

Title	雑音環境下における音源分離を認識規範とした音声認識に関する研究
Author(s)	羽二生, 篤
Citation	
Issue Date	2004-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1768
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

雑音環境下における音源分離を認識規範とした 音声認識に関する研究

羽二生 篤 (210070)

北陸先端科学技術大学院大学 情報科学研究科

2004年2月13日

キーワード: 分離過程の妥当性、分離結果の妥当性、音声認識、音源分離、知識、聴覚情景解析、二波形分離モデル.

雑音環境下では、目的音と背景音が、周波数的にも時間的にも予測不可能な重なりをもって存在する。雑音環境下での音声認識は、このような混合音から目的音の内容を認識する必要がある。過去の研究としては、雑音除去を前処理として認識モデルに付加したり、音響モデルを変形する手法がある。しかし、現時点では雑音環境下で精度よく認識を行う音声認識システムはない。

妥当な振る舞いをする音源分離手法を使用したときに、知識を用いて分離を行う過程が妥当であり、かつ、分離の結果が目的音として尤もらしいといえるのであれば、入力音中に目的音が存在したといえる。なぜならば、その分離音の出所は入力音以外あり得ないからである。このことから、妥当な分離手法の分離過程と分離結果の妥当性を評価できるのであれば、入力音中の目的音を認識できると考えられる。そこで、本論文では、分離過程や分離音の妥当性を評価することにより認識を行う、音源分離を認識規範とした音声認識手法を提案し、本手法の有効性について検討する。

本研究のモデルの大きな枠組みとなるのは、窪らにより提案された音源分離モデルである。このモデルは、鷓木らが提案した聴覚情景解析にに基づく音源分離モデルに対して、目的音の知識を利用して分離を行うようにしたものである。

本研究のモデルは、信号解析部、基本周波数 (F0) 推定部、知識制御部、波形分離部、認識部の5つの処理からなる。入力信号は、はじめに信号解析部において時間領域の波形が、時間と周波数領域の表現に変換される。F0推定部において、入力音のF0を推定し調波関係にある周波数を算出する。波形分離部では振幅と位相、F0

推定部からの調波の周波数、さらに、知識制御部に必要とされる知識を要求して、Bregman の 4 つの発見的規則に基づく立上り立下りの同期、調波関係、漸近的变化、振幅包絡間の相関に加え、知識と周波数領域での相関を用いて目的音を分離する。分離の際は、入力音の時刻と分離に用いる知識の時刻をどれだけ進めるのが妥当なのかを検討しながら分離を行う。知識制御部では、目的音を表す音素表記を受け取り、あらかじめ格納されている目的音の周波数領域でのスペクトル形状の集合を知識群として選択する。知識制御部は波形分離部から要求された知識を波形分離部に送り出す。認識部では、F0 推定部で推定された F0 の値が話声の F0 として妥当かを判断する。次いで、波形分離の過程を監視し、分離の過程が妥当であるかどうかを判断する。さらに、分離された波形の周波数領域での形状と知識との間で相関をとり、その平均値で分離結果の妥当性を判断する。最後に、認識部は認識の結果を記号 (列) にて出力する。

背景音として定常な白色雑音が存在する状況で 1 名の話者が日本語単母音を発話している状況を想定し、提案手法の核となる波形分離部と認識部のみの有効性を確認するシミュレーションを行った。SNR は、-20、-10、0、10、20 dB をした。比較には、DTW (Dynamic Time Wrapping) による音声認識システムを用いた。シミュレーションの結果、提案手法は -10 dB から 20 dB まで正しく認識を行うことができた。また、分離結果の確からしさを表す指標から、DTW による手法より常に確からしい認識結果を得られることがわかった。

実装したシステムの性能評価を行った。条件は先ほどのシミュレーションと同じである。その結果、実装したシステムは 20 dB から 0 dB までは正しく認識を行うことができた。-10 dB においては、F0 推定値が妥当でないために認識に失敗した。このシミュレーションから本手法を有効に利用するためには、現在用いている自己相関法による F0 推定にかわって、雑音に対してよりロバストな手法を取り入れる必要があることがわかった。雑音に対してロバストな F0 推定法に関しては数多く行われている、この手法を取り入れることにより、本研究のコンセプトを有効に活用できるモデルを構築できると考えられる。

以上のことから、白色雑音と単母音が空間的に加算された状況において本手法が有効であることが確認できた。

今後の展望として、雑音に対してロバストな F0 推定法の組み込み、知識の検討、妥当性の検討を行う。