韻律情報を利用した
雑音重畳音声の認識に関する研究

川崎 真護

北陸先端科学技術大学院大学 情報科学研究科

1998 年 2 月 13 日

キーワード： 白色雑音, 単語音声認識, ピッチパターン,F_0生成モデル, 韻律尤度.

1 はじめに

音声認識技術を実用化するにあたっては、実環境下において頑健であることを想定しなければならない。しかし、音声などに含まれる音韻情報は雑音の影響を受けやすく、雑音除去、あるいは雑音適応などの技術が必要である。一方、代表的な韻律特徴量であるピッチパターンは音声の基本周波数情報であるので、ある程度の大きさの白色雑音下では影響を受けない。

そこで、本研究ではこのピッチパターン情報を従来からある音韻認識と併用したシステムを構築し、雑音環境下における評価を行う。

2 韻律情報を用いた単語音声認識システム

このシステムは HMM を利用した音韻認識部とピッチパターン情報を利用した韻律認識部で構成される。音韻認識部には音韻 HMM と単語辞書を備えている。また、韻律認識部にはピッチテンプレートと韻律辞書を備えている。入力単語に対して、音韻認識部では音韻尤度 S^{ph} を計算し、韻律認識部ではピッチパターン整合法により韻律尤度 S^{pr} を計算する。この 2 つの認識結果を結合係数 α を用いて結合し、総合スコア $S = (1.0 - \alpha)S^{ph} + \alpha S^{pr}$ を計算する。

3　韻律認識部の学習と認識処理

ピッチテンプレート作成のために、まず、自動抽出されたピッチパターンに対して F_0 生成モデルパラメータを推定することによりピッチパターンを再構成し、これを学習資料とする。この際に、話者に依存した最低基本周波数 F_bを引くことにより高さの正規化、および長さの正規化を行う。

次に、LBG 法によるクラスタリングによって全学習資料を J個のクラスに分類し、これをピッチテンプレートとする。韻律辞書には、複数の話者によって発声された同一単語の各クラスへの出現頻度を記載する。

韻律辞書に記載されている単語 i がクラス jに出現する確率を p_{ij}とし、入力単語のピッチパターンとテンプレート jとの距離を d_jとする。この時、入力単語が単語 i であると仮定した場合の韻律尤度は $S_i^{pr} = -\sum_{j=1}^{J} p_{ij} d_j$ によって計算される。

4　雑音環境下におけるシステムの評価

実験には ATR の多数話者データ (男性 20 話者) による最重要単語 520 語を用いる。この内、雑音の重畳していない男性 15 話者のデータを学習データとして用い、雑音を重畳した残りの男性 5 話者を評価用データとして用いる。雑音としては、SNR が 45dB,25dB,5dB となる白色雑音を重畳する。

実験では SNR の値を変化させ、誤認識単語数の変化よりピッチパターンの雑音重畳単語音声認識への有効性を検討する。

5　結論

音韻尤度と韻律尤度の結合係数 α の値を変化させた場合、どの SNR の単語音声においても音韻尤度のみの認識率と比べて、認識率が改善されることが確認された。結合係数 α を全入力において固定値として適用した場合には、 $\alpha = 0.90$ の時にすべての SNR において認識率がもっとも改善することが確認された。この時の HMM のみによる誤認識単語数からの改善率を調べてみると、SNR=∞ から 25dB までの区間において平均 7%弱の改善率が得られた。

また、上の実験では、すべての話者に共通な値の結合係数 α を用いたが、話者ごとに最適な値に設定すると、さらなる認識率の改善結果が得られることが確認された。

さらに、結合係数 α を入力毎に可変値 $(0 < \alpha < 1)$ とし、ピッチパターンによって改善できる単語数の上限値を調べた結果、SNR=∞ から 25dB までの区間において平均 50%強の改善率が得られることが判明した。これは何らかの基準で結合係数 α を自動制御することができれば、この区間の SNR においてピッチパターン情報が有効であることを示唆している。