2007－ＳＬＰ－69（５）
2007／12／20

音声GMMと雑音重み推定を用いた雑音除去

三宅　信之†　　滝口　哲也†　　有木　康雄†

† 神戸大学大学院工学研究科　〒657-8501 神戸市灘区六甲台町１－１
E-mail: †miyake@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

あらまし　本稿では突発性の雑音除去について述べる．突発性の雑音は短時間しか起こらないため推定しにくいが，音声に混入することで音声認識率が下がることは多い．以前，我々はそれらの突発性雑音の検出と識別手法を提案した．本稿ではより多くの雑音を識別できるように拡張し，さらに音声GMMを用いた雑音の除去手法についても述べる．Seguraらによって提案されたGMMを利用した雑音除去法は加法性の雑音を精度よく除去できる．本稿においてもこの手法を利用するが，我々のタスクにおいては，識別された段階ではSNR(信号対雑音比)は未知であるため，除去時に使用する雑音のデータと実際に重畳している雑音はミスマッチを引き起こすことが多い．そこで我々はSeguraらのGMMによる雑音除去法にこのSNRの推定を雑音の重みという形で加え除去を行う．重み推定にはGMMの尤度が最も大きくなるようにEMアルゴリズムを用いる方法，GMMの混合ごとに尤度が最大になるように決める方法の２つを用いる．実験結果より比較的低SNRである5 dB以下の雑音に対して認識率の改善が見られ，また重みの推定の効果も確認できた．

キーワード　突発性雑音除去，GMM, EMアルゴリズム，重み推定

Sudden Noise Reduction Based on GMM with Noise Power Estimation

Nobuyuki MIYAKE†, Tetsuya TAKIGUCHI†, and Yasuo ARIKI†

† Graduate School of Engineering, Kobe University　Nada, Kobe 657-8501, Japan
E-mail: †miyake@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

Abstract　This paper describes a method for reducing sudden noise using a noise detection and classification methods, and a noise power estimation. Sudden noise detection and classification have been dealt with in our previous study. In this paper, noise classification is improved to classify more kinds of noises based on k-means clustering, and GMM-based noise reduction, which was proposed by Segura et al, is performed using the detection and classification results. As a result of classification, we can know the kind of noise, but the power is unknown. In this paper, this problem is solved by combining an estimation of noise power with the noise reduction method. In our experiments, the proposed method achieved good performance for recognition of utterances overlapped by sudden noises.

Key words　Sudden noise reduction, GMM, EM algorithm, Noise power estimation

1. はじめに

音声認識技術を使用するとき、発話に雑音が重畳することで誤認識を引き起こすことが少なくない．そのためスペクトルサブトラクションを始めとした雑音を除去する研究が数多くなされている[1],[2]．近年では音声の特徴を混合ガウス分布(GMM)といったモデルで保持しておき，その情報を利用して除去する手法[3],[4]，マイクロホンアレーを利用して雑音を除去するといった手法などが多く見られる[5]．従来の雑音除去は基本的に雑音を推定し，その推定された雑音を雑音重畳音声から減算するという手順が多い．雑音の推定には発話直前の雑音のみの区間や，その情報を確率的に追跡していくものが用いられる．雑音は時間的に緩やかに変化するものだと考えると，発話付近の雑音の情報を使用することで雑音抑圧は高い効果が得られると期待できる．

例えば家の中のような実環境で音声認識を使用することを考えるとき，雑音にはドアの開閉音や電話の音など中には突然発生するものも少なくない．発話中に雑音が発生した場合，そのデータから雑音の情報のみを取り出すことは困難である．このような突発的に発生する雑音の除去に関する研究はいくつか存在する[6],[7]．しかしこれらの手法が対象としているもののほとんどは雑音重畳時間が非常に短いものであったり，比較的

長いものでも，対応している種類が少なかった．また，マイクロフォンの数を増やすことで除去することもできるが，これはハードウェアが複雑になるという問題点がある．

我々は以前この突発性雑音の検出と識別について提案を行った[8]．本稿ではこれを以前より多くの雑音を検出・識別できるように改良し，さらにその除去法を提案する．除去法にはSegura らが提案した手法を用いている[3]．この手法はクリーン音声の GMM を基にして雑音除去を行う手法である．加法性雑音に対して効果的であり，改良案も多い[9],[10]．しかしながらこの手法を利用するためには雑音の推定値が必要である．本稿では雑音の推定値は識別結果に応じて用意することができるが，SNR が未知であるため，雑音の強さを推定しなければ実際の雑音とのミスマッチを引き起こす可能性が高い．そこで強さの推定を GMM による雑音除去に組み込み使用する．重み推定には GMM の尤度が最も大きくなるように EM アルゴリズムを用いる方法，GMM の混合ごとに尤度が最大になるように決める方法の２つを用いる．

2. 雑音のクラスタリング

雑音には様々なものがあるが，本稿では RWCP 非音声データベース[11] のすべての雑音を取り扱うものとする．このデータベースには 105 種類の雑音がそれぞれ 100 データ入っている．この種類が非常に多く，そのままでは識別器の学習や識別そのものに非常に時間がかかる．そのためあらかじめクラスタリングしておくことで，似たような特徴を持つ雑音はひとつにまとめて考え，高速化したい．しかし多くの雑音をひとつにまとめてしまうと，除去時は雑音のデータとしてそのクラスの平均ベクトルを使用するために，雑音の特徴を捉えられず除去がうまく働かないと考えられる．そこでクラスタリングを用いながらツリーを構成しておくことで，高速な識別を行うことができ，雑音も細かく分類する．

2.1 クラスタリング手法

クラスタリング手法として k-means 法を使用する．k-means 法をそのまま適用すると，クラス数は手動で与えねばならないため，ツリーを構成するときに各ノードでどのような値を設定すればいいかわからない．そこでそれぞれのデータからクラスの中心までの距離の最大値を与えることで，クラス数を自動的に決定しながらクラスタリングを行うように改良する．

まずクラスの中心から距離の最大値 θ を手動で決める．k-means 法を用いてクラスタリングする．その後１つずつデータとデータが属すクラスの中心との距離 d を測り，距離が指定した値以上 $(d > \theta)$ ならばそのデータが属すクラスを２つに分割する．そして再び k-means 法を行う．すべてのデータとクラスの中心の距離が θ を下回るまでこれを繰り返す．この θ の値を小さく設定することで，クラス数を自動決定し，よく似た雑音だけが同じクラスに集まるようなクラスタリングができる．

2.2 ツ リ ー

上記の手法をそのまま適用すると，距離の最大値 θ の値を小さくしたときにクラス数が大きくなりすぎる．そこで上記のクラスタリング手法を用いて図１のようなツリーを形成する．

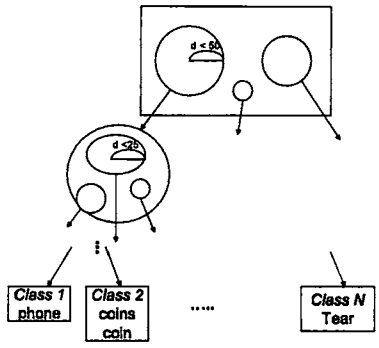

図１ 雑音の種類によるツリーの例
Fig. 1 An example of tree of noises

上段ではこの中心までの距離の最大値の値を大きくすることで荒くクラスタリングを行う．下段では，上段で分けられたクラスをより小さな θ を設定することでより細かく分ける．θ の値は何段目かによって決めておく．こうすることで各クラス内の距離が最終的に設定した値以下になるツリーを形成する．

3. 雑音の検出と識別

本稿で扱う雑音はそのほとんどが短時間しか継続せず，またいつ起こるかわからないものである．そのためまず除去を行う前に雑音が重畳しているかどうか判定を行い，またそれがどのような雑音であるか識別する必要がある[8]．

3.1 雑音の検出

雑音の検出には AdaBoost を用いる．AdaBoost は Boosting の一種であり，多数の弱識別器を使うことで，非線形な識別器を作成することができる[12]．同じように非線形な識別器として SVM などがあげられるが，それらの手法に比べると，パラメータの調整が少なく，また高速で動作するというメリットがある．学習する雑音データの雑音重畳音声を作成し，それらすべてとクリーン音声を用いて識別器の学習を行う．その時の学習アルゴリズムを図２に示す．ひとつの発話には複数のフレームが存在するため，フレームごとに特徴量が得られるのでその特徴量すべてを使用して学習する．次に，この識別器を用いてフレーム単位で雑音の検出を行う．具体的には特徴量を入力したときの式 (2) の正負でクリーン音声か雑音が重畳しているかの判定を行う．そしてこの識別器によって雑音が重畳していると判定されたフレームに対して雑音の識別が行われる．

3.2 雑音の識別

検出において雑音と判定されたフレームに対し雑音の識別を行う．雑音のみのデータであれば，あらかじめ用意したデータと距離を取ることでクラスを判定できるが，実際には音声に重畳した状態の雑音を識別しなければならない．このため単純に距離をとるといった手法は使うことができない．2.2 節で作成したツリーの各ノード毎に，AdaBoost を用いて識別器を作成し，上段から順に識別を行う．この識別器の作成は各クラスに属す雑音重畳音声を使用して行う．AdaBoost は二値判別の識

Input: n examples $Z = \{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_n, y_n)\}$

Initialize:

$$w_1(\mathbf{z}_i) = \begin{cases} \frac{1}{2m}, & \text{if} \quad y_i = 1 \\ \frac{1}{2l}, & \text{if} \quad y_i = -1 \end{cases}$$

where, m is the number of positive data, and l is the number of negative data.

Do for $t = 1, \cdots, T$,

(1) Train a base learner with respect to weighted example distribution w_t and obtain hypothesis $h_t : \mathbf{x} \mapsto \{-1, 1\}$

(2) Calculate the training error ϵ_t of h_t:

$$\epsilon_t = \sum_{i=1}^{n} w_t(\mathbf{z}_i) \frac{I(h_t(\mathbf{x}_i) \neq y_i) + 1}{2}.$$

(3) Set

$$\alpha_t = log \frac{1 - \epsilon_t}{\epsilon_t}$$

(4) Update example distribution w_t:

$$w_{t+1}(\mathbf{z}_i) = \frac{w_t(\mathbf{z}_i) \exp\{\alpha_t I(h_t(\mathbf{x}_i) \neq y_i)\}}{\sum_{j=1}^{n} w_t(\mathbf{z}_i) \exp\{\alpha_t I(h_t(\mathbf{x}_i) \neq y_i)\}}. \quad (1)$$

Output: final hypothesis:

$$f(\mathbf{x}) = \frac{1}{||\alpha||} \sum_t \alpha_t h_t(\mathbf{x}). \quad (2)$$

図 2 AdaBoost algorithm

Fig. 2 AdaBoost algorithm

Input: m examples $\{(\mathbf{x}_1, y_1), \cdots, (\mathbf{x}_m, y_m)\}$

$y_i = \{1, \cdots, K\}$

Do for $k = 1, \cdots, K$

1. Set labels

$$y_i^k = \begin{cases} +1, & \text{if} \quad y_i = k \\ -1, & \text{otherwise} \end{cases} \quad (3)$$

2. Learn k-th classifier $f^k(\mathbf{x})$ using AdaBoost for data set $Z^k = (\mathbf{x}_1, y_1^k), \cdots, (\mathbf{x}_m, y_m^k)$

Final classifier:

$$\hat{k} = \underset{k}{\text{argmax}}\, f^k(\mathbf{x}) \quad (4)$$

図 3 one-vs-rest 法による AdaBoost のマルチクラス化

Fig. 3 one-vs-rest multi-class algorithm for AdaBoost

別器しか作成できないため，実際には他クラスを識別するために図 3 のように one-vs-rest 法を用いて他クラス判別ができるように拡張している [13]．これは AdaBoost によって作られる識別器である式 (2) を 1 つのクラスとその他のクラスを分離できるように作成し，その中で $f(\mathbf{x})$ が最大になるものを識別結果とする手法である．この識別器を用いて入力されたフレームを上段から順に分類していく．図 4 は検出・識別の例である．識別された結果，最終的に重畳している雑音はフレームごとに 1 つのクラスに分類される．雑音除去はこのクラスに存在する雑音の平均ベクトルを使用して行われる．

4. 雑音重畳音声

本稿では対数メルフィルタバンク特徴量を使用して雑音の除去を行う．そこで雑音重畳音声の対数メルフィルタバンク特徴量について考える．時刻 τ において観測信号である雑音重畳音声 $x(\tau)$ は，クリーン音声 $s(\tau)$ と雑音 $n(\tau)$ を用いて以下のように書くことができる．

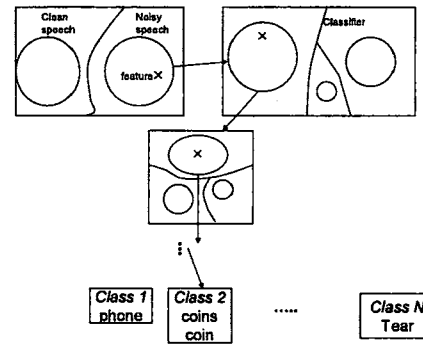

図 4 検出・識別の例

Fig. 4 An example of detection and classification

$$x(\tau) = s(\tau) + n(\tau) \quad (5)$$

この時フレーム t において，b 次元目の雑音重畳音声のメルフィルタバンク特徴量 $X_b(t)$ は以下のように近似することができる．

$$X_b(t) \approx S_b(t) + N_b(t) \quad (6)$$

ここで，$S_b(t)$，$N_b(t)$ はそれぞれクリーン音声，雑音の b 次元目のメルフィルタバンク特徴量である．我々のタスクでは雑音の識別まで行うため，この雑音の特徴量の分布はある程度わかっていると考える．しかしながら雑音の強さは未知である．そのためこの雑音に強さを表す未知の定数 α を掛け合わせ以下のように表す．

$$X_b(t) \approx S_b(t) + \alpha \cdot N_b(t) \quad (7)$$

さらにこれを対数メルフィルタバンクに変換することを考える．観測信号，音声，雑音の b 次元目の対数メルフィルタバンク特徴量を $x_b(t)$，$s_b(t)$，$n_b(t)$ と表すと，$x_b(t)$ は $s_b(t)$，$n_b(t)$ を使って，

$$\begin{aligned} x_b(t) &= \log\left(X_b(t)\right) \\ &\approx \log\left(S_b(t) + \alpha \cdot N_b(t)\right) \\ &= \log\left[\exp\left(s_b(t)\right) + \alpha \exp\left(n_b(t)\right)\right] \\ &= s_b(t) + \log\left[1 + \alpha \exp\left(n_b(t) - s_b(t)\right)\right] \\ &= s_b(t) + g_b(\mathrm{n}(t), \mathrm{s}(t), \alpha) \end{aligned} \quad (8)$$

と書くことができる．ここで n，s は $n_b(t)$，$s_b(t)$ のベクトルである．またこの先特に断りがない限りフレーム番号 t は省略する．観測信号から音声の対数メルフィルタバンク出力を求めるならば，式 (8) の第 2 項 $g_b(\mathrm{n}(t), \mathrm{s}(t), \alpha)$ を推定し x_b から引けばよい．

5. 音声 GMM を用いた雑音除去

クリーン音声の GMM を使用する雑音除去法は Segura らによって提案された [3]. これはクリーン音声の対数メルフィルタバンクの GMM をあらかじめ用意したときに, 式 (8) の第 2 項を推定し, 観測された特徴量から減算することで音声特徴量を推定する手法である.

まずクリーン音声の GMM を次のように与える.

$$p(\mathrm{s}) = \sum_{m}^{M} Pr(m) N(\mathrm{s}; \mu_{s,m}, \Sigma_{s,m}) \tag{9}$$

ここで, M は混合数, $\mu_{s,m}$ は混合 m の音声の平均ベクトル, $\Sigma_{s,m}$ は分散共分散行列で, 本稿では対角行列としている. これを用いて混合数, 混合重みが共に式 (9) の GMM と同様の雑音重畳音声のモデルを考える.

$$p(\mathrm{x}) = \sum_{m}^{M} Pr(m) N(\mathrm{x}; \mu_{\mathrm{x},m}, \Sigma_{\mathrm{x},m}) \tag{10}$$

この時, 雑音重畳音声は式 (8) のように表すことができることから平均 $\mu_{x,m}$ は以下のように近似する. この時, 共分散行列は同じ行列とする.

$$\mu_{\mathrm{x},m} \approx \mu_{s,m} + \mu_{\mathrm{g},m} \tag{11}$$

$$\mu_{\mathrm{g},m} \approx \mathrm{g}(\mu_{\mathrm{n}}, \mu_{\mathrm{s,m}}, \alpha) \tag{12}$$

$$\Sigma_{\mathrm{x},m} \approx \Sigma_{s,m} \tag{13}$$

ここで μ_n は重畳雑音の特徴量の平均ベクトルであり, 本稿では, 識別結果のクラスに属する雑音の平均ベクトルになっている.

この GMM を使ってクリーン音声の特徴量を推定する. $g(\mathrm{n},\mathrm{s},\alpha)$ の推定値 $\hat{g}(\mathrm{n},\mathrm{s},\alpha)$ は以下のように $\mu_{g,m}$ の各混合の尤度による重み付き平均とする.

$$\hat{g}(\mathrm{n},\mathrm{s},\alpha) = \frac{\sum_m p(\mathrm{x},m)\mu_{\mathrm{g},m}}{\sum_m p(\mathrm{x},m)} \tag{14}$$

$$p(\mathrm{x},m) = Pr(m) N(\mathrm{x}; \mu_{\mathrm{x},m}, \Sigma_{\mathrm{x},m}) \tag{15}$$

この推定値 $\hat{g}$ を雑音重畳音声の特徴量 x から減算することで音声特徴量の推定値を求める.

$$\hat{\mathrm{s}} = \mathrm{x} - \hat{g}(\mathrm{n},\mathrm{s},\alpha) \tag{16}$$

6. 雑音重みの推定

上記の手法は加法性雑音の除去に有効であるが, 本タスクでは式 (16) における雑音の重み α が未知の値であるため計算することができない. そのためこの手法を適用するにはなんらかの方法を用いて雑音重み α を推定する必要がある. 本稿ではこれを 2 つの方法で解き, 両方について検証する. 1 つは m を隠れ変数とした EM アルゴリズムを用いて, 雑音重畳音声の尤度 $p(\mathrm{x})$ が最大になるように α を決定する手法であり, もう 1 つは各混合ごとの尤度 $p(\mathrm{x},m)$ を最大化, 各混合ごとに α を決定し推定値を出す手法である. 前者は α を尤度を最大化するような α がただひとつ求まる. 後者は α は混合毎に最大化するので α が複数出現する. そのためそれらを使って音声信号を推定する方法を別に用意する必要がある.

この重み推定を行うことによって, 誤検出したフレームにおいて, 除去時に音声信号を歪ませることを防ぐことができる. これは誤検出したフレームであっても, 音声尤度は重みが 0 に近いほど高くなると期待できるからである.

6.1 EM アルゴリズムを用いた雑音重みの推定

EM アルゴリズムを利用して, 最適な α の値を求める. まず, E ステップとして, ある $\alpha^{(k)}$ が与えられたとき,

$$Q(\alpha^{(k)}, \overline{\alpha}) = \sum_m p(\mathrm{x},m,\alpha^{(k)}) \log p(\mathrm{x},m,\overline{\alpha}) \tag{17}$$

という Q 関数を作成する. ここで

$$\begin{aligned} p(\mathrm{x},m,\alpha) &= Pr(m) N(\mathrm{x}, \mu_{\mathrm{x},m}, \Sigma_{\mathrm{x},m}) \\ &= Pr(m) N(\mathrm{x}, \mu_{s,m} + \mathrm{g}(\mu_{s,m}, \mu_{\mathrm{n}}, \alpha), \Sigma_{\mathrm{x},m}) \end{aligned} \tag{18}$$

とする. 次に, M ステップとして, 式 (17) を最大化する $\overline{\alpha}$ を求める. そして求まった $\overline{\alpha}$ を $\alpha^{(k+1)}$ に代入し, 再び Q 関数を作成する. すなわち

$$\alpha^{(k+1)} = \operatorname*{argmax}_{\overline{\alpha}} Q(\alpha^{(k)}, \overline{\alpha}) \tag{19}$$

この Q 関数を最大化する $\overline{\alpha}$ は偏微分

$$\frac{\partial Q(\alpha^{(k)}, \overline{\alpha})}{\partial \overline{\alpha}} = 0 \tag{20}$$

を解くことで求まる. この 2 つのステップを収束するまで繰り返すことで最適な α の値を決定する.

しかしながら方程式 (20) の左辺を展開すると, 行列が対角行列でその要素 $\sigma^2_{b,b}$ のみと考えても,

$$\begin{aligned} \frac{\partial Q(\alpha^{(k)}, \overline{\alpha})}{\partial \overline{\alpha}} &= \frac{\partial}{\partial \overline{\alpha}} \sum_m p(\mathrm{x},m,\alpha^{(k)}) \log p(\mathrm{x},m,\overline{\alpha}) \\ &= \sum_m p(\mathrm{x},m,\alpha^{(k)}) \cdot \\ &\sum_b \frac{x_b - \mu_{s,m,b} - \log(1+\exp(\mu_{n,b} - \mu_{s,m,b}))}{\sigma^2_{b,b}(1+\overline{\alpha}\exp(\mu_{n,b} - \mu_{s,m,b}))} \end{aligned} \tag{21}$$

と複雑な式になり, そのままでは解くことができない. よって方程式 (20) をニュートン法を用いて解く. ニュートン法とは

$$f^1 = \frac{\partial Q(\alpha^{(k)}, \overline{\alpha})}{\partial \overline{\alpha}} \tag{22}$$

$$f^2 = \frac{\partial^2 Q(\alpha^{(k)}, \overline{\alpha})}{\partial \overline{\alpha}^2} \tag{23}$$

とし,

$$\overline{\alpha} = \overline{\alpha} - \frac{f^1}{f^2} \tag{24}$$

を収束するまで繰り返すことで近似解を求める手法である.

このように EM アルゴリズムを使うことで雑音重み α が求まる. 求めた α を用いて式 (15), (16) を計算し, クリーン音声の特徴量を抽出できる.

6.2 混合毎の重み推定

混合毎の尤度を最大化するように α の値を決定する．すなわち求まる α は混合数と同じ数だけ存在し，それぞれを α_m とおく．この時，

$$\alpha_m = \arg\max_{\alpha} p(\mathbf{x}, m, \alpha) \tag{25}$$

となる．この値を求めるためには

$$\frac{\partial \log p(\mathbf{x}, m, \alpha)}{\partial \alpha} = 0 \tag{26}$$

を解けばよいが，式 (21) 同様に複雑な式となるためニュートン法を用いて近似値を求める．この α_m を用いる時，g の推定値は式 (15) の代わりに

$$\hat{g}(\mathbf{n}, \mathbf{s}, \alpha) = \frac{\sum_m p(\mathbf{x}, m, \alpha) \mathbf{g}(\mu_{\mathbf{n}}, \mu_{\mathbf{s},\mathbf{m}}, \alpha_m)}{\sum_m p(\mathbf{x}, m, \alpha_m)} \tag{27}$$

として求める．これを観測信号 $\mathbf{x}$ から引くことで音声の特徴量を推定する．

この手法はが混合ごとに最適な重みが出現するため，本来は尤度が低く，推定に影響しないような混合の尤度が高くなり，推定に影響を及ぼすことも考えられる．しかしながら，たいていの場合はそのような混合は尤度を最大化しても他に比べて尤度が低く，あまり考慮されないことが多い．

7. 実　　験

7.1 実験条件

ATR の特定話者単語データベースを用いて実験を行った．男性話者 2 名，女性話者 2 名を用い，各話者 2,720 発話を学習に，500 発話をテストに使用した．雑音は RWCP 非音声ドライソースを使用した [11]．このデータベースには 105 種類の雑音が各種類 100 データ存在し，そのうち 50 データを学習用に残りの 50 データをテストデータとして利用した．このデータベースの雑音重畳時間は 20〜300 msec 程度になっている．この 105 種類の雑音のパワーをそろえ，24 次元の対数メルフィルタバンクを作成，種類ごとに平均ベクトルを算出し，クラスタリングを行った．この時作成したツリーはクラスとの中心が上段から順に 50,25,12,6 以下の距離になるように分けられており 5 段のツリーになっている．この時，クラス数は 45 クラスとなった．

検出・識別時の識別器に使用する学習データはこの 2 つのデータベースから各クラスの雑音重畳音声を作成し，対数メルフィルタバンクに変換したものを用いた．この時 SNR は-5〜5 dB になるように調整した．

テストデータに対しては，発話ごとに SNR を調整して雑音を 1〜5 つ重畳させた．この時の SNR は 5 dB, 0 dB, -5 dB になっている．このデータを 24 次元の対数メルフィルタバンクに変換し，雑音の検出・識別・除去を行った．除去時に使用する GMM は学習データのクリーン音声から作成し，16, 32, 64 混合の 3 つを用意し，それぞれの場合において実験を行った．この時 EM アルゴリズム，ニュートン法の α の初期値は 0 としている．そして対数メルフィルタバンクを 12 次元の MFCC に変換し，認識実験を行った．認識モデルにはクリーン音声で作成した 4 混合 5 状態の音素 HMM を使用した．

表 1　検出・識別結果
Table 1　The results of detection and classification

	再現率	適合率	識別率
-5 dB	0.542	0.875	0.649
0 dB	0.497	0.882	0.636
5 dB	0.474	0.881	0.591

7.2 実験結果

まず，フレーム単位の検出・識別の結果を表 1 に示す．結果を見てみると再現率が低く，正しく検出できていないフレームも多いことがわかる．また識別率もあまりよくない．しかしながら，部分的な SNR が低いフレームが多く，そのようなフレームは検出はしづらくとも，認識にはあまり影響を及ぼさないと考えられる．また識別結果も正解ラベルとして与えたクラスに分類される必要はなく，距離の近いクラスに分類されるならば，除去の精度に大きな影響は与えないと考えられる．よって本稿はこの結果をそのまま利用した．その時の認識実験の結果を図 5 に示す．

結果を見ると-5 dB の場合，最大で 23 %程度改善していることがわかる．ここで重み推定なしとは 6. 節で行う雑音重みを推定せずに $\alpha = 1$ として除去した場合の結果であるが，こちらはミスマッチが大きくなり，認識率が下がるという結果も見られた．重み推定を利用した場合はそちらと比べると高い認識率になっている．重み推定で EM アルゴリズムを利用した場合と，混合ごとに重み推定を行った場合の結果を比べると，わずかながら混合ごとに推定した場合のほうがいいが，大きな差は見られなかった．また，雑音除去時の混合数の変化させたときは，混合数を増やすほど高い認識率になった．

7.3 未知雑音に対する実験

上記の実験はデータはオープンであるが，学習データの雑音の種類は等しかった．そこで 10 fold クロスバリデーションを用いて未知の雑音に対しての実験を行う．まず 105 種類を 10 セットに分割し，そのうち 9 セットを学習用に，1 セットをテスト用に使用する．重み推定は混合ごとの推定のみを 64 混合の GMM に対してのみ行った．そのほかの条件は上記の実験と同様である．その時の各セットの除去前と除去後の認識率は表 2 のようになる．

認識率は結果にはセット毎にばらつきが多い．それでも改善率を見てみると，すべてのセットに対して認識率が改善していることがわかる．どの程度改善したかはセットごとに様々であるため，除去しにくい，あるいは全く学習データがないために除去できていない雑音も存在していると思われる．しかしながら，認識率が改悪しているものはなく，また大きく改善する場合も多いため未知の雑音に対しても本手法を使用することは有効であることがいえる．

8. おわりに

本稿では Segura らの手法に雑音の重み推定を組み込み，以

表 2　10 fold クロスバリデーションのセットごとの結果
Table 2　The results of 10fold cross-validation

SNR -5 dB	Set1	Set2	Set3	Set4	Set5	Set6	Set7	Set8	Set9	Set10	平均
除去なし	61.2	55.4	53.1	48.3	56.6	67.1	53.8	70.4	52.8	49.6	56.8
提案手法	78.0	67.8	72.0	66.2	75.5	75.3	79.9	83.4	68.2	59.4	72.6

SNR 0 dB	Set1	Set2	Set3	Set4	Set5	Set6	Set7	Set8	Set9	Set10	平均
除去なし	68.0	61.6	60.9	55.6	63.8	71.5	59.0	76.3	58.9	56.4	63.2
提案手法	83.8	72.1	77.3	70.7	78.1	79.6	83.9	86.0	73.2	65.4	77.0

SNR 5 dB	Set1	Set2	Set3	Set4	Set5	Set6	Set7	Set8	Set9	Set10	平均
除去なし	74.0	67.8	69.8	63.8	70.9	77.2	65.9	80.9	65.7	65.3	70.1
提案手法	87.1	75.3	80.2	73.0	80.1	84.5	84.3	87.3	77.4	70.6	80.0

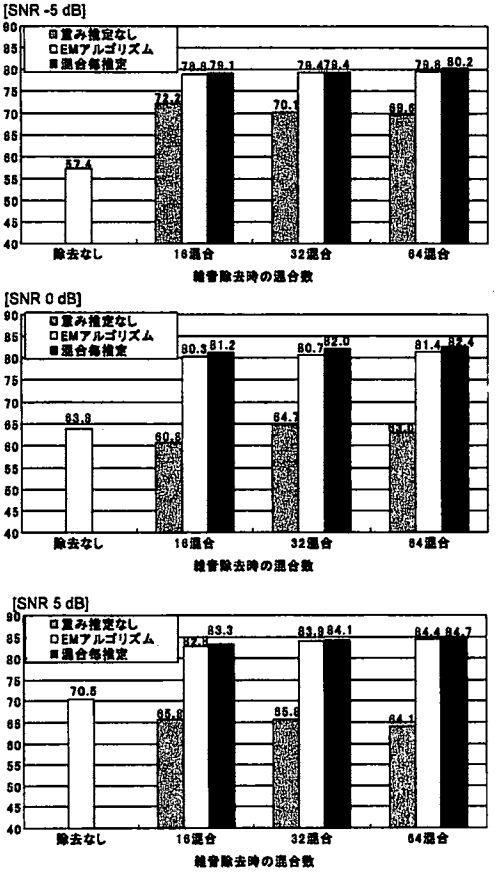

図 5　The results of word recognition
Fig. 5　単語認識実験の結果

前提案した雑音の検出と識別を組み合わせることで SNR が未知の場合においても可能な突発性雑音の除去法を提案した．重み推定は EM アルゴリズムを用いた場合と，混合ごとに推定した場合の 2 つを用い，どちらの場合においても認識率の改善があることを確認した．また未知の雑音に対しても実験を行い，有効性を確認した．今後は，検出の精度を上げることによる認識率の変化を調べ，不特定話者・大語彙の音声認識に対して実験を行う予定である．

文　献

[1] H. Xu, et al., "Spectral Subtraction with Full-wave rectification and Likelihood Controlled Instantaneous Noise Estimation for Robust Speech Recognition," in Proc. Interspeech, 2004, pp. 2085-2088.

[2] N. W. D. Evans et al., "An Assessment on the Fundamental Limitations of Spectral Subtraction," Proc. ICASSP, pp. I-145-I-148 ,2006.

[3] J. C. Segura, et al., " Model-based compensation of the additive noise for continuous speech recognition. Experiments using the AURORA II database and tasks, " Eurospeech, pp. 221-224, 2001.

[4] L. Deng, et al., " Enhancement of log Mel power spectra of speech using a phase-sensitive model of the acoustic environment and sequential estimation of the corrupting noise," IEEE Trans. SAP, vol. 12, pp. 133-143, 2004.

[5] S. Araki, et al., "Blind Speech Separation in a Meeting Situation," ICASSP2007, vol. I, pp. 41-45, 2007.

[6] 野口貴一ら，"通信会議における 1 チャネル突発性雑音抑圧，" 電子情報通信学会技術研究報告. EA, Vol.105, No.403, pp. 31-36 , 2005.

[7] K.Manohar, "Speech enhancement in nonstationary noise environments using noise properties," Speech Communication, 48(1), pp. 96-109, 2006.

[8] 三宅ら，"AdaBoost を用いた雑音の検出と識別，" 日本音響学会 2007 年春季研究発表会，pp.141-142, 2007-03.

[9] M.Fujimoto et al., "Combination of Temporal Domain SVD Based Speech Enhancement and GMM Based Speech Estimation for ASR in Noise - Evaluation on the AURORA2 Task -," Proc. Eurospeech'03, pp. 1781-1784, 2003.

[10] 實廣ら，"複数雑音合成モデルによるマルチパス探索に基づく雑音抑圧，" 日本音響学会 2007 年秋季研究発表会，pp.151-154, 2007-03.

[11] S. Nakamura, et al., " Acoustical Sound Database in Real Environments for Sound Scene Understanding and Hands-Free Speech Recognition," 2nd ICLRE, pp. 965- 968, 2000.

[12] Freund. Y, et al., " A decision-theoretic generalization of on-line learning and an application to boosting, " Journal of Comp. and System Sci., 55, pp. 119-139, 1997.

[13] Ethem ALPAYDIN, " Introduction to Machine Learning, " The MIT Press, October 2004