環境音モデルを用いた頑健な音声認識に関する研究（0221036）

Robust Speech Recognition Using Non-Speech Models

山田武志　筑波大学
Takeshi Yamada　University of Tsukuba

研究期間　平成 14 年度～平成 15 年度

概要　本研究の目的は、環境音の隠れマルコフモデル（HMM）を精密に生成し、雑音環境下での音声認識精度を改善することである。まず、環境音モデルの単位や構造を決定する手法として、逐次状態分割法による環境音モデルの生成法、及びエルゴディックモデルからの環境音モデルの再生成法を考案し、評価実験によりその有効性を確認した。また、このような環境音モデルを最大限に活用するために、音声と環境音の重畳区間情報を推定する手法、及び推定した重畳区間情報を HMM 合成法に基づく音声認識において利用する方法を開発し、認識実験によりその有効性を示した。さらに、尤度に基づいて音声・非音声判別、既知・未知環境音判別を行うシステムを構築し、その性能を評価すると共に環境音データを収集した。

Abstract　The purpose of this research is to generate accurate non-speech models by HMM (Hidden Markov Model) and then to improve the performance of noisy speech recognition by using them. This research has proposed two methods for the decision of the unit/architecture/parameter of the non-speech models, which are based on a successive state splitting algorithm and an ergodic non-speech model. To make the best use of the non-speech models, a recognition algorithm based on HMM composition and noisy frame detection has been proposed. Furthermore, a recording system, which always monitors unknown noise, has been developed. Experimental results confirmed the effectiveness of the proposed methods.

1　研究内容

近年の音声認識技術の発展は目覚しく、ディクテーションシステム（音声ワープロ）などのアプリケーションが市販されるまでに至っている。しかし、これらの音声認識システムには、周囲雑音の影響によって認識精度が低下するという問題がある。現状では接話マイクロホンの使用によってこの問題を回避しているが、このままでは音声インタフェースとしての利便性を十分に生かすことができない。よって、マイクロホンから離れた位置での発話を可能にする技術（ハンズフリー音声認識の技術）が必要不可欠である。

従来、ハンズフリー音声認識を実現するために様々な研究がなされている。これらの研究では、認識対象である音声以外の音を一律的に雑音とみなしていることが多い。しかし、実世界に存在する多種多様な音（環境音）を雑音として一括りに扱うことには無理があり、雑音の種類によっては十分な認識精度が得られないという問題が生じる。よって、広範な雑音環境下で頑健な音声認識を実現するためには、個々の環境音の特性を十分に考慮する必要があると考えられる。

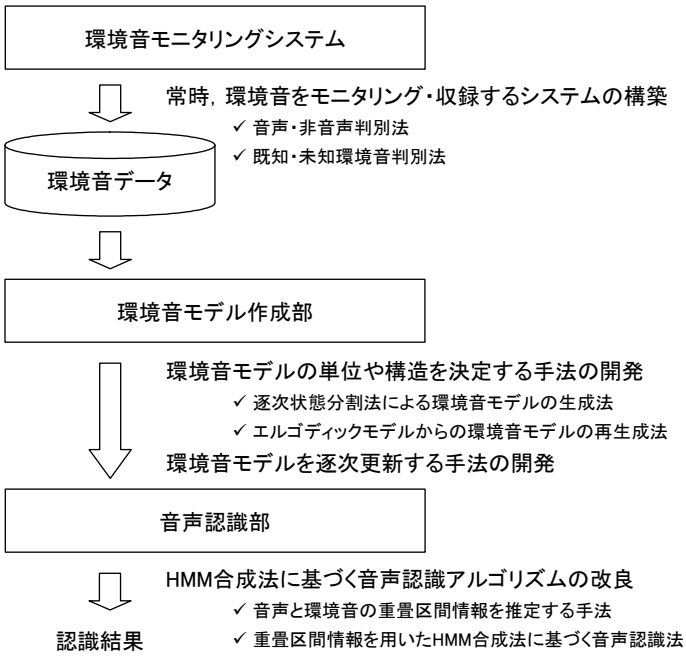

図 1　提案システムの概要

以上から、本研究では、環境音の隠れマルコフモデル（HMM）を精密に生成し、このような環境音モデルを用いて雑音環境下での音声認識精度を改善することを目的とする。具体的には、図 1 に示すように、環境音モデルの単位や構造を決定する手法を開発すると共に、このような環境音モデルを最大限に活用するために、HMM 合成法に代表される音声認識アルゴリズムの改良に取組む。さらに、無数に存在する環境音をあらかじめ全てモデル化しておくことは現実的ではないので、未知の環境音を常時モニタリング・収録するシステムを構築し、環境音モデルを逐次更新していく手法を開発する。

2　研究結果

2.1　環境音モデルの単位や構造を決定する手法の開発

2.1.1　逐次状態分割法による環境音モデルの生成法

環境音認識においては、モデルの単位を、その音源の意味内容に基づいて決めることが多い。一方、音声認識においては、必ずしもその必要はないと考えられる。そこで、環境音の音響的・時間的特長を反映するモデルの構造に基づいて環

境音をクラスタリングし、このようにして得られた各クラスをモデルの単位とすることを考え、1993 年に鷹見らにより提案された、逐次状態分割法による隠れマルコフ網の自動生成法に着目した。遂次状態分割法は、学習データに基づいて自動的に隠れマルコフ網の構造を決定する方法である。その利点は、隠れマルコフ網の構造の決定やモデルパラメータの推定を、出力尤度最大化という共通の基準の下で自動的に行うことにある。また、複数のコンテキスト間で状態を共有することにより、統計的に安定した学習が行える。これは、学習データの量が制限されることが多い環境音にとって非常に有効である。

逐次状態分割法により、RWCP 実環境音声・音響データベースに含まれる 92 種類の環境音に対してモデルの単位と構造を決定し、環境音認識実験により初期評価を行った。その結果、逐次状態分割法により環境音の音響的・時間的特長に応じたクラスが概ね形成されているものの、クラスの分割に失敗しているケースもあることが分かった。今後は、その原因を調査すると共に、2.2 の音声認識アルゴリズムを用いて評価する予定である。

2.1.2 エルゴディックモデルからの環境音モデルの再生成法

2.1.1 の手法は、事前に想定した環境音に特化したモデルを生成するため、そのモデルは未知の環境音に対しては必ずしも適していないという問題がある。この問題を解決するために、多種の環境音で学習したエルゴディック構造のモデルから、認識対象の音声に重畳している環境音の特性を適切に表すモデルをその都度再生成する手法を考案した。

提案法の有効性を評価するために、雑音下連続数字認識タスクである AURORA-2J を用いて認識実験を行なった。その結果、環境音の種類毎に十分なデータで学習したモデルを用いた場合と同程度の認識性能が得られること、様々な構造のモデルを柔軟に生成できることが分かった。今後は、環境音の特性に応じた構造を自動的に決定する手法を開発する予定である。

2.2 HMM 合成法に基づく音声認識アルゴリズムの改良

環境音が重畳している音声を精度良く認識するための手法として、HMM 合成法が提案されており、その有効性が広く示されている。しかし、HMM 合成法を適用する際には、音声と環境音が重畳している区間、音声に重畳している環境音の種類とその SN 比、といった情報をあらかじめ推定しておく必要がある。そこで、音声と環境音の重畳区間情報を推定する手法、及び推定した重畳区間情報を HMM 合成法に基づく音声認識において利用する方法を提案した。

生活環境音データベースに含まれる 7 種類の環境音を音声データに重畳し、連続単語認識実験を行った。その結果、提案法の単語認識率は重畳区間情報の正解を与えた場合と同程度であることを確認した。その一方で、認識性能の改善度は環境音の種類に依存していることが分かった。今後は、2.1 の環境音モデルを用いることにより、認識性能のさらなる改善を図る予定である。

2.3 環境音を常時モニタリング・収録するシステムの構築

従来の HMM 合成法に基づく音声認識では、環境音モデルをあらかじめ作成しておく必要があるため、未知の環境音に対してはその効果が十分に得られないことがある。よって、音声認識の利用時に、環境音データを自動的に収集し、環境音モデルを逐次学習する機能が必要である。さらに、その環境音が既知のものか、未知のものかを判別し、学習の高効率化・高精度化を図る必要がある。そこで、音声・非音声判別、既知・未知環境音判別を行うシステムを提案した。提案システムは、話者認識・話者照合の技術を応用したものであり、尤度に基づいて各種判別を行う。

実環境で収録したデータを用いて性能を評価した結果、音声・非音声判別、既知・未知環境音判別共に、80%以上の判別率が得られた。一方、判別率と判別数の間にはトレードオフの関係があり、判別率を優先すると学習のためのデータ量が不足することになる。今後は、尤度に対する閾値を適切に決定する方法について検討する予定である。また、システムのリアルタイム化に取組む予定である。

3. 今後の課題

今後は、上述した課題の解決を図ると共に、環境音モデルを逐次更新していく手法の開発と評価を行っていく予定である。さらには、全てを統合したシステムを構築し、広範な雑音環境下でその性能を評価していきたい。

音声認識において、入力装置であるマイクを特に意識せずに、人に話しかけるような感覚での自然な発話を可能とするためには、雑音の問題を解決せねばならない。本研究は、そのためのアプローチの一つとして提案・実施したものである。本研究の成果が、雑音下音声認識の研究開発や実世界に存在する多種多様な環境音の特性の系統的な分類のための一助となることを願っている。

口頭発表リスト

[1] Takeshi Yamada, Naoto Isaka, Hiroshi Osuka, Nobuhiko Kitawaki, Futoshi Asano, “Noise robust speech recognition based on HMM composition and noisy frame detection,” Proc. International Congress on Acoustics, ICA2004, Vol. IV, pp. 2835-2838, Apr. 2004.

[2] 井坂直人, 山田武志, 北脇信彦, “HMM 合成法を用いた音声認識のための環境音モデルの生成の検討,” 日本音響学会春季研究発表会, pp. 163-164, Mar. 2004.

[3] 井坂直人, 山田武志, 北脇信彦, 浅野太, “隠れマルコフ網と逐次状態分割法による環境音モデル化の検討,” 日本音響学会秋季研究発表会, pp. 179-180, Sep. 2002.

他 3 編