

フーリエ合成を利用する話速制御の一方法

1N-01

関本 智美
東京電機大学大学院
工学研究科
電子工学専攻

斎藤 博人
東京電機大学
情報環境学部
情報環境デザイン学科

中村 尚五
東京電機大学
情報環境学部
情報環境デザイン学科

1 はじめに

言語学習においてネイティブスピーカの発声は、他言語を母国語とする人にとって聞き取れない部分があることがある。このような場合、話速を遅くすると聞き取りやすくなるため、テープレコーダ等の再生速度を遅くすることなどがよく使われていた。しかしながら、この方法では原音声のピッチ周波数が変化するため、原音声にくらべ明瞭度が落ち、声質が不自然なものになる。そのため、明瞭度、声質を変化させずに再生速度を遅くするには、ピッチ周期を変化させずに発声速度を遅くする必要がある。話速変換の方法として、有声音部分のピッチ構造を変化させずに発声時間を変化させる手法が一般的である。本稿では、フーリエ合成における係数を時間的に変化させることにより、再生速度を可変する方法を提案する。

2 システムの構成

提案した話速制御システムの構成を図 1 に示す。

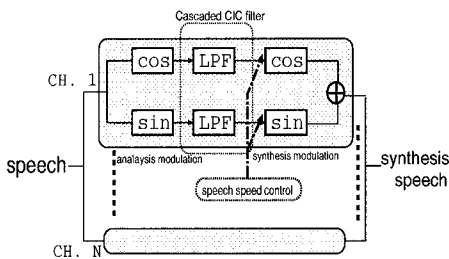


図 1: 提案した話速制御システム

2.1 Band-pass filter Bank

原音声信号 (サンプリング周波数 $f_s = 11\text{kHz}$) をバンドパスフィルタを用い 128 チャンネルに分割する。

各帯域のバンドパス信号は、各バンドの中心周波数を直流分まで変調し、狭帯域のローパスフィルタを通過させるより得る。

このローパスフィルタを実現するには正規化周波数で 0.0025Hz で遮断周波数を設定する必要がある。FIR フィルタで実現した場合、高次なものになる。ハードウェア化を考慮し、システムの簡便さから畳み込み演算を必要としない CIC (Cascaded Comb and Integrator) フィルタを用いて狭帯域フィルタを実現した。

2.2 CIC フィルタの構成

CIC フィルタは図 2 に示すように、遅延と加算のみで構成できるシステムが簡便である。バンドパスフィルタの仕様は、 $N = 128$ とし CIC フィルタ 4 段のカスケード接続構成で実現できる。このときのローパスフィルタを 4(a) に示す。また、周波数 $80 \sim 5080\text{Hz}$ での全バンド総合特性を図 4(b) に示す。

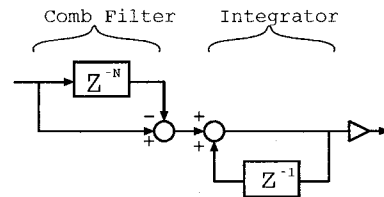


図 2: CIC フィルタの構成

CIC フィルタの伝達関数 $B(z)$ は式 (1) となる。

$$B(z) = \frac{1}{128} \cdot \frac{1 - z^{-128}}{1 - z^{-1}} \quad (1)$$

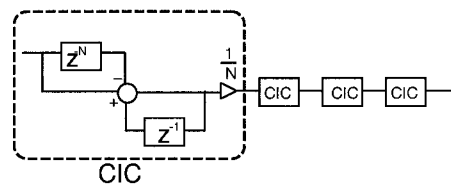


図 3: カスケード接続した CIC フィルタ

2.3 話速制御

合成音声は、BPF の中心周波数で変調された帯域通過信号の出力を足し合わせたものである。N チャンネルの BPF で合成された信号 $s(n)$ は、次のように示される：

$$s(n) = \sum_{k=1}^N x_k(n) \sin(\omega_k(n) + \phi_k)$$

$x_k(n)$ は k 番目の帯域通過信号、 $\omega_k(n)$ は k 番目の BPF の中心角周波数、 ϕ_k は k 番目の帯域通過信号に

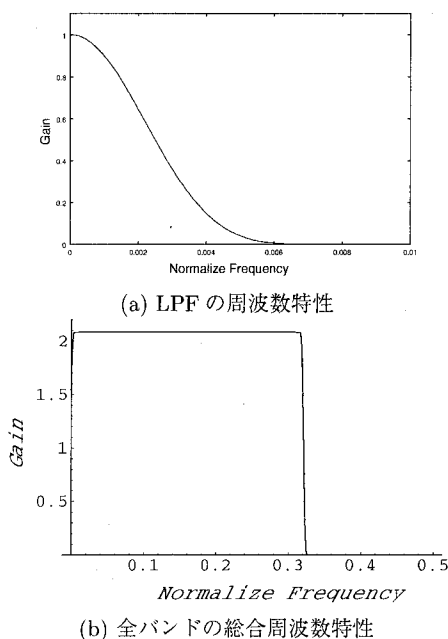


図 4: カスケード CIC フィルタの周波数応答

よる位相をそれぞれ示す。帯域通過信号 $x_k(n)$ はある帯域を示す包絡線で、その包絡線の長さを変えて話速制御を行なった。帯域通過信号 $x_k(n)$ の長さは、デシメーション (N) とインターポレーション (M) の有理数比 M/N によって決めた。図 5 に構成を示す。帯域通過信号 $x_k(n)$ が、実数 M によってアップサンプリングされ、アンチエイリアシングフィルターを通過した後、実数 N によってダウンサンプリングされたのち、帯域通過信号 $x_k(n)$ は有理数比 M/N で指定した長さに変わる。



図 5: データの長を可変にするための構成

3 シミュレーション

音声信号に対してフーリエ合成を用いたこの提案方法で話速を遅くしたとき、合成信号の音質は、混変調の影響で低下してしまう。オリジナルの BPF は、同じ周波数の間隔で BPF を構成している。そこで、混変調の影響を減少させるために BPF の中心周波数はランダムに摂動させた。その場合の、フーリエ合成の理論式を修正すると以下ようになる：

$$s(n) = \sum_{k=1}^N x_k(n) \sin((\omega_k + \Delta_k)n + \phi_k)$$

Δ_k は中心角周波数 $\omega_k(n)$ をランダムに摂動させた部分である。図 6 にはオリジナル音声、図 7 中心角周波数を摂動させない場合のオリジナル音声と話速 3/2 で合成した音声波形、図 8 中心角周波数を摂動させてオリジナル音声と話速 3/2 で合成した音声波形を示した。図 7、8 より中心角周波数を摂動させた場合混変調の影響が減少しているのが確認できる。

4 まとめ

正弦波モデルによる話速制御の方法を提案した。この提案方法はフーリエ合成に基づいているので、母音部分の抽出や母音と子音との滑らかなつながりを考える必要はなく、摂動フーリエ合成方法を提案したことにより混変調の影響を減少させることが確認できた。提案する手法では、CIC フィルタを用いる事により、変調以外では乗算を必要としない。そのために、今後はハードウェア化の実現も考えている。

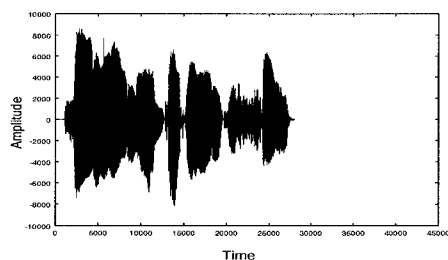


図 6: オリジナル音声波形

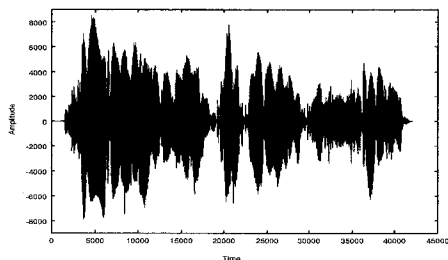


図 7: フーリエ合成法の波形

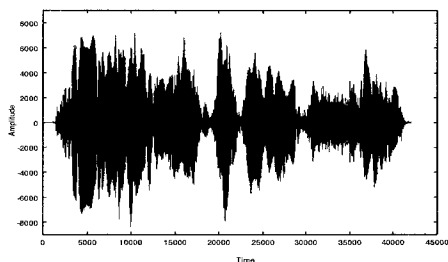


図 8: 摂動させたフーリエ合成法の波形