様式 C-19

科学研究費補助金研究成果報告書

平成２３年５月３０日現在

機関番号：１２６０８
研究種目：若手研究（Ｂ）
研究期間：２００９～２０１０
課題番号：21700188
研究課題名（和文）　目的音モデル尤度を用いた高速な耐雑音音声認識フロントエンドの研究

研究課題名（英文）Efficient noise robust front-end based on target speech model likelihood for automatic speech recognition

研究代表者
篠崎　隆宏　（Shinozaki Takahiro）
東京工業大学　大学院情報理工学研究科　助教
研究者番号：80447903

研究成果の概要（和文）：
雑音の補償操作を雑音の定式化に最も適した短時間スペクトル領域で行い、補償のためのパラメタ推定に音声の性質を表すのに最も適した音声特徴量に対する最尤基準を用いる目的音 GMM スペクトル補正法(TGSC 法)の提案を行った。構成のバリエーションやパラメタの推定法等について最適な条件の探索を行い、音声認識実験により効果を示した。また実時間動作が可能であることを確認した。

研究成果の概要（英文）：
To improve speech recognition performance in adverse conditions, a noise compensation method is proposed and investigated that applies a transformation in the spectral domain whose parameters are optimized based on likelihood of speech GMM modeled on the feature domain. Experimental results show that the proposed method is able to work in real-time and it is effective to reduce noise effects.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2009 年度	1,800,000	540,000	2,340,000
2010 年度	1,500,000	450,000	1,950,000
年度			
年度			
年度			
総　計	3,300,000	990,000	4,290,000

研究分野：総合領域
科研費の分科・細目：情報学　知能情報処理・知能ロボティクス
キーワード：音声情報処理、音声認識、耐雑音処理

１．研究開始当初の背景

音声認識技術は人が話す音声に周囲の雑音（加算性の雑音）や、音声が伝搬するチャネル特性の変化（乗算性の雑音）が加わると、認識性能が大幅に低下してしまう問題がある。音声認識システムは音声波形から認識に適した特徴量を抽出するフロントエンドと特徴量から単語列への変換を行うデコーダから構成される。様々な環境で頑健に動作する音声認識システムを実現するためには、環境の変化に対して適応的に動作する特徴抽出フロントエンドを実現することが重要であると考えられる。

２．研究の目的

様々な雑音に対して適応的に動作するためには、様々な雑音の知識を網羅的にモデル化するか、あるいは雑音の知識を用いずに耐雑音処理を行う必要がある。本研究では後者の立場をとる一方で、音声に関する知識をガウ

ス混合モデルとしてモデル化し雑音抑圧へ利用することで効果的な雑音抑圧を実現することを目的とする。

３．研究の方法

電話やインターネットを介した音声対話システム、あるいはポータブルなデバイスのマイクを通した音声インタフェースなどにおいて、様々な雑音環境に対して頑健に動作するオンライン認識システムを実現することは非常に重要である。マイクから音声認識システムに入力される音声には、図１に示すように、加算性の雑音と乗算性の雑音が重畳される。これらの雑音の影響は、スペクトル領域ではそれぞれ音声に対する加算項および乗算項として表現される。もしこれら加算項および乗算項が推定できれば、元の音声を雑音重畳音声から周波数ごとにアフィン変換を行うことで求めることができる。

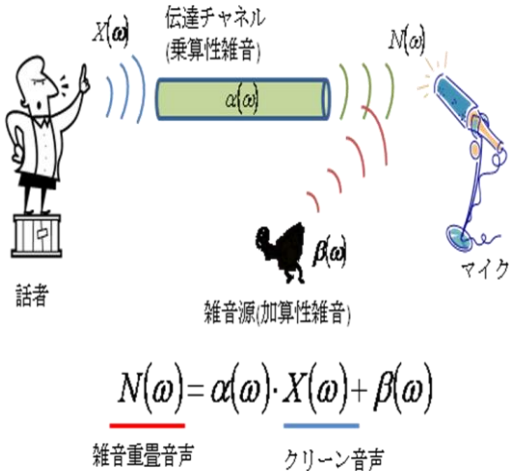

図１　発声された音声信号 X(ω)に対し、マイクにより録音される音声は乗算性雑音 α(ω)と加算性雑音 β(ω)が重畳された雑音重畳音声 N(ω)である。

ここで問題となるのが、如何にして乗算性雑音と加算性雑音を、雑音重畳音声から推定するかということである。本研究では、予め雑音の重畳の無いクリーンな音声から音声のモデルを作成しておき、そのモデルと観測された雑音重畳音声を元に一定区間ごとに雑音係数の推定を行うアプローチをとる。音声のモデルとしては、ガウス混合分布を用いる。

ガウス混合モデルによる音声のモデル化は通常、音声のモデル化に適したメルケプストラム係数などの領域で行われる。一方で雑音抑圧の観点からは、加算性および乗算性雑音の影響はスペクトル領域で最も簡潔に表現され、音声特徴量に変換してしまうと非線形に歪んでしまい対応が難しくなる。そこでスペクトル領域での雑音抑圧のための逆線形変換処理と特徴量領域での雑音抑圧効果の評価をニューラルネットに似た微分のチェーンルールで連結した形のアルゴリズムを提案し、構成のバリエーションやパラメタの初期化法等について最適な条件の探索を行う。

４．研究成果

音声認識における一般的な特徴量抽出は、まず数十ミリ秒ごとに短時間スペクトルを求め、フィルタバンクによりスペクトルの微細構造を取り除き、対数スペクトル領域でコサイン変換することにより行われる。音声認識に必要な音声の特徴はこれら一連の操作を経て抽出された音声特徴量により最も的確に表現され、隠れマルコフモデルやガウス混合モデル(GMM)などによりモデル化される。他方で音声に対する雑音の影響はこれらフィルタ演算や対数演算等により定式化が難しくなる。

そこで、雑音の補償操作を雑音の定式化に最も適した短時間スペクトル領域で行い、補償のためのパラメタ推定に音声の性質を表すのに最も適した音声特徴量に対する最尤基準を用いる目的音 GMM スペクトル補正法(TGSC 法)の提案を行った。提案法における処理の流れを図２に示す。

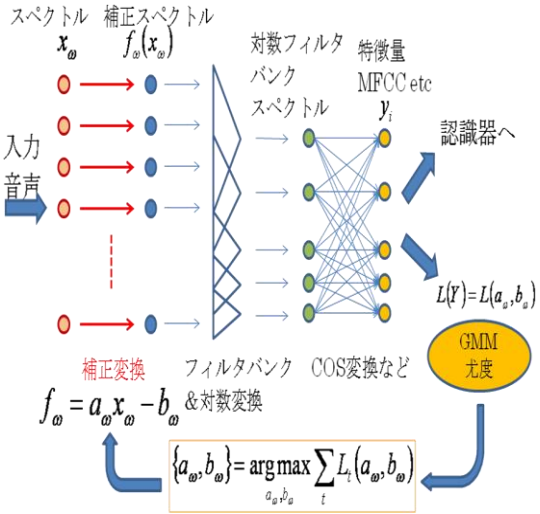

図２　提案法における雑音補正処理。スペクトル分析された雑音が重畳した入力音声に対し、アフィン変換により雑音補正を行う。補正の際に、加算性雑音と乗算背雑音に対応した補正係数が必要となるが、これらは補正後の音声が、事前に作成した音声モデルを基準に、最も音声として尤もらしくなるように決定する。

さらに、ケプストラム領域でその長時間平均を差し引くことで音声伝達チャネルの影響を取り除く技術である CMS を TGSC 法に組み込むことで拡張を行った。また、TGSC 法におけるパラメタの最適化について、使用する GMM の混合数の最適化や繰り返し最適化法における繰り返し数の最適化などを行い、耐雑音性能への影響を抑えながら計算量を抑制できる条件について調べた。

図３に提案法により雑音補正を行った場合の、音声認識精度の向上の様子を示す。提案法においては、係数ベクトルの最適化の際に使用する初期値の決定が重要である。初期値を定数とした場合よりも、雑音区間から推定したノイズベクトルを用いた場合に特に大きな認識精度の向上が見られた。

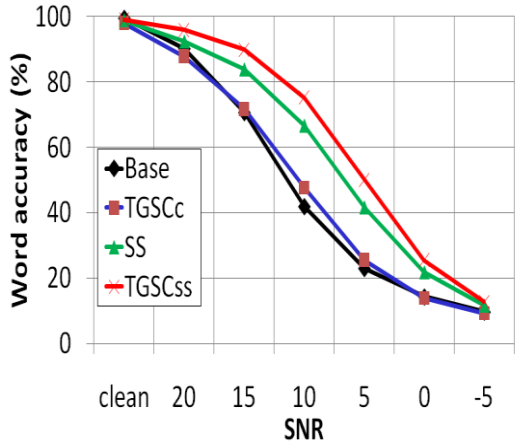

図３　雑音条件(SNR)と単語正解精度(Word accuracy)の関係。”Base”は雑音補正なし、”SS”はスペクトルサブトラクション手法により ノイズ補正を行った場合、”TGSCc”は提案法において定数を初期値とした場合、”TGSCss”はノイズベクトルを初期値として用いた場合である。

図４に、補正係数を求める際に必要な繰り返し最適化における繰り返し数を変えた時の、処理時間と単語認識精度の関係を示す。繰り返しの１回目における認識精度の向上が大きく、その後５回目までは繰り返しを増やすことで単語認識精度が向上しているのが分かる。繰り返し数を 10 まで増やすと、過学習のため若干の精度低下がみられる。また図より、繰り返し数を１とした場合、認識率向上の大きな効果を得ながら、実時間動作が可能であることが分かる。

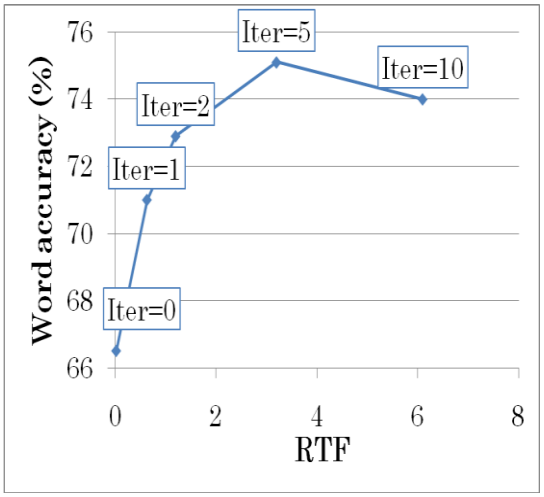

図４　実時間ファクター(RTF)と単語正解精度の関係。RTF は１秒間の音声を処理するのに何秒かかったかを表す。

図５に、音声モデルとして使用するＧMMの混合数と単語正解精度の関係を示す。混合数が低い間は混合数の増加とともに単語認識精度が大きく向上するが、100 以上になると伸びは緩やかになりやがて飽和する。混合数が大きくなると、必要なメモリサイズや計算量が増加するが、概ね混合数を 200 程度にとれば十分であることが分かる。

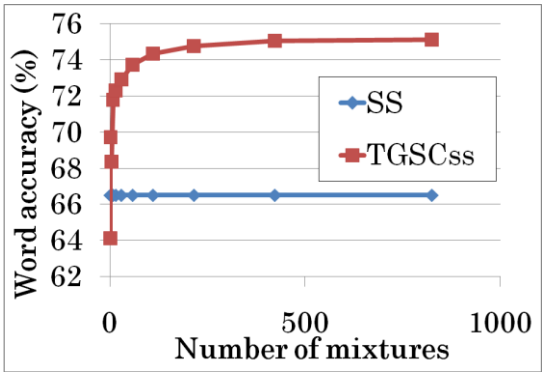

図５　混合ガウス分布の混合数と単語正解精度の関係。混合数は 200 程度にとれば、それ以上大きくしても単語認識精度には余り影響しないことが分かる。

またこの他に、ＧMMをもとに最尤推定を行う TGSC 法の派生と関連して、ＧMMの最尤基準により推定された変換パラメタを、音声の発話者の特徴とみなす手法について検討を行った。話者による声質の違いは一種の伝達チャネルの雑音とみなすことが出来、その特徴から発話者の年齢を推定することができる。この場合識別器としてサポートベクタ

一回帰が優れていることを示した。

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２件）
① Toshiya Wada, Takahiro Shinozaki, and Sadaoki Furui ″Investigations of Features and Estimators for Speech-based Age Estimation,″ Proc. APSIPA, Vol.1, pp.470-473, 2010.12. 査読あり

② Takahiro Shinozaki and Sadaoki Furui, ″Target Speech GMM-based Spectral Compensation for Noise Robust Speech Recognition,″ Proc. INTERSPEECH, Vol.1, pp. 1255-1258, 2009. 査読あり

〔学会発表〕（計２件）
① 和田 俊也, 篠崎 隆宏, 古井 貞煕 「年齢推定のための音声特徴量および推定器の検討」電子情報通信学会 2010年6月17日 九州大学（福岡県春日市）

② 篠崎 隆宏, 古井 貞煕 「目的音 GMM 尤度基準スペクトル補正法の諸評価」日本音響学会、2009年9月15日 日本大学(福島県郡山市)

〔その他〕
ホームページ等
http://www.furui.cs.titech.ac.jp/~shinot/

６．研究組織
(1)研究代表者
篠崎 隆宏 (Shinozaki Takahiro)
東京工業大学 大学院情報理工学研究科
助教
研究者番号：80447903

(2)研究分担者 該当なし

(3)連携研究者 該当なし