社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

ガウス分布の補正による突発性雑音に頑健な音声認識

山本　　仁　　篠田　浩一　　嵯峨山茂樹

東京大学大学院情報理工学系研究科
〒 113–0033 東京都文京区本郷 7–3–1
E-mail: {yamamoto,k-shino,sagayama}@hil.t.u-tokyo.ac.jp

あらまし　音響 HMM の出力確率分布にガウス分布を用いた音声認識において，認識計算時の音響尤度の補正により雑音への頑健性を高める方法について報告する．近年，音声認識技術を実環境でも用いるための研究が盛んに行なわれているが，その重要な課題の一つに未知の非定常雑音への対処がある．音響 HMM の出力確率分布にガウス分布を用いるとき，対数音響尤度は分布平均からのずれの 2 乗に従い低下する．このとき雑音等の影響で分布の裾部に相当する音響的外れ（outlier）が発生すると，それは著しい尤度の低下とともにモデル間に過大な尤度差を生む．この現象は認識誤りの一因になると考えられる．本稿では，この問題に対処するための手法として，ガウス分布から求まる音響尤度の補正を提案する．クリーンな音声の音響モデルと提案手法を使用して，計算機上で突発性雑音を加算した音声を入力とした大語彙連続音声認識実験を行ったところ，最大 46.9%の誤り削減率が得られ，効果が確認された．
キーワード　突発性雑音，未知雑音，ガウス分布，音響尤度

Compensated Gaussian Distribution for Robust Speech Recognition against Non-Stationary Noise

Hitoshi YAMAMOTO, Koichi SHINODA, and Shigeki SAGAYAMA

Graduate School of Information Science and Technology, University of Tokyo
7–3–1 Hongo, Bunkyo-ku, Tokyo, 113–0033 Japan
E-mail: {yamamoto,k-shino,sagayama}@hil.t.u-tokyo.ac.jp

Abstract　This report describes a compensation method of acoustic likelihood from Gaussian distribution to achieve robustness in noisy environments. Recently, research for speech recognition in adverse environments have been done extensively, and one of most important issues is to cope with unknown non-stational noise. When using a Gaussian distribution as an output probability density function for HMMs, the logarithmic acoustic likelihood of a sample decreases according to the second power of distance between the sample and the mean of the distribution. Therefore, if a sample from noise-overlapped speech stays on the tail of distribution(an acoustic outliers), excessive likelihood differences among models are generated. These differences may cause recognition errors. In this report, a method of compensating the acoustic likelihood calculated from Gaussian distribution is proposed. Large vocabulary continuous speech recognition experiments were conducted using speech data to which unknown non-stationary noise added artificially. The experimental results showed 46.9% error reduction rate.
Key words　Unknown Noise, Non-Stationary Noise, Gaussian Distribution, Acoustic Likelihood

1. はじめに

本稿は，音響 HMM の出力確率分布にガウス分布を用いた音声認識において，音響尤度を認識計算時に補正することによって，雑音への頑健性を向上させる方法について報告する．

近年の音声認識技術の進歩により，静かな環境においては精度の高い認識ができるようになった．だが，実環境では，雑音の存在や伝達特性などの影響で認識性能が劣化するという問題があり，中でも未知の非定常雑音への対処は重要な課題のひとつである [1].

本稿では，非定常雑音のうち特に短時間区間の加法性の突発性雑音を扱う．加法性雑音に対処する方針として代表的なものには，スペクトル減算法などの雑音成分の除去と，MLLR 法や PMC 法などのモデルの適応化・合成がある．いずれの方法で

も雑音情報を必要とするが，通常それらは環境に頻繁に存在するものや対処すべきものとしてあらかじめ想定されており，未知雑音に対しては対処が難しい．また，パワーやスペクトル特徴が時間的に大きく変動する非定常雑音も対処が困難である．この問題に対し、これまでにフレーム単位でのモデル選択 [2] や適応化モデルの SN 比別マルチパスモデル [3] が提案されているが，いずれもモデルの学習データに含まれない雑音に対しては対処が困難である．

さて，音響 HMM の出力確率分布にはガウス分布が広く用いられている．このとき，ガウス分布の分布形状により，雑音等の影響を受けた音声に与える各モデルの対数音響尤度に大きな差が生じる．ガウス分布裾部のもたらすこの尤度差は認識誤りの一因になると考えられる．ガウス分布と異なる分布形状を用いた音響モデリング手法として，一般化ラプラス混合分布 [4] や Richter 分布 [5] が提案されているが，これらは裾部の音響尤度の改善を直接の目的とはしていない．

本稿では，雑音等の影響を受けたとみなされる入力に対し，ガウス分布から得られる音響尤度を補正する計算法を提案する．フレームごとに補正を行なうことで，短時間の突発性雑音に対する頑健性の向上が期待できる．また，SN 比の変動や雑音の種類に関わらず効果が期待できる．

以下では，まず 2 章で音響 HMM の出力確率分布に着目した認識誤り発生の仮説を立て，次に 3 章でその仮説を踏まえた提案手法について述べる．4 章で大語彙連続音声認識実験による評価結果を報告する．

2. 認識誤り発生の仮説

音声認識は音声特徴量空間上のベクトル時系列のパターン認識として捉えられる．特徴量空間において，音声の振る舞いは軌跡として，音響モデルは超楕円球状の分布として現れる．音声が雑音等の影響を受けると，もとの音声のものとかけ離れた特徴量の点が現れ，軌跡も大きく変動すると考えられる．本来はこのような現象も音響モデルに反映されるべきであるが，一般に非定常雑音はその多様性のため，すべてを学習することはできず，モデルにとって未知の特徴量点が生じる．このような，音声から大きく外れて音響モデルに学習されない特徴量点を音響的外れ（outlier）と呼ぶことにする．

音響的外れを含む音声が誤認識される仕組みについて考える．入力音声のある次元の特徴量を y とすると，音響 HMM の出力確率分布にガウス分布を用いるとき y の分布は (1) 式で表される．すると，対数音響尤度は (2) 式のように y の 2 次関数となり，分布平均からのずれの 2 乗にしたがって低下する．

$$N(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{(y-\mu)^2}{2\sigma^2}\Big) \tag{1}$$

$$\log N(y;\mu,\sigma^2) = \log\frac{1}{\sqrt{2\pi}\sigma} - \frac{(y-\mu)^2}{2\sigma^2} \tag{2}$$

よって，雑音等の影響で音声特徴量が大きく変化すると，その特徴量は多くの場合分布の裾部に相当するため，尤度の著しい低下に伴いモデル間に通常の音声入力よりも大きな尤度差を生むと考えられる．この尤度差は，仮説間尤度差の過剰な拡大や仮説順位の逆転を引き起こし，認識誤りの一因になると考えられる．

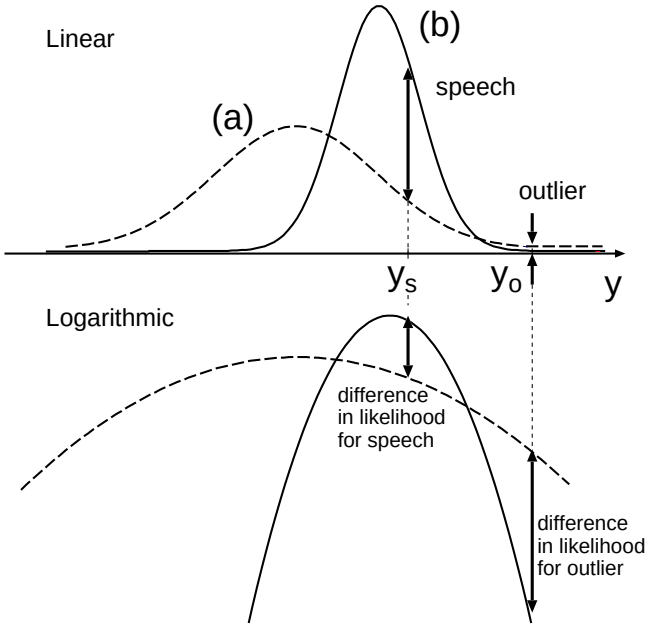

図 1　認識誤り発生の仮説
Fig. 1　Hypothesis of speech recognition error

例として図 1 を用いて説明する．図 1 は 2 つのモデル（a）・（b）の y と同次元についての模式図である．入力音声の特徴量が y_s のとき，分布によれば y_s はモデル（a）よりもモデル（b）からの出力とするのが妥当と考えられる．そして，対数音響尤度で見ても（b）の尤度が（a）の尤度より大きい．認識計算部ではこのような尤度差を各次元各フレーム積み重ねていくことで，結果的に最適な状態遷移を得る．次に，y_s が雑音等の影響を受けた y_o を考える．このとき y_o は音響的外れである．分布によると，y_o は（b）からの出力とも（a）からの出力ともとれるので，対数音響尤度において大きな差を持たないことが望ましい．しかし，図からもわかる通りそうはなっておらず，(a) と（b）の尤度が逆転しているだけではなく，2 次曲線によってその間に通常の音声によるものよりも大きな尤度差を生じている．この尤度差は，仮説間の尤度の過剰な拡大や順位の逆転を引き起こし，他の次元や前後のフレームでの正しい評価を覆す可能性がある．

以上より，音響モデルでは学習できない音響的外れに対してガウス分布はモデル間に過大な尤度差を生むという現象が仮説順位に影響して認識誤りを起こす，という仮説を立てる．

3. ガウス分布の補正

前章の仮説によると，雑音等の影響により現れた音響的外れに対して各モデルのガウス分布の与える対数音響尤度に大きな差があることが認識誤りの一因である．音響的外れはすでにどのカテゴリの分布中心からも遠い点であり，明確なカテゴリ分類ができないという状況を考えると，それは尤度計算に大きく影響させるべきでなく，カテゴリ間で大きな差が生じないように尤度を与えることが現実的である．その手段としてガウス分

布の補正を考える．音響的外れにより音声特徴量ベクトルの空間軌跡は大きく変動すると仮定したが，その変動がどの次元にどのように現れるかは予測できないため，音響的外れに対し同程度の尤度を一律に与えることにする．

ガウス分布 $N(y)$ に正の微小な定数 ϵ を加えた $N'(y)$ を考える ((3) 式)．この ϵ を補正定数と呼ぶことにする．すると $N'(y)$ の対数は，図 2 下図のように分布中心部ではほぼ $\log N(y)$ に近く，分布裾部ではほぼ $\log \epsilon$ に近いという形状になる ((4) 式)．

$$N'(y;\mu,\sigma^2) = N(y;\mu,\sigma^2) + \epsilon \tag{3}$$

$$\begin{aligned}\log N'(y;\mu,\sigma^2) &= \log(N(y;\mu,\sigma^2) + \epsilon) \\ &\simeq \begin{cases} \log\epsilon & \text{if } |y-\mu| \gg 0 \\ \log N & \text{else} \end{cases}\end{aligned} \tag{4}$$

この (4) 式はすなわち，$N'(y)$ が通常の音声入力に対してはガウス分布に基づく尤度を，音響的外れに対しては補正した尤度をそれぞれ出力することを意味する．そして，全分布に対して等しい補正定数を用いることで，音響的外れに同程度の尤度を一律に与えることができる．ここでは 1 次元のガウス分布について述べたが，多次元のガウス分布や混合ガウス分布でも同様に補正を行なうことができる．

ここでひとつ議論を要するのは，(3) 式で表現した $N'(y)$ をそのまま確率分布モデルと理解すると，全空間での積分値が発散するために矛盾が生じる点である．しかし，音声および音響的外れが覆う空間のみを積分範囲とすると，その積分値は $N(y)$ とほぼ同一に見なせると考えられる．このように，音響的外れは学習データに出現しないために存在する範囲が未知であるが，出現した場合にその尤度を $N'(y)$ により求めることには矛盾は生じない．その積分範囲 (分布範囲) を事前に特定できないからこそ，上式を計算法として事前に用意するわけである．

なお，上記の $N'(y)$ は，混合成分が m 個の混合ガウス分布において，第 $m+1$ 番目の混合成分を付加し，その分岐確率を ϵ，分散を ∞ としたとの理解も可能で，実際に通常の混合ガウス分布モデルの枠組の中で計算することもできる．

この原理に基づく補正をフレームごとに行なうことで特に突発性雑音に対しての頑健性が期待できる．なぜなら，入力が音響的外れか否かによって尤度補正が行なわれるため，音響的外れの現れた数フレームだけの尤度低下を抑えることができるからである．また同様に，雑音の種類やその SN 比の未知性によらず汎用的な対処ができる．モデルそのものは変化させないため，雑音のない部分に関しては従来と変わらない認識性能が期待できる．逆に定常的な雑音や常に存在する雑音などへの効果は他手法に比べ小さいが，その場合には従来の耐雑音音声認識手法と組み合わせて用いることもできる．

4. 認識実験による評価

提案手法を検証するため，大語彙連続音声認識実験を行った．認識システムとしては IPA の日本語ディクテーション基本ソフトウェア [6], [7] を用いた．以後，本稿で提案する補正を行う手法を「提案法」，これを行わない従来の標準的な手法を「従来法」と呼ぶことにする．

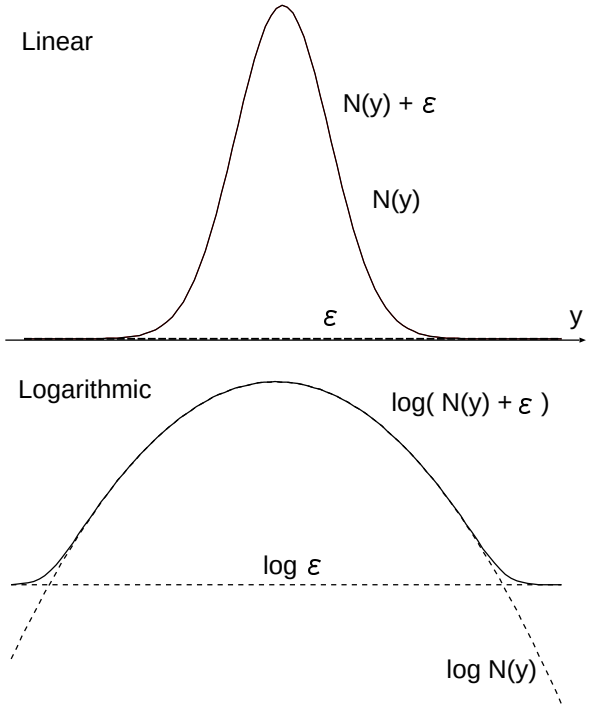

図 2　ガウス分布より求まる音響尤度の補正
Fig. 2　Compensation of acoustic likelihood

4.1 実験条件

評価用の音声は，計算機上で音声波形と雑音波形を加算して作成した．音声データには，IPA-98-TestSet から男女それぞれ 23 名の合計 200 文の新聞朗読音声を用いた．また，雑音データには，RWCP 実環境音響音声データベース [8] の環境音のうち突発性雑音と考えられるものを用いた．雑音はそれぞれの種類について数十のサンプルがある．雑音の加算方法を述べる．まず各サンプルから雑音を含む区間を 1 秒切り出し，各雑音間に 1 秒の無音を挟んでつないだデータを作成した．次にこの長い雑音データから各音声データと同時間分切り出し，その区間の SN 比に基づいて加算した．SN 比には，音声・雑音それぞれにおいて時間幅 30m 秒のパワーの移動平均を求め，区間内におけるそれぞれの最大値の比を用いた．各音声データに加算した雑音データは全て異なる部分を切り出している．また，雑音が音声の部分に重なるとは限らない．表 1，2 に使用した音声データと雑音データの概要をそれぞれ示す．

特徴量ベクトルは，12 次 MFCC と ΔMFCC，および Δ 対数パワーの計 25 次元とした．表 3 に音声分析条件を示す．音響モデルには，IPA のソフトウェア付属のものの中から monophone モデルと 2000 状態の triphone モデルを用いた．それぞれ混合数 16 の性別依存（GD）モデルとした．これらは，クリーンな音声で学習した対角共分散の連続混合分布 HMM である．表 4 に音響モデルの概要を示す．言語モデルは，IPA のモデルのうち，語彙サイズ 20000 で新聞 75ヶ月分の学習データから構築されたもの（75month cutoff-1-1）を用いた．これはデコーダに合わせた前向き 2-gram 後向き 3-gram の単語 N-gram である．デコーダには Julius 3.1p2 の高速版を用い，音響尤度計算部を書き換えて提案手法を実装した．

表 1　音声データ
Table 1　Speech data

性別	話者数	文数	単語数	時間/文 (秒)
女性	23	100	1575	6.27
男性				5.84

表 2　雑音データ
Table 2　Noise data

種類	内容	総数	時間/雑音 (秒)
whistle3	ホイッスルを吹く	50	0.1–0.2
cup1	ガラスコップを叩く	100	0.1–0.4
cap1	キャップを勢いよく開ける	100	0.02–0.03
bell1	糸で鈴を吊し，引いて鳴らす	50	0.4–0.9
cracker	クラッカーを破裂させる	18	0.04–0.05

表 3　音声分析条件
Table 3　Conditions of speech ayalysis

標本化周波数	16kHz
量子化ビット数	16bit
高域強調	$1-0.97z^{-1}$
フレーム幅	25ms ハミング窓
フレーム間隔	10ms
特徴量	25 次元: 12 次 MFCC, ΔMFCC, ΔlogPower

表 4　音響モデル
Table 4　Acoustic model

HMM	対角共分散連続混合分布 left-to-right HMM
状態数	5 状態 3 ループ
混合数	16
音素数	43

前章でも述べたように，補正は 1 次元のガウス分布に対してだけでなく多次元分布や混合分布に対しても可能である．本実験では混合ガウス分布に対して補正を行なった．認識計算時には対数音響尤度を扱うため，補正には (5) 式のように addlog 関数を用いた．この関数は，以下に示す通り (6) 式または (7) 式により和の対数をそれぞれの対数から求めるものである．

$$\log N'(y) = \mathrm{addlog}(\log N(y), \log\epsilon) \quad (5)$$

$$\begin{aligned}\log(a+b) &= \mathrm{addlog}(\log a, \log b)\\ &= \log a + \log(1+\frac{b}{a}) \quad (a>b)\\ &= \log a + \log(1+\exp(\log b - \log a)) \quad (6)\\ &\simeq \log a \quad (a \gg b) \quad (7)\end{aligned}$$

4.2 実験結果

前節の条件のもとで行った実験結果を示す．評価基準には次式より求まる単語認識精度 (Word Accuracy;WA) の，23 話者の平均を用いた．

$$\mathrm{WA} = \frac{N-S-D-I}{N} \times 100(\%)$$

N：全単語数，S：置換誤り，D：削除誤り，I：挿入誤り

4.2.1　提案法の効果

図 3 に提案法の補正定数 ϵ の変化に対する認識率を示す．雑音は whistle3，SN 比は 0dB，モデルは女性の triphone である．広範囲の ϵ で従来法を上回る認識率を得た．また，ϵ が大きくなるに従って 0 に近づき，小さくなるにつれて従来法と同等の認識率に近づいた．これは，ϵ が大きくなると音響的外れとは言えない部分についてまで尤度を補正するためかえって認識誤りを起こすためであると考えられる．逆に，ϵ が小さくなると補正がほとんど行なわれなくなり，効果が現れないと考えられる．従来法は $\epsilon=0$（図 3 では $x=\infty$）に等価である．

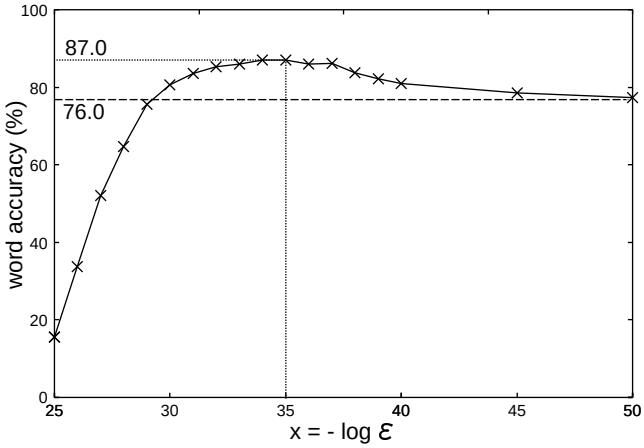

図 3　さまざまな補正定数 $\epsilon=10^{-x}$ に対する認識率（WA，%）（雑音：whistle3，SN 比：0dB，モデル：女性 triphone）
Fig. 3　Recognition rates(WA, %) for different compensation constants $\epsilon=10^{-x}$ (noise: whistle3, SNR: 0dB, model: female triphone)

表 5　クリーンな音声の認識率（WA，%）
Table 5　Recognition rates for clean speech(WA, %)

	model	baseline	$\epsilon=10^{-35}$	$\epsilon=10^{-36}$	$\epsilon=10^{-37}$
女性	monophone	85.6	85.4	85.9	86.3
男性	monophone	82.6	81.3	82.1	82.6
女性	triphone	94.6	94.5	94.5	94.5
男性	triphone	93.5	93.0	93.3	93.4

次に，提案法による認識率向上の大きかった部分を誤り削減率とともに表 6 および図 4 に示す．音響モデルの種類によらず従来法に比べ性能向上が見られ，女性の triphone と $\epsilon=10^{-35}$ のときは 40%を越える誤り削減率を得た．また，このときの各話者の認識率を図 7 に示す．ほとんどの話者で認識率が向上し，特に従来法で認識率の低かった話者に対して大きく上昇した．傾向としては monophone よりも triphone で，また男性よりも女性で，より高い誤り削減率が得られた．また，同じ条件でクリーンな音声を認識したときの結果を表 5 に示す．クリーンな音声の認識率は従来法の場合と同等であった．

以上より，提案法によってクリーンな音声の認識性能を損なうことなく突発性雑音が加わった音声に対する頑健性が高まることが確認された．

4.2.2　SN 比の変動や雑音の種類に対する頑健性

さまざまな SN 比で突発性雑音を加えた音声の認識率を表 7 と図 5 に示す．雑音は whistle3 で，これを–10dB から 20dB

表 6　各音響モデル時の認識率（WA，%）．() 内は従来法に対する誤り削減率（%）（雑音：whistle3，SN 比：0dB）．
Table 6　Recognition rates(WA, %) for various acoustic models. Error reduction rates(%) are in () (noise: whistle3, SNR: 0dB).

	model	baseline	$\epsilon=10^{-35}$	$\epsilon=10^{-36}$	$\epsilon=10^{-37}$
女性	monophone	70.0	75.9(19.6)	75.1(17.0)	73.7(12.3)
男性	monophone	65.9	69.8(11.4)	70.3(12.9)	69.4(10.3)
女性	triphone	76.0	87.0(45.7)	86.1(42.1)	86.1(42.1)
男性	triphone	76.8	84.1(31.2)	82.4(24.1)	82.9(26.3)

表 7　SN 比を変化させたときの認識率（WA，%）．() 内は従来法に対する誤り削減率（%）（雑音：whistle3，モデル：女性 triphone）．
Table 7　Recognition rates(WA, %) for various SN ratio. Error reduction rates(%) are in () (noise: whistle3, model: female triphone).

	–10dB	0dB	10dB	20dB	clean
baseline	68.0	76.0	84.1	89.4	94.6
$\epsilon=10^{-35}$	83.0(46.9)	87.0(45.8)	90.4(39.6)	91.4(18.9)	94.5
$\epsilon=10^{-36}$	82.0(43.8)	86.1(42.1)	89.8(35.8)	91.2(17.0)	94.5
$\epsilon=10^{-37}$	81.0(40.6)	86.1(42.1)	89.4(33.3)	90.9(14.2)	94.5

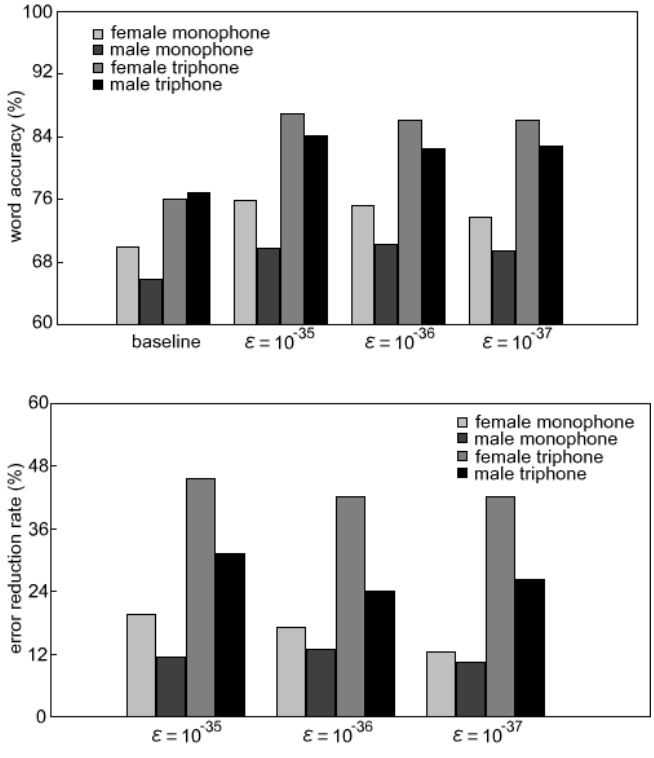

図 4　各音響モデル時の認識率（WA，%，上段）と従来法に対する誤り削減率（%，下段）（雑音：whistle3，SN 比：0dB）
Fig. 4　Recognition rates(WA, %) and error reduction rates(%) for various acoutic models (noise: whistle3, SNR: 0dB)

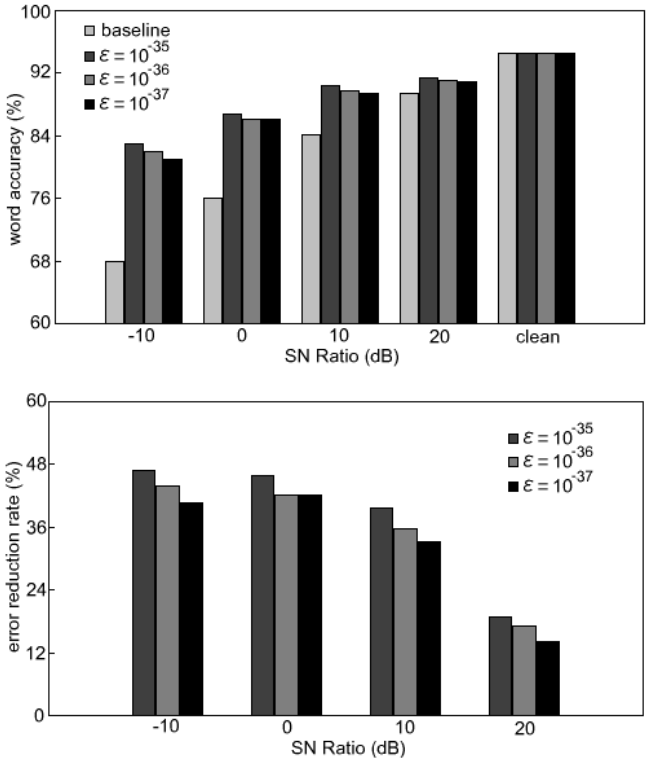

図 5　SN 比を変化させたときの認識率（WA，%，上段）および従来法に対する誤り削減率（%，下段）（雑音：whistle3，モデル：女性 triphone）
Fig. 5　Recognition rates(WA, %) and error reduction rates(%) for various SN ratio (noise: whistle3, model: female triphone)

まで 10dB 間隔で変化させた．モデルは女性の triphone モデルである．提案法では従来法に比べ，SN 比を減少させたときの認識率低下が少なかった．また，–10dB のとき $\epsilon=10^{-35}$ で 46.9%の誤り削減率を得た．この結果から，同一の補正定数で幅広い SN 比に対処が可能であること，すなわち未知の SN 比に対して頑健性が向上することが確認された．

次に，雑音の種類を変えたときの結果を表 8 と図 6 に示す．SN 比は，bell1 が 0dB の他はすべて–10dB である．補正定数は whistle3 で最も性能向上が見られた $\epsilon=10^{-35}$ とした，モデルは女性の triphone モデルである．誤り削減率が 10–20%前後と whistle3 の結果に及ばなかったことから，雑音によって最適な補正定数が異なる可能性が考えられる．また，bell1 のようにやや時間長の長い雑音では誤り削減率が低かったことから，時間長によっても補正の効果が異なると考えられる．このような点を今後明らかにする必要があるが，いずれの雑音においても認識率が上昇したという結果は補正により未知雑音に対しての頑健性が高まる可能性を示している．

5. おわりに

本稿では，音声認識の雑音への頑健性を高める手法として，ガウス分布から求まる音響尤度の補正を提案した．これは，モデルのガウス分布と補正定数の和から尤度を得るという補正によりガウス分布裾部での 2 次曲線のよるモデル間尤度差の抑制を試みるものである．実際にこの補正をフレームごとに行うことにより，大語彙連続音声認識実験において突発性雑音加算音声に対する認識性能の向上が確認され，最大で 46.9%の単語誤り削減率を得た．また，幅広い SN 比や数種類の雑音といった雑音の未知性に対して対処が可能であることがわかった．これ

表 8　雑音の種類を変えたときの認識率（WA，%）．() 内は従来法に対する誤り削減率（%）（モデル：女性 triphone）．

Table 8　Recognition rates(WA, %) for various kind of noises. Error reduction rates(%) are in () (model: female triphone).

	whistle3	cup1	cap1	bell1	cracker	clean
baseline	68.0	83.3	91.8	83.6	92.0	94.6
$\epsilon=10^{-35}$	83.0(46.9)	87.0(22.2)	93.5(20.7)	84.7(6.7)	92.9(11.3)	94.5
$\epsilon=10^{-36}$	82.0(43.8)	86.2(17.4)	93.2(17.1)	84.6(6.0)	93.0(12.5)	94.5
$\epsilon=10^{-37}$	81.0(40.6)	86.5(19.2)	93.3(18.3)	85.4(11.0)	92.8(10.0)	94.5

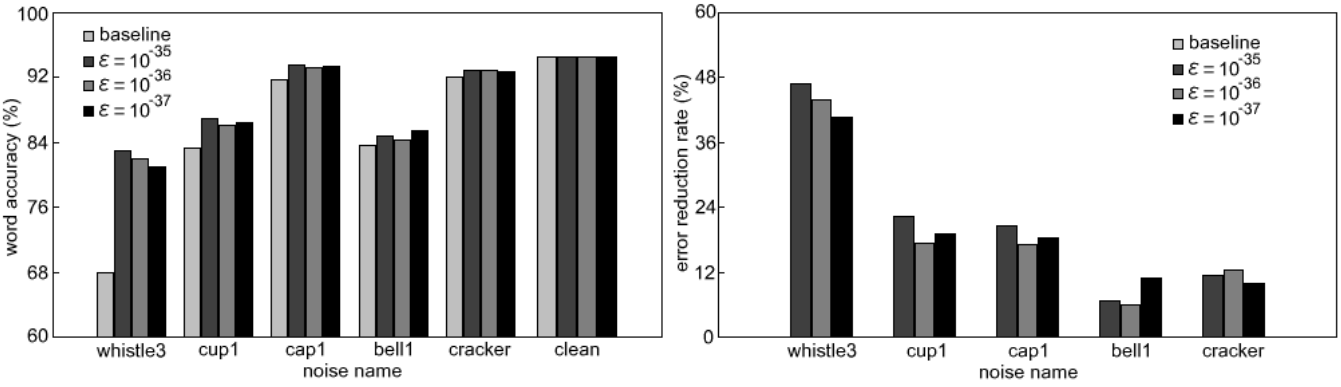

図 6　雑音の種類を変えたときの認識率（WA，%，左）および従来法に対する誤り削減率（%，右）（モデル：女性 triphone）

Fig. 6　Recognition rates(WA, %) and error reduction rates(%) for various kinds of noises (model: female triphone)

らの結果から，本手法は，突発的な雑音に対し，従来の枠組でモデルの頑健性を簡単に向上させられ，かつ副作用がない有用な手法であると考えられる．

今後の展開としては，雑音の種類や時間長などと本手法の関係をさらに明らかにすること，また，SN 比や雑音の種類に応じた最適な補正定数を自動で与えること，などを考慮中である．

文　献

[1] 中村哲, “実音響環境に頑健な音声認識を目指して,” 電子情報通信学会技術研究報告, SP 2002-12, 2002.

[2] 滝口哲也, 西村雅史, “フレーム単位でのモデル選択による突発性雑音下での音声認識,” 日本音響学会 2002 春季講演論文集, pp.57-58, 2002.

[3] 伊田政樹, 中村哲, “雑音 DB を用いたモデル適応化 HMM の SN 比別マルチパスモデルによる雑音下音声認識,” 電子情報通信学会技術研究報告, SP 2001-92, 2001.

[4] 中村篤, “一般化ラプラス混合分布に基づく音声認識用音響モデリング,” 電子情報通信学会論文誌 D-II, Vol. J83-D-II, No.11, pp.2118-2127, 2000.

[5] M. J. F. Gales and P. A. Olsen, “Tail Distribution Modeling Using the Richiter and Power Exponential Distribution,” Proc. Eurospeech99, pp.1507-1510, 1999.

[6] 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄, “音声認識システム,” オーム社, 2001.

[7] 河原達也, 李晃伸, 小林哲則, 武田一哉, 峯松信明, 嵯峨山茂樹, 伊藤克亘, 伊藤彰則, 山本幹雄, 山田篤, 宇津呂武仁, 鹿野清宏, “日本語ディクテーション基本ソフトウェア (99 年度版) の性能評価,” 情報処理学会研究報告, SLP 31-2, pp.9-16, 2000.

[8] S. Nakamura, K. Hiyane, F. Asano, T. Nishimura, T. Yamada, “Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition,” Proc. International Conference on Language Resources and Evaluation, pp. 965-968, 2000.

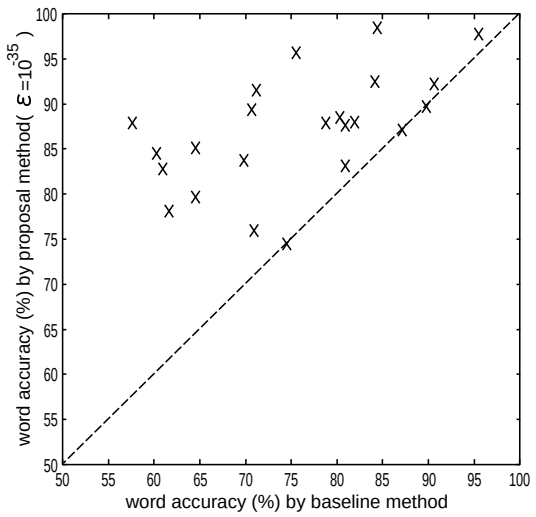

図 7　各話者の認識率（WA，%）の従来法との比較（$\epsilon=10^{-35}$，雑音：whistle3，SN 比：0dB，モデル：女性 triphone）

Fig. 7　Comparison of the recognition rates(WA, %) by proposal method and those by baseline method in each speeker ($\epsilon = 10^{-35}$, noise: whistle3, SNR: 0dB, model: female triphone)