論　文

雑音下音声認識のための重み付け分散拡大に基づく単語 HMM の耐雑音性の改善

漢野　救泰[†]　　船田　哲男[††]

Improved Noise Robustness of Word HMMs Based on Weighted Variance Expansion for Noisy Speech Recognition

Sukeyasu KANNO[†] and Tetsuo FUNADA[††]

あらまし　工場のような実環境下では，非定常な高騒音の発生など周囲の状況の変化により，雑音のスペクトルや SN 比が急激に変動することがある．認識時の雑音の種類や SN 比が学習時や適応時と異なる場合，音声認識性能は著しく低下するため，雑音の変動に対して頑健な HMM が必要となる．本論文では，既存の単語 HMM の耐雑音性を向上させるため，各状態あるいは各分布のパワーによる重み付け分散拡大を行い，雑音の影響を受けやすい状態からの出力確率を，観測ベクトルの違いにより大きく変動しないように制御する手法を提案する．語彙数 50 の単語音声と 2 種類の工場の雑音を使用した不特定話者単語認識実験により，クリーン音声 HMM と雑音付加音声 HMM の学習時とは異なる雑音状況（加法性雑音の種類，SN 比）における本手法の有効性を評価した．実験の結果，いずれの HMM の場合も，雑音条件の変動に対して広範囲の SN 比で認識率が向上し，雑音下小語彙音声認識における単語 HMM の耐雑音性の改善を確認した．特に，学習時より低い SN 比に対しては，重み付け分散拡大は拡大率一定よりも認識性能を顕著に改善できた．

キーワード　耐雑音性，HMM，雑音環境，音声認識，分散拡大

1. まえがき

雑音環境下の音声認識については，これまでに多くの研究が報告されている．認識時の雑音条件（加法性雑音のスペクトル，SN 比）が，HMM の学習時あるいは適応時の雑音条件と同じで安定していれば，その条件を想定した学習 [1], [2] あるいは HMM 合成 [3], [4] 等の適応化手法が有効である．また，推定雑音の減算に基づくスペクトルサブトラクション [5] も効果がある．しかし，実環境下においては，周囲の状況により騒音の種類やレベルの変化，話者の発声音入力レベルの変動が原因で，雑音条件が急激に変動することがある．認識時に雑音条件が変動した場合には，音声認識性能が著しく低下する [1], [6]．認識時での未知の雑音の混入や急激な SN 比の変動に対しては，前述の適応化手法，正規化手法による対処は極めて困難である．

このため，雑音 HMM の分散を拡大して HMM 合成を行うことで雑音条件の急激な変動に対処する方法 [7] が提案されている．ただし，HMM 合成による手法では，非線形変換における近似的処理に基づいて合成された HMM が使用されるため，その認識性能は想定した雑音条件に対しても決して高くはない．その上，発声前の適応時（雑音 HMM の学習・合成処理時）に非定常雑音が発生し，認識時が定常雑音の場合には，適応化が性能低下を招いてしまう．例えば，工場のような騒音環境では，通常は比較的定常な雑音条件が存在し，一時的な非定常雑音（機械加工音，クレーン動作音等）の発生あるいは騒音レベル低下（機械の停止等）による雑音条件変動が頻繁に生じる．そこで，定常時に最も性能を発揮できる HMM をあらかじめ用意しておき，非定常雑音発生時でも従来の認識性能低下を改善できる方式が有効かつ実用的と考えられる．

定常雑音のように認識時の雑音条件が推定できる場合，その SN 比で環境雑音をクリーン音声に付加させた雑音付加音声でモデルを学習する（以下，雑音

[†] 石川県工業試験場，金沢市
Industrial Research Institute of Ishikawa, Kanazawa-shi, 920–8203 Japan
[††] 金沢大学工学部，金沢市
Faculty of Engineering, Kanazawa University, Kanazawa-shi, 920–8667 Japan

付加音声 HMM）ことで，その SN 比周辺では最も優れた認識性能が得られる [1], [2], [8], [9]．雑音付加音声 HMM の学習には時間を要するが，あらかじめ学習しておくことで定常時では常に使用できる利点がある．そして，雑音付加音声 HMM は，HMM 合成やスペクトルサブトラクション等の他の処理よりも性能が高く，その認識率は目標値となっている [1], [2], [8], [10]．したがって，静環境においてはクリーン音声 HMM，騒音環境においては雑音付加音声 HMM が，各々，定常な環境条件に対して最適であるといえる．そして，周囲雑音・SN 比の急激な変動に対してこれらの音声 HMM を頑健にしておく必要がある．

本論文では，単語音声 HMM の各状態あるいは各分布のパワーによる重み付け分散拡大を行い，雑音の影響を受けやすい状態からの出力確率を，観測ベクトルの違いにより大きく変動しないように制御する手法を提案する．これにより，認識時の雑音条件の変動に対して，HMM の耐雑音性を向上させる．本手法による認識性能の改善効果は，単語認識実験で評価する．そして，定常時が比較的静かな環境であればクリーン音声 HMM，騒音環境では雑音付加音声 HMM を使用することで，その雑音条件（マッチド条件）での高い認識性能を維持するとともに，異なる雑音条件となった場合でも従来の性能低下を顕著に改善できることを示す．

以下，本論文では，**2.** で従来の HMM 分散制御手法と本手法との相違点について述べ，**3.** で提案する重み付け分散拡大手法について説明する．**4.** で提案法の評価のための実験方法について述べ，**5.** で各種 HMM における実験結果について示す．最後に **6.** において本論文の結論について述べる．

2. 従来の HMM 分散制御手法と本手法の概念

雑音下音声認識における HMM の分散制御に関しては，これまでに HMM 合成における改善方法として，前述の雑音 HMM 分散拡大法 [7] や分散縮小後の合成に基づく方法 [11] が提案されている．文献 [11] では，音声，雑音の各 HMM の分散を $1/\alpha$ で縮小して各分布の中心部を強調してから HMM 合成を行い，合成後に分散拡大（α 倍）によりその強調を解除するもので，HMM の各状態・分布で分散の縮小率と拡大率は SN 比に関係なく常に一定である．一方，文献 [7] では，音声 HMM の分散は変更せずに，雑音 HMM で設定された分散拡大率も音声 HMM の各状態・分布に対して一定である．ただし，合成時の SN 比が低下すると，合成後の分散に対する雑音の分散拡大の寄与度合が増す．しかしながら，同 SN 比であっても，音声の分散値によっては，合成後の分散が雑音の分散拡大により大きく影響を受ける場合（音声の分散が小さい場合）もあれば，さほど変わらない場合（音声の分散が大きい場合）もある．また，HMM 合成に基づく手法では，分散の変更は平均値ベクトルにも影響を及ぼすという問題がある．更に，HMM 合成では，非線形変換処理が複雑で，動的特徴パラメータを扱うのは容易ではなく, [7], [11] では，Δ 成分の分散制御評価はなされていない．そして, [7], [11] は合成処理の枠組みであるため，前述 **1.** のとおり，定常時での性能は雑音付加音声 HMM と比べると低くなる．

これに対して，本論文では，定常時のマッチド条件で学習された音声 HMM の高性能を利用することを目的としている．このため，合成処理の枠組みを用いることなく，それら音声 HMM 自身の重み付け分散拡大により，異なる雑音条件での耐雑音性を向上させる手法を検討する．まず，本手法は音声 HMM の各状態・分布でパワーに基づいて明確に分散拡大率を設定するもので，音声の分散値の大小に関係なく所定の分散拡大が実行され，その結果は直接，音声 HMM の各特徴パラメータに反映される．また，本手法では平均値ベクトルに影響を与えず分散のみを制御できる．更に，本手法では Δ 成分の分散制御も容易に実行可能で，Δ 成分を含めた最適な分散拡大条件を求めることを目指している．

3. パワーによる重み付け分散拡大

本章では，提案手法について述べる．

3.1 重み付け分散拡大手法

HMM で表現されるモデルによって，観測ベクトル系列が生起するゆう度は，状態遷移確率と出力確率の積和で表される．このうちの出力確率が適切かどうかは，出力確率分布が認識時のそれに近いかどうかに依存している．定常時であれば，認識時の雑音状況が推定できるため，確率分布を適応させる効果が期待できるが，非定常時の未知の雑音状況に対しては対応できない．そして，認識時の雑音の種類や SN 比が学習時や適応時と異なる場合，連続出力分布型 HMM の各状態の確率分布は，認識時の観測ベクトルに対して適切な分布とはいえなくなる．とりわけ，雑音の影響を受

けやすい確率分布（雑音条件の変動により学習時と認識時で差が生じやすい確率分布，以下，変動性が高い確率分布）による不適切な出力確率を，変動性が低い確率分布による出力確率と同様にモデルゆう度計算に用いることにより，誤認識に寄与してしまうと考えられる．

そこで，分散拡大により，このような変動性の高い確率分布を広げて，その出力確率を観測ベクトルの違いにより大きく変動しないように制御する．これにより，異なる観測ベクトル間で出力確率の比を小さくするとともに，雑音が混入し平均値から離れた観測ベクトルに対して出力確率が極端に小さくならないようにする．そして，雑音に対して比較的変動性が低い確率分布からの出力確率はできるだけゆう度に反映させるのに対して，変動性が高い確率分布からは，分散拡大による出力確率の前記制御により，ゆう度算出に関する寄与を低減させる．このようにして，学習時や適応時とは異なる未知の雑音が付加されたり雑音レベルが変動した場合にも対応できる確率分布を表現する．

単語 HMM の分散拡大方法としては，以下のような重み付けを行う（図 1）．一般に，パワーの低い音声区間は，パワーの高い区間と比べて雑音が変動あるいは新たに付加されたときの影響を受けやすい．このため，HMM の各状態あるいは各分布のパワーが各確率分布の雑音に対する変動性を示す尺度になるという考

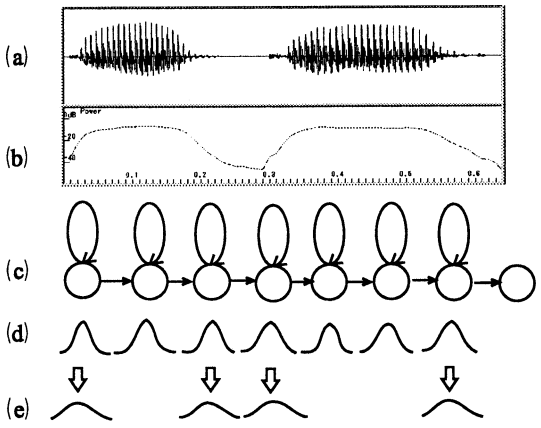


図 1　パワーによる重み付け分散拡大の概念図 (a) 音声波形 ‘東京’, (b) パワー, (c) 単語 HMM, (d) 出力確率分布, (e) 低パワーの状態での分散拡大

Fig. 1　Concept of weighted variance expansion. (a) speech waveform ‘Tokyo’, (b) time series of the power, (c) a word HMM, (d) output probability distributions, (e) variance expansion at the states with low power.

えに基づき，変動性の低い確率分布と比べて，パワーが低い状態に属する確率分布の分散拡大率，あるいはパワーが低い確率分布の分散拡大率を大きくする．

ここで，各状態・各分布のパワーとしては，その状態における分布のパワー（平均値ベクトルのパワー項）が利用できる．混合数が 2 以上の混合分布では，パワーによる分散拡大の重み付けを，各状態ごとに行う場合と各分布ごとに行う場合が考えられ，その両方について検討する．分布ごとでは各分布のパワーを使用する．状態ごとにおいては，i 番目の状態のパワー P_i を以下のように算出する．

$$P_i = \sum_{j=1}^{M} \lambda_{ij} \cdot P_{ij}, \qquad \sum_{j=1}^{M} \lambda_{ij} = 1, \tag{1}$$

ここで，λ_{ij}，P_{ij} は，状態 i の j 番目の分布の混合重み係数，パワーである．また，M は混合数である．

3.2　分散拡大率の設定

本手法では，平均値ベクトルは学習で得られたベクトルを変更せずに使用する．分散の拡大率 K は，学習済みの単語 HMM について，全モデル共通に以下のように設定する．混合分布を構成する一つの分布で，ケプストラム及び Δ ケプストラムの各々1 次以上については，対角共分散の各要素（分散）の K を一定とする．そして，全特徴パラメータについて，一つのモデルで状態（または分布）のパワー P によりパワーの低い状態（または分布）に対して K を重み付けする（以下，K-weighted）．また，比較のため，一つのモデルの全状態で K を一定とする場合（以下，K-const）についても検討する．

K-weighted は，以下①～⑤のように各種方法で検討する．

①：全状態（または分布）中のパワー最大値 P_{max} を基準に，P が 10 dB 低下するごとに 4 段階で，その状態（または分布）の K を設定する．

②：P_{max} とパワー平均値 P_{av} から同様に，P_{max} を基準に P が P_s 低下するごとに 4 段階で K を設定する．ここで，$P_s = (P_{max} - P_{av})/2$ である．

③：P_{max} とパワー最小値 P_{min} から同様に，P_{max} を基準に P が P_t 低下するごとに 4 段階で K を設定する．ここで，$P_t = (P_{max} - P_{min})/4$ である．

①～③に対して④，⑤では，K が P_{max} において最小値 K_{min}，P_{min} で最大値 K_{max} となるように，(P_{max}～P_{min}) 間で K を連続的に設定する．

④：

$$K = a/P + b \tag{2}$$

$$a = \frac{P_{max}P_{min}(K_{max} - K_{min})}{P_{max} - P_{min}} \tag{3}$$

$$b = \frac{P_{max}K_{min} - P_{min}K_{max}}{P_{max} - P_{min}}. \tag{4}$$

⑤ ：

$$K = a \cdot P + b \tag{5}$$

$$a = \frac{K_{min} - K_{max}}{P_{max} - P_{min}} \tag{6}$$

$$b = \frac{P_{max}K_{max} - P_{min}K_{min}}{P_{max} - P_{min}}. \tag{7}$$

各拡大率の表記は以下のようにする．HMM の全状態で分散を拡大しない場合を all(1) で表す．K-const の場合は，$K = 3$ では all(3) と表記する．K-weighted の場合は，①〜③については，① (1,2,3,4) のように表記し，P が P_{max}〜$(P_{max} - 10)$ の範囲の状態（または分布）では $K = 1$，以下 10 dB ごとに K を変更し，P が $(P_{max} - 30)$ 以下で $K = 4$ を表す．④，⑤では，④ (1–3) のように表記し，P が P_{max} で $K = 1$，P_{min} では $K = 3$ を表す．なお，混合数が 2 以上の混合分布においては，状態ごとに重み付けした場合を S，分布ごとの場合を D で表す．

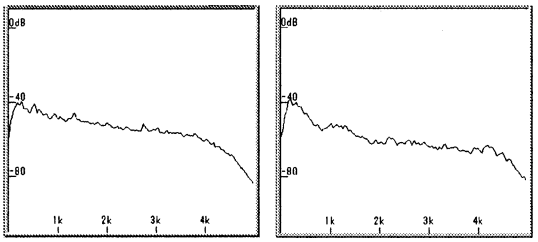
4. 実験方法

本章では，重み付け分散拡大の有効性を評価するための単語認識実験方法について述べる．

音声資料は男性話者 10 名が都道府県名，都市名など 50 種類の単語を静環境で各 5 回発声した音声データを使用した．そのうちの 8 名分を学習，他の 2 名を認識に用い，学習と認識の話者を 2 名ずつ 5 回入れ換えることにより，10 名分の不特定話者単語認識実験で評価した．雑音資料は二つの工場（A，B）の雑音を使用した．A は日本電子協（JEIDA）の騒音データベース [12] の No.7（板金工場），B は別に収録した機械加工工場の雑音である．雑音 A，B の平均スペクトルを図 2 に示す．これらの雑音資料は，コンピュータ上で所定の SN 比となるように学習用，認識用の前記音声資料に付加し，雑音付加音声を作成した．また，雑音付加音声 HMM の学習には，各音声ごとに雑音 A の異なる区間を使用した．

本論文では，SN 比（SNR）は次のように定義している．

$$SNR = 10\log_{10}\left(\frac{P_{speech}}{P_{noise}}\right) \tag{8}$$



(a) Noise A　(b) Noise B

図 2　雑音の平均スペクトル
Fig. 2　Averaged spectra of noises.

ここで，P_{speech}，P_{noise} はそれぞれ単語発声区間ごとの音声，その区間の付加雑音のパワーである．

各データは，サンプリング周波数 10 kHz，16 ビットでディジタル化し，フレーム長 25.6 ms（ハミング窓），フレーム周期 10 ms で LPC 分析（12 次）を行った．特徴パラメータは，1〜10 次 LPC メルケプストラム（cep），正規化した対数パワー（pow）及びそれらの動的特徴パラメータ（Δcep，Δpow）を使用した．単語 HMM は 10 状態 9 出力の混合正規分布型（混合数は 1 及び 2 を使用）で対角共分散を用いた．

そして，比較的静かな環境で非定常雑音が発生する場合を想定して，クリーン音声 HMM を用いたときの未知の雑音の混入に対する認識性能の改善を，K-const と K-weighted とで比較して評価する．また，非定常高騒音環境での急激な雑音条件の変動を想定して，雑音付加音声 HMM を用いたときの学習時とは異なる雑音条件での性能改善を，同様に K-const と K-weighted とで比較して評価する．

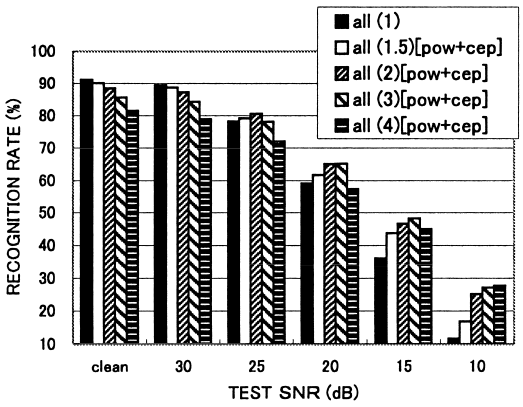
雑音付加音声 HMM としては，15 dB-HMM と 5SNR-HMM を使用した．15 dB-HMM は，SN 比が 15 dB の雑音付加音声を用いて学習した．5SNR-HMM は，各話者・各単語ごとに，同一単語 5 回発声の各音声に対して 5 種類の SN 比（25，20，15，10，5 dB）の雑音付加音声を作成して学習した．

5. 単語認識実験結果

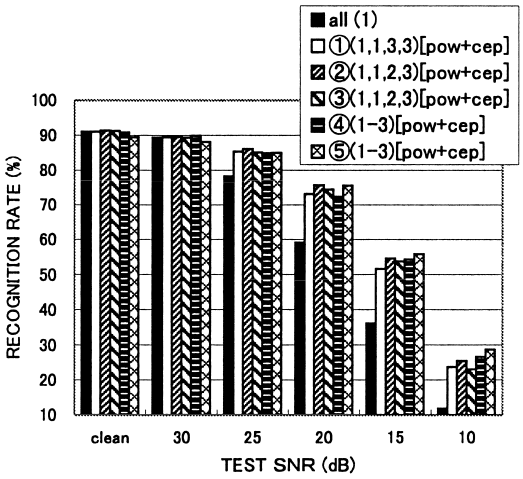
本章では，クリーン音声 HMM と雑音付加音声 HMM を使用して，不特定話者単語認識実験が行われる．実験結果は，種々の雑音条件での単語認識率を求めることにより得られる．

5.1 K-const と各種 K-weighted との比較実験結果

まず，クリーン音声 HMM（混合数 1）を使用して，



(a) Recognition rates in the case of K-const.



(b) Recognition rates in the case of K-weighted.

図 3 特徴パラメータとして cep と pow のみを用いたクリーン音声 HMM での認識性能［評価雑音：A］
Fig. 3 Recognition results by clean speech HMMs with cep and pow using A as test noise.

K-const と，K-weighted（①〜⑤）とで，認識性能の傾向を比較した．認識評価用雑音は A，特徴パラメータは cep と pow だけを用いた．実験結果を図 3 に示す．図で [·] は分散拡大したパラメータを表す．

図 3 (a) より K-const の場合は，低 SN 比で改善効果が見られたが，高 SN 比では K を大きくすると性能が著しく低下した．これは，認識時の雑音条件が学習時に近い高 SN 比（マッチド条件）の場合は，$K = 1$ が適しているのに，分布の分散拡大によって，その出力確率が $K = 1$ のときのそれと異なり，ゆう度に正しく反映されなくなると考えられる．

これに対して，K-weighted（図 3 (b)）では，高 SN 比でも性能を低下させることなく，低 SN 比においては K-const 以上の改善効果が確認できる．①〜⑤の各 K の値は，各々の設定条件での予備実験において効果のあった値である．中でも②が，全 SN 比で比較的性能が良い．雑音の影響を受けやすくなる P_{av} 以下の状態についてのみ分散拡大することにより，30 dB 以上の高 SN 比で all(1) 以上を維持し，25〜15 dB では all(1) に対する誤り削減率が 29〜40%程度であった．⑤は 25 dB 以下で改善率が高いが，高 SN 比では all(1) と比べて 1%以上低下する．

以後の評価実験では，重み付け設定方法として②を用いた．また，図 3 (a) より，K を大きくすると学習時とは異なる SN 比（ミスマッチド条件）で改善率が高くなるが，学習時の SN 比（マッチド条件）では性能が低下する傾向がある．このため，マッチド条件での高認識性能を維持するという本研究の目的から，K-weighted と K-const との以後の比較は，学習時の SN 比での性能低下が 1%以内となる各々の K の値を用いた．

5.2 Δ 成分を含むクリーン音声 HMM での実験結果

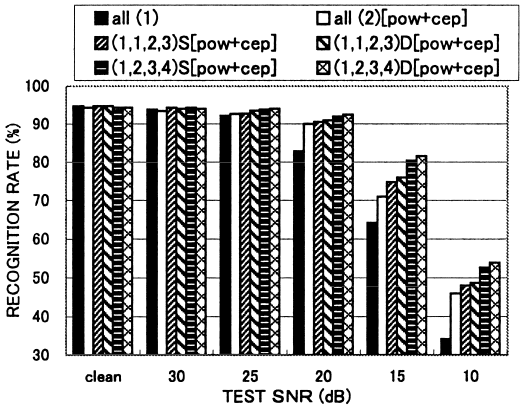
クリーン音声 HMM（混合数 2）を使用した．認識評価用雑音は A，特徴パラメータは cep，pow，Δcep，Δpow を使用した．実験結果を図 4 に示す．

図 4 (a) から，この実験でも，K-weighted により性能低下が低減された．all(3) 以上の K-const ではクリーンにおける性能低下が 1%を超えた（all(3) で 1.3%低下）ため，all(2) と比較したが，広範囲の SN 比で K-weighted の方が高い認識率を示している．また，S と D の比較では，25 dB 以下で D の方が上回った．

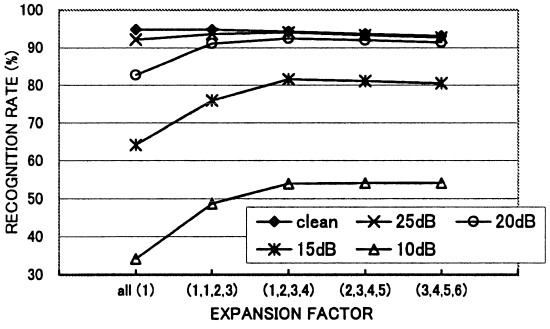
図 4 (b) は，D[pow+cep] を使用して K を大きくしたときの認識性能の変化を表している．10 dB 以外では，(1,2,3,4) より K を大きくすると性能が低下する．25 dB 以下では，(1,2,3,4)D が比較的改善率が高く，20 及び 15 dB で all(1) に対する誤り削減率は 48%以上になる．

次に，(1,2,3,4)D で，Δcep，Δpow も含めて分散拡大した場合及び pow だけの場合と比較した．その結果（図 4 (c)）より，Δ 成分の分散拡大を加えても効果は見られなかった．また，pow のみ分散拡大のときの all(1) に対する性能向上はわずかで，結果として [pow+cep] が最も改善率が高い．

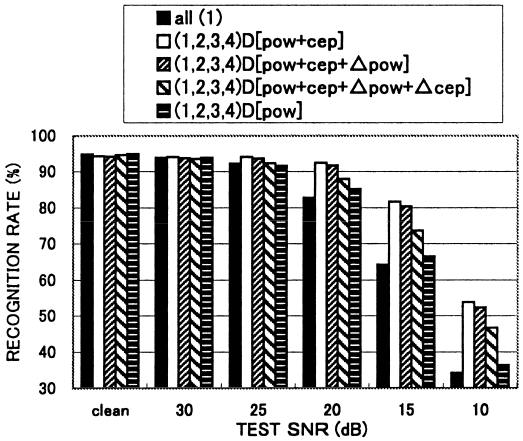
クリーンから 10 dB までの範囲での平均認識率は，all(1) の 77.0%，K-const で最も性



(a) Comparison of recognition rates between S (weighting K at every state) and D (weighting K at every distribution), as well as, between K-const and K-weighted.



(b) Dependency of recognition rates on increase of K (expansion factor).



(c) Recognition rates for feature parameters with variance expansion.

図 4　全特徴パラメータを用いたクリーン音声 HMM での実験結果［評価雑音：A，重み付け：②］

Fig. 4　Recognition results by clean speech HMMs with all feature parameters using A as test noise and ② as K-weighted.

能が高かった all(2)[pow+cep] の 81.2%に対して，(1,2,3,4)D[pow+cep] は 85.1%であった．K は (1,2,3,4) が適していたが，これは，パワーが比較的高い分布は高 SN 比では変動性が低く，その分散拡大は高 SN 比での性能低下となるため，それ以外の分布の重み付け分散拡大が有効であることを示している．

5.3　15 dB-HMM での耐同種雑音の実験結果

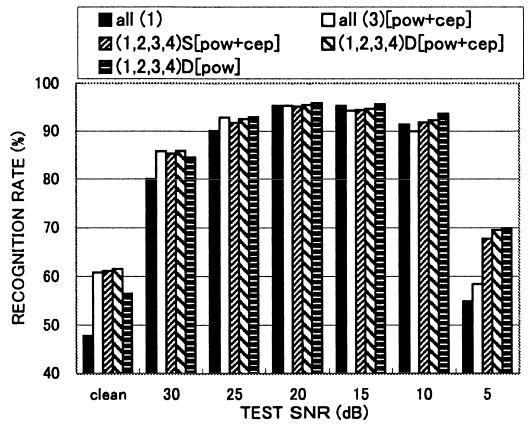
15 dB-HMM（混合数 2）を使用した．学習時と同環境の雑音である A の学習時と異なる区間を，認識評価用雑音として使用した．特徴パラメータは cep，pow，Δcep，Δpow を使用した．実験結果を図 5 に示す．

図 5 (a) では，15 dB での性能低下が 1%以内であった all(3)[pow+cep] と (1,2,3,4)[pow+cep] の S 及び D で比較した．30 及び 25 dB では，K-weighted は，all(1) に対しては改善されるが，all(3) と比べると改善されていない．しかし，10 dB 以下の低 SN 比においては K-weighted が優位性を示し，特に 5 dB では (1,2,3,4)D は all(3) に対して認識率が 10%以上高く，改善効果が大きいことが明らかである．S と D の比較では，すべての SN 比で D による認識率の方が高く，分布ごとに詳細に重み付けした方がわずかながら効果があった．また，高 SN 比以外では，[pow] が [pow+cep] を上回っており，学習時と同種の雑音の SN 比変動に対しては，pow の分散拡大効果が大きいことが分かる．

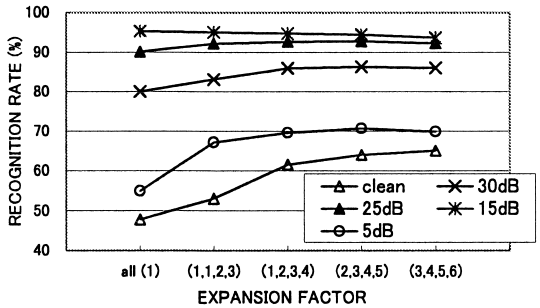
図 5 (b) より，D[pow+cep] の使用では，(2,3,4,5) が 15 dB で all(1) に対して 0.9%低下するが，異なる SN 比での改善効果が高い．そして，クリーン以外では K を更に大きくしてもこれ以上の性能向上は得られない．

図 5 (c) は，(2,3,4,5)D で Δcep，Δpow も含めて分散拡大した場合を示している．全パラメータを分散拡大した場合は，高 SN 比で改善効果が非常に大きいが，20 dB 以下の低 SN 比では逆に最も性能が低い．これに対して，[pow+cep+Δpow] では高 SN 比での改善だけでなく，全範囲で [pow+cep] 以上の性能が得られ，全般的に比較的良好な認識特性を示した．ただし，20〜10 dB では [pow+Δpow] の方が上回った．なお，[pow] は 20 dB 以下で [pow+Δpow] と同程度，25 dB 以上では [pow+Δpow] より性能は低かった．

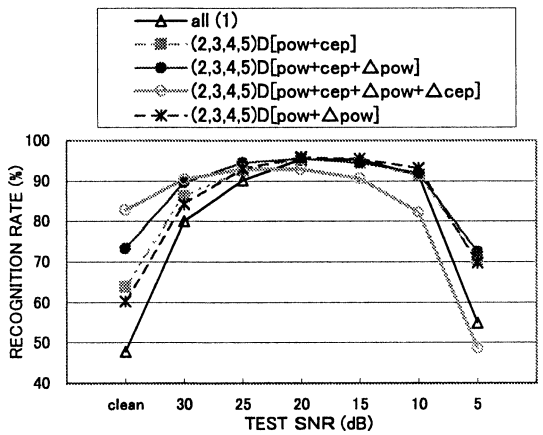
クリーンから 5 dB までの範囲での平均認識率は all(1) の 79.3%，K-const では最も性能が高かった all(3)[pow+cep+Δpow] の 84.0%に対して，(2,3,4,5)D[pow+cep+Δpow] では 87.4%であった．た



(a) Comparison of recognition rates between S (weighting K at every state) and D (weighting K at every distribution), as well as, between K-const and K-weighted.



(b) Dependency of recognition rates on increase of K (expansion factor).



(c) Recognition rates for feature parameters with variance expansion.

図 5　15 dB-HMM（SN 比 15 dB の雑音付加音声で学習）の雑音 A（学習時と同環境の雑音）での認識性能［重み付け：②］

Fig. 5　Recognition results by 15 dB-HMMs, which were trained with noisy speech of 15 dB SNR(noise A), using A as test noise and ② as K-weighted.

だし，20〜10 dB では，(2,3,4,5)D[pow+cep+Δpow] 及び all(1) の 94.0%，K-const では最も性能が高かった all(2)[pow+Δpow] の 94.8%に対して，(1,2,3,4)D [pow+Δpow] が 95.3%（クリーンから 5 dB までは 85.4%）を示した．

以上より，学習時と同種の雑音の付加に対しては，認識時の SN 比が学習時に近い場合は [pow+Δpow]，SN 比が大きく異なる場合には [pow+cep+Δpow] が最も適している．このことから，認識時の雑音状況が学習時に近いときは，雑音レベルの変動は cep よりも主に pow に影響し，レベル変動が大きくなると pow だけでなく cep への影響度合も増してくると推察できる．

K に関しては，以下のことが確認できた．同種雑音の学習時 SN 比の周辺については，(1,2,3,4) が適していた．クリーン音声 HMM の場合と同様，パワーが比較的高い分布は学習時 SN 比の周辺では変動性が低く，それ以外の分布の重み付け分散拡大が有効と考えられる．これに対して，広範囲 SN 比では (2,3,4,5) が最高性能を示した．これは，雑音条件の変動が大きい場合，パワーの高い分布も変動性が高くなるためと考えられ，全分布について重み付け分散拡大を適用することで改善性能が高くなる．

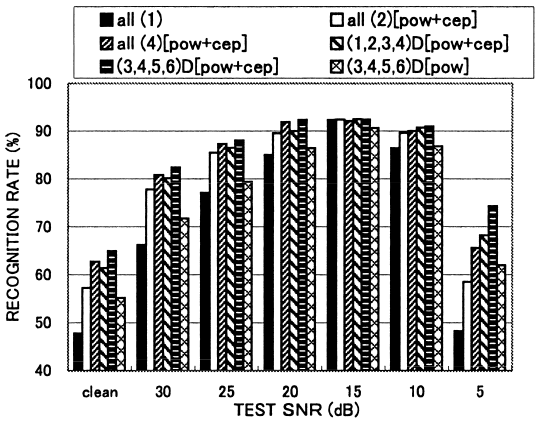
5. 4　15 dB-HMM での耐異種雑音の実験結果

前節と同じ 15 dB-HMM（混合数 2）を使用し，認識評価用雑音として学習時とは異環境の雑音である B を使用した．特徴パラメータは cep，pow，Δcep，Δpow を使用した．実験結果を図 6 に示す．

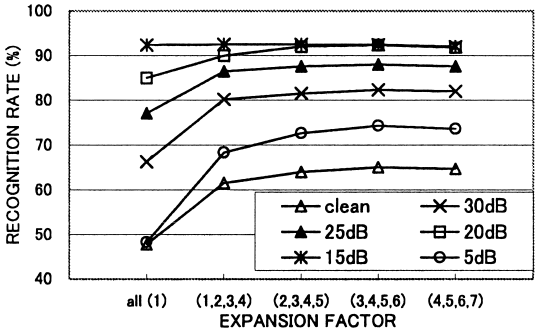
図 6 (a) では，15 dB での認識性能が 0.3%以内で近い all(2) と (1,2,3,4)D，all(4) と (3,4,5,6)D で比較した．低 SN 比では K-const と比べて K-weighted の改善効果が大きく，高 SN 比でも K-weighted の方が高い認識率を示した．(3,4,5,6)D[pow+cep] では 20〜30 dB 及び 5 dB で，all(1) に対する誤り削減率が 47%以上である．また，pow のみ分散拡大は，all(1) に対する改善率が低く，同種雑音の場合と異なり，異種雑音に対しては pow よりも cep の分散拡大の方が効果があることが分かる．これは，異種雑音の混入が cep に対して大きな影響を及ぼすためと考えられる．

図 6 (b) より，異種雑音の混入では，前節の同種雑音の場合と比べて更に K が大きい方が効果的である．広範囲の SN 比では (3,4,5,6) が最も改善率が高い．

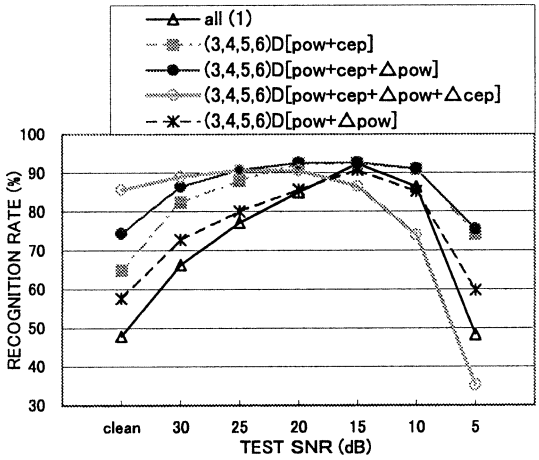
図 6 (c) より，Δ 成分も含めて全パラメータを分散拡大したときの傾向は前節と同様である．高 SN 比で改善



(a) Comparison of recognition rates between K-const and K-weighted.



(b) Dependency of recognition rates on increase of K (expansion factor).



(c) Recognition rates for feature parameters with variance expansion.

図 6　15 dB-HMM（SN 比 15 dB の雑音付加音声で学習）の雑音 B（学習時とは異環境の雑音）での認識性能

Fig. 6　Recognition results by 15 dB-HMMs, which were trained with noisy speech of 15 dB SNR(noise A), using B as test noise and ② as K-weighted.

効果が非常に大きく，(3,4,5,6)D の使用により，all(1) に対する誤り削減率はクリーン及び 30 dB で 67%以上に達する．ただし，15 dB 以下の SN 比では劣化する．これに対して，[pow+cep+Δpow] では 15 dB でも all(1) を 0.3%上回り，その他の広範囲の SN 比で良好な改善特性が得られた．一方，[pow+Δpow] の改善性能は低く，特に 20〜10 dB では平均認識率が all(1) 以下であり，前節とは全く異なる特性を示している．

以上の結果より，異種雑音の混入に対しては，パワー成分のみの分散拡大では改善効果が小さく，[pow+cep+Δpow] の分散拡大が優れていることが分かる．クリーンから 5 dB までの範囲での平均認識率は all(1) の 71.9%，K-const では最も性能が高かった all(4)[pow+cep+Δpow] の 83.2%に対して，(3,4,5,6)D[pow+cep+Δpow] は 86.2%で，all(1) に対する誤り削減率は 50.9%であった．

K としては，同種雑音の場合より大きい (3,4,5,6) が最高性能を示した．これは，雑音条件の変動が大きい場合，パワーの高い分布も変動性が高くなるが，特に異種雑音に対しては，学習時 SN 比においても分布の変動性が高くなるためであり，全分布について大きい K による重み付け分散拡大が性能改善に適している．

5.5　5SNR-HMM での実験結果

5SNR-HMM（混合数 2）を使用した．特徴パラメータは cep，pow，Δcep，Δpow を使用した．認識評価用雑音として，学習時と同環境の雑音 A を使用した場合の実験結果を図 7，異環境の雑音である B を使用した場合の実験結果を図 8 に示す．

図 7 より，同種雑音の場合は，[pow+cep+Δpow] の使用では 25〜5 dB で性能改善が得られず，学習時以下の SN 比では [pow+Δpow] が改善に適していることが分かる．特に，0 dB における改善が顕著であった．クリーンから 0 dB までの範囲での平均認識率は，all(1) で 85.5%，K-const では最も性能が高かった all(2)[pow+Δpow] で 86.4%，15 dB-HMM の学習時 SN 比の場合に最適であった (1,2,3,4)D[pow+Δpow] では 87.0%であった．これに対して，(1,1,1,4)D[pow+Δpow] が 87.7%を示した．

異種雑音の場合では，分散拡大の有効なパラメータは **5.4** と同様 [pow+cep+Δpow] であった．そして，K-weighted により広範囲 SN 比で性能改善が得られ，特に低 SN 比においては，K-const と比べて改善性が高いことが明らかである（図 8）．K は（2,3,4,5）が

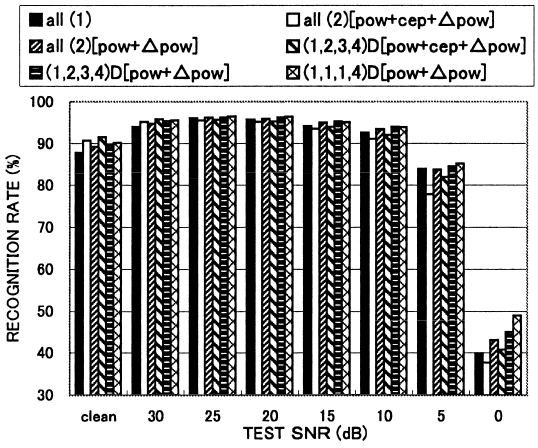


図 7　5SNR-HMM（5 種類の SN 比の雑音付加音声で学習した HMM）の雑音 A（学習時と同環境の雑音）での認識性能

Fig. 7　Recognition rates by 5SNR-HMMs, which were trained with noisy speech of five kinds of SNRs(noise A), using A as test noise and ② as K-weighted.

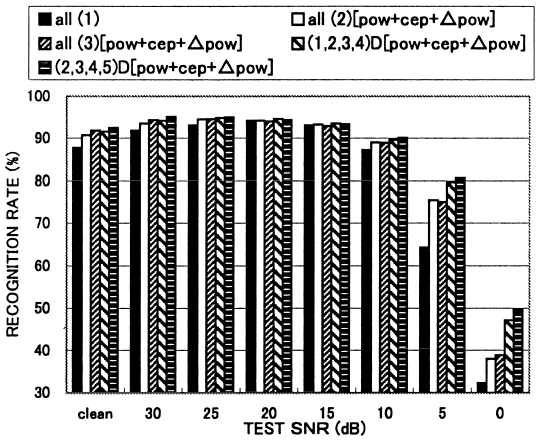


図 8　5SNR-HMM（5 種類の SN 比の雑音付加音声で学習した HMM）の雑音 B（学習時とは異環境の雑音）での認識性能

Fig. 8　Recognition rates by 5SNR-HMMs, which were trained with noisy speech of five kinds of SNRs(noise A), using B as test noise and ② as K-weighted.

最高性能を示し，更に K を大きくしても，性能向上は得られなかった．クリーンから 0 dB までの範囲での平均認識率は，all(1) で 80.4%，K-const では最も性能が高かった all(3)[pow+cep+Δpow] で 83.8%であった．これに対して，(2,3,4,5)D[pow+cep+Δpow] では 86.3%に向上した．

同種雑音と異種雑音のいずれも，広範囲の SN 比においては 15 dB-HMM の場合と比べて K は小さい方が適していた．また，同種雑音では，よりパワーの低い分布についてのみ重み付け分散拡大した場合に最高性能を示した．以上は，5SNR-HMM の分布が，あらかじめ広範囲の SN 比に対応できるように学習されており，広範囲の SN 比で 15 dB-HMM の分布と比べて変動性が低くなるためと考えられる．

6. むすび

雑音条件の急激な変動に対する単語 HMM の性能低下を改善するため，雑音の影響を受けやすい低パワーの状態，分布について分散拡大率を重み付けすることで，その状態の出力確率を，観測ベクトルの違いにより大きく変動しないように制御する手法を提案した．そして，学習時とは異なる加法性雑音の種類，SN 比における認識性能の改善効果を，語彙数 50 の単語音声と 2 種類の工場の雑音を使用した不特定話者単語認識実験で検討した．HMM としては，クリーン音声 HMM と雑音付加音声 HMM を使用した．

評価実験の結果，いずれの HMM においても，以下の確認ができた．

(1) 本手法により，加法性雑音の条件の変動に対して広範囲の SN 比で平均認識率が向上し，雑音下小語彙音声認識における単語 HMM の耐雑音性が改善できた．

(2) 特に，学習時より低い SN 比の雑音付加音声に対しては，重み付け分散拡大は拡大率一定よりも認識性能を顕著に改善できた．

以上のことは，モデルのゆう度計算の観点からは，雑音付加・変動に対して変動性が高い出力確率分布を有する状態からの出力確率のモデルゆう度への寄与を低減させる効果が現れていると考えられる．

また，分布ごとと状態ごとの重み付けの比較では，わずかに分布ごとの方が効果が見られた．

更に，分散拡大が有効なパラメータとしては，以下のことが確認できた．

- クリーン音声 HMM

パワー (pow)+ケプストラム (cep) の分散拡大が SN 比の低下に対して有効である．

- 雑音付加音声 HMM

認識時の雑音の種類が学習時と同じか異なるかで改善性能に違いが生じる．同種雑音に対しては，学習時の SN 比 ±5 dB 以内では pow+Δpow，それ以上に SN 比が異なる場合及び異種雑音に対しては

pow+cep+Δpow の分散拡大による改善効果が高い．

分散拡大率の設定に関しては，以下の知見が得られた．すなわち，認識環境が，学習時に使用した雑音と同種の雑音でかつ SN 比変動が学習時 SN 比の周辺と推定できる場合は，パワーの高い分布は分散を変更せず，その他の分布についての 2〜4 倍程度の重み付け分散拡大が有効である．広範囲の SN 比の同種雑音付加音声を学習に使用できた場合は，パワーの低い分布のみの分散拡大の方が性能が向上する．一方，学習時に使用できない様々な雑音の混入あるいは SN 比の大きな変動が想定される場合は，パワーの高い分布も含めた重み付け分散拡大が効果的である．特に，異種雑音に対しては比較的大きい拡大率（3〜6 倍程度）の設定により高い改善性が期待できる．以上の分散拡大率設定の考え方は，他の環境雑音においても適用可能と推察できる．ただし，最適値は対象となる認識環境や語彙により異なると考えられるため，最適拡大率の決定方法は今後の検討課題である．

本論文では，単語 HMM について有効性を確認したが，今後，他の HMM（音素 HMM 等）に対しても本手法の適用と有効性を評価する必要がある．また，パワー以外のパラメータによる重み付け分散制御方法についても検討する．

文　献

[1] Y. Gong, “Speech recognition in noisy environments: A survey,” Speech Communication, vol.16, pp.261–291, 1995.

[2] S.V. Vaseghi and B.P. Milner, “Noise compensation methods for hidden Markov model speech recognition in adverse environments,” IEEE Trans. Speech Audio Process., vol.5, no.1, pp.11–21, Jan. 1997.

[3] M.J.F. Gales and S.J. Young, “Cepstral parameter compensation for HMM recognition in noise,” Speech Communication, vol.12, no.3, pp.231–239, 1993.

[4] F. Martin, K. Shikano, and Y. Minami, “Recognition of noisy speech by composition of hidden Markov models,” Proc. Eurospeech, pp.1031–1034, 1993.

[5] S.F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoust. Speech Signal Process., vol.27, no.2, pp.113–120, Feb. 1979.

[6] B.-H. Juang, “Speech recognition in adverse environments,” Computer Speech Language, vol.5, pp.275–294, 1991.

[7] H. Matsumoto and H. Ubukata, “Robust HMM to variation of noisy environments based on variance extension of noise models,” Proc. Eurospeech, pp.2387–2390, 1999.

[8] J.-W. Hung, J.-L. Shen, and L.-S. Lee, “Improved robustness for speech recognition under noisy conditions using correlated parallel model combination,” Proc. ICASSP, pp.553–556, 1998.

[9] D. Matrouf and J.-L. Gauvain, “Model compensation for noises in training and test data,” Proc. ICASSP, pp.831–835, 1997.

[10] 中川聖一，“音声認識研究の動向,” 信学論（D-II），vol.J83-D-II，no.2, pp.433–457, Feb. 2000.

[11] T.-H. Hwang and H.-C. Wang, “Weighted parallel model combination for noisy speech recognition,” Proc. ICSLP, pp.1527–1530, 1998.

[12] S. Itahashi, “Creating speech copora for speech science and technology,” IEICE Trans., vol.E74, no.7, pp.1906–1910, July 1991.

（平成 15 年 6 月 25 日受付，12 月 2 日再受付）

漢野　救泰　（正員）

昭 55 金沢大・工・電子卒．昭 57 東工大大学院修士課程了．同年日本電装（株）入社．昭 59 石川県工業試験場入所．博士（工学）．音声，画像の認識処理に関する研究に従事．日本音響学会会員．

船田　哲男　（正員）

昭 41 金沢大・工・電子卒．昭 46 名大大学院博士課程了．昭 46 金沢大・工・講師．現在同大教授．工博．生体情報処理，音声情報処理の研究に従事．共著「数値解析の基礎」,「音声情報処理」など．IEEE，日本音響学会，日本 ME 学会，情報処理学会各会員．