

有声音検出に基づくロンバード音声認識と工場における 検反システムへの応用

漢野 救泰[†] 船田 哲男^{††}

Lombard Speech Recognition Based on Voiced Sound Detection and Application
to the Fabric Inspection System in Factories

Sukeyasu KANNO[†] and Tetsuo FUNADA^{††}

あらまし 工場のような非定常高騒音下においては、雑音の混入及び音声特徴の変形により、音声認識性能が低下する。この課題に対して、本論文では、工場の騒音下において有効な有声音区間検出に基づく実用に適した雑音混入ロンバード音声認識手法について検討した。男性話者3名がそれぞれ発声した語彙数120を用いた単語認識実験より、WGD尺度が騒音下発声音を標準パターンとして利用することで雑音混入ロンバード音声に対する認識性能が高いことを明らかにした。そして、本論文で提案する有声音に基づく単語区間利用の有効性及び有声音区間重み付けの効果を確認した。また、これらの評価結果をもとに、特定話者音声認識による検反システムを開発した。検査工場（騒音レベル：75～90 dBA）において、音声により織物欠点名（語彙数61）を入力する動作実験を行った結果、認識率93.1%を達成し、操作性向上が可能なことがわかった。

キーワード 音声認識，ロンバード効果，有声音，検反システム，騒音環境

1. ま え が き

生産現場での技術開発の中で、自動化は最も重要な課題の一つである。本研究は、音声認識技術の利用により、織布業の生産工程効率化に大きく寄与する織物検査装置の自動化を目指すものである。この検査装置は検反システムと呼ばれている。織物検査の工程は、従来から、熟練した検査作業者がこのシステムを使用して、織物巻取り機械を稼働させて目視による検査を行い、欠点発見ごとにその結果をキーボードやタッチパネルなどにより手で入力して品質管理することにより行われている。このため、目視検査と結果入力の二つの工程を別々に繰り返し行っており、非効率的である。そこで、目視検査で発見した織物欠点の名称を、手入力から音声入力に切り換えることで、目視と同時に目をそらすことなく欠点入力を可能にし、操作の効

率化・作業者の負担軽減を図る。しかしながら、工場の騒音のため、近年の音声認識技術の向上[1]にもかかわらず工場内での実用化は進んでいない。工場内での音声認識性能低下の原因として、非定常高騒音による雑音の混入及び騒音下での発声における音声特徴の変形（ロンバード効果）[2]～[4]が考えられる。また、騒音環境では音声区間の検出誤りが生じやすく、これに起因する認識率の低下も知られている。したがって、工場内での実用化を考えた場合には、音声区間検出とロンバード効果の二つの重要な問題が存在する。

音声の検出については、これまでにピッチ対応型低域LPC分析手法とLPC適合度を使用した有声音区間検出手法を提案し、工場でのロンバード音声に適していることを示した[5]。有声音区間に限定したのは、高騒音下においては無声音を含めた区間検出方法では誤検出が避けられず、有声音区間を確実に検出する方が有効と考えたためである。そして、単語の最初の有声音区間の始端から最後の有声音区間の終端までを単語区間として定義し、以下、この有声音に基づく単語区間を有声音単語区間と記述する。そして、文献[5]では、LPC適合度に基づく本有声音単語区間検出手法が、高騒

[†] 石川県工業試験場，金沢市
Industrial Research Institute of Ishikawa, Kanazawa-shi,
920-0223 Japan

^{††} 金沢大学工学部，金沢市
Faculty of Engineering, Kanazawa University, Kanazawa-shi,
920-8667 Japan

音下での発声音の検出実験により、SIFT 法に基づく検出手法よりも優れていることを明らかにした。ただし、この文献では、有声音検出率、有声音単語区間検出率で評価したもので、音声認識性能では評価していなかった。このため、騒音下音声認識における本手法による有声音単語区間利用の有効性に関しては、これまで明らかにされていなかった。また、従来のロンバード音声認識の研究 [6] ~ [14] では、標準パターンとして認識時と同様な騒音条件での発声音を用いた場合に性能が良く、静環境発声音を用いた場合には、別に収録した模擬騒音下ロンバード 音声を用いた適応処理が必要となることが示されている。しかし、実的に必要な実環境下ロンバード 音声の区間検出手法を含めた認識性能評価はなされていない。そこで、本研究では、有声音単語区間検出手法を利用した実環境下ロンバード 音声認識手法を検討するとともに、最も性能が期待できる方式として、標準パターンに実環境下発声音を使用できる認識システムを開発した。

本研究では、音声認識系として、2. で述べるようにパターンマッチングを用いており、まず、雑音混入ロンバード 音声に対する認識性能から本システムに適した距離尺度の頑健性について評価する。そして、標準パターンとして、有声音単語区間を利用した場合と有声音単語区間前後の無声音も含めた音声区間（一般的な音声区間）を利用した場合とで比較を行い、本研究提案の有声音単語区間利用の有効性について明らかにする。また、高騒音下において、パワーの強い有声音は無声音と比べて雑音混入に対する信頼性が高いと推定されることから、有声音区間で照合度の重み付けを行う手法を提案し、その効果を示す。そして、これらの評価結果をもとに、工場の検査工程で利用できる音声入力検反システムの開発を行い、騒音環境下で織物欠点名を音声で入力する動作実験により、このシステムの実用性を評価した。

以下、本論文では、2. で作業環境での実用に適した雑音混入ロンバード 音声認識方式について述べ、3. で有声音単語区間の検出手法について説明し、4. で使用した距離尺度と有声音区間重み付け手法について示す。5. で騒音下での有声音単語区間に基づく特定話者単語認識実験方法について述べ、6. でその認識実験結果について、特に距離尺度の実環境下ロンバード 音声に対する性能、有声音単語区間利用の有効性及び有声音区間重み付けの効果について示す。そして、7. で 6. までの検討結果をもとに開発した音声入力検反シス

テムについて述べる。

2. 作業環境に適したロンバード 音声認識方式

本章では、ロンバード 音声認識方法に関する研究について概観し、工場の作業者が認識システムを使用する場合での実用に適した方式について検討する。

騒音下ではロンバード 効果による音声認識性能の低下が顕著 [6] であるとともに、この効果は話者ごと・音韻ごとに異なる [2] ~ [4], [7] ことが報告されている。このため、静環境発声音で学習したモデル、パターンに対して各種適応処理などによりロンバード 音声を認識する手法が提案されている [4], [7] ~ [14]。ただし、そのほとんどは、ヘッドホン聴取による模擬騒音下発音により、静環境音声に対する特徴変形が抽出しやすい無雑音のロンバード 音声データを大量に収集して適応に用いるなど、研究室段階で性能改善がなされている。しかし、このような発音の実環境での発音と特徴変形が一致するかどうかは定かではない。実際の高騒音環境下での発音による周囲の雑音が混入したロンバード 音声に対しては、実的に必要な認識評価がほとんどなされていない。また、実環境下での使用を考えた場合には、作業者にヘッドホン聴取発音を強いるのは非現実的であり、作業者から発音を収集できるのは通常の騒音下作業場所のみである。そして、認識性能としては、認識時と同じ騒音条件での発音による学習が最も優れている [14]。したがって、作業工程・認識性能の観点から、実的には認識時と同じ環境での発音を学習に用いる方法が最善である。この場合、HMM に基づく認識システムの枠組みでは、大量の学習データを収集する必要があるが、作業者の負担が大きく、現状では困難である。一方、パターンマッチング手法では、特定話者に限定すれば、標準パターンとして少ない発音回数で使用可能であり、雑音に頑健な距離尺度も開発されている。そして、織物検査工程では検反システム 1 台につき作業者 1 名が固定の特定話者認識であり、かつ認識対象語彙も織物欠点名称 30 ~ 100 種類程度で固定である。

以上より、特定話者で限られた語彙数という条件で、作業者の負担軽減、高騒音下発音変形の影響低減を考慮し、検反システムとしての実用に適した認識方式として、認識時と同環境での発音を標準パターンとして利用できるシステムを構成し、認識系に DP マッチングを使用する。

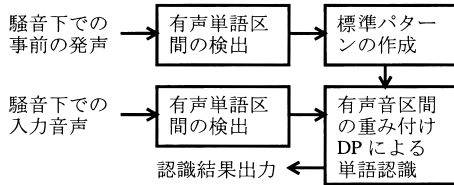


図 1 工場での本ロンバード音声認識方式の構成図
Fig. 1 Block diagram of a proposed method for Lombard speech recognition in factories.

図 1 に本方式の構成を示す．1. で述べた二つの課題に対して，音声の区間検出問題については，次章で述べるピッチ対応型低域 LPC 分析手法と LPC 適合度を使用した有声音区間検出手法を使用する．また，騒音下ロンバード効果問題については，実環境下発声音の有声音区間を標準パターンとして使用し，入力音声の有声音区間について有声音区間を重み付けして認識を行う点が本方式の特徴である．

3. ピッチ対応型低域 LPC 分析手法と LPC 適合度による有声音区間検出

本章では，工場の非定常騒音下における有声音検出に適したピッチ対応型低域 LPC 分析手法と LPC 適合度及び有声音区間検出手法について述べる．

3.1 ピッチ対応型低域 LPC 分析手法

ピッチ対応型低域 LPC 分析手法は，工場騒音下での雑音の重畳した有声音を効率的に抽出できるように，高域と比べて雑音の影響の少ない低域に着目した，狭帯域 LPC 分析手法である．この手法では，声帯振動によるピッチ周波数とその高調波に対応するスペクトルピークを，全極型モデルの極とみなして分析を行い，その適合の度合から有声音を検出する．一般に無声音はパワーが弱いため，その信号が埋もれる高騒音下では雑音との区別が非常に困難である．この場合，無声音を含む区間検出では誤検出が多く実用的でないため無声音を検出対象外とする．この手法での分析帯域は，音声パワーが工場雑音に対して比較的優勢な帯域で，かつ多くの話者に共通してピッチ成分のピークが良好に観測された帯域である 600 Hz 程度以下とした．また，ロンバード効果では，表 1 に示すようにピッチ周波数が上昇する傾向があり，この手法では低分析次数化が可能となる．表 1 は後述 (5.1) の 3 名の話者の静環境での通常の発声と騒音環境での発声におけるピッチ周波数の測定結果であり，いずれの話者も騒音環境発声では静環境発声の約 1.6 倍となっている．例

表 1 話者のピッチ周波数 (Hz)
Table 1 Speakers and pitch frequencies (Hz).

話者		A	B	C
静環境	範囲	110~180	100~140	90~120
	平均	160	125	110
騒音環境	範囲	170~280	140~220	120~190
	平均	260	200	175

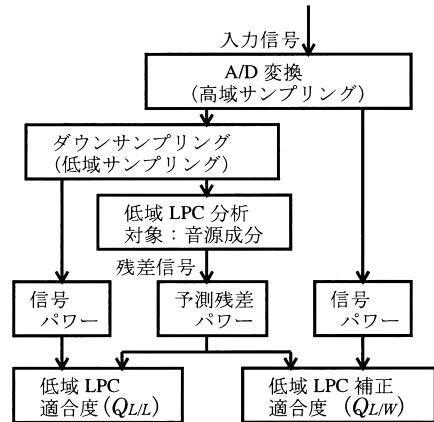


図 2 LPC 適合度算出ブロック図
Fig. 2 Block diagram of calculation for LPC conformity degrees.

えば，話者 B ではピッチ周波数が 75 Hz 上昇 (125 200 Hz) し，本研究での実験による 640 Hz 以下の範囲では，スペクトルピークは静環境発声で 5 個，騒音環境発声で 3 個であるため，有声音に適合した分析次数は 4 次程度低くできる．分析次数が高くなると雑音に対する適合の度合が高くなり，有声音区間の検出性能が幾分低下する傾向があることがわかっているため [5]，分析次数を低くできる方が効果的である．

3.2 LPC 適合度

図 2 に，有声音検出用特徴パラメータである LPC 適合度の算出ブロック図を示す．

入力信号 (広帯域信号) x_n をダウンサンプリングして，低域信号 y_n を生成する． y_n をハミング窓 (フレーム長: L) で切り出し，分析次数 p で LPC 分析し，第 i フレームの線形予測係数 $\{a_{ij}\}$ ($i = 1, 2, 3, \dots, I; j = 1, 2, 3, \dots, p$) を算出する．ここで I は全フレーム数であり， $n = (i - 1)L + k$ である．フレーム周期はフレーム長と同じ L としている．残差信号時系列 z_n は，第 i フレーム k 番目 ($1 \leq k \leq L$) の残差データを次式のように求めることで作成する．

$$z_n = y_n + \sum_{j=1}^p a_{ij} y_{n-j}, \quad (n > p) \quad (1)$$

$$z_n = 0, \quad (n \leq p) \quad (2)$$

有声音と雑音を分離する特徴パラメータとして、LPC 残差パワー（以下 R ）を入力信号パワー（以下 P ）で正規化した正規化残差パワー（ R/P ）の逆数の対数を LPC 適合度として定義する．この値が大きいほど、モデルの適合度が高いことを意味する．

通常の広帯域（おおむね 5 kHz 以下）における入力信号のパワーを P_W で表し、低域信号のパワーを P_L ，その LPC 残差パワーを R_L で記述すると、低域 LPC 適合度 $Q_{L/L}$ は、

$$Q_{L/L} = -10 \log(R_L/P_L) \quad (3)$$

である．この $Q_{L/L}$ に対して P_L/P_W による補正を施した低域 LPC 補正適合度 $Q_{L/W}$ は、次式で定義される．

$$Q_{L/W} = -10 \log(R_L/P_W) \quad (4)$$

$Q_{L/W}$ は、高調波構造の LPC 適合の度合とパワー値が考慮されており、工場騒音下での有声音検出の特徴パラメータとして、優れていることがわかっていて [5]． $Q_{L/W}$ の時系列は、以下のように算出できる．

$$Q_{L/W}(m) = -10 \log\{R_L(m)/P_W(m)\} \quad (5)$$

$$P_W(m) = \sum_{k=1}^{L_1} x_{(m-1)M_1+k}^2 / L_1 \quad (6)$$

$$R_L(m) = \sum_{k=1}^{L_2} z_{(m-1)M_2+k}^2 / L_2 \quad (7)$$

ここで、 L_1 、 L_2 は各々広帯域、低域でのパワー算出に使用するフレーム長（サンプル数）であり、 M_1 、 M_2 は各々フレーム周期である．LPC 適合度算出に使用した分析条件を表 2 に示す．

3.3 有声音区間検出

非定常高騒音環境下での孤立単語発声に対して、 $Q_{L/W}$ の時系列を用いて、有声音区間検出を行った．高騒音下では無声音始端は検出困難なため、1. で述べたように単語の最初の有声音区間の始端から最後の有声音区間の終端までを有声音区間と定義している．また、実用的には騒音環境下で話者の $Q_{L/W}$ 分布を事前に求めることは困難であるため、雑音の

表 2 分析条件
Table 2 Analysis conditions.

	広帯域	低域
サンプリング	16 bits 10.24 kHz	16 bits 1.28 kHz
LPC フレーム 長及び周期		29.7 ms 38 points
LPC 分析次数		6 次
パワー算出の フレーム長	29.7 ms 304 points (L_1)	29.7 ms 38 points (L_2)
パワー算出の フレーム周期	3.1 ms 32 points (M_1)	3.1 ms 4 points (M_2)

$Q_{L/W}$ 分布のみが既知であるという条件で、有声音単語区間の始端検出を以下に行った．雑音の $Q_{L/W}$ の平均 μ_N ，標準偏差 σ_N より設定したしきい値 $Q_N = \mu_N + 2\sigma_N$ をもとに、有声音フレームを検出し、その連続性から有声音区間を抽出する．そして、同一単語内とみなせるすべての有声音区間より、前述した定義に基づき有声音単語区間の始端を検出する． $Q_{L/W}$ を使用した有声音単語区間検出手法は、同じく低域を対象とした SIFT 法に基づく検出手法よりも工場騒音下での検出精度が高いことが確認されている [5]．

4. 距離尺度と有声音区間重み付け

本認識方式の騒音下ロンバード音声に対する性能を評価するにあたり、まず、距離尺度の頑健性について確認する必要があるため、使用する距離尺度について述べる．そして、有声音区間で距離の重み付けを行う手法を提案する．

4.1 距離尺度

距離尺度としては、一般的な LPC ケプストラム距離尺度（CEP）、騒音環境での効果が期待できる尺度としてスペクトルのピークを強調したスペクトル傾斜距離尺度（RPS）[15] 及びピーク重み付けを付加した重み付け群遅延距離尺度（WGD）[16] を用いた．耐雑音性が確認されている多数の距離尺度の中で、WGD は標準パターンとテストパターンで SN 比が異なった場合でも、比較的性能が高いことが報告されている [16]．ただし、ロンバード音声に対しては頑健性が予想されている [17] が、検証はなされていない．

各距離尺度の計算式は以下のとおりである．以下の計算で、標準パターン（ f ）、テストパターン（ g ）の LPC ケプストラム係数を $C_n^{(f)}$ 、 $C_n^{(g)}$ 、自己相関係数を $r_n^{(f)}$ 、 $r_n^{(g)}$ で表し、打ち切り次数を N としている．

(1) LPC ケブストラム距離尺度

基本的な距離尺度である LPC ケブストラム距離尺度は、ケブストラム係数を用いてユークリッド距離を計算することにより以下で与えられる。

$$d_{CEP} = 2 \sum_{n=1}^N (C_n^{(f)} - C_n^{(g)})^2 \quad (8)$$

(2) スペクトル傾斜距離尺度 [15]

スペクトル傾斜距離尺度 (Root-Power Sums: RPS) の定義は、対数スペクトルの周波数微分のユークリッド距離で与えられ、その計算式は次式で表される。

$$d_{RPS} = 2 \sum_{n=1}^N \{n(C_n^{(f)} - C_n^{(g)})\}^2 \quad (9)$$

この距離尺度は、スペクトルのピークすなわちホルマントに対する感度が高く、スペクトルの全体的傾斜成分の変動に耐性をもつため、低 SN 比においても広帯域雑音に強いという特徴がある。

(3) 重み付け群遅延距離尺度 [16]

重み付け群遅延距離尺度 (Weighted Group Delay Spectrum Distance: WGD) は、RPS と同様にスペクトルの傾斜変動に強い群遅延スペクトルに加えて正規化パワースペクトルでピーク重み付けした尺度であり、次式で近似される。

$$d_{WGD} = \sum_{n=1}^N n(C_n^{(f)} - C_n^{(g)})(r_n^{(f)} - r_n^{(g)}) \quad (10)$$

この距離尺度は、スペクトルの全体的傾斜成分の変化に強く、重み付けによりパワーの強いピークに感度をもっている。

4.2 有声音区間の距離重み付け

有声音区間で距離の重み付けをしたパターンマッチング手法を提案し、その効果について検討する。

騒音環境での発声音では、無声音やパワーの弱い有声音の信号は雑音成分の占める割合が大きく、そのフレームの SN 比は一般に単語全体の平均 SN 比より低く、その照合度の信頼性も必然的に低くなる。これに対して、パワーが強い有声音は騒音下においてもその SN 比は比較的高く、そのフレームの照合度も無声音フレームと比べて信頼性が高くなる。このため、マッチングにおける各フレームの距離の算出において、パワーの強い有声音フレームの距離が距離総和による結果に反映されやすくなるように考慮する。有声音区間

の距離重み付け方法は、以下のとおりである。

あらかじめ評価用発声音以外で単語認識実験を行い、正しく認識された 10 単語の 1 フレーム当りの平均距離を d_a とする。

評価用テストパターンの認識では、有声音と判定されたフレームの距離が d のとき、そのフレームの距離を以下のように重み付けする。

$$d_y = d \left(\frac{d}{ad_a} \right)^s \quad (11)$$

有声音以外のフレームは、重み付けを行わず、 $s = 0$ すなわち $d_y = d$ とする。つまり、有声音フレームの距離を他のフレームのそれと比べて大小関係をより顕著にする。有声音フレームの検出パラメータとしては、 $Q_{L/W}$ を用い、実験的に定める Q_V をしきい値として $Q_{L/W} \geq Q_V$ のフレームに対して重み付けを行う。 Q_V に対する認識率を算出することで、重み付けに最適な Q_V を求める。

5. 有声音単語区間に基づく単語認識実験方法

本章では、工場騒音環境でのロンバード音声の認識において、本研究提案の有声音単語区間利用の有効性、有声音区間重み付けの効果、及びどの距離尺度の使用が適しているかを評価するための実験方法について述べる。最初に、使用した音声資料について示した後、実験条件について述べる。認識実験は、DP マッチングによる特定話者単語認識で行い、局所的な傾斜は有効な範囲として知られている $1/2 \sim 2$ に制限し、テストパターンのフレームに同期した実時間計算の観点から非対称型を使用した。

5.1 工場騒音環境と発声音

認識実験で使用した工場の騒音と発声音について述べる。工場では機械のモータや油圧ポンプの動作音がほぼ定常な雑音であり、プレス加工衝撃音、金属接触音などの非定常雑音が頻繁に発生する。パワーレベルでは、パワーの変動が比較的緩やかな区間（ほぼ定常な雑音区間）と衝撃音などにより急激に変動する区間（非定常高雑音区間）が存在する。音声認識実験で使った工場の騒音は、騒音計 A 特性で、ほぼ定常な雑音区間は 75～85 dBA、非定常高雑音は 85 dBA 以上で最大 100 dBA に達する。

音声資料は、3 名の成人男性各々が 120 語彙をこの騒音環境で 2 回、静環境で 1 回及びヘッドホン騒音聴取による模擬騒音下で 1 回の各発声により得られた合計 1,440 サンプルを使用した。模擬騒音下発声は、収

録した上記工場騒音をヘッドホンで再生し、イアパッド面が 85 dBA となるように設定して、静環境において収集した。騒音環境での発声音の SN 比は、有声音区間で平均して約 9 dB である。静環境での通常の発声と比べて騒音環境及び模擬騒音下での発声では、表 1 で示したピッチ周波数のほか、パワー、ホルマント周波数、スペクトル傾斜などで変形が確認され、ロンバード効果と考えられる。

120 語彙を付録に示す。実際に工場で使用されている用語を中心に数字も含めて、特定の語彙に依存しないように、各種音素を始端・終端とする 120 語彙を使用した。

5.2 認識実験条件

音声波形は、サンプリング周波数 10.24 kHz, 16 ビットでデジタル化し、フレーム長 29.7 ms (ハミング窓)、フレーム周期 12.5 ms で LPC 分析を行った。LPC 分析次数、係数打ち切り次数はいずれも 16 である。また、プリエンファシスによる高域強調を行った場合と行わない場合について評価した。プリエンファシスについては、1 次の適応プリエンファシス ($1 - \alpha z^{-1}$) の方が係数固定 ($1 - 0.98z^{-1}$) よりわずかに性能が高かったため、適応プリエンファシスを使用した。ロンバード効果では、スペクトルの全体傾斜の変動が観測されるが、係数固定と比べて傾斜変動に対応できていないためと考えられる。

単語認識実験は、騒音環境での各語彙 1 回の発声音を評価用テストパターンとし、標準パターンとして騒音環境での別の 1 回の発声音を用いた場合 (1)、静環境で発声された 1 回の音声を用いた場合 (2) 及び模擬騒音下での 1 回の音声を用いた場合 (3) について行い、標準パターンの違いによる性能を比較した。(1) については、標準パターンがテストパターンと同一の場所での収録であるが、音声信号に付加された非定常雑音及びロンバード効果には違いが見られる。

テストパターンは有声音区間検出により得られた有声音区間を用い、その始末端を端点固定及びフリーの両方で使用した。また、標準パターンの始末端は、視察により切り出した有声音区間の始末端を端点固定で使用し (2) (3) については同様に音声区間を端点固定で使用的場合についても評価実験を行った (1) については有声音区間前後の無声音区間が雑音に埋もれており、視察においても抽出が困難であったため、有声音区間利用と音声区間利用との比較は (2) (3) について行った。2. で述べたよう

に、実用上は (3) は困難であるが、有声音区間利用と音声区間利用との比較評価のため行った。

以上の条件で、まず 4.1 で述べた各距離尺度による雑音混入ロンバード音声の認識性能を評価する。そして、始末端近傍の無声音を考慮しない有声音区間利用による認識性能と、通常の音声区間を利用した認識性能について、各距離尺度で比較評価する。ここで、テストパターンの有声音区間検出手法としては、 $Q_{L/W}$ を使用した本検出手法と SIFT 法に基づく検出手法 [5] の両方を使用し比較する。SIFT 法に基づく有声音区間検出は以下のように行った。 $Q_{L/W}$ と同じ 1.28 kHz サンプリング信号ではスペクトル平坦化が不十分で検出性能が低かったため、この手法による騒音下有声音の検出性能が高かった 2.56 kHz サンプリング信号を用いた。そして、LPC 分析 (4 次) によりスペクトルを平坦化し、自己相関係数ピーク値を算出した。フレームごとの自己相関係数ピーク値による有声音区間検出は、3.3 の $Q_{L/W}$ による検出と同様である。

また、最も認識性能の高い距離尺度を使用して、 $Q_{L/W}$ により検出された有声音区間に対して式 (11) の重み付けを行った場合の効果について評価する。

6. ロンバード音声認識実験結果

本章では、前章で述べた実験方法により得られた雑音混入ロンバード音声の認識実験結果について述べる。実験結果はすべて、3 名の話者の平均認識率で求めた。

6.1 距離尺度による認識性能

$Q_{L/W}$ を使用した有声音区間検出で得られた始末端をもとに端点フリーで使用し、標準パターンが有声音区間使用の場合は、始端で ± 25 ms、終端で ± 50 ms の範囲内で、音声区間使用の場合は、始端で -100 ms ~ 0 、終端で $0 \sim +100$ ms の範囲内で、各々マッチングにおける距離が最小となるように最適始末端点を求めた。

(1) 標準パターンが騒音環境で発声された場合

実験結果を表 3 に示す。標準パターンは有声音区間を使用した。標準パターンには、有声音の一部が埋もれてしまう衝撃音を含んだ発声単語が、3 名で合計 76 サンプル見られたが、これらを含めた認識率 (衝撃音あり) とこれらを除いて算出した認識率 (衝撃音なし) を表している。衝撃音のほとんどはプレス加工音であり、1 kHz 以下のパワーが強い。

騒音環境では一般的な距離尺度 (CEP) よりも、音

表 3 各距離尺度の単語認識率 (%) の比較 (騒音環境での発声音を標準パターンとして用いた場合)

Table 3 Comparison of word recognition rates (%) between the distance measures when using utterances under noisy environments as reference pattern.

標準パターンに衝撃音の有無		あり	なし
CEP	ブリエンファシスなし	78.1	89.1
	ブリエンファシスあり	79.7	89.4
RPS	ブリエンファシスなし	85.0	94.0
	ブリエンファシスあり	86.1	94.0
WGD	ブリエンファシスなし	81.4	90.1
	ブリエンファシスあり	89.4	96.5

声スペクトル中の雑音に埋もれにくい周波数成分を強調した距離尺度 (RPS, WGD) が適していることが明らかである。テストパターンの一部は衝撃音を含んでいるが、標準パターンとして認識時と同じ環境での発声音を使用することで認識性能は高い。特に WGD では、標準パターンとして衝撃音を含まない発声単語を使用すれば、単語認識率で 96.5% が得られた。ただし、WGD ではブリエンファシスを用いない場合はその効果が小さいのに対して、RPS ではブリエンファシスの有無に関係なく比較的高い性能を維持しており、雑音の混入したロンバード音声に対するピーク強調処理が効果的であることがわかる。

(2) 標準パターンが静環境で発声された場合

ブリエンファシスなしでは性能が低かったため、すべてブリエンファシスを使用し、標準パターンとして、有声音語区間を使用した場合を表 4(a) に、音声区間を使用した場合を表 4(b) に示す。音声区間長と有声音語区間長の差は、全発声音の 3 名の平均で 92.8 ms であり、300 ms を超える発声もあったが、200 ms 以下が 94.4% を占めている。また、標準パターンを無雑音で用いた場合とコンピュータ処理により波形上で雑音を付加させた場合について評価した。付加した雑音の種類は、5.1 で述べた工場騒音であり、発声時とは異なる雑音区間を使用し、工場での発声音と同程度の SN 比 (有声音区間で平均 9 dB) となるように付加した。

騒音環境ではロンバード効果によるスペクトル変形が生じるため、標準パターンに騒音環境での発声音を用いた場合と比べて、全般的に認識性能は低下する。ただし、この場合でも RPS, WGD は CEP と比べて効果があった。表 4(a), (b) とともに、静環境発声音に雑音を付加させた方が無雑音の場合より性能が高く、実環境に近い標準パターンを用いると効果があること

表 4 各距離尺度の単語認識率 (%) の比較 (静環境での発声音を標準パターンとして用いた場合)

Table 4 Comparison of word recognition rates (%) between the distance measures when using utterances under clean environments as reference pattern.

(a) In the case of voiced sound based word period.

騒音環境雑音の付加	無雑音	雑音付加
CEP	56.4	69.4
RPS	71.7	75.8
WGD	75.0	76.7

(b) In the case of speech period including unvoiced sound.

騒音環境雑音の付加	無雑音	雑音付加
CEP	57.5	68.6
RPS	69.7	73.6
WGD	73.9	75.3

を示している。とりわけ、CEP は雑音付加の有無により性能が大きく異なる。これに対して、WGD では差は小さく、標準パターンに雑音を付加させない場合でも比較的性能が高い。

以上より、WGD はブリエンファシスを使用すれば、SN 比が異なる場合や、発声変形が生じた場合でも他の距離尺度と比べて認識性能が高く、雑音混入ロンバード音声認識に最も優れていることがわかる。また、本節の端点フリー探索範囲内では、(a) での各距離尺度の最高認識率は、(b) でのそれらより高く、有声音語区間利用の音声区間利用に対する優位性を示唆している。ただし、端点フリー範囲により認識性能が異なることが考えられるため、次節において、端点フリーにおける最適な探索範囲を求めることで検証する。また、次節においては、模擬騒音下発声によるロンバード効果を含む音を標準パターンとして用いた場合についても、有声音語区間利用と音声区間利用との認識性能を比較する。

6.2 有声音語区間に基づく最適探索範囲 (音声区間との比較)

すべてブリエンファシスを用い、標準パターンとして有声音語区間利用と音声区間利用の各々に対して、テストパターンの最適な探索範囲を求めて比較した。端点固定では有声音語区間検出で得られた始終端を使用し、端点フリーでは各設定範囲内で最適始終端点を求めた。

(1) 標準パターンが騒音環境で発声された場合

実験結果を表 5 に示す。標準パターン (衝撃音を含む) は有声音語区間を使用した。 $Q_{L/W}$ による有声音語区間検出では、各距離尺度とも、始終端を $\pm 25 \sim$

表 5 各種端点フリー範囲での単語認識率(%) (騒音環境での発声音を標準パターンとして用いた場合 Q:本検出手法 S:SIFT 法)

Table 5 Relationship between word recognition rates (%) and the limits of unconstrained endpoints when using utterances under noisy environments as reference pattern. Q: $Q_{L/W}$, S:SIFT

始端 終端	固定	±25 ms	±25 ms	±50 ms	±50 ms	±100 ms	±100 ms
CEP Q	78.3	79.4	79.7	78.3	74.7	71.9	
CEP S	74.7	75.8	76.4	75.8	74.2	70.8	
RPS Q	85.8	86.4	86.1	85.8	85.6	81.1	
RPS S	81.1	82.2	82.5	82.5	84.4	81.1	
WGD Q	88.3	88.9	89.4	88.6	87.5	84.2	
WGD S	82.8	83.3	84.4	84.2	85.0	83.1	

±50 ms の範囲で端点フリーとした場合に認識率が高くなるが、いずれの距離尺度も始終端を端点固定とした場合と比較して大きな性能差はなく、テストパターンの有声単語区間検出精度が高いことが推測できる。これに比べて、SIFT 法に基づく検出では、端点固定で性能が低く、端点フリーの範囲が広い場合に $Q_{L/W}$ との性能差が小さくなるが、各距離尺度ともその最高認識率は $Q_{L/W}$ に及ばない。WGD では、 $Q_{L/W}$ の使用による最高認識率が SIFT 法の使用によるそれに対して 4.4%上回っている。

(2) 標準パターンが静環境で発声された場合

標準パターンとして、有声単語区間を使用した場合を表 6(a) に、音声区間を使用した場合を表 6(b) に示す。いずれも、工場雑音を付加 (SN 比は有声音区間で平均 9 dB) させて評価した。有声単語区間を使用した場合、RPS、WGD ではテストパターンの始端・終端ともに ±50 ms 程度の範囲で端点フリーとした場合に認識率が高くなり、騒音下発声音を標準パターンとしたときと同様の傾向を有する。音声区間を使用した場合では、有声単語区間前後の無声音も考慮されるため、端点フリーの探索範囲を始端・終端で各々、 $Q_{L/W}$ 使用では 100 ms、SIFT 法使用では 150 ms まで拡張したときに最も認識率が高い。そして、 $Q_{L/W}$ の使用は各距離尺度において SIFT 法の使用より認識性能が高いことが確認できる。

有声単語区間と音声区間との認識率比較では、有声単語区間を使用した各距離尺度の最高認識率 (CEP:69.4%, RPS:76.7%, WGD:77.5%) は音声区間を使用した最高認識率 (CEP:68.6%, RPS:73.6%, WGD:75.3%) を上回っており、有声単語区間の方が性能が高い。

表 6 各種端点フリー範囲での単語認識率(%) (静環境での発声音を標準パターンとして用いた場合 Q:本検出手法 S:SIFT 法)

Table 6 Relationship between word recognition rates (%) and the limits of unconstrained endpoints when using utterances under clean environments as reference pattern. Q: $Q_{L/W}$, S:SIFT

(a) In the case of voiced sound based word period.

始端 終端	固定	±25 ms	±25 ms	±50 ms	±50 ms	±100 ms	±100 ms
CEP Q	67.8	68.3	69.4	66.9	64.7	61.7	
CEP S	60.3	62.5	65.0	62.8	61.9	58.9	
RPS Q	73.6	75.6	75.8	76.7	71.9	67.8	
RPS S	67.8	70.0	71.9	72.5	72.2	66.1	
WGD Q	74.7	76.4	76.7	77.5	73.9	71.1	
WGD S	70.3	72.2	72.5	73.3	71.4	69.2	

(b) In the case of speech period including unvoiced sound.

始端 終端	固定	-50 ms ~ 0	-100 ms ~ 0	-150 ms ~ 0	-200 ms ~ 0
CEP Q	57.8	66.1	68.6	65.3	64.4
CEP S	52.8	59.7	63.9	62.8	62.2
RPS Q	63.1	72.8	73.6	72.8	70.3
RPS S	57.2	65.8	68.6	69.4	68.9
WGD Q	62.8	73.6	75.3	74.2	73.6
WGD S	56.9	66.7	71.4	71.7	71.1

(3) 標準パターンが模擬騒音下で発声された場合
認識時と同環境での発声音を標準パターンとして用いたときの有声単語区間利用と音声区間利用との比較を模擬するため、模擬騒音下発声音を使用した。標準パターンとして、有声単語区間を使用した場合を表 7(a) に、音声区間を使用した場合を表 7(b) に示す。いずれも、工場雑音を付加 (SN 比は有声音区間で平均 9 dB) させて評価した。音声区間長と有声単語区間長の差は、全発声音の 3 名の平均で 96.6 ms であり、200 ms 以下が 93.3%を占めている。

この実験では、すべての距離尺度、探索範囲において、 $Q_{L/W}$ の使用は SIFT 法の使用より性能が高く、ここでも、有声単語区間検出手法として本手法の方が優れていることを示している。

$Q_{L/W}$ では、標準パターンとして有声単語区間を使用した場合、テストパターンの始端・終端を ±25 ~ ±50 ms 程度の範囲で端点フリーとした場合に認識率が高くなり、音声区間を使用した場合では、始端・終端で各々150 ms まで探索範囲を拡張したときに最も認識率が高くなる。そして、有声単語区間を使用した各距離尺度での最高認識率 (CEP:79.4%, RPS:85.3%, WGD:87.8%) は、音声区間を使用した各最高認識率 (CEP:77.5%, RPS:83.3%, WGD:85.6%) を上回っ

表 7 各種端点フリー範囲での単語認識率 (%) (ヘッドホンによる 85 dBA 騒音聴取での発声音を標準パターンとして用いた場合 Q:本検出手法 S:SIFT 法)

Table 7 Relationship between word recognition rates (%) and the limits of unconstrained endpoints when using utterances at 85 dBA through headphone as reference pattern. Q: $Q_{L/W}$, S:SIFT

(a) In the case of voiced sound based word period.

始端	固定	±25 ms	±25 ms	±50 ms	±50 ms	±100 ms
終端	固定	±25 ms	±50 ms	±50 ms	±100 ms	±100 ms
CEP Q	76.7	78.6	79.4	78.6	78.1	73.9
CEP S	71.1	72.2	74.7	74.2	73.1	71.1
RPS Q	83.3	83.9	84.4	85.3	83.3	79.4
RPS S	77.2	78.3	79.2	80.3	80.6	76.9
WGD Q	85.8	86.9	86.9	87.8	87.2	84.7
WGD S	79.4	80.8	81.7	81.7	83.9	81.1

(b) In the case of speech period including unvoiced sound.

始端	固定	-50 ms ~ 0	-100 ms ~ 0	-150 ms ~ 0	-200 ms ~ 0
終端	固定	0 ~ 50 ms	0 ~ 100 ms	0 ~ 150 ms	0 ~ 200 ms
CEP Q	65.6	74.2	76.9	77.5	76.4
CEP S	60.6	69.2	72.5	73.3	73.1
RPS Q	71.4	79.2	82.5	83.3	82.5
RPS S	65.6	73.3	77.8	80.3	80.0
WGD Q	71.6	80.3	85.3	85.6	84.7
WGD S	66.7	74.7	81.1	83.1	82.8

ており、有声音語区間利用の有効性が確認できる。また、音声区間を使用して最高認識率を示した場合に誤認識となった語彙で、有声音語区間を用いて最高認識率を達成した場合に正解となった語彙は「型冷却」、「3」、「寸動」、「テスト」、「なし」、「バラサ圧」、「非常停止」、「プラス」など、始端・終端の音素や発声長が種々の用語であり、有声音語区間利用の有効性が特定の語彙に依存していないことを示している。

本節の実験結果より、有声音語区間前後の無声音を含めて考慮しても認識率の向上は見られなかった。これは、低 SN 比の騒音下無声音区間は雑音重畳により雑音区間に近くなり、他のパターンの雑音区間と類似することがあるため、有声音語区間前後の無声音を含むパターンマッチングが認識誤りに寄与してしまうという作用が考えられる。また、有声音語区間を利用した場合でも、端点フリーの範囲を長く設定しすぎると認識性能は低下する。したがって、高騒音下では、有声音語区間を利用し、端点フリーの範囲を前述のように適度に制限する方が効果があることがわかる。

6.3 有声音区間の距離重み付けの効果

前節より、 $Q_{L/W}$ を使用した有声音語区間が有効であり、騒音下発声音を標準パターンに使用した場合に性能が高いことから、騒音下発声音（衝撃音を含む）

表 8 有声音区間重み付けによる単語認識実験結果 (%) (標準パターンが騒音環境で発声された場合: WGD 使用)

Table 8 Experimental results (%) of word recognition based on weighted voiced sound period in the case of WGD when using utterances under noisy environments as reference pattern.

Q_V	$Q_N - 1$	Q_N	$Q_N + 1$	$Q_N + 2$	$Q_N + 3$
$a = 1.5, s = 1$	89.4	90.6	90.8	91.4	90.6
$a = 1.5, s = 2$	89.4	90.3	90.6	91.1	90.3
$a = 2, s = 1$	89.2	90.3	90.3	90.6	90.3
$a = 2, s = 2$	89.7	90.6	90.6	91.1	90.8
$a = 2.5, s = 2$	89.7	90.6	90.6	90.6	90.6

を標準パターンとして、有声音語区間内における有声音区間について重み付けする効果を検討した。

距離尺度として WGD を用い、テストパターンについて有声音フレームの距離重み付けを行った場合の結果を表 8 に示す。有声音検出のしきい値 Q_V として、3.3 で述べた Q_N 及びその周辺の値を用いた。その結果、 $Q_V = Q_N \sim (Q_N + 3)$ dB の範囲で、 $Q_{L/W} \geq Q_V$ の入力フレームに対して距離の重み付けを行うことで、重み付けがない場合（認識率：89.4%）と比べて、認識率の向上が確認できた。特に $Q_V = (Q_N + 2)$ dB で最大 2% の認識率の向上が見られ、有声音の度合いが比較的高いフレームでの重み付けに効果があることがわかる。

本章の実験結果より、WGD 尺度の工場騒音下ロンバード音声認識における優位性、 $Q_{L/W}$ による有声音語区間利用の有効性及び有声音区間重み付けの効果が確認できた。また、標準パターンとして、衝撃音を含む場合であっても、認識時と同環境での発声音が利用できるシステムが実用的であることが明らかである。

7. 音声入力検反システムの開発

前章までの評価結果をもとに、WGD 尺度を用いて、 $Q_{L/W}$ による有声音語区間検出手法と有声音区間重み付け手法を使用した音声認識により、工場における織物欠点名の音声入力が可能なる検反システムを開発した。

7.1 システムの概要

音声認識対象語彙は織物欠点名称であり、その数は工場により異なるが最大で約 100 種類程度である。ただし、作業員全員が欠点名をすべて把握しているわけではなく、特に初心者には欠点名リストを見ながら判断して入力するという様子も見られる。したがって、欠点発見後すぐに発声できる欠点名の数には個人差があ

り、多くすると作業には逆に負担になりかねない。しかし、欠点のほとんどは頻度の高い欠点であり、その種類は限られており、頻度の高い欠点名のみを音声入力化しても能率の向上が期待できる。そして、頻度が低く判断に時間のかかる欠点名の入力は従来どおり手入力で行う方が操作性が良い場合もある。そこで、操作性向上の観点から、手入力・音声入力の両入力が常に可能なシステムとした。作業ごとの標準音声データベースの構築においては、作業者の熟練度に応じて、頻度の高い欠点名を優先的に登録可能とし、音声入力の対象となる欠点名・数も作業者が自由に選択できるようにした。

7.2 システム構成

システムの構成を図3、音声による欠点名入力の手順を図4に示す。

作業者は初期設定で自分のデータベースを選択し、

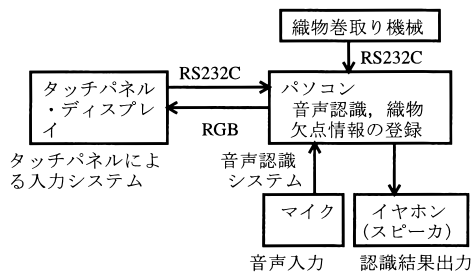


図3 音声認識による検反システムの構成図

Fig. 3 Block diagram of the fabric inspection system using speech recognition.

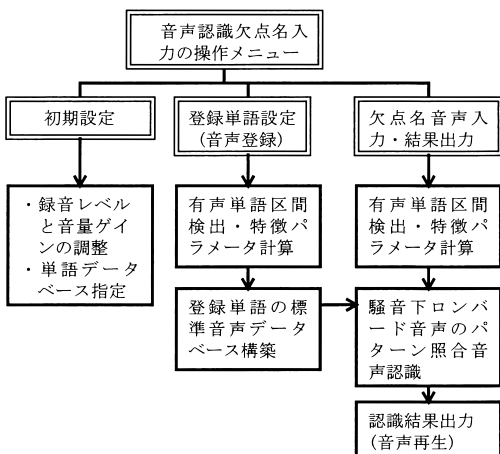


図4 検反システムでの音声認識による欠点名入力の手順

Fig. 4 Flow of feeding the name of defect by speech recognition in the fabric inspection system.

音声入力対象とする欠点名をあらかじめすべて発声して標準音声データベースに登録しておく。目視検査中は、欠点発見時にその名称を発声して入力する。認識結果は音声でその名称が出力されるので、作業者は目視検査中に目をそらすことなく入力の確認ができ、検査を中断せずに続けることができる。ただし、誤認識時での修正は、初期設定と同様、現状ではタッチパネルまたはキーボードによる手入力である。なお、織物検査では欠点の位置情報も重要な要素であるが、図3でRS232Cを介して織物巻取り機械から欠点発見時の位置情報を自動的に入力できる。

音声認識部は、図4中の(1)有声単語区間検出・音声特徴パラメータ計算、(2)パターン照合音声認識の各ブロックで構成される。各々のブロックでは、(1)有声単語区間を検出し、音声スペクトルの特徴パラメータ(LPC ケプストラム係数と自己相関係数)を算出する。(2)登録単語音声(標準パターン)と入力音声の各特徴パラメータを用いて、有声音区間重み付け端点フリー DP マッチングにより、工場内で発声されたロンバード音声の認識を行う。

7.3 動作実験結果

織物検査工場において、騒音下発声による音声入力検反システムの動作実験を行い、その実用性について評価した。実験を行った場所は、織物巻取り機械前の通常の作業者位置で、近くで他の機械も稼動しており、騒音レベルは75~90 dBAであった。話者は5.1で述べた3名とは別の成人男性3名で、まず各々が上記場所で欠点名61単語を標準パターン用に1回発声して、各話者の標準音声データベースを作成した。次に、各話者が同場所で61単語を各3回発声して認識評価実験を行った。有声単語区間検出条件は3.と同様で、標準パターンは有声単語区間検出による始終端を端点固定で使用し、評価用入力音声は検出された始終端をもとに始端で ± 25 ms、終端で ± 50 msの範囲内で端点フリーとした。その他の音声認識条件は5.と同様で、WGD尺度を使用し、有声音フレームの重み付けしきい値は $Q_V = (Q_N + 2)$ dB、また $a = 1.5$ 、 $s = 1$ とした。実験の結果、3名の平均で単語認識率93.1%を達成し、認識結果も音声で確認できた。

現状の作業において、作業者が欠点発見から、コンピュータに向きを変えて欠点名を入力し、再び機械を稼動して検査を再開するまでに要する入力時間の平均を t_1 秒とする。音声入力方式では、欠点発見から音声入力し認識結果出力による確認までの時間を t_2 秒

とすると、入力時間は認識成功で t_2 秒である。誤認識時では、誤入力訂正に要する時間を最大 2 秒として $(t_1 + t_2 + 2)$ 秒必要となり、認識率が p では、平均入力時間は、 $t_3 = pt_2 + (1 - p)(t_1 + t_2 + 2)$ となる。したがって、操作時間の観点からは、 $t_3 < t_1$ を満たす認識性能が、操作の効率化に寄与できることになる。現状の作業における最も操作性の良い入力方式はタッチパネル方式であり、この場合の熟練した作業者の動作解析の結果、 $t_1 = 7.0$ 秒であった。これに対して、音声入力方式では $t_2 = 6.0$ 秒以下で十分可能であった。この場合、上記条件式を満たす認識率は 88.9% となる。したがって、認識率 93.1% は操作の効率化に寄与できることを示している。その上、音声入力方式は、欠点発見後から認識結果確認までの間も、検査物から視線をそらすことなく目視検査を続けることができるため、入力時間が同じであっても検査に要する総時間は削減できることになる。また、作業動作の軽減が可能になり、欠点見逃しの回避も期待できる。ただし、作業によってはすべての欠点名を発声するのは困難な場合もあるため、手入力・音声入力の両方を常に使用できるシステムであることが重要である。

8. む す び

工場騒音下において有効な低域 LPC 補正適合度 $Q_{L/W}$ による有声音区間検出を利用して、騒音環境で発声された雑音混入ロンバード音声の認識手法を検討し、この評価結果をもとに音声認識による検反システムを開発した。そして、特定話者単語認識実験により以下の結果が得られた。

(1) 工場騒音下での発声音の有声単語区間検出手法として、 $Q_{L/W}$ による本手法が SIFT 法に基づく手法より優れていることが認識性能から明らかになった。そして、距離尺度としては、重み付け群遅延距離尺度 WGD が最も高い認識性能を示し、実用的には、標準パターンとして認識環境下で衝撃音を含まないように発声音を有声単語区間で登録する方法が最も有効で、96.5% の認識率が得られた。

(2) 高騒音下発声音の認識では、音声区間を利用するよりも有声単語区間を利用の方が認識性能が高く、有声単語区間利用の有効性が確認できた。そして、有声単語区間利用では、端点フリーの探索範囲を $\pm 25 \sim \pm 50$ ms 程度に制限する方が、それ以上長くするより効果がある。

(3) 有声音区間に対して距離の重み付けをするこ

とにより、認識性能が最大で 2% 向上し、その効果が確認できた。

(4) 音声入力検反システムを開発し、織物検査工場において欠点名入力の動作実験を行った結果、認識率 93.1% を達成し、検査工程での操作性向上が可能なことがわかった。

今回の認識手法の検討、検反システムの開発では、実用に耐え得る認識性能維持のため、標準パターンに認識時と同じ環境での同一話者の発声音を使用した。このため、作業者に事前の発声が必要である。事前の発声を極力少なくするためには、既存のモデル、パターンを使用して不特定話者に対しても実用に耐え得る認識性能が得られるような環境適応・発声適応技術の開発が必要である。今後は、雑音混入ロンバード音声に有効な環境適応・発声適応による認識システムの実用化について検討する。

謝辞 本研究を進めるにあたり、御協力して頂いた (株) 北村製作所の浜崎圭佑氏、森脇達也氏に感謝します。また、本研究の一部は、平成 12 年度科学技術振興事業団の RSP 事業可能性試験により実施した。

文 献

- [1] 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.
- [2] W.V. Summers, D.B. Pisoni, R.H. Bernacki, R.I. Pedlow, and M.A. Stokes, “Effects of noise on speech production: Acoustic and perceptual analyses,” J. Acoust. Soc. Amer., vol. 84, no. 3, pp. 917–928, 1988.
- [3] J.-C. Junqua, “The Lombard reflex and its role on human listeners and automatic speech recognizers,” J. Acoust. Soc. Amer., vol. 93, no. 1, pp. 510–524, 1993.
- [4] J.H.L. Hansen and O.N. Bria, “Lombard effect compensation for robust automatic speech recognition in noise,” Proc. ICSLP, pp. 1125–1128, 1990.
- [5] 漢野敦泰, 下平 博, “低域スペクトルの予測残差を利用した非定常高騒音環境での有声音区間の検出,” 信学論(D-II), vol. J80-D-II, no. 1, pp. 26–35, Jan. 1997.
- [6] P.K. Rajasekaran, G.R. Doddington, and J.W. Picone, “Recognition of speech under stress and in noise,” Proc. ICASSP, pp. 733–736, 1986.
- [7] C.E. Mokbel and G.F.A. Chollet, “Automatic word recognition in cars,” IEEE Trans. Speech and Audio Proc., vol. 3, no. 5, pp. 346–356, 1995.
- [8] D.B. Roe, “音声認識における高雑音中の Lombard 効果の適応化,” 信学技報, SP86-66, 1986.
- [9] 服部浩明, 吉田和永, “雑音環境下で発声した音声の認識方法,” 信学技報, SP88-11, 1988.
- [10] 滝沢由実, 浜田正宏, “雑音下での発声変形に対する補正法の検討,” 音響学会講演論文集, no. 1-3-9, pp. 17–18, March 1990.
- [11] 新田恒雄, 南 重信, 中山 昭, 小野木智宏, “高騒音下

- おける音声認識方式の検討,” 信学技報, SP94-20, 1994.
- [12] 若尾 淳, 武田一哉, 板倉文忠, “種々の定常雑音下における Lombard 音声の認識法の検討,” 信学論 (D-II), vol.J80-D-II, no.7, pp.1643-1650, July 1997.
- [13] 鈴木 忠, 阿部芳春, 中島邦男, “発声変形モデルを用いた騒音環境下音声認識,” 音響学会講演論文集, no.1-5-1, pp.1-2, March 1995.
- [14] M. Afify, Y. Gong, and J.-P. Haton, “A general joint additive and convolutive bias compensation approach applied to noisy Lombard speech recognition,” IEEE Trans. Speech and Audio Proc., vol.6, no.6, pp.524-537, 1998.
- [15] B.A. Hanson and H. Wakita, “Spectral slope distance measures with linear prediction analysis for word recognition in noise,” IEEE Trans. Acoust., Speech, and Signal Proc., vol.ASSP-35, no.7, pp.968-973, July 1987.
- [16] 松本 弘, 三井洋和, “雑音下音声認識のための重み付け群遅延スペクトル距離尺度,” 信学論 (A), vol.J74-A, no.8, pp.1257-1266, Aug. 1991.
- [17] 松本 弘, 三井洋和, “単語認識における重み付き群遅延距離尺度の耐雑音・耐話者性—各種尺度との比較評価,” 信学論 (D-II), vol.J75-D-II, no.2, pp.297-305, Feb. 1992.

付 録

【120 語彙】

安全柵, アップロード, アイドル, あり, ID (アイディ), 異常メッセージ, 1, 一括モニタ, 運転, 上型クランパ, うしろ, ウィンドウ, ADC, MB 走行, オープン, オン角, カウンター, 型データ, 画面, 確認, ガード, 加算, 型冷却, 9 (きゅう), キャンセル, 許可, 切換, グリッパージョー, クローズ, クロス MB, 減算, 現型エリア, 5, 合計, 個別, コイルライン, 3, 最高速時, 再生, 4 (しー), 7 (しち), 自動化, 次型確認, 下型クランパ, 自動, 状態確認, 使用, じゃない, 消灯, 消去, スライド ノックアウト, 寸動, スタート, ステージ, ステーション, 千, 生産, 設定, 選択, 説明, ゼロ, その他, 操作, ダイクッション圧, ダイビンリフタ, ダイライト, タイマー, 単機, チャンネル, 中断, ツー, 取消し, テスト, 点灯, テレスコ, 低速スピード, データ, データ設定, ディスタッカー, 止まれ, 登録, 同一, 時計変更, 7 (なな), ナンバー, なし, 2, 日時変更, 8, 排出コンベア, バランサ圧, パーコネクタ, ハヌケ, 百, 非常停止, ピン折れ確認, 左, プッシャー, フィードマスタ, プラス, プリロード, 変更, ボルスタクランパ, 方法, マイナス, ミスフィード, メニュー, 元画面, やめ, 4, リターン, リセット, 連続, 連動, 6, ロータリカム, ロットカウ

タ, 録音, 割込み, ワン

(平成 13 年 7 月 12 日受付, 11 月 22 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. 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. 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.