音声認識技術の開発と検反システムへの応用

製品科学部　○漢野救泰

１．目　的

近年，各種装置の操作性向上に音声認識の利用が望まれているが，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，非定常雑音の混入及び騒音下発声における音声特徴の変形（ロンバード効果）が考えられる。この課題に対して，本研究では，有声音に基づく単語区間（有声単語区間）検出手法を利用した実環境下ロンバード音声認識方式を提案するとともに，この方式に基づく認識システムを開発することを目的とする。本報では，まず距離尺度の頑健性について述べ，有声単語区間利用の有効性について示すとともに有声音区間で照合度を重み付けする手法の効果を明らかにする。更に，これらの評価結果を基に音声入力検反システムについて検討し，動作実験により，このシステムの実用性を評価する。

２．内　容

2.1 音声認識方式

図１に本方式の構成を示す。音声の検出については,有声単語区間検出手法を使用する。そして，ロンバード効果問題に対しては，騒音下発声音の有声単語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を重み付けしてDPマッチングにより認識を行う方式で対処する。

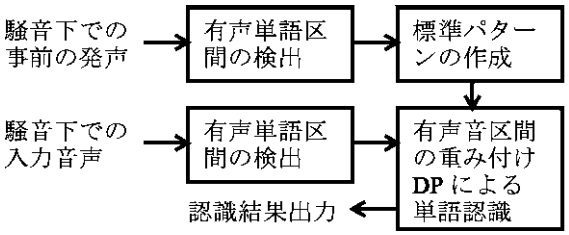

図１　音声認識方式の構成図

2.2 有声単語区間検出

本節では，ピッチ（声帯振動の基本周波数）対応型の低域LPC（線形予測）分析手法とこれに基づくLPC適合度及び有声単語区間検出手法について述べる。

2.2.1 ピッチ対応型低域LPC分析手法

本分析手法は，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。声帯振動による基本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，その適合の度合いから有声音を検出する。

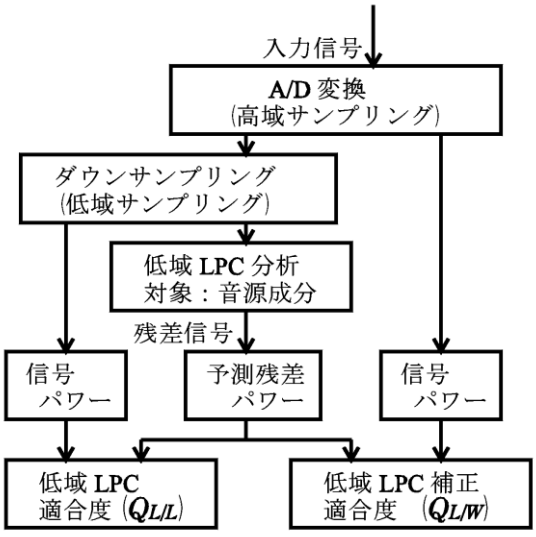

図２　LPC適合度の算出ブロック図

2.2.2 LPC適合度

図２に，有声音検出用の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。通常の広帯域（概ね5kHz以下）におけるLPC分析に使われる入力信号のパワーを P_W で表し，低域（600Hz程度以下）分析のためにダウンサンプリングされた入力信号のパワーを P_L，そのLPC残差パワーを R_L で記述すると，低域におけるLPC適合度 $Q_{L/L}$ は，$Q_{L/L} = -10 \cdot \log(R_L / P_L)$，こ

れに 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，$Q_{L/W} = -10 \cdot \log(R_L/P_W)$で表される。

2.2.3 有声単語区間検出手順

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，有声単語区間検出を行う。高騒音下では無声音始終端は検出困難なため，本研究では，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単語区間と定義している。また，雑音の $Q_{L/W}$ 分布のみが既知であるという条件で，有声単語区間の始終端検出を以下のように行う。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 $Q_N = \mu_N + 2\sigma_N$ を基に，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単語区間の始終端を検出する。本有声単語区間検出手法は，従来の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。

2.3 距離尺度と有声音区間の距離重み付け

2.3.1 距離尺度

パターン・マッチングにおける距離尺度としては，一般的なLPCケプストラム距離尺度（CEP），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）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）を用いた。

2.3.2 有声音区間の距離重み付け

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度合いが大きく，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。そこで，マッチングにおける各フレームの距離の算出では，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，有声音区間の距離重み付けを以下のように行う。

あらかじめ評価用発声音以外で単語認識実験を行い，正しく認識された単語の１フレームあたりの平均距離を 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の時，そのフレームの距離 d_yを以下のように重み付けする。

$$d_y = d(d/(a \cdot d_a))^s$$

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。

2.4 評価実験

2.4.1 実験条件

実験で使用した工場の騒音レベルは，ほぼ定常な雑音区間は75～85dBA，非定常高雑音は85dBA以上である。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回の各発声により得られた合計1,080サンプルを使用した。このうち，騒音環境での各１回の発声について有声単語区間で検出し，評価用入力パターンとして用いた。一方，標準パターンには，騒音環境での別の発声または静環境発声を用いた。

音声波形は，サンプリング周波数10.24kHz，16ビットでディジタル化し，フレーム長29.7ms（ハミング窓），フレーム周期12.5msでLPC分析を行った。LPC分析次数，係数打切り次数はいずれも16である。認識実験は，始終端点フリーDPマッチングによる特定話者単語認識で行った。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。

2.4.2 実験結果１

(1)　標準パターンが騒音環境発声の場合

実験結果を表１に示す。標準パターン（有声単語区間）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存在するが，これらを含めた認識率（衝撃音あり）と除いた認識率（同なし）を表している。 騒音環境では，音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。

表１　騒音下発声音を標準パターンに使用した時の単語認識結果(認識率：%)

衝撃音の有無		あり	なし
CEP	ﾌﾟﾘｴﾝﾌｧｼｽなし	**78.1**	**89.1**
	〃　　あり	**79.7**	**89.4**
RPS	ﾌﾟﾘｴﾝﾌｧｼｽなし	**85.0**	**94.0**
	〃　　あり	**86.1**	**94.0**
WGD	ﾌﾟﾘｴﾝﾌｧｼｽなし	**81.4**	**90.1**
	〃　　あり	**89.4**	**96.5**

(2)　標準パターンが静環境発声の場合

すべてプリエンファシスを使用し，標準パターンとして，有声単語区間を使用した場合(a)と，音声区間を使用した場合(b)を表２に示す。ここで，標準パターンを無雑音で用いた場合と雑音を付加させた場合について評価した。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下発声音を用いた場合(1)と比べて，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。(a)，(b)ともに，雑音を付加させた方が無雑音より性能が高く，実環境に近い標準パターンを用いると効果があることを示している。中でも，CEPは雑音付加の有無で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。

表２　静環境発声音を標準パターンに使用した時の単語認識結果（認識率：%）

	(a)単語区間利用		(b) 音声区間利用	
雑音の付加	なし	付加	なし	付加
CEP	**56.4**	**69.4**	**57.5**	**68.6**
RPS	**71.7**	**75.8**	**69.7**	**73.6**
WGD	**75.0**	**76.7**	**73.9**	**75.3**

以上より，有声単語区間検出に基づくWGDは，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より認識性能が高いことがわかる。また，(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，有声単語区間利用の音声区間利用に対する優位性を示している。

2.4.3 実験結果２

騒音下発声音（衝撃音を含む）を標準パターンとして用い，距離重み付けの効果を検討した。距離尺度としてWGDを用い，入力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。その結果，有声音検出のしきい値 Q_Vとして，$Q_V = Q_N$ ～(Q_N+3)dBの範囲で，$Q_{L/W} \geqq Q_V$ のフレームに対して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，$a = 1.5$，$s = 1$で91.4%と２%の認識率の向上があり，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った。

2.5 音声入力検反システム

前節までの評価結果を基に，WGD尺度を用いて，$Q_{L/W}$による有声単語区間検出手法と有声音区間重み付け手法を使用した特定話者音声認識による検反システムを検討し，実証化を行った。

2.5.1 システムの構想

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30～100種類程度である。また，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，頻度が低く判断に時間のかかる欠点名の入力は従来通り手入力で行う方が操作性が良い場合もある。そこで，

手入力・音声入力の両入力が可能なシステムとした。作業者毎の標準音声データベースの構築では，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，音声入力対象の欠点名・数も作業者が自由に選択できるようにした。

2.5.2 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，音声入力対象の欠点名をあらかじめすべて発声して登録しておく。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，作業者は検査中に目をそらすことなく入力の確認ができ，検査を続けることができる。ただし，誤認識時での修正は，タッチパネルまたはキーボードによる手入力である。

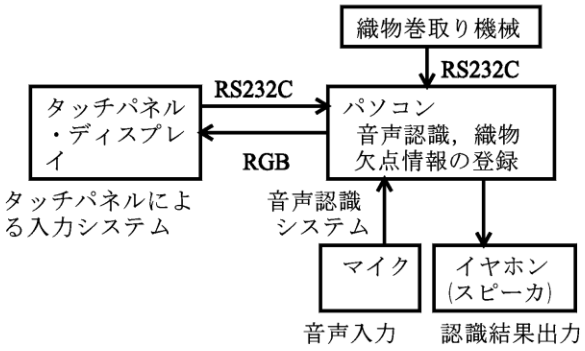

図３　システムの構成

2.5.3 動作実験

織物検査工程において，騒音下発声による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。実験の結果，単語認識率93.1%を達成し，認識結果も音声で確認できた。そして，従来の手入力による入力方式と比べて，音声入力方式では作業動作の軽減が可能になり，操作の効率化に寄与できることがわかった。その上，音声入力方式は，欠点発見から入力までの間も，検査を続けることができるため，検査に要する総時間は削減できることになる。

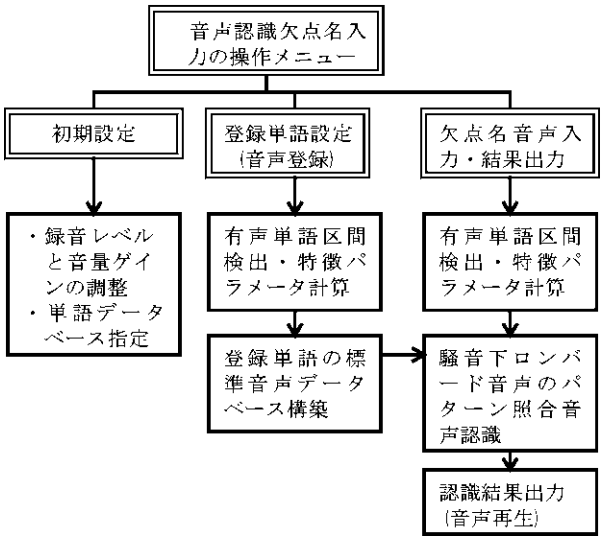

図４　欠点名入力の手順

３．結　果

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ロンバード音声の有声音区間重み付け認識手法を提案し，特定話者音声認識システムを開発した。そして，雑音混入ロンバード音声に対するWGD尺度の頑健性，有声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。更に，この認識方式を検反システムに応用し，欠点名入力の動作実験を行った結果，検査工程での操作性が向上することがわかった。

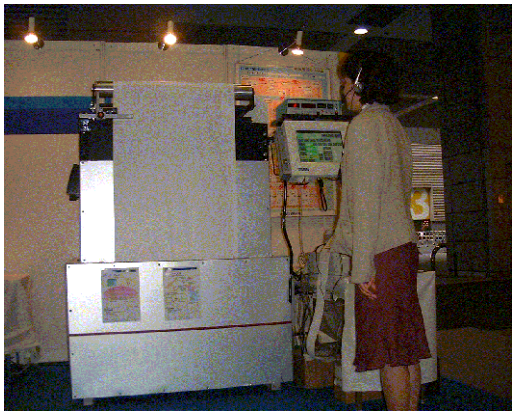

図５ 音声入力検反システム

本研究に基づく音声入力検反システムは，㈱北村製作所で製品化が進められ，第７回大阪国際繊維機械ショー（2001年10月，インテックス大阪）に出展された（図５）。