様式 C-19

科学研究費補助金研究成果報告書

平成２１年５月８日現在

研究種目：若手研究(B)
研究期間：2007～2008
課題番号：１９７００１７０
研究課題名（和文）　シングルマイクロフォンによる発話区間検出及び音源方向推定の研究

研究課題名（英文）　Voice activity detection and estimation of sound source direction using a single microphone

研究代表者
滝口　哲也（TAKIGUCHI TETSUYA）
神戸大学・自然科学系先端融合研究環都市安全研究センター・講師
研究者番号：４０３９７８１５

研究成果の概要：従来の音声認識システムでは，背景雑音や残響の影響を抑圧するために，ユーザはマイクロフォンの前で（マイクスイッチを押してから），音声入力を行なう必要がある．そのような音声認識装置では，音声を使うメリットの一つである“ハンズフリー”なインターフェースを提供しているとは言えない．本研究課題では，マイクスイッチレスな音声認識の実現を目指し，雑音に頑健な音声特徴量抽出法，雑音除去手法，音源方向推定の研究を行い，その有効性を示した．

交付額

（金額単位：円）

	直接経費	間接経費	合　計
２００７年度	1,700,000	0	1,700,000
２００８年度	1,600,000	480,000	2,080,000
年度			
年度			
年度			
総　計	3,300,000	480,000	3,780,000

研究分野：音声認識
科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス
キーワード：音声情報処理，音声認識

１．研究開始当初の背景

近年，音声認識は不特定話者に対する認識が実現するという飛躍的な発展を遂げてきた．これに伴い，音声認識の実社会での利用が期待されている．ところが実際には，社会での音声認識の利用はそれほどみられない．これは，音声認識システムが，音響的に静的である環境下での理想的な発話データを基にして構築されているため，実際に音声認識が利用されると考えられる環境においては，認識精度が著しく低下するためである．

現状の音声認識装置では，背景雑音の影響を受けにくくするために，ユーザはマイクロフォンの前で（マイクロフォンスイッチを押してから）発話をする必要がある．そのような音声認識装置では，音声を使うメリットの一つである“ハンズフリー”なインターフェースを提供しているとは言えない．今後，音声認識装置が人の生活環境下において自然に使われるためには，その装置を人が意識せずに使えるようにする必要がある．

２．研究の目的

本研究課題では，上記のような研究背景に基づき，マイクスイッチレスな音声認識の実現を目指し，各要素技術の研究を進めた．具体的には，以下の４つの要素技術に取り組んだ．

（1）３次キュムラント特徴量を用いた発話区間検出の研究

雑音下において音声認識を行う際，音声非音声の判定により音声区間検出(VAD: Voice Activity Detection)を行う必要がある．静かな状況ではゼロクロッシング法などにより区間検出を行うことが可能である．しかし雑音下，特に音声の大部分が雑音に埋もれてしまっているような状況においては，従来の手法では十分な結果を得ることができない．本研究では，雑音に対するロバストな音声区間検出の手法として，高次統計量として知られている 3 次キュムラント（ 3rd order cumulant ）の Bispectrum を用いて，PCA による次元圧縮後，MFCC (Mel Frequency Cepstrum Coefficient) との初期統合を行う方法を提案する．

（2）スペクトル平面における勾配ヒストグラムに基づく音声特徴量抽出の研究

時間－周波数平面上における対数パワースペクトルの勾配情報に基づく特徴量を用いた音声特徴量抽出手法について検討を行う．現在，音声特徴量として MFCC が広く用いられているが，時間特徴が表現されていないという問題がある．また，ΔMFCC や ΔΔMFCC は線形回帰係数であるため，時間特徴の直接的な表現でないと言える．これに対し，本研究では，より直接的に時間特徴を表現するため，時間－周波数平面上の局所領域から勾配情報に基づく音声特徴量を抽出する手法を提案する．

（3）音響モデルを用いた突発性雑音除去の研究

家の中のような実環境で音声認識を使用することを考えるとき，雑音にはドアの開閉音や電話の音など中には突然発生するものも少なくない．発話中に雑音が発生した場合，そのデータから雑音の情報のみを取り出すことは困難である．そこで，本研究課題では，突発性雑音の検出と識別手法を提案し，さらにその除去手法を提案する．

（4）音響モデルを利用したシングルチャネル音源方向推定の研究

これまでに提案されてきた音源位置の推定方法は，マイクロフォンアレーにおける各観測信号の時間差を用いた手法が多く，複数のマイクロフォンが必要であった．コストが重要視される車載機器や，様々な環境下で計算機が使われるユビキタス社会においては，シングルマイクロフォンだけで方向推定を行なう手法は，非常に重要であると考えられる．そこで，本研究課題では，音響モデルを利用することにより，時間差という情報を用いずに単一マイクロフォンで音源方向を推定する方法を提案する．

３．研究の方法

本研究課題にて取り組んだ４つの要素技術の研究方法について述べる．

（1）３次キュムラント特徴量を用いた発話区間検出の研究

本研究課題では，音声に非常に強い雑音が重畳している様な環境においても，音声と非音声を精度良く分離できるような音声特徴の定式化を行なった．

一般的に，雑音は音声に比べ正規分布から発生した乱数に近い．そのため 3 次以上のキュムラントについては，音声であれば大きな値となり，雑音であれば小さな値になると考えられる．すなわち，3 次以上のキュムラントには音声と雑音を区別する能力が存在する．そこで計算コストも考慮に入れ，音声特徴抽出に 3 次のキュムラントを用いることを考えた．また，3 次キュムラントのバイスペクトルによって得られる音声特徴はフレーム間での相関であり，従来用いられる音声特徴量 MFCC によって得られる情報は，各フレーム内での音声情報である．これらは相互に補完しあっていると考えられるので，統合することを考える．

（2）スペクトル平面における勾配ヒストグラムに基づく音声特徴量抽出の研究

本研究課題では，HOG (Histogram of Oriented Gradients)による局所特徴量について研究を行なった．提案手法では，まず時間－周波数平面における勾配強度 m(t,f)と勾配方向 θ (t,f)を次のように求める．

$$m(t,f) = (d_t(t,f)^2 + d_f(t,f)^2)^{1/2}$$

$$\theta(t,f) = \tan^{-1}(d_f(t,f)/d_t(t,f))$$

d_t は，時間方向の変化量，d_f は周波数方向の変化量を表している．求めた局所領域における勾配強度 m(t,f)と勾配方向 θ (t,f)から重み付き方向ヒストグラムを作成する．図１に局所勾配特徴量と重み付方向ヒストグラムを示す．この重み付方向ヒストグラムを算出する参照点を周波数方向に等間隔で配置し，得られたヒストグラムをフレーム内で縦に並

べたベクトルを勾配特徴量として用いて音声認識を行なう．

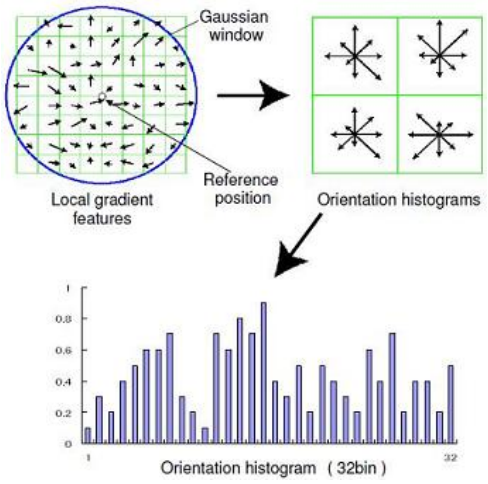

図 1. 局所勾配特徴と方向ヒストグラム

（3）音響モデルを用いた突発性雑音除去の研究

本研究で扱う雑音はそのほとんどが短時間しか継続せず，またいつ起こるかわからないものである．そのためまず除去を行う前に雑音が重畳しているかどうか判定を行い，またそれがどのような雑音であるか識別する必要がある．雑音の検出には AdaBoost を用いる．AdaBoost は Boosting の一種であり，多数の弱識別器を使うことで，非線形な識別器を作成することができる．同じように非線形な識別器として SVM などがあげられるが，それらの手法に比べると，パラメータの調整が少なく，また高速で動作するというメリットがある．学習する雑音データの雑音重畳音声を作成し，それらすべてとクリーン音声を用いて識別器の学習を行う．この識別器を用いて雑音と判定されたフレームに対し，雑音の除去を行なう．

本研究では対数メルフィルタバンク特徴量を使用して雑音の除去を行なう．メルフィルタバンク領域において，雑音重畳音声は，クリーン音声 S と雑音 N を用いて以下のように表される．

$$X(t) = S(t) + N(t)$$

我々の手法では，雑音の識別まで行なうため，雑音の特徴量の分布はある程度わかっていると考えられる．しかしながら，雑音の強さは未知である．そのため，この雑音に強さを表す未知の定数 α を掛け合わせ以下のように表す．

$$X(t) = S(t) + \alpha \cdot N(t)$$

上式を対数領域に変換し，混合多次元正規分布による音声モデルを用いて，ＥＭアルゴリズムに基づきパラメータ α，及び除去後の音声信号 $\hat{S}$ を求める．

（4）音響モデルを利用したシングルチャネル音源方向推定の研究

本研究では，音響モデルを利用することにより，時間差という情報を用いずに単一マイクロフォンで音源方向を推定する方法を提案する．

ある場所（位置）で発声されたクリーン音声 s は，音響伝達特性 h の影響を受ける．この時，観測信号 o はケプストラム領域にて，以下のように表される．

$$O(i;n) = H(i) + S(i;n)$$

ここで i は，ケプストラムの次元を表し，n はフレーム番号を表す．ケプストラムは音声情報を効率よく表現出来るパラメータの一つであり，音声認識ではよく使われる．本研究では，このケプストラム領域にてクリーン音声モデルを作成する．上式より，O が観測されれば，後は S が分かれば音響伝達特性 H を推定することが出来る．しかしクリーン音声 S を観測することは出来ないので，本研究では，S の代わりにクリーン音声モデルを用い，ケプストラム領域にて尤度最大基準に基づいて O から H を分離する．具体的には，あらかじめ事前にクリーン音声の音響モデルを作成しておき，各方向から到来する数単語の音声から EM アルゴリズムを用いることにより，音響伝達特性を推定する．これにより得られた音響伝達特性の時系列データから，各方向における音響伝達特性モデルを作成する．そして実際の入力音声から同様にして音響伝達特性を推定し，これらのモデルとの尤度を求めることで音源方向の決定を行う．

４．研究成果

本研究課題にて取り組んだ４つの要素技術の研究成果について述べる．

（1）３次キュムラント特徴量を用いた発話区間検出の研究

提案手法の評価データとして，高速道路走行時にて録音された発話データを用いた．男性 4 名，女性 4 名，各話者 100 発話で計 800 発話からなる．発話内容は日本各地の地名である．SN 比は高速道路走行時でおよそ 0～7 dB，平均約 4 dB である．高速道路走行時には背景雑音として排気音，走行音等が含まれる．図 2 に従来手法である MFCC+ΔMFCC と，提案手法を用いた場合の音声，雑音識別率を示す．実験結果より，本提案手法が従来手法よりも上回る識別結果を得られることを国内外で初めて示した．今後の予定として，現在，波形から算出している３次キュムラントバイスペクトル特徴を時間－周波数平面など別の空間上から算出する方法を検討し，

さらに雑音に音楽などを加えた状況下での実験を行なっていく．

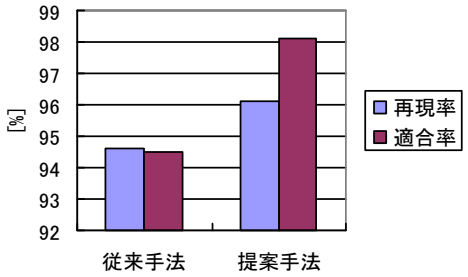

図 2. 高速道路走行時における音声区間検出結果

（2）スペクトル平面における勾配ヒストグラムに基づく音声特徴量抽出の研究

実際の雑音環境下（学生食堂，高速道路付近）にて，提案手法の有効性を検討した．図 3 に音声認識結果を示す．雑音が存在しないクリーン（Clean）環境下では，従来手法と比較して，大きな認識改善は得られなかった．一方，雑音下では，MFCC 特徴量と勾配ヒストグラム特徴量を統合することにより，従来手法から 2～3%弱までの認識精度の改善が得られた．この結果より，提案手法では，従来の MFCC 特徴量が持たない識別に寄与する情報を含んでいると言える．

しかし，提案手法は主成分分析による次元削減後でも 50 次元のベクトルであり，MFCC と比べると高次元であることから，勾配特徴ベクトルの持つ情報を損なわずにより低次元へと次元を削減することが求められる．また，より雑音にロバストな性質を持たせるためには，ケプストラム減算法のような正規化処理が必要と考えられる．今後は，これらの問題解決に取り組む予定である．

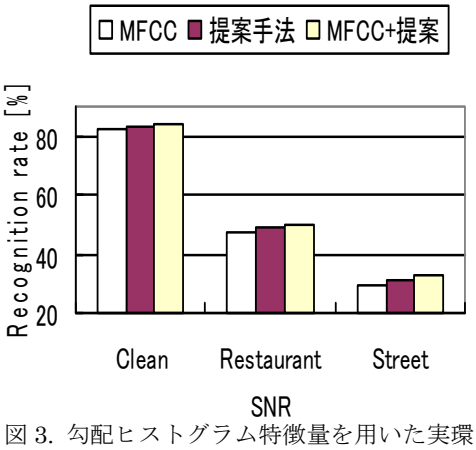

図 3. 勾配ヒストグラム特徴量を用いた実環境下での音声認識実験結果

（3）音響モデルを用いた突発性雑音除去の研究

雑音の検出，識別を組み合わせた突発性雑音除去手法の有効性を音声認識実験より示す. 音声認識率を図 4 に示す. テスト話者は，男性２名，女性２名に対し各話者 500 発話を用いた．突発性雑音は 105 種類とし，SNR は，-5, 0, 5 dB とした．音声認識実験結果より，従来手法の重みα（雑音の強さ）を考慮しない場合と比べて, 提案手法は全ての SNR 環境下において音声認識精度の改善が得られることを国内外にて初めて示した．今後は，突発性雑音の検出精度を改善することによる認識率の変化を調べていく予定である．

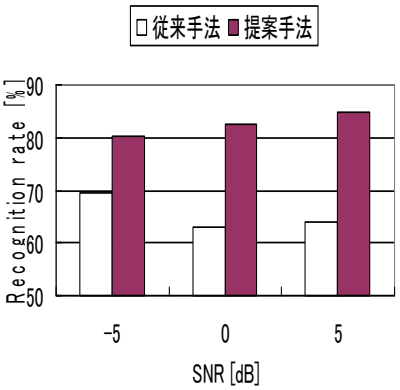

図 4. 突発性雑音除去後の音声認識結果

（4）音響モデルを利用したシングルチャネル音源方向推定の研究

単一マイクロフォンによる音源位置推定法の有効性を示す．音源は男性話者 1 名とし，クリーン音響モデルの作成には，2,620 単語を用いた. 音源とマイクロフォンの距離は 2m, 残響時間は 300 msec とした．評価データには，1,000 単語を使用し，30，90，130 度の位置から到来する音声信号を用いた．音響伝達特性, 観測信号の学習データには, 10 単語, もしくは 50 単語を使用した．方向推定結果を図 5 に示す．学習データ数に限らず，提案手法のほうが従来手法より良い結果であり，単一マイクロフォンでの音源方向推定手法の有効性を国内外にて初めて示した．今後の課題として，話者が不特定になった場合や雑音が混入した場合，未知の方向から音源が到来する場合や学習と評価で部屋が違う場合などに対応していくことが挙げられる．

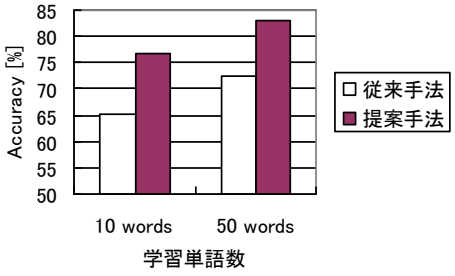

図 5. 音源位置推定結果

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計１１件）

① Nobuyuki Miyake, Tetsuya Takiguchi, Yasuo Ariki, “Sudden Noise Reduction Based on GMM with Noise Power Estimation,” Interspeech, pp. 403-406, 2008, 査読有.
② Tetsuya Takiguchi, R. Takashima, Y. Ariki, “Active Microphone with Parabolic Reflection Board for Estimation of Sound Source Direction,” Joint workshop on Hand-free Speech Communication and Microphone Arrays, pp. 65-68, 2008, 査読有.
③ Tetsuya Takiguchi, Y. Ariki, “PCA-Based Speech Enhancement for Distorted Speech Recognition,” Journal of Multimedia, Vol. 2, Issue 5, pp. 13-18, 2007, 査読有.
（他８件）

〔学会発表〕（計２０件）

① 高島遼一, “音響伝達特性モデルを用いたシングルチャネル音源位置推定の検討” 日本音響学会 2009 年春季研究発表会, 2009 年 3 月 18 日, 東京.
② 室井貴司, “スペクトル平面における勾配ヒストグラムに基づく音声特徴量の検討” 第 10 回音声言語シンポジウム, 2008 年 12 月 10 日, 東京.
③ 三宅信之, “音声の動的特徴のモデルを使った突発性雑音の除去” 第 10 回音声言語シンポジウム, 2008 年 12 月 10 日, 東京.
（他１７件）

６．研究組織
(1)研究代表者
滝口　哲也（TAKIGUCHI TETSUYA）
神戸大学・自然科学系先端融合研究環都市安全研究センター・講師
研究者番号：４０３９７８１５

(2)研究分担者

(3)連携研究者