提案する。また、Aurora2<sup>3)</sup>を用いた実験を通して、従来の二つ

<sup>†1</sup> 日本電気株式会社 情報・メディアプロセッシング研究所  
NEC Information and Media Processing Laboratories

の方式よりも高い認識性能を得る抑圧係数と適応係数が存在することを明らかにし、提案法の有効性を示す。

## 2. 雑音抑圧法とモデル適応法

従来の二つの耐雑音方式である雑音抑圧法とモデル適応法の説明を行う。ここでは特に、本稿で提案法の評価実験を行う上で雑音抑圧法として用いた WF 法と、モデル適応法として用いた VTS 法についてそれぞれ述べる。

### 2.1 雑音抑圧法 (WF 法<sup>1)</sup>)

一般的な雑音抑圧法は、入力信号に含まれる雑音を逐次的に推定し、入力信号から推定雑音を抑圧し、入力信号をクリーン信号で学習した音響モデルに近づける。このような手法には、SS 法や WF 法などがあるが、以下では特に、WF 法について述べる。

時刻  $t$  にて、WF 法で抑圧されたケプストラム領域での抑圧信号  $\hat{x}_t$  は以下の式で計算される：

$$\hat{x}_t = y_t - \hat{g}_t(y_0^t, \hat{n}_0^t). \quad (1)$$

ここで、 $y_t$  はケプストラム領域における入力信号を示す。また  $\hat{g}_t(y_0^t, \hat{n}_0^t)$  はケプストラム領域における Wiener ゲインを示す。Wiener ゲイン  $\hat{g}_t(y_0^t, \hat{n}_0^t)$  は時刻  $t$  までの入力信号  $y_0^t$  と入力信号から推定された推定雑音  $\hat{n}_0^t$  で計算される。なお、雑音抑圧に用いる推定雑音  $\hat{n}_t$  の推定方法として、例えば WiNE 法<sup>4)</sup>がある。WiNE 法は、SNR の推定値に応じて重みづけした観測信号を用いて、継続的に雑音推定値を更新する。

WF 法の特徴は、雑音環境に応じて逐次的に計算された推定雑音を用いて、雑音抑圧することである。従って、雑音推定精度が高い場合は、クリーン信号に近い抑圧信号を得ることができるため高い耐雑音性能を発揮することができる。しかし、雑音推定精度が低い場合は、耐雑音性能を低下させてしまう。ここで雑音推定の難易は雑音の性質に大きく依存し、例えば、短時間で激しく変動するパブル雑音の推定は難しくなる。結果として、WF 法の耐雑音性能は、対象とする雑音推定の難易に大きく依存する。なお、この性質は、WF 法に限らず、逐次的に計算された推定雑音を利用する雑音抑圧法全般に当てはまる。

### 2.2 モデル適応法 (VTS 法<sup>2)</sup>)

一般的なモデル適応法は、無音区間などで観測された雑音の平均や分散の統計量を音響モデルに反映する適応化を行い、クリーン信号で学習した音響モデルを入力信号の分布に近づける。このような手法には、HMM 合成法やヤコビ法、VTS 法などがあるが、以下では特に、VTS 法について述べる。

VTS 法では、ケプストラム領域におけるクリーン信号  $x$ 、雑音  $n$ 、観測信号  $y$  の間の非線形の関係：

$$\begin{aligned} y &= x + D \log[1 + \exp\{D^+(n - x)\}] \\ &= x + g(x, n) \end{aligned}$$

をある点  $(x_0, n_0)$  を中心に線形近似する：

$$y \approx x_0 + g(x_0, n_0) + (I - G(x_0, n_0))(x - x_0) + G(x_0, n_0)(n - n_0). \quad (2)$$

これにより、線形演算のみで雑音環境下の信号で学習した音響モデルを模した音響モデルを作ることができる。なお、 $D$  は DCT 行列を示し、 $D^+$  はその一般逆行列を示す。また、 $I$  は単位行列を示し、 $G(x_0, n_0)$  は、 $x = x_0, n = n_0$  における  $n$  に関する  $g(x, n)$  のヤコビアンを示す。

クリーン学習した音響モデルの HMM における第  $i$  状態、第  $j$  ガウシアン平均を  $\mu_{x,ij}, \mu_{\Delta x,ij}, \mu_{\Delta\Delta x,ij}$  とおき、分散を  $\Sigma_{x,ij}, \Sigma_{\Delta x,ij}, \Sigma_{\Delta\Delta x,ij}$  とする。なお、ガウシアン平均と分散には、ケプストラム  $x$  の平均と分散だけでなく、一次動的ケプストラム  $\Delta x$ 、および二次動的ケプストラム  $\Delta\Delta x$  の平均と分散を含めている。このとき、式 (2) の下で (ここでは  $x_0 = \mu_{x,ij}, n_0 = \mu_n$  とする)、雑音に適応した HMM における第  $i$  状態、第  $j$  ガウシアン平均は、以下のように表わされる：

$$\mu_{y,ij} = \mu_{x,ij} + g(\mu_{x,ij}, \mu_n), \quad (3)$$

$$\mu_{\Delta y,ij} = (I - G_{ij})\mu_{\Delta x,ij}, \quad (4)$$

$$\mu_{\Delta\Delta y,ij} = (I - G_{ij})\mu_{\Delta\Delta x,ij}. \quad (5)$$

また分散は以下のように表わされる：

$$\Sigma_{y,ij} = (I - G_{ij})\Sigma_{x,ij}(I - G_{ij})^T + G_{ij}\Sigma_n G_{ij}^T, \quad (6)$$

$$\Sigma_{\Delta y,ij} = (I - G_{ij})\Sigma_{\Delta x,ij}(I - G_{ij})^T + G_{ij}\Sigma_{\Delta n} G_{ij}^T, \quad (7)$$

$$\Sigma_{\Delta\Delta y,ij} = (I - G_{ij})\Sigma_{\Delta\Delta x,ij}(I - G_{ij})^T + G_{ij}\Sigma_{\Delta\Delta n} G_{ij}^T. \quad (8)$$

ここで、 $\mu_n$  は雑音  $n_0^t$  の平均であり、 $\Sigma_n, \Sigma_{\Delta n}, \Sigma_{\Delta\Delta n}$  は雑音  $n_0^t$  およびその動的成分の分散である。また、 $G_{ij} = G(\mu_{x,ij}, \mu_n)$  とした。

以上、VTS 法による音響モデルのパラメータの推定式を示した。VTS 法では、比較的長い時間の雑音信号から計算した雑音の平均や分散を用いることで、逐次的な推定が困難な雑音 (短時間の変化が激しい雑音) に対しても、安定した耐雑音性能を示す。しかし、雑音の逐次的な推定値を用いないため、雑音推定が正しく動作する場合は、雑音の逐次推定に基づく雑音抑圧法の方が有利である。なお、この性質は、VTS 法に限らずモデル適応法全般に当てはまる。

### 3. 提案法

提案法では、雑音抑圧法とモデル適応法にそれぞれの寄与度を規定する係数、すなわち抑圧係数と適応係数を導入し、両者を同時に実行することで、雑音抑圧法とモデル適応法の連続的な切り替えを実現する。まず、抑圧係数を導入した WF 法と適応係数を導入した VTS 法を述べる。

#### 3.1 抑圧係数を導入した WF 法

雑音抑圧法において、寄与度を規定する抑圧係数  $\alpha \in [0, 1]$  を導入する。例えば、前節で説明した WF 法 (1) に抑圧係数  $\alpha$  を適用すると、抑圧係数  $\alpha$  に従う抑圧信号  $\hat{z}_t(\alpha)$  は以下の式で表わされる：

$$\hat{z}_t(\alpha) = y_t - \alpha \hat{g}_t(y_0^t, \hat{n}_0^t). \quad (9)$$

抑圧係数を導入した WF 法は、 $\alpha = 0$  では雑音抑圧を全くしない場合と等価になり、 $\alpha = 1$  で従来の WF 法と等価になる。

#### 3.2 適応係数を導入した VTS 法

モデル適応法において、寄与度を規定する適応係数  $\beta \in [0, 1]$  を導入する。例えば、前節で説明した VTS 法 (3)–(8) に適応係数  $\beta$  を適用した場合、雑音に適應した HMM における第  $i$  状態、第  $j$  ガウシアン平均は次式で与えられる：

$$\mu_{z,ij}(\beta) = \mu_{x,ij} + \beta g(\mu_{x,ij}, \mu_n), \quad (10)$$

$$\mu_{\Delta z,ij}(\beta) = (I - \beta G_{ij}) \mu_{\Delta x,ij}, \quad (11)$$

$$\mu_{\Delta \Delta z,ij}(\beta) = (I - \beta G_{ij}) \mu_{\Delta \Delta x,ij}. \quad (12)$$

また、分散は次式で与えられる：

$$\Sigma_{z,ij}(\beta) = (I - \beta G_{ij}) \Sigma_{x,ij} (I - \beta G_{ij})^T + \beta^2 G_{ij} \Sigma_n G_{ij}^T, \quad (13)$$

$$\Sigma_{\Delta z,ij}(\beta) = (I - \beta G_{ij}) \Sigma_{\Delta x,ij} (I - \beta G_{ij})^T + \beta^2 G_{ij} \Sigma_{\Delta n} G_{ij}^T, \quad (14)$$

$$\Sigma_{\Delta \Delta z,ij}(\beta) = (I - \beta G_{ij}) \Sigma_{\Delta \Delta x,ij} (I - \beta G_{ij})^T + \beta^2 G_{ij} \Sigma_{\Delta \Delta n} G_{ij}^T. \quad (15)$$

抑圧係数を導入した VTS 法は、 $\beta = 0$  では、モデル適応を全くしない場合と等価になり、 $\beta = 1$  で従来の VTS 法と等価になる。

#### 3.3 雑音抑圧法とモデル適応法の組み合わせ

提案法では、雑音環境などに応じて決まる抑圧係数  $\alpha$  と適応係数  $\beta$  の下で、雑音抑圧法とモデル適応法を組み合わせる音声認識を行う。ここで、雑音抑圧法とモデル適応法の寄与度は独立ではなく、両者が相補的に働くことが望ましいと考えられる。提案法では、 $\alpha$  と  $\beta$  の間に次の関係を仮定する：

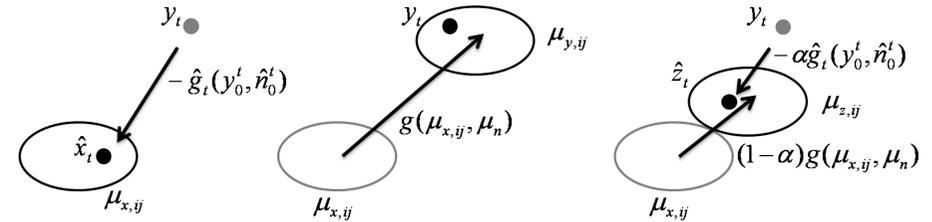


図 1 各耐雑音手法のデータとモデルの動き方  
左：雑音抑圧法，中：モデル適応法，右：提案法

$$\alpha + \beta = 1. \quad (16)$$

上式は、抑圧係数  $\alpha$  を導入した WF 法 (9) をあるガウシアン平均の周りに展開、線形近似し、期待値をとった式と、適応係数を導入した VTS 法 (10) とのアナロジーから導出される。上述の関係式を仮定することにより、独立なパラメータは抑圧係数  $\alpha$  のみとなる。従来法の雑音抑圧法と、モデル適応法、提案法のデータとモデルの動き方の違いの概略を図 1 に示す。提案法では、従来の WF 法、VTS 法にはないパラメータ  $\alpha$  を導入したことにより、データとモデルをより効果的に近づけることができると考えられる。

次節では、従来の二つの方式よりも高い認識性能が得られるような抑圧係数および適応係数の存在を検証する実験を行い、提案法の有効性を調べる。

## 4. 実験

提案法の有効性を調べるために、 $\alpha + \beta = 1$  の条件の下で、抑圧係数  $\alpha$  を 0 から 1 の間で段階的に変化させ、それぞれの場合における認識率を調べた。

### 4.1 実験方法

実験では、抑圧係数  $\alpha$  を導入した WF 法 (9) で得られる抑圧信号を認識対象とする。そして、適応係数  $\beta$  を導入した VTS 法については、

- 音響モデルのガウシアン平均のみに適用し (式 (10)–(12))、分散は雑音適応しない
- 音響モデルのガウシアン平均と分散に適用する (式 (10)–(15))

の 2 通りを考え、それぞれ性能を評価する。 $\alpha$  には、0, 0, 0.2, 0.5, 0.8, 1.0 の 5 つの値を与えた。

### 4.2 実験環境

実験は、Aurora2<sup>3)</sup> による連続数字認識による。Aurora2 のデータには SN 比が -5[dB] が

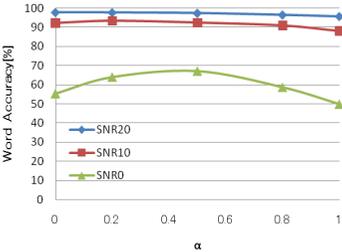


図2 ガウシアン平均パラメータのみモデル適応した場合の抑圧係数  $\alpha$  と認識率の関係。左: subway 雑音, 右: restaurant 雑音

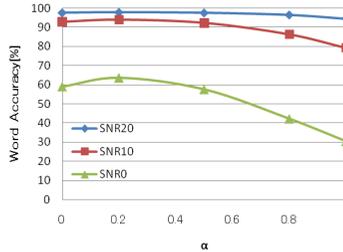


図3 ガウシアン平均と分散をモデル適応した場合の抑圧係数  $\alpha$  と認識率の関係。左: subway 雑音, 右: restaurant 雑音

ら 20[dB] までの 8 種類の雑音のいずれかが付加されている。各 SN 比, 各雑音に対して, 男性話者 50 人, 女性話者 50 人, 一人につき約 10 発声の, 計 1000 発声のテストデータがある。本稿では, SN 比を 0[dB], 10[dB], 20[dB] の 3 パターンとし, 8 種類全ての雑音に対して, 提案法の効果を調べた。8 種類の雑音の内訳は, subway, babble, car, exhibition, restaurant, street, airport, train-station である。

音響モデルは, Aurora2 のクリーン学習データにより HTK<sup>5)</sup> を用いて作成した。音響モデルである HMM を left-to-right 型として一つの数字毎に作成した。一つの数字に対して 16 状態とし, 一状態のガウシアン数を 3 とした。なお, 無音モデルに関しては, 3 状態とし, 一状態のガウシアン数を 6 とした。特徴量は,  $C_0$  を含んだ 13 次元 MFCC とし, それに一次の動的成分と二次の動的成分も加え, 39 次元のベクトルとした。

WF 法で用いる雑音は WiNE 法<sup>4)</sup> により逐次推定し, VTS 法で用いる雑音は, 無音区間から計算した。なお, モデル適応は約 10 発声毎に行った。

4.3 平均のみをモデル適応した場合

subway 雑音, restaurant 雑音に関する結果を図 2 に示す。そして, 全ての雑音に関する結果を表 1 に示す。図 2 によると, subway 雑音, restaurant 雑音の双方の場合において, 認識率 (W.A.) は概ね上に凸の曲線を描いた。すなわち, 提案法により, 従来の雑音抑圧法, モデル適応法単体よりも高い認識率が得られた。表 1 からほとんどの場合においても従来手法よりも認識率を向上させる  $\alpha$  の存在を確認した。

4.4 平均と分散をモデル適応した場合

subway 雑音, restaurant 雑音に関する結果を図 3 に示す。図 3 によると, subway 雑音, restaurant 雑音の双方の多くの場合において, 認識率は上に凸の曲線とはならなかった。図

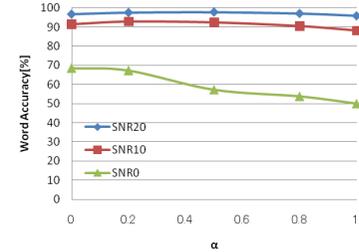


図4 ガウシアン平均のみをモデル適応した場合の抑圧係数  $\alpha$  と認識率の関係。左: subway 雑音, 右: restaurant 雑音

3 を図 2 と比較しても, ガウシアン分散もモデル適応することによって, 認識率が低下してしまう場合が多い。従って, 式 (13)–(15) のガウシアン分散の適応式は適切ではないと考えられる。

この原因は, 式 (13)–(15) と, 抑圧係数  $\alpha$  を導入した WF 法によって得られた信号の分

表 1 ガウシアン平均のみをモデル適応した場合の提案法の認識結果。 $\alpha = 0.0$  のとき,  $\alpha = 1.0$  のときと最大認識率を得られた  $\alpha$  のときの認識率のみを示している (括弧内にその時の  $\alpha$  の値を示している)。

		subway	babble	car	exhibition
SNR20	VTS 法 ( $\alpha = 0.0$ )	97.79	97.49	97.58	97.32
	WF 法 ( $\alpha = 1.0$ )	95.70	95.28	95.79	94.34
	提案法	97.79(0.0)	<b>97.76(0.2)</b>	<b>97.91(0.2)</b>	<b>97.38(0.2)</b>
SNR10	VTS 法 ( $\alpha = 0.0$ )	92.20	92.20	90.93	90.03
	WF 法 ( $\alpha = 1.0$ )	88.06	81.29	86.31	84.48
	提案法	<b>93.31(0.2)</b>	<b>93.20(0.2)</b>	<b>93.44(0.2)</b>	<b>92.13(0.2)</b>
SNR0	VTS 法 ( $\alpha = 0.0$ )	55.45	50.85	36.74	52.18
	WF 法 ( $\alpha = 1.0$ )	49.92	30.99	32.09	38.85
	提案法	<b>67.21(0.5)</b>	<b>58.40(0.5)</b>	<b>55.56(0.5)</b>	<b>61.25(0.5)</b>
		restaurant	street	airport	train-station
SNR20	VTS 法 ( $\alpha = 0.0$ )	97.57	97.13	98.09	98.40
	WF 法 ( $\alpha = 1.0$ )	94.54	95.01	95.05	94.79
	提案法	<b>97.79(0.2)</b>	<b>97.22(0.2)</b>	<b>98.18(0.2)</b>	<b>98.40(0.0)</b>
SNR10	VTS 法 ( $\alpha = 0.0$ )	92.78	90.21	93.89	92.63
	WF 法 ( $\alpha = 1.0$ )	79.24	84.52	80.85	83.06
	提案法	<b>93.86(0.2)</b>	<b>91.84(0.2)</b>	<b>94.90(0.2)</b>	<b>93.83(0.2)</b>
SNR0	VTS 法 ( $\alpha = 0.0$ )	59.01	46.13	56.78	45.73
	WF 法 ( $\alpha = 1.0$ )	30.43	37.61	33.88	31.47
	提案法	<b>63.65(0.2)</b>	<b>58.95(0.5)</b>	<b>63.61(0.5)</b>	<b>57.14(0.5)</b>

散とのずれであると考えられる．元来，雑音抑圧法で得られた信号には抑圧誤差  $\epsilon$  が含まれる．そのため抑圧誤差  $\epsilon$  の分散  $\Sigma_\epsilon$  をガウシアン分散に足すことで，抑圧誤差を補償するという研究が，Deng<sup>6)</sup> やデルクロア<sup>7)</sup> らによってなされている．これらの研究では，抑圧誤差の分散  $\Sigma_\epsilon$  を考慮しないときと比較して認識率の改善がみられている．著者らは，式(13)–(15)のガウシアン分散と抑圧信号の分散との不一致は，この抑圧誤差  $\epsilon$  による影響であるという仮説を設け，式(13)–(15)を以下のように書き換えた：

$$\Sigma_{z,ij}(\alpha, \beta) = (I - \beta G_{ij})\Sigma_{x,ij}(I - \beta G_{ij})^T + \beta^2 G_{ij}\Sigma_n G_{ij}^T + \alpha^2 \Sigma_\epsilon, \quad (17)$$

$$\Sigma_{\Delta z,ij}(\alpha, \beta) = (I - \beta G_{ij})\Sigma_{\Delta x,ij}(I - \beta G_{ij})^T + \beta^2 G_{ij}\Sigma_{\Delta n} G_{ij}^T + \alpha^2 \Sigma_{\Delta\epsilon}, \quad (18)$$

$$\Sigma_{\Delta\Delta z,ij}(\alpha, \beta) = (I - \beta G_{ij})\Sigma_{\Delta\Delta x,ij}(I - \beta G_{ij})^T + \beta^2 G_{ij}\Sigma_{\Delta\Delta n} G_{ij}^T + \alpha^2 \Sigma_{\Delta\Delta\epsilon}. \quad (19)$$

なお， $\Delta\epsilon$  や  $\Delta\Delta\epsilon$  はそれぞれ抑圧誤差の一次の動的成分，二次の静的成分を示す．また，抑圧誤差分散には，抑圧係数  $\alpha^2$  がかかる．式(17)–(19)の正当性を確認するために，以下の追加実験を行った．

#### 4.5 追加実験（雑音抑圧誤差分散の付加）

ここでは，抑圧誤差分散の必要性を実験的に検証するために， $\Sigma_\epsilon$ ， $\Sigma_{\Delta\epsilon}$ ， $\Sigma_{\Delta\Delta\epsilon}$  を，Aurora2のクリーン信号と抑圧信号の差の分散として，10発声毎に計算した．実際応用上は，クリーン信号は未知であるため，ある種の理想的な実験となる．

subway 雑音，restaurant 雑音に関する結果を図4に示す．そして，全ての雑音に関する結果を表2に示す．図4によると，subway 雑音では，各SNRのほとんどの場合で，認識率は上に凸の曲線を描いた．一方，restaurant 雑音では，ほとんどの場合において（上に凸ではあるが）局大点の存在しない曲線となった．図4を図3と比較すると， $0.0 < \alpha < 0.8$  で抑圧誤差分散による効果が現れた．しかし， $0.8 \leq \alpha$  では，抑圧誤差分散が認識率に悪影響を与えてしまっている．これは，すべてのガウシアン分散を一律に広げてしまっていることによる．すなわち，本来広げる必要のないガウシアン分散についても広げて過ぎてしまい，識別誤りを大きくしてしまったと考えられる．

表2に関しては，subway，car，exhibition，street，train-station に関して，従来法よりも認識率を高める  $\alpha$  の存在を確認した．一方，babble，restaurant，airport に関しては，VTS法による認識率が最も高くなった．なお，babble，restaurant，airport は，バブル雑音を多く含む雑音であり，雑音推定の難易度が高いと考えられる．したがって，逐次推定された雑音はむしろ音声認識に悪影響を与え，それゆえ， $\alpha = 0.0$  のときが一番良い認識率が得られたと考えられる．

この実験により， $0.0 < \alpha < 0.8$  において，ガウシアン分散に抑圧誤差分散を加えるこ

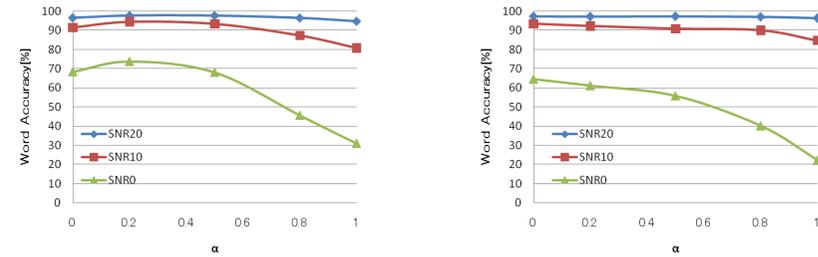


図4 抑圧誤差分散を付加した場合の抑圧係数  $\alpha$  と認識率の関係．左：subway 雑音，右：restaurant 雑音

とで，抑圧係数  $\alpha$  を導入した WF 法による抑圧信号の分散に近づけることができたと考えられる．ただし，今回は，実際応用上は未知であるクリーン信号を用いて抑圧誤差分散を計算している．雑音が重畳した入力信号からの抑圧誤差分散の計算は，今後の課題であるが，例えば，デルクロアらによる，“抑圧誤差分散は観測信号と抑圧信号の差の分散に比例する”という仮説<sup>7)</sup>の下で抑圧誤差分散を推定する方法などが考えられる．

## 5. おわりに

本稿では，雑音抑圧法とモデル適応法にそれぞれ寄与度を規定する係数，すなわち抑圧係数と適応係数を導入し，両者を同時に実行することで，雑音抑圧法とモデル適応法の連続的な切り替えを実現する方法を提案した．実験では，Aurora2の8種類の雑音を用いて，雑音抑圧法にWF法を，モデル適応法にVTS法を使用し，提案法を評価した．

適応係数を導入したVTSにおいて，音響モデルの平均のみをモデル適応をした場合，いずれの従来手法よりも高い認識率が得られる抑圧係数，適応係数が存在し，提案法の有効性が確認できた．一方，音響モデルの平均と分散の両方をモデル適応した場合には，従来手法よりも高い認識率は得られなかった．これは，抑圧信号の分散と雑音適応したモデルのガウシアン分散の不一致から生じると考えられる．そこで，適応後のモデルに不一致を補償するための抑圧誤差分散を付加することにより，認識率を高め，従来法を上回る抑圧係数，適応係数の存在が確認できた．

今後，抑圧誤差分散の推定方法や，尤度最大化基準による抑圧係数，適応係数の自動推定法について検討を行う予定である．

参 考 文 献

- 1) Hansen J.H.L., "Speech Enhancement," Encyclopedia of Electrical and Electronic Engineering, John Wiley and Sons inc., Publishers, 1999.
- 2) Moreno, P., et al., "A Vector Taylor Series Approach for Environment-Independent Speech Recognition," Proc. ICASSP, pp. 733-736, Atlanta, 1996.
- 3) H.G. Hirsch, et al., "The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions," Proc ISCA ITRW ASR, 2000.9
- 4) Kato, M., et al., "Noise Suppression with High Speech Quality Based on Weighted Noise Estimation and MMSE STSA", Proc. IWAENC2001, pp.183-186, 2001.
- 5) Young, S., et al., "HTK Book (for HTK Version 3.1)," 2006.
- 6) Deng, L., et al., "Dynamic compensation of HMM variances using the feature enhancement uncertainty computed from a parametric model of speech distortion," Proc IEEE Trans. SAP, vol. 2, no.2, pp.245-257, 2005.
- 7) デルクロア . 他 . '動的分散適応に基づく音声強調と音声認識の統合手法の提案 , ' 研究報告 - 音声言語情報処理 ( S L P ) , pp 55-60. 2007.

表 2 雑音抑圧誤差分散を付加した場合の提案法の認識結果 .  $\alpha = 0$  のとき ,  $\alpha = 1$  のときと最大認識率を得られた  $\alpha$  のときの認識率のみを示している (括弧内にその時の  $\alpha$  の値を示している) .

		subway	babble	car	exhibition
SNR20	VTS 法 ( $\alpha = 0.0$ )	96.59	97.28	97.11	95.56
	WF 法 ( $\alpha = 1.0$ )	94.81	95.62	95.74	94.88
	提案法	<b>97.73(0.2)</b>	<b>97.55(0.2)</b>	<b>98.36(0.2)</b>	<b>97.10(0.5)</b>
SNR10	VTS 法 ( $\alpha = 0.0$ )	91.46	93.74	89.98	86.55
	WF 法 ( $\alpha = 1.0$ )	80.72	83.74	80.38	76.64
	提案法	<b>94.38(0.2)</b>	93.74(0.0)	<b>95.23(0.2)</b>	<b>92.38(0.2)</b>
SNR0	VTS 法 ( $\alpha = 0.0$ )	68.41	61.28	66.12	60.60
	WF 法 ( $\alpha = 1.0$ )	30.92	18.20	11.69	14.96
	提案法	<b>73.78(0.2)</b>	61.28(0.0)	<b>72.79(0.2)</b>	<b>71.18(0.2)</b>
		restaurant	street	airport	train-station
SNR20	VTS 法 ( $\alpha = 0.0$ )	97.30	95.89	97.88	96.54
	WF 法 ( $\alpha = 1.0$ )	96.41	95.34	96.30	95.87
	提案法	97.30(0.5)	<b>96.89(0.5)</b>	97.88(0.0)	<b>96.88(0.5)</b>
SNR10	VTS 法 ( $\alpha = 0.0$ )	93.34	88.78	94.15	90.74
	WF 法 ( $\alpha = 1.0$ )	84.74	79.41	86.16	83.37
	提案法	93.34(0.0)	<b>92.29(0.2)</b>	94.15(0.0)	<b>91.14(0.5)</b>
SNR0	VTS 法 ( $\alpha = 0.0$ )	64.60	60.94	69.97	63.25
	WF 法 ( $\alpha = 1.0$ )	22.26	20.28	23.38	16.29
	提案法	64.60(0.0)	<b>66.29(0.2)</b>	69.97(0.0)	<b>65.07(0.2)</b>