

LPCによる歌声のピッチ補償

5Z-9

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

近年、カラオケ等の普及により、我々が歌を歌う機会が非常に増えてきている。ほとんどの場合、我々はあらかじめ録音された伴奏にあわせて歌うわけであるが、その際、(1)歌い手のもつ音域が曲の本来の音域と適切にかみあわない、(2)歌う間にテンポがずれてしまう、(3)本来の曲の音程からずれた状態で歌ってしまう、(4)歌の音域が歌い手のそれより広い、などの問題がしばしば発生する。(1)の問題はカラオケのキーチェンジャーのように伴奏のキーを変更することで容易に解決される。また、(2)の問題点については歌い手のテンポに伴奏が合わせる適応型カラオケシステムなどの研究が現在なされている[1][2]。一方、(3)と(4)の問題については、現在のところ有効な解決法が存在しない。市販製品でピッチ補正を行うと、歌声が変質してしまい、ひどく不自然な音声になってしまう。そのため個人の声質情報を保存したままピッチのみを修正する必要がある。最近、これまでのサンプリングレートの変調によるピッチ補正に対しピッチ情報と個人特性情報を分離した上でピッチ変更、再合成する方法が声質を保存するピッチ変更法として提案されている。たとえば、複素ケプストラム分析合成による方式[3]、スペクトル領域で歪みを修正する方式[4]などがあるが、我々は音声の合成、符号化等で広く用いられている線形予測符号化方式(LPC)を人の歌声に適用してピッチ補償を試みた[5]。ここではその最近の成果について報告する。

2 ピッチ補償アルゴリズム

2.1 母音区間の抽出

明確なピッチの存在する区間とピッチが不明瞭な区間の判定がまず必要となる。ピッチが存在するのは主として母音部分である。本研究では音声信号にローパスフィルタをかけPeak-to-Peakの検出でピッチ周期を確認できた部分を処理の対象としている。

Pitch Compensation of Singing Voice Using LPC

SASAHIRA Yoshinari HASHIMOTO Shuji

Dept. of Applied Physics, School of Science and Engineering, Waseda Univ.

2.2 線形予測分析

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は強い相関関係がある。したがって S_n は、過去のサンプル値から予測することができる。このように S_n をその近傍のサンプル値との線形一次結合の形で予測する方法を線形予測法という。過去の p 個のサンプル値から S_n を予測する式は S'_n をその予測値として、

$$S'_n = - \sum_{i=1}^p \alpha_i S_{n-i}$$

で与えられる。また、予測値とサンプル値との誤差 e_n は、

$$e_n = S_n - S'_n = S_n + \sum_{i=1}^p \alpha_i S_{n-i}$$

で与えられ、よく知られているように、線形予測係数 α_i の値は、過去のデータの自己相関を用いて e_n の自乗平均が最小になるように決定される。

2.3 ピッチ変更の手法

現在、ピッチ変更を目的とした市販製品のほとんどはサンプリングレートの変調を基本としたハーモナイザーであって、各音素のピッチ補償を目的とするものではない。更に、これらの製品を用いて変調された音声は、ひどく不自然なものとなる。これは、これらの製品が歌声の基本周波数のみならず、全体のスペクトル包絡まで変えてしまうためである。人間の歌声において、スペクトル包絡は歌い手の声の特性情報を有していることはよく知られている。そのため、ピッチ変更の際は特性情報を保存するために、スペクトル包絡を変化させてはならない。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線の情報が与えられ、線形予測誤差により音声のピッチ情報が与えられる点に注目して、切り出された母音部分の波形を用いて得られた残差波形の打ち切りやゼロつめを行うことでピッチ周期の変更を行った。

2.4 線形予測係数の更新

人間の音声は一般的には非定常的であるが、母音に限っていえば、20ms程度の短い時間であれば、ほぼ定常とみなせる。そこで本研究では、処理時

間短縮のために各サンプル値ごとでなく、母音の数周期毎に(約20ms)に線形予測誤差の修正とあわせて線形予測係数の更新を行った。

2.5 音程情報の獲得と音程のずれの補償

歌い手側と伴奏とのずれにおいて、市販のカラオケのピッチチェンジャーは歌い手側の音程、テンポはあっているが、キー全体が伴奏のキーからずれているという状況を想定している。しかし、実際は歌声の音程が一時的に伴奏の音程からはずれる場合が圧倒的に多い。そこでここでは、歌声からピッチ検出処理の際得られた基本周波数と時間的に対応する楽譜情報とを比較し、音程にずれがある場合は、音程A1=440Hzとする十二平均率にしたがい、ピッチ補正を行うことにした。図1にピッチ修正の処理の概略を示す。

3 システム構成

実験では、Microsoft Windows 3.1に付属のサウンドレコーダーを用いて、サンプリング周波数22050Hz、量子化ビット数16bitでマイクより録音した歌声を用いた。また、線形予測係数の次数は20、自己相関の積分範囲は350サンプルである。母音及び、ピッチ検出、線形予測、予測誤差の修正、線形予測係数の更新は全て分析側で行われ、合成側へは、線形予測係数とピッチ補正された予測誤差のみが送られ、それらをもとに正しい楽譜に従ってピッチ変更された歌声が合成される。システムの一連の流れを図2に示す。

4 実験結果とまとめ

本研究では、歌い手側は本来の曲と比べてテンポは正しいが音程が異なっている、という状況を想定した。そこで、今回我々はあらかじめ与えられた伴奏にテンポをあわせて朗読調で歌詞を歌ってもらい、その後、ピッチ補償を行うという実験を

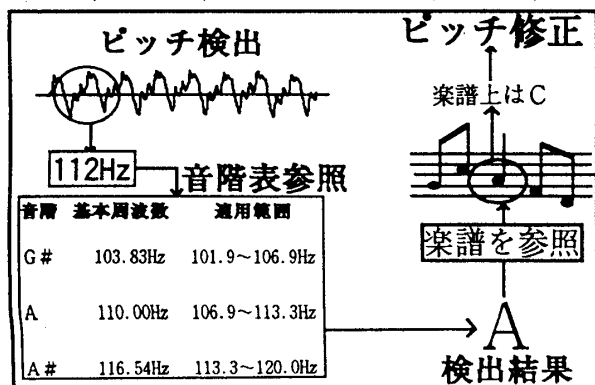


図1 音程情報に基づくピッチ修正方法

行った。

提案手法を用いてピッチ同期で音程のずれの補正を行った歌声は、市販されているピッチチェンジャーを用いてピッチ補正した歌声と比較してかなり自然な声であるといえる。しかし(1)聴感上雑音が存在する、(2)音程のずれと同様に起きる可能性の高いテンポのずれについて考慮をしていない、などいくつかの問題が残った。今後これらの問題点について改良を重ねて、最終的には歌声で駆動する新しいタイプの楽器へと発展させていきたいと考えている。

参考文献

- [1] Haruhiro KATAYOSE, Tsutomu KANAMORI, et.al "Virtual Performer", Proceeding of the 1993 International Computer Music Conference, pp.138-pp.145
- [2] Wataru INOUE, Shuji HASHIMOTO, Sadamu OHTERU, "A Computer Music System for Human Singing", Proceeding of the 1993 International Computer Music Conference, pp.150-pp.153
- [3] 清山信正 都木徹 梅田哲夫 宮坂栄一, "複素ケプストラム分析合成によるピッチ変換法" 信学論(A), J73-A, No. 4pp.694-pp.702, (1992)
- [4] 都木徹 梅田哲夫, "ピッチ変更時のひずみをスペクトル領域で修正する声質変換方式とその品質の心理評価", 信学論(A), J73-A, No. 3pp.387-pp.396(1990)
- [5] Yoshinari SASASHIRA, Shuji HASHIMOTO, "Voice Pitch Changing by Linear Predictive Coding Method to Keep the Singer's Personal Timbre", Proceeding of the 1995 International Computer Music Conference, pp.404-407

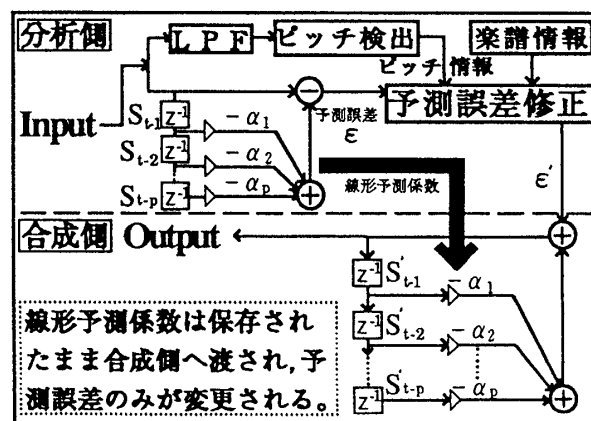


図2 ピッチ補償の流れ