

# 論文内容の要旨

## 博士論文題目

Automatic Acoustic Modeling using Tied-Mixture Successive State Split Algorithm  
(逐次状態分割アルゴリズムに基づくタイドミクスチャー音韻 HMM の学習)

氏 名

Alexandre Girardi

## (論文内容の要旨)

この博士論文では、音声認識における HMM 音響モデルのトポロジーの自動学習アルゴリズムについて述べる。この新しいアルゴリズムは、従来の最大尤度基準に基づく状態分割アルゴリズム(ML-SSS)と音韻木作成アルゴリズム(PDT)の拡張である。

ML-SSS と PDT は、どちらも音響モデルのトポロジー決定のためのトップダウン手法である。ML-SSS と PDT の問題点は、音韻の分割や HMM の状態の分割の基準として一個の正規分布を仮定していることである。これは、音声の特徴量の分布は複雑で、一個の正規分布では表現できないことが知られていることと矛盾している。この論文では、音韻分割、HMM 状態分割の基準としての混合正規分布を扱える実現可能な計算量のアルゴリズムを導出した。

通常の混合正規分布 HMM では、計算量が膨大となり、基準として混合正規分布は利用できない。この論文では、タイドミクスチャーHMM を利用することにより、音韻分割、HMM 状態分割の基準として、混合正規分布を用いるアルゴリズムを導出した。このアルゴリズムは、ML-SSS と同様に、最大尤度基準で音韻分割、HMM 状態分割を行なうことができる。しかも、混合正規分布であるので、実際の特徴量空間の性質をよりよく表している。

このタイドミクスチャーHMM を用いた音韻モデルのトポロジーの自動生成アルゴリズム(TM-SSS)の評価を、従来の ML-SSS と PDT との比較して行なった。まず、一人の話者が発声した音声データベースを用いて音韻モデルのトポロジーの自動生成を実行した。5千単語音声の認識実験を行ない、従来の ML-SSS の単語誤り率を約半減できることを確認した。この実験により、TM-SSS アルゴリズムの有効性が確認できた。

つぎに、現在要請されている不特定話者音韻モデルの作成に TM-SSS を適用することを試みた。不特定話者音韻モデルの学習には、少なくとも300人程度の膨大な音声データを用いる必要がある。現在の計算機の能力では、TM-SSS をそのまま用いるには無理があり、計算量を大幅に削減する必要がある。いくつかの仮定を導入することによって、認識性能の劣化を抑えながら、かつ計算量を大幅に削減した不特定話者音韻モデルのトポロジー学習が可能なアルゴリズム(FTM-SSS)を考案した。300人の音声データベースを用いて、FTM-SSS による音韻モデルのトポロジー学習を行なった。生成した音韻モデルの認識性能評価を、大語彙連続音声認識によって行ない、従来からのアルゴリズム ML-SSS と PDT との認識性能の比較も行なった。その結果、FTM-SSS は、ML-SSS および PDT よりも優れていることが確かめられた。

氏名	Alexandre Girard!
----	-------------------

## (論文審査結果の要旨)

この博士論文は、音声認識における HMM 音響モデルのトポロジーの自動学習アルゴリズムについて述べたものである。音声認識における音韻モデルの研究は、もっとも重要な研究項目である。この論文で提案した新しいアルゴリズムでは、音響モデルのトポロジー決定のための基準を画期的に精密化し、かつ、実行可能な計算量で実現している。従来の音韻モデルトポロジー決定の代表的なアルゴリズムである ML-SSS と PDT は、音韻の分割や HMM の状態の分割の基準として一個の正規分布を仮定している。しかしながら、音声の特徴量の分布は複雑で、一個の正規分布では表現できないことが知られている。この論文では、音韻分割、HMM 状態分割の基準として、タイドミクスチャー-HMM を用いた音韻モデルに着目して、分割基準として混合正規分布を扱えるようにしたことが大きな発展である。

この音韻モデルトポロジー決定アルゴリズム(TM-SSS)の有効性を確かめるために、一人の話者が発声した音声データベースによる音韻モデルのトポロジーおよびモデルの学習を行ない、5千単語音声の認識実験で、従来の ML-SSS の単語誤り率を約半減できることを確認している。この実験により、TM-SSS アルゴリズムの有効性が確認できている。

さらに、現在要請されている不特定話者音韻モデルの作成に TM-SSS を適用することを試みている。大量の音声データからの音韻トポロジーの自動生成を可能にするために、いくつかの仮定を用いて大幅に計算量を減らし、不特定話者音韻モデルのトポロジー学習が可能なアルゴリズム(FTM-SSS)を考案している。300人の音声データベースを用いて、FTM-SSS による音韻モデルのトポロジー学習を行ない、大語彙連続音声認識において、従来からのアルゴリズム ML-SSS と PDT より認識性能で優れていること示している。

本博士論文は、HMM の音韻トポロジーの自動生成アルゴリズムの導出とその有効性を大規模な信頼のおける実験で確かめており、大語彙連続音声認識の研究に大いに貢献している。とくに、本博士論文で提案された TM-SSS および FTM-SSS は、従来の音韻モデルのトポロジーの自動生成手法を大幅に改善しており、今後の理論およびさらなる実際の応用への発展が期待される。

これらの研究成果は、情報科学、音情報処理学の分野における学術的な貢献として高く評価できる。平成13年1月16日開催の公聴会の結果も参考にして、本博士論文の審査を行なった結果、本論文は博士(工学)の学位論文として十分に価値のあるものと判断した。