

声質を保存した歌声のピッチ補償

7 J-8

笹平 宜誠 橋本周司

早稲田大学理工学部応用物理学科

1 はじめに

より手軽にプロの歌手のように歌を歌いたいという我々の欲求に答えるかたちでカラオケは誕生した。初期の頃は伴奏と歌詞を流すだけであったが最近では音程の一時的なずれを修正するものも登場してきている。これらは我々の、より原曲に近いイメージ、音域で歌いたいという欲求に対して提案された解決方法の一つといえよう。音程がはずれた部分だけを正しい音程に修正する機能は計算機の高機能化、音声のデジタル化により初めて実現したといえるが、現在商用化されているシステムで大幅な音程のずれを修正するとひどく不自然な音声になってしまう。これはピッチ変更の際、個人の音声の特質まで変化させるためであり、個人の声質情報を保存したままでピッチのみを変更する必要がある。最近、ピッチ情報と個人特性情報を分離した上でピッチ変更、再合成する方式が声質を保存するピッチ変更手法として提案されている。例えば複素ケプストラム分析合成による方式[1]、時間周波数領域での補間を用いた方式[2]等があげられるが、我々は実時間処理を最優先する意味で、音声の分析合成、符号化等で広く用いられている線形予測分析方式(LPC)を人の歌声に適用してピッチ補償を試みてきた[3][4][5][6]。ここではその最近の成果について報告する。

2 ピッチ補正アルゴリズム

2.1 線形予測分析を用いた声質情報の保存

線形予測分析(LPC)とは、音声信号の時刻 n におけるサンプル値 S_n とその近傍のサンプル値との間に強い相関関係があることを利用して S_n を近傍のサンプル値の線形一次結合の形で予測する手法である。LPCでは線形予測係数により音声信号の個人情報を有するスペクトル包絡の情報が与えられ、線形予測誤差により音声のピッチ情報が与えられるため、本研究では線形予測誤差(残差波形)を合成“Pitch Compensation of Singing Voice to Keep the Personal Phonetic Feature”, Yoshinari SASAHIRA Shuji HASHIMOTO Dept. of Applied Physics, School of Science and Engineering, WASEDA Univ.

側で変更することで個人の声質情報を保存した歌声のピッチ補正を行っている。

2.2 原音声の基本周波数の抽出

不特定多数の話を対象とするピッチ検出システムの場合、ピッチの正確な検出は容易ではなく、多くの計算量、計算手順を必要とする。本システムでは、サンプリング周波数が8.3kHzと低めであることから、低域フィルタリングとよりPeak-to-Peakの検出を基底とした検出手法で基本周波数を抽出した。

2.3 線形予測係数の更新

人間の音声は定常的であれば、一度線形予測係数を求めれば更新の必要はないが実際は非定常的であり、ほぼ定常とみなせるのは20ms程度の短い時間の範囲である。そこで本研究では約20ms毎に線形予測係数の更新を行っている。1データだけサンプリングする間に線形予測係数を求めるための計算のすべてを処理することはできないため、計算過程を約150過程に分割して線形予測誤差から音声を合成する処理と同期して1過程ずつ行った。これにより実時間でピッチ変更した音声を出力すると同時に全ての係数予測のための計算処理を実行する。この流れを図1に示す。

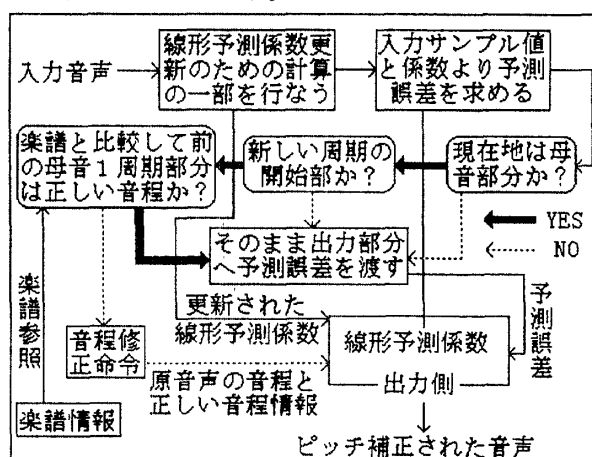


図1 線形予測係数更新とピッチ補正の流れ

2.4 ピッチ変更と時系列上の整合性

歌い手の音声は分析側に入り、予測誤差の算出、ピッチ判定、楽譜の参照等が行われる。原音声の音程と楽譜上の音程が異なる場合、合成側で一周期単位

で予測誤差の伸縮が行われるため時間上の誤差が生じる。ここではこの誤差時間を一定時間内におさえるため適当に波形のカット、複写を行っている。図2に遅延が生じる場合の調整の様子を示す。

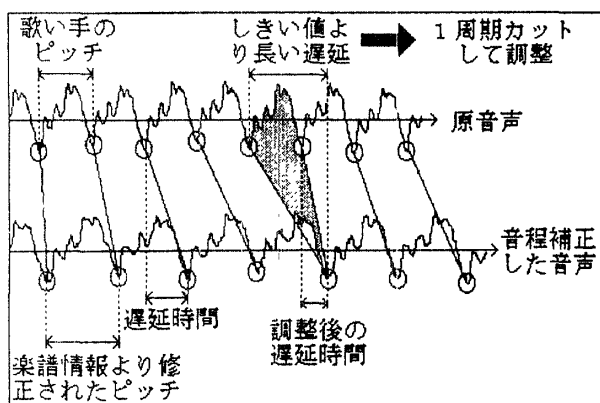


図2 ピッチ変更と誤差時間の調整

3 システム構成

マイクロサイエンス社のAD、DAボードをNEC製PC-9821 Ap2(Intel 486DX2 66MHz)に装着して実験を行った。サンプリング周波数は8.33kHz、量子化ビット数は12bitである。また、線形予測係数の次数は計算量削減のため7次、自己相関の積分範囲は150サンプルで行なった。

4 実験結果と今後の課題

本研究では歌い手は本来の曲と比べてテンポは正しいが音程をまちがえて歌っているという状況を想定している。そこで今回我々は歌い手にあらかじめ与えられた伴奏にテンポをあわせて適当にずれた音程で(ただし楽譜とのずれは上下1オクターブ以内)自由に歌詞を歌ってもらい楽譜情報にしたがってピッチの修正を行うという実験をおこなった。各音声のスペクトルの様子を図3～6に示す。

原音声(図3)を、従来の手法を用いた市販機器を用いてピッチ補正した歌声(図4)はスペクトルの形が原音声と比べて大きく変化しているのがわかる。これに対して提案手法を用いてピッチ同期で音程のずれの補償を行った声(図5)は、原音声と同一人物による1オクターブ高い音声(図6)と比較しても従来手法のものよりかなり自然な声であることがわかる。しかし(1)聴感上雑音が存在する、(2)音程の一時的なずれ同様に頻繁に起こりうるテンポのずれに対して考慮をしていない、等の問題点が依然として残っている。今後、実時間処理を第一に考え、マシンの処理能力に応じたアルゴリズムを随

時試して改良を重ね最終的にはカラオケの高音質なピッチ補償のみならず、歌声で駆動する新しい形式の楽器への拡張も考慮していきたい。

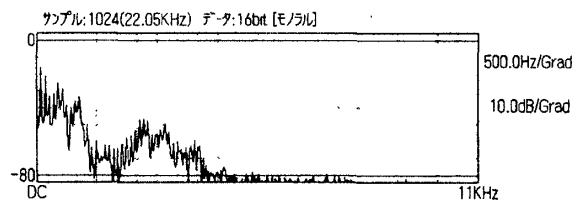


図3 原音声の(母音あ 男性 110Hz)スペクトル

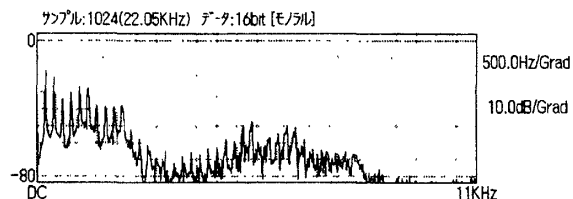


図4 原音声を従来の手法で1オクターブ上昇させた音声のスペクトル

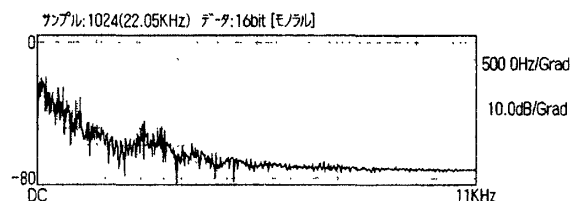


図5 原声を提案手法により1オクターブ上昇させた音声のスペクトル

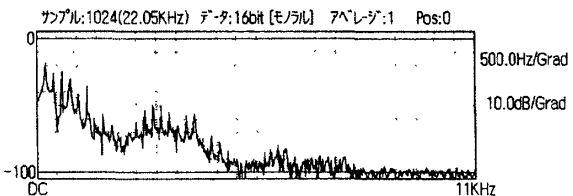


図6 原音声と同一人物の1オクターブ高い音声(母音あ 男性 220Hz)のスペクトル

5 参考文献

- [1] 清山信正、都木徹、梅田哲夫、宮坂栄一 一信学論(A), J73-A, No. 3 pp. 387-396(1990)
- [2] 河原英紀、増田郁代 一信学技報 EA96-28, (1996)
- [3] Yoshinari SASAHARA, Shuji HASHIMOTO - Proceeding of the 1995 ICMC, pp. 404-pp. 407
- [4] 笹平宜誠、橋本周司 一信学技報 HIP95-21, (1995)
- [5] 笹平宜誠、橋本周司、一第52回情報処理学会全国大会分冊1 pp. 463-pp. 464(1996)
- [6] SASHIRA Yoshinari, HASHIMOTO Shuji - Proceeding of the 1996 ICMC, pp. 124-pp. 125(1996)