

電子・情報工学専攻	学籍番号	015705	指導教員氏名	新田 恒雄 堀川 順生
申請者氏名	Muhammad Ghulam			

論文要旨(博士)

論文題目	A Study on Auditory Based Feature Extraction and HMM/SM Based Classification for Robust Speech Recognition (頑健な音声認識を目的とする聴覚に基づいた特徴抽出およびHMM/SMに基づく識別に関する研究)
------	---

音声は最も自然なコミュニケーション手段である。コンピュータ性能の最近の進歩は、音声技術の向上と相まって音声によるコンピュータと人間との対話の実用化を可能にしつつある。自動音声認識 (Automatic Speech Recognition: ASR) は、実環境での動作に課題が残されているものの、今日、多様な応用分野で利用されている。一方、雑音環境下でのASRは実用に耐えない程度に性能が急激に低下し、人間の能力と比較するとまだ大きく劣る。ASRシステムの性能低下は、学習時と認識時の環境条件のミスマッチにより生じる。この問題を改善する方法の一つは、人間の聴覚のメカニズムをASRに組み込むことが考えられる。特徴抽出器と隠れマルコフモデル (Hidden Markov Model: HMM) の音響モデルを使用した分類器から構成される従来のASRシステムでは、学習時と認識時の環境条件差を特徴領域とモデル領域において減らすことが大変重要な問題であった。

本論文では、環境の違いから生じるASRシステムの性能劣化を、特徴領域とモデル領域双方で軽減する手法を提案する。最初に特徴領域では、聴覚に基づく特徴抽出方法を提案する。この方法は、聴覚システムが雑音に対して非常にロバストであることに研究の動機を得ている。聴覚神経系はピッチ同期機構を持っており、これが音声検出に有効であることが知られている。しかし、現在の聴覚に基づく特徴抽出モデルの中には、ピッチ情報を特徴抽出の重要な要素として利用しているものはみられない。本論文では、聴覚に基づくピッチ同期特徴抽出方法を提案し、雑音に対する頑健性を示す。

ピッチの自動追跡は、音声符号化や合成を含む音声処理技術分野で多くの応用を持っている。しかし、その多くは雑音の少ない音声に動作が限定され、雑音環境での実用には向かない。我々は本論文で、雑音にロバストで遅延の無い新規のピッチ抽出アルゴリズムを導入する。提案する特徴抽出方法では、(i) ピッチ同期のゼロ交差ピーク振幅に基づく手法 (Pitch-Synchronous Zero-Crossing Peak-Amplitude: PS-ZCPA) と、設計上より効率的な (ii) ピッチ同期ピーク振幅 (Pitch-Synchronous Peak-Amplitude: PS-PA) に基づく方法の二つを提案し、雑音に頑健なASRを目指す。

特徴抽出の主な目的は認識に必要な話者特性、チャンネル特性、付加雑音などを除き、重要な言語情報を欠落させないことにある。音声認識で用いられる言語情報は、帯域フィルタの各チャンネル出力信号をband-passフィルタ (modulationフィルタとも呼ばれる) を通すことで抽出される(変調周波数強調処理)。他の重要な聴覚効果の一つにマスキングがある。本論文では、変調周波数強調と聴覚マスキングの双方を提案特徴抽出手法に組み込み、性能向上に寄与することを示す。また、提案手法にはウィナーフィルタ (Wiener Filter: WF) を基にした雑音除去手法も組み込み評価している。

人間の脳が音声を知覚する高次の処理については、まだほとんど解明されていない。一方、確率論的統計に基づいた単語や文音声のモデルは、数式の理解し易さから広く利用されている。HMMはこれまで、音声認識の様々な応用分野に亘るタスクに適合してきた。しかし、音声には不確定要素が多く、それが要因となり発話の音響特性に影響を与える。我々は、異なる尤度正規化手法をHMMシステムで検討した後、モデル領域での新しい音声認識手法としてHMM/SM (Subspace Method) 法を提案した。この手法は、SMが音素などマイクロなレベルで高い性能を示す一方、文発話などマクロな時間変動が大きいレベルではHMMが高い性能を示すという事実ヒントを得ている。本論文では、発話内の音響品質の違い(文頭や文末では発声器官の影響で品質が劣る)による音響特性の劣化を、HMMとSMを組み合わせることで解決している。具体的には、HMMから与えられる音素境界情報を利用して、モノフォン単位の固有ベクトルを学習し、SMで音素を尤度正規化しながら高精度に識別する。