

GMM 変換とスペクトルピークの直接変換による声質変換法

Voice conversion method by direct conversion of the spectral peak and GMM conversion

遠藤一輝
Kazuki Endo

岡山大学 大学院自然科学研究科
Graduate school of Natural Science and Technology, Okayama University

1. はじめに

近年、音声合成技術の発展により音声合成は身の回りの様々な場面で用いられるようになった。合成音声の話者性に関する研究の1つとして、ある話者が発声した音声を他の話者が発声した音声のように変換する声質変換技術についての研究がある[1][2]。この分野の研究で近年の主流となっているのが分析合成方式による声質変換技術であり、中でも統計処理に基づく手法が盛んに研究されている。しかし、統計処理に基づく声質変換技術では変換スペクトルを求める際にスペクトルが過剰に平滑化されてしまい、音質が劣化するという問題点がある。

そこで本研究では、統計処理に基づく声質変換技術の1つである混合正規分布(Gaussian Mixture Model; GMM)に基づく変換[3]を用いてスペクトルを変換し、それをターゲットとして音声分析処理によって得られた元話者のスペクトルのピークを直接変形させることにより変換スペクトルを求める方式を提案する。提案方式では元話者のスペクトルを直接変形するので、スペクトルの過剰な平滑化による音質の劣化を防ぐことが期待できる。本稿では、ピークの直接変形を行うためのピークの抽出法などについて記述する。

2. GMM 変換とスペクトルピークの直接変換による声質変換法

声質変換の処理の概要を図1に示す。音声分析処理にはSTRAIGHT[4]を用いる。STRAIGHT分析により音声信号からスペクトルと基本周波数と非周期性指標という3つの特徴量を抽出する。この内、スペクトルについてはスペクトルから求めたケプストラムを用いてGMM変換を行い、それをターゲットとして音声分析処理によって得られたスペクトルのピークを直接変形させることで変換スペクトルを求める。基本周波数の変形については、戸田ら[5]の手法を用いる。

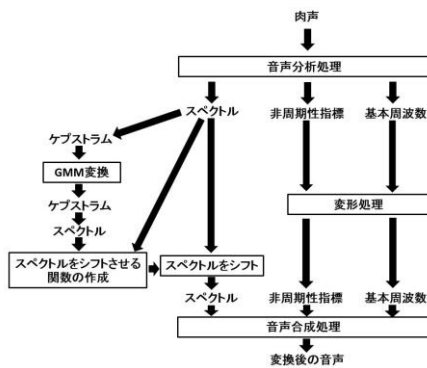


図1 声質変換の概要図

3. スペクトルのピークの直接変形について

本稿ではスペクトルのピークのパワーの大きい部分が特徴量として使用できると判断し、ピークのパワーの大きいものから4つを抽出し、変形に使用する。スペクトルピークのグループ作成の概要を図2に示す。

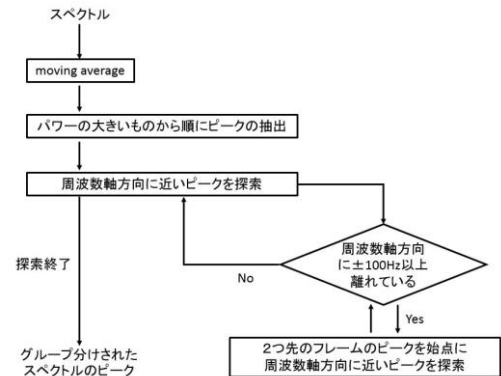


図2 スペクトルピークのグループ作成概要図

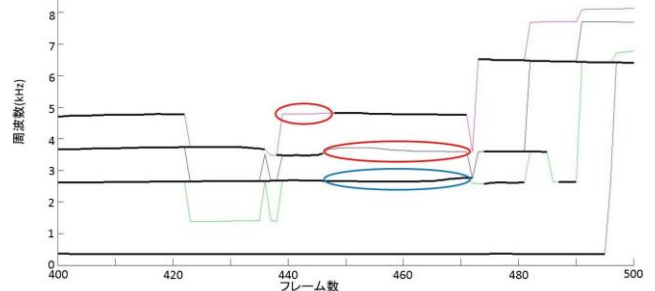


図3 あるフレーム間におけるパワーの大きいものから4つ抽出したものと再度グループ分けを行ったピーク

図3にあるフレーム間における、パワーの大きいものから4つ抽出したピーク(色線)と再度グループ分けを行った結果を重ねたもの(黒線)を示す。図3から、ある程度は正しくグループ分けが行えているが、一部グループとして抽出されていない箇所(赤丸部分)や、グループ分けしたものが重なってしまった箇所(青丸部分)が存在している。

4. まとめ

本稿では、GMM変換とスペクトルピークの直接変換による声質変換法を実装するためのピークの抽出法とそのグループ分けについて検討した。今後はピークのグループ分けの改善と、ピークを直接変形させるための関数の作成を行い提案手法の実装を目指す。

5. 参考文献

- [1] H.Kuwabara and Y.Sagisaka, "Acoustic characteristics of speaker individuality: Control and conversion," *Speech Communication*, vol.16, no.2, pp.165-173, 1995.
- [2] 戸田智基, "確率モデルに基づく声質変換技術," *日本音響学会誌*, vol.67, no.1, pp. 34-39, 2011.
- [3] Y.Stylianou, O.Cappe, and E.Moulines, "Statistical methods for voice quality transformation," in *Proc. of EUROSPEECH*, pp. 447-450, Madrid, Spain, Sept. 1995.
- [4] Hideki Kawahara, "STRAIGHT-TEMPO: A Universal Tool to Manipulate Linguistic and Para-Linguistic Speech Information," in *Proc. of SMC-97*, Orlando, Florida, USA, Oct. 1997.
- [5] 戸田智基, 陸金林, 猿渡洋, 鹿野清宏, "周波数軸伸縮を用いた混合正規分布モデルに基づく声質変換法" *電子情報通信学会論文誌 D-II*, vol.J84-D-II, no.10, pp.2181-2189, Oct. 2001.