特定話者音声認識と検反システムへの応用

漢野救泰* 林克明* 米沢裕司*

高騒音下においては，非定常な雑音の混入や音声特徴の変形(ロンバード効果)により，音声認識性能が顕著に低下する。この課題に対して，本研究では，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するとともに，有声音に基づく単語区間利用の有効性及び有声音区間の距離重み付けの効果を確認した。また，これらの手法を，特定話者音声認識による検反システムに応用した。検査工場において，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，高認識率を達成し，操作性を向上できることがわかった。

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

Speaker-dependent Speech Recognition and Application to the Fabric Inspection System

Sukeyasu KANNO, Katsuaki HAYASHI and Yuji YONEZAWA

The performance of speech recognition degrades remarkably due to the non-stationary noise and Lombard effect under heavy noisy environments. This paper describes a practical method to recognize noisy Lombard speech , based on the detection of voiced sound periods in factories. The effectiveness of three techniques, i.e., WGD measure, word periods and weighted distances in voiced sound periods for noisy Lombard speech, was confirmed through the word recognition experiments using a 120-word vocabulary uttered by three male speakers. Then, the approach was applied to the fabric inspection system using speaker-dependent speech recognition. As experimental results, this system achieved high recognition performance for a 61-word vocabulary and proved to be available for an improvement of inspection efficiency in the factory.

Keywords：speech recognition, Lombard effect, voiced sound, fabric inspection system, noisy environment

１．緒　　言

本研究は，音声認識技術の利用により，織物検査装置(検反システム)の操作性向上を目指すものである。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，織物巻き取り機械を稼動させて目視による検査を行い，織物欠点を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替えることで，目視と同時に目をそらすことなく欠点名の入力を可能にし，操作の効率化と作業者の負担軽減を目指した。

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音声認識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。このため，実環境下ロンバード音声に適した認識方式の実現が望まれているが，これまでに達成されていない。

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

*製品科学部

実環境下発声音を使用できる認識システムを開発した。ここで，音声認識系としては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，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の重み付けを行う手法を検討し，その効果を明らかにする。そして，これらの評価結果を基に，工場の検査工程で利用できる音声入力検反システムを検討し，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。

２．音声認識方式

本章では，ロンバード音声認識方法に関して，工場の作業者が認識システムを使用する場合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であるとともに，この効果は話者毎・音韻毎に異なることが報告されている[2)]。そして，認識性能としては，認識時と同じ騒音条件での発声音による学習が最も優れている。従って，作業工程・認識性能の観点から，実用的には認識時と同じ環境での発声音を学習に用いる方法が最善である。DPマッチング手法は，特定話者に限定すれば標準パターンとして少ない発声回数で使用可能である。また，織物検査工程では検反システムに対して作業者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で固定である。

図1に本方式の構成を示す。音声の検出については，次章で述べる有声単語区間検出手法を使用する。そして，騒音下ロンバード効果問題に対して，実環境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。

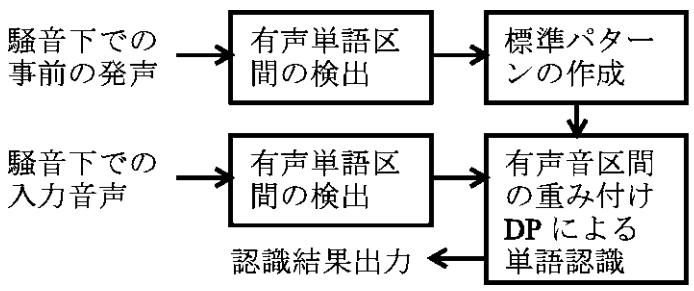

図1　音声認識方式の構成図

３．有声単語区間検出

本章では，非定常騒音下における有声音検出に適したピッチ（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。

３．１　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。この手法では，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。

３．２　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，

$$Q_{L/L} = -10 \log(R_L / P_L) \qquad (1)$$

で表わされる。

これに対して，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，

$$Q_{L/W} = -10 \log(R_L / P_W) \qquad (2)$$

で表される。

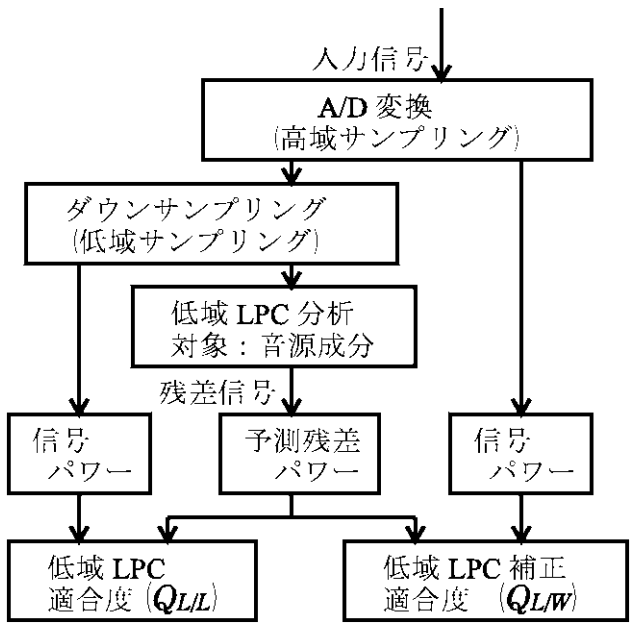

図2　LPC適合度の算出ブロック図

３．３　有声単語区間検出手法

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，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 $Q_{L/W}$ 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。$Q_{L/W}$ を使用した有声単語区間検出手法は，同じく低域を対象とした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[3]。

４．距離尺度と有声音区間重み付け

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検討する。

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)[4]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)[5]を用いた。

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，打切り次数をNとしている。

４．１．１　LPCケプストラム距離尺度

基本的な距離尺度であるLPCケプストラム距離尺度は，ケプ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。

$$d_{CEP} = 2\sum_{n=1}^{N}\left(C_n^{(f)} - C_n^{(g)}\right)^2 \qquad (3)$$

４．１．２　スペクトル傾斜距離尺度

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，次式で表される。

$$d_{RPS} = 2\sum_{n=1}^{N}\left(n\left(C_n^{(f)} - C_n^{(g)}\right)\right)^2 \qquad (4)$$

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持つため，低SN比においても広帯域雑音に強いという特徴がある。

４．１．３　重み付け群遅延距離尺度

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。

$$d_{WGD} = \sum_{n=1}^{N} n\left(C_n^{(f)} - C_n^{(g)}\right)\left(r_n^{(f)} - r_n^{(g)}\right) \qquad (5)$$

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持っている。

４．２　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10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を以下のように重み付けする。

$$d_y = d\left(d / (a \cdot d_a)\right)^s \qquad (6)$$

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$Q_{L/W} \geqq Q_V$のフレームに対して重み付けを行う。

５．認識実験結果

５．１　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を**３章**の方法で検出し，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，各々，視察により切り出した有声単語区間または無声音を含む一般的な音声区間を使用した。

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い，入力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。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。

５．２　実験結果1

(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1に示す。標準パターンは有声単語区間を使用した。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。

騒音環境では一般的な距離尺度(CEP)よりも，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。ただし，WGDではプ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ではプ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，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。

表1　騒音環境発声音を標準パターンに使用した時の単語認識実験結果(認識率：%)

標準パターンに衝撃音の有無		あり	なし
CEP	プリエンファシスなし	**78.1**	**89.1**
	〃　　あり	**79.7**	**89.4**
RPS	プリエンファシスなし	**85.0**	**94.0**
	〃　　あり	**86.1**	**94.0**
WGD	プ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で発声された場合

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を表2(a)に，音声区間を使用した場合を表2(b)に示す。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工場騒音であり，工場での発声音と同程度のSN比(有声音区間で平均9dB)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(1)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。(a)，(b)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。

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実験結果(認識率：%)

	(a)有声単語区間利用		(b) 音声区間利用	
雑音の付加	無	付加	無	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と比べて認識性能が高く，雑音混入ロンバード音声の認識に最も優れていることがわかる。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も示している。

５．３　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，有声単語区間内の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，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，$a = 1.5$，$s = 1$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。

６．音声入力検反システム

前章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り，工場における織物欠点名の入力が可能な検反システムを検討し，実証化を行った。

６．１　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。ただし，作業者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。従って，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，多くすると作業者には逆に負担になりかねない。しかし，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，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。

６．２　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。また，欠点の位置情報は，図3でRS232Cを介して織物巻取り機械から自動的に入力できる。

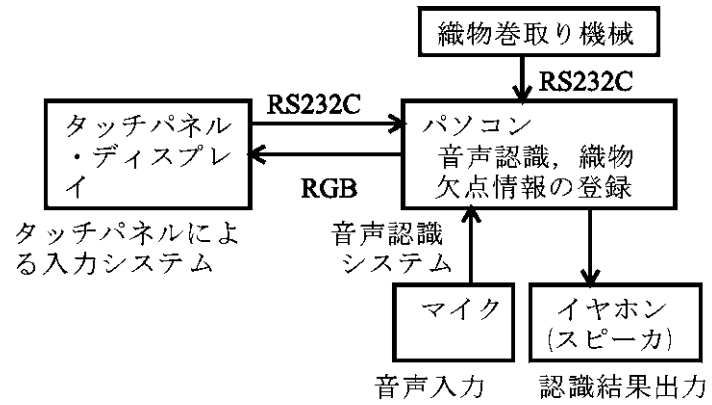

図3　システムの構成

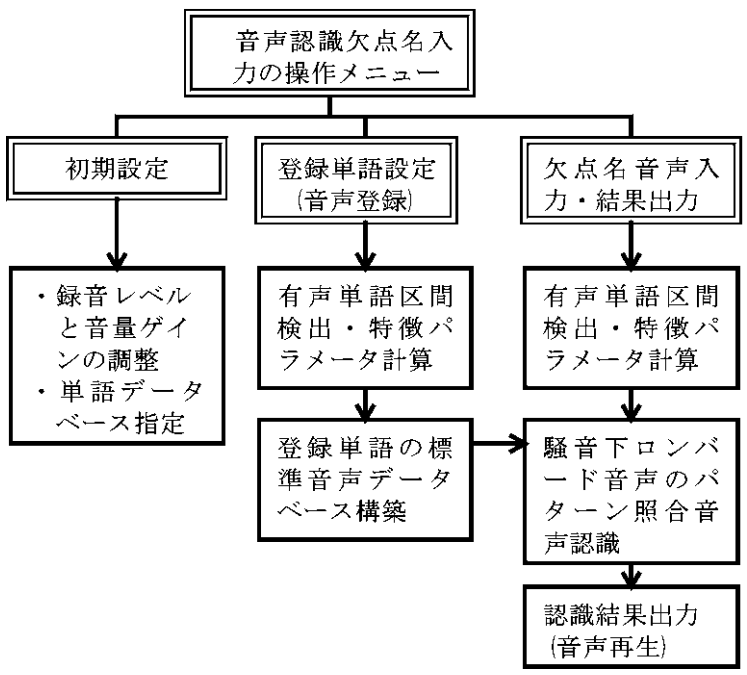

図4　欠点名入力の手順

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。

６．３　実験結果

織物検査工場において，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75～90dBAであった。認識評価実験は，欠点名61単語を使用して以下のように行った。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，評価用入力音声は同様に検出された始終端を基に始端で±25ms，終端で±50msの範囲内で端点フリーとした。音声分析・認識条件は**５．１**節と同様で，WGD尺度を使用した。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，単語認識率93.1%を達成し，認識結果も音声で確認できた。現状の作業で最も操作性の良い入力方式はタッチパネル方式であるが，音声入力方式では作業動作の軽減が可能になり，さらに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，欠点発見から欠点名入力完了までの間も，目視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。

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)で出展された(図5)。

図5 大阪繊維機械ショーでの音声入力検反システム

７．結　　言

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検討し，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た。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，検査工程での操作性が向上することがわかった。

謝　　辞

本研究の遂行に当たり，適切なご助言を頂いた金沢大学教授船田哲男氏に感謝します。

本研究の遂行に当たり，協力して頂いた㈱北村製作所の浜崎圭佑氏，森脇達也氏に感謝します。

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されたものです。

参考文献

1) 中川聖一:音声認識研究の動向,電子情報通信学会論文誌,Vol.J83-D-Ⅱ,No.2,p.433-457(2000)

2) J.H.L.Hansen and O.N.Bria:Lombard effect compensation for robust automatic speech recognition in noise,Proc.ICSLP,p.1125-1128(1990)

3) 漢野救泰,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,電子情報通信学会論文誌,Vol.J80-D-Ⅱ，No.1,p.26-35(1997)

4) 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.,ASSP-35,No.7,p.968-973(1987)

5) 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,電子情報通信学会論文誌,Vol.J74-A,No.8,p.1257-1266(1991)