

1Q-7

時間軸の圧縮複製および 伸長削除による声質変換

加藤誠巳 藤原ひろみ 鵜飼敏之
(上智大学理工学部)

1. まえがき

人間の会話音声の声質変換はテレビ放送にあっては話者の特定を回避するために、電話にあっては悪戯電話を撃退する等の目的で実際に使用されている。本稿ではA/D変換された音声信号をサンプル単位で間引く(複製する)ことにより時間軸を圧縮(伸長)してピッチを上昇(下降)させ、しかる後に不連続点を生じさせないようにある一定ブロック長の信号を複製(削除)して、元の時間長に戻すことにより声質変換を行う手法について述べる。

2. 声質変換の手法

対象とする音声信号は一定速度(ここでは10kHz)でサンプリングされ、符号化されたものとする。以下の説明は予め符号化して記憶された信号を一括処理する場合に関して行うが、数10msから100ms程度の時間遅れを導入すればリアルタイムでの処理も可能である。

(1) ピッチ上昇による声質変換

ピッチ周波数を p 倍($p > 1$)にするには原信号に対しサンプル単位での間引きを行い、第 $[i/p]$ サンプル($i = 0, 1, \dots$)のみを取り出せばよい。ここで $[x]$ はガウス記号であり、 x を越さない最大の整数値を表すものとする。この操作により波形の時間長は $1/p$ 倍になるため、ある一定ブロック長の信号を複製して元の時間に戻す必要があるが、これは次のように行われる。即ち、図1(b)に示すように、ピッチ変換された信号をサンプル数 N なる一定長の切り出し窓を用いて切り出すのであるが、第 $k+1$ 切り出し窓の始点は第 k 切り出し窓の始点から第 N/p 番目サンプル点となるようにする。

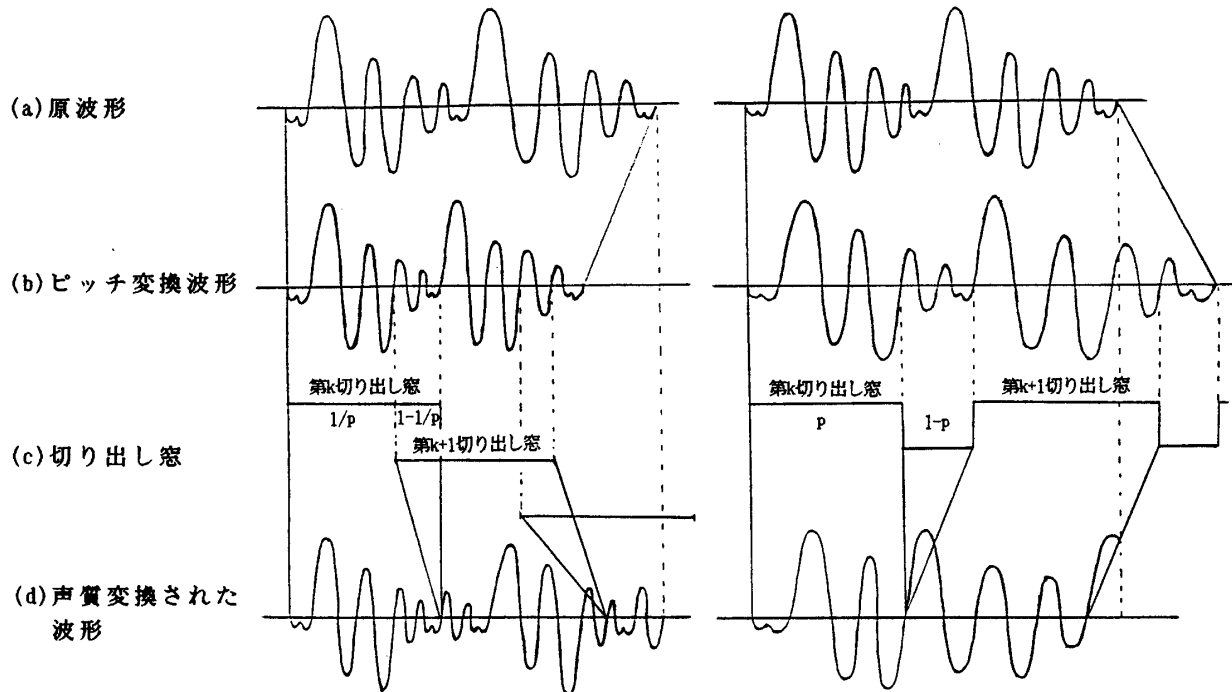


図1 ピッチ上昇変換($p > 1$)

図2 ピッチ下降変換($p < 1$)

この様な単純な切り出し窓でゲートすると両端で不連続を生じ、耳障りな雑音の原因となるので微係数が正である零交差点に一番近いサンプル点で切り出すものとする。この様な切り出しを行うと得られた信号の時間長は原信号の時間長と一致しなくなるので、切り出し窓の前縁および後縁において夫々前後方向に対して切り出し点を探索し、累積誤差ができるだけ少なくなる方の点を採用する。

(2) ピッチ下降による声質変換

ピッチ周波数を p 倍 ($p < 1$) にするためには原信号に対し、サンプル単位での複製を行い、第 i サンプル ($i = 0, 1, \dots$) として、原信号の第 $[ip]$ サンプルを用いればよい。この操作により波形の時間長は $1/p$ 倍になるため、ある一定ブロック長の信号を一括間引きして元の時間に戻す必要がある。即ち、図 2 (b) に示す波形をサンプル数 N なる一定長の切り出し窓を用いて切り出し、後半の $N(1-p)$ のサンプルから成るブロックを一括削除する。このとき波形の連続性を保つよう切り出し点を選ぶことは(1)の場合と同様である。

3. 切り出し窓長の決定

ピッチ上昇変換の場合にブロック複製されるサンプル数およびピッチ下降変換の場合にブロック削除されるサンプル数は、原信号換算で夫々 $Np(p-1)$ 、 $Np(1-p)$ 個であり、この値は時間長にして数 ms から数 10 ms が適当であることが実験的に確認された。これからサンプリング速度、ピッチ倍率 p が与えられれば適当な窓長 N が計算される。

4. スペクトログラムによる評価

図 3～5 に夫々原信号、ピッチ上昇変換された信号、ピッチ下降変換された信号のスペクトログラムの例を示す。

5. むすび

ピッチ上昇または下降された後に元の時間長に戻すことにより声質変換を行う手法について述べた。本方式は簡単であるため数 10 ms の時間遅れを許容すればリアルタイム処理が可能である。終わりに御討論戴いた本学マルチメディア・ラボの諸氏に謝意を表す。

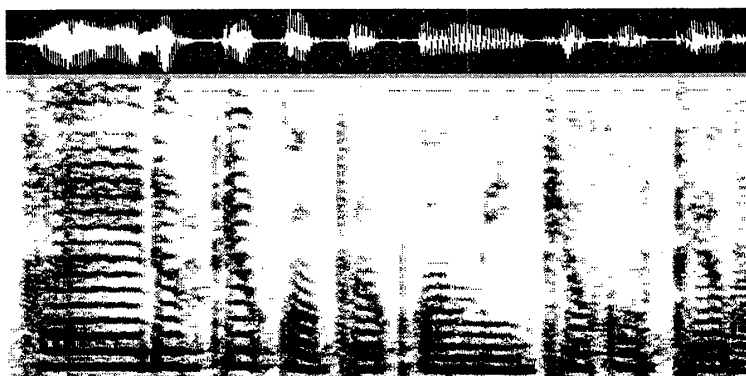


図 3 原信号

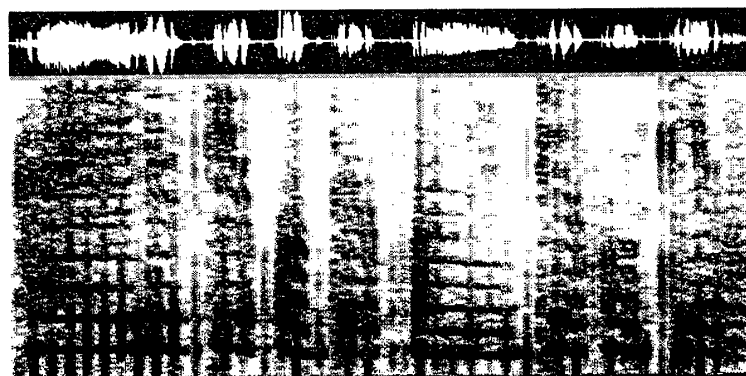


図 4 ピッチ上昇変換された信号 ($p=2.0$ $N=300$)

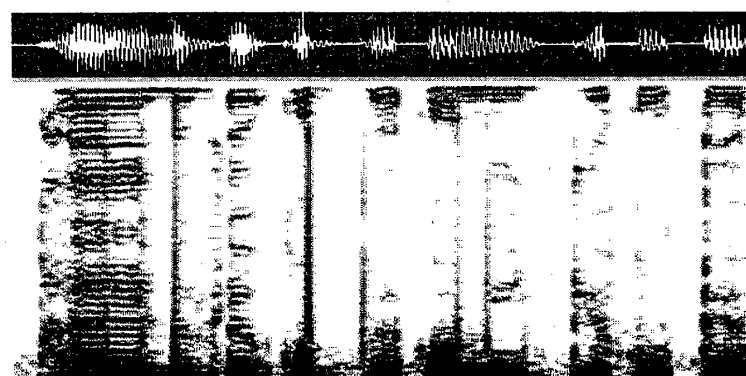


図 5 ピッチ下降変換された信号 ($p=0.5$ $N=1000$)