

ケプストラム領域ピッチ操作とスペクトル位相復元を組み合わせた音声加工*

松原 聖人† 林 耕平† 光本 大記† 濱田 康弘† 小野 順貴‡ 嵯峨山茂樹†

(†明治大学総合数理学部, ‡国立情報学研究所)

1. はじめに

本稿では、ケプストラム分析と位相復元を用いた音声分析合成の可能性を検討する。

ケプストラム分析[1]は、音声スペクトルの包絡と微細構造の積がケプストラム領域では和として扱え、両者がケフレンシ軸上で（ほぼ）分離できることが利点で、ピッチ加工などの音声操作を施して歌声合成などに有用である可能性があり、パラメトリックなモデルに基づかぬため、原音の情報を保持できる可能性がある。

ケプストラムを介した分析合成系（ボコーダ）は、Oppenheimらの複素ケプストラムに基づく準同型処理[2]を別にすると、パワースペクトルを経由するために位相が失われ、（加工した）ケプストラムから波形領域に変換することができなかった。そのため、今井らの対数振幅特性近似フィルタ[3]を用いる手法が取られてきたが、線形予測分析合成(LPC)などと同様に、音源モデルと少数のスペクトル包絡パラメータを用いて巡回型フィルタにより信号を生成する source-filter モデルにより波形を生成する方式であるため、合成音声からは人工的な印象が感じられ、高いQ値の共振が生じる場合などでは、振幅異常や時間遅れなどを生じやすい。そこで、スペクトル位相復元技術[4]を組み合わせることで、ケプストラムが豊富な情報を活かせる可能性があるフィルタによらない音声合成法を検討する。

2. 非フィルタケプストラム分析合成系

2.1 ケプストラムドメインでの音声加工

音声のケプストラムは、ケフレンシ軸の低次部分は主に声道特性に対応するスペクトル包絡を反映し、高次部分は声帯振動に対応するスペクトル微細構造を反映し、特に基本周波数(F_0)に対応してピークを生じるものと考えられて来た。この F_0 成分の位置に関して操作を行えば、音声に対してピッチ変更が可能と考えられる。つまり、入力音声をケプストラム分析し、それによって取得した情報を F_0 成分に関して操作したあと、パワースペクトルから音声波形に戻すこと

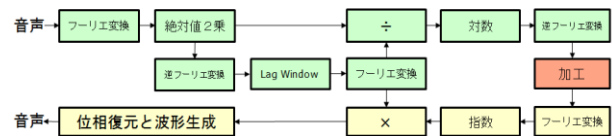


Fig. 1. ケプストラム非フィルタ分析合成系

ができれば、ピッチ変更などの変化を入力音声に与えることができる。

しかし、ケプストラムは対数パワースペクトル領域からの逆フーリエ変換なので、位相情報を持たないため、ケプストラムを操作できたとしても、音声信号に復元することができない。本稿ではそこにスペクトル位相復元技術[4]を用いて、スペクトラムに対して無矛盾な位相を付加することにより、ケプストラム領域からの波形生成を可能としている。ケプストラム領域で音声を加工することと、それを位相復元で音声波形領域に戻して可聴とすることによって、音声を加工する。

2.2 Lag 窓によるスペクトル平滑化の利用

ケプストラムは、リフタによって、低ケフレンシ部分の声道特性と高ケフレンシ部分の音源特性に分離することがよく行われるが、低ケフレンシ部分はパワースペクトルの包絡を表現しているわけではなく、謂わば微細構造の山谷の中心を辿る成分を表している。そこで、包絡成分の分離の精度を向上させるために、パワースペクトルの自己相関関数に Lag 窓[5,6]を掛けることにより微細構造の谷を埋めて、パワースペクトルの包絡成分を直接求めることができる（Wiener-Khinchin の定理と畳み込み定理の応用）。これによりパワースペクトルを割って微細構造を得ることで、パワースペクトルから直接ケプストラムを求める方法に比べより正確な分離が可能になる。

2.3 ケプストラム非フィルタ分析合成処理の流れ

以上により、 F_0 移動の分析合成系の流れは以下の通りである。(Fig.1)

- (1) 時間信号のフーリエ変換の絶対値の2乗によりパワースペクトルを得る。
- (2) 逆フーリエ変換をし、短時間自己相関関数を得る。
- (3) 上記に Lag 窓をかける（今回はガウス窓）。

*Speech processing combining pitch manipulation in cepstrum domain and phase reconstruction from power spectrum

†Masato Matsubara, Kouhei Hayashi, Daiki Mitsumoto, Yasuhiro Hamada, Shigeki Sagayama (Meiji University)

‡Nobutaka Ono (National Institute of Informatics)

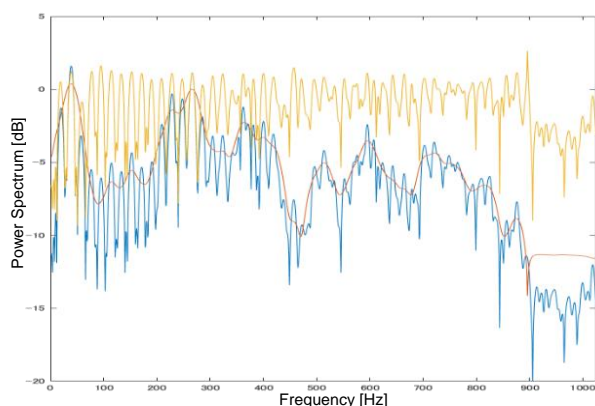


Fig. 2. 音声のスペクトルと音源スペクトル

- (4) これをフーリエ変換することで、パワースペクトルの包絡成分を得る. パワースペクトルをこの成分で除して微細構造を得る(Fig.2. ここでは「 F_0 成分スペクトル」と呼ぶ).
- (5) 上記の対数を取り逆フーリエ変換して, F_0 成分ケプストラムを得る(Fig.3).
- (6) このケプストラムをケフレンシ軸で線形補間により伸縮操作する(Fig.4).
- (7) 上記をフーリエ変換し指数関数変換し, (4)の包絡成分と掛け合わせることでパワースペクトル領域へ戻す.
- (8) 位相復元により, 時間信号を得る.

3. F_0 加工実験と定性的評価

3.1 F_0 加工の方法

まず, 予備実験として, ケフレンシ軸において低周波領域と F_0 成分のピークを示す部分以外を0にしたところ, 著しく品質が損なわれ, ケフレンシ軸のピーク以外の細かな上下も音声を加工する際に必要な情報であるということがわかった. そこで, ケプストラム分析によって得られた情報をなるべく保持するため, F_0 成分を伸縮する際には, ケフレンシ軸において, 微細構造に対応する部分を線形圧縮・伸張した.

3.2 検証結果と考察

ケフレンシ軸で F_0 成分に対応する部分の伸縮操作を行うことで, ピッチ操作が行えた. 但し, 原音と比べると全体的に濁りが感じられたが, source-filter 方式による機械的な印象は少なく, より肉声に近づけることができる原理的可能性を持っていると考えられる. 合成したケプストラムからの音声合成[7]では濁りが感じられないことから, フレーム間の揺らぎを押しさえることで濁りを軽減できる可能性がある.

4. 結論

本稿では, ケプストラム領域でノンパラメト

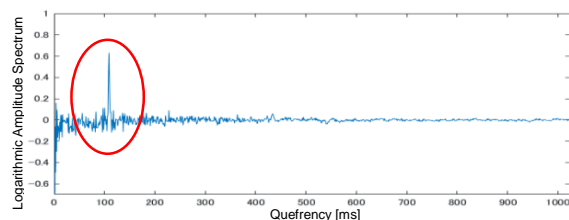


Fig. 3. F_0 成分ケプストラム

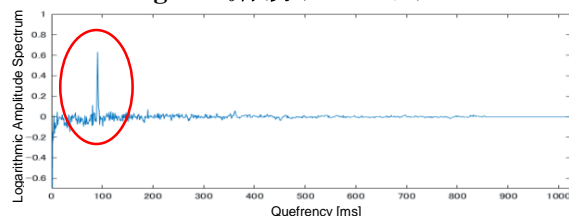


Fig. 4. 伸縮した F_0 成分ケプストラム

リックな非フィルタ方式の音声加工法を実現するために, ケプストラム領域での操作とパワースペクトルの位相復元を組み合わせる. また, その加工をするにあたっては, 自己相関関数に lag 窓をかけることによって, 包絡成分と微細構造の分離の向上を目指した. 線形補間を用いた F_0 成分の操作によって, 原音の豊かな情報を保持して音声ピッチを変更することを可能にしたが, 結果の音声にはフレーム間の揺らぎと見られる濁りが聴かれた. 今後の展望として, ピッチ同期のケプストラム分析を行うことで, 不必要なフレーム間の揺らぎを軽減できる可能性を追求したい.

文献

- [1] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The quefrequency alanysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking," Proceedings of the Symposium on Time Series Analysis (M. Rosenblatt, Ed) Chapter 15, 209-243. New York: Wiley, 1963.
- [2] B. ゴールド, C. M. レイダー (石田訳), 電子計算機による信号処理, 共立出版, 1972.
- [3] 今井, 北村, "対数振幅特性近似フィルタを用いた音声の分析合成系," 電子通信学会論文誌, J61-A(6): 527-534, 1978.
- [4] J. Le Roux, N. Ono, and S. Sagayama, "Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction," in Proc. SAPA, Sep. 2008.
- [5] Y. Tohkura, F. Itakura, S. Hashimoto, "Spectral Smoothing Technique in PARCOR Speech Analysis-Synthesis," IEEE Trans. ASSP, Vol. 26 , No.6, pp.587-596, 1978.
- [6] 嵯峨山, 古井, "ラグ窓を用いたピッチ抽出の一方方法," 電子情報通信学会全国大会予稿集, 1235, Vol. 5, p. 263, 1978.
- [7] 濱田, 小野, 嵯峨山, "無矛盾位相復元を用いたケプストラム特徴量からの音声合成," 本論文集所収.