31 ウェーブレット変換を用いた話速変換

芹沢 恵 菊池久和 中静 真 阿部淑人 佐々木重信 渡辺弘道
新潟大学工学部

1. はじめに

音声の有声区間、および無声区間を伸長することにより、老人性難聴者における聞こえが向上することが示され、老人性難聴者のために音程に変化を与えずに音声の速度(話速)を変更する研究がされている[1]。

本研究では、聴覚の周波数特性に類似した特性を持つウェーブレット変換を用いた話速変換法を提案する。

2. 信号の伸長法

信号$s(t)$のヒルベルト変換により解析信号を

$$a(t)=s(t)+jH[s(t)] \tag{1}$$

と定義する。これを位相

$$\phi(t)=\tan^{-1}\frac{H[s(t)]}{s(t)} \tag{2}$$

と振幅

$$\alpha(t)=\sqrt{s^2(t)+(H[s(t)])^2} \tag{3}$$

に分け、

$$a(t)=\alpha(t)\exp\left(j\phi(t)\right) \tag{4}$$

とかく。瞬時周波数を

$$f(t)=\frac{1}{2\pi}\cdot\frac{d\phi(t)}{dt} \tag{5}$$

と定義する。そして、聴覚ではある時刻tでの音の大きさが$\alpha(t)$に、高さが$f(t)$に対応していると仮定し、解析信号$a(t)$の振幅$\alpha(t)$と瞬時周波数$f(t)$をそれぞれ

$$\alpha(t/m) \tag{6}$$

$$f(t/m) \tag{7}$$

に変更すれば、時間的にm倍に伸長した信号

$$b(t)=\beta(t)\exp\left(j\theta(t)\right) \tag{8}$$

ただし、

$$\beta(t)=\alpha(t/m) \tag{9}$$

$$\theta(t)=2\pi\int_0^t f(t/m)\,dt \tag{10}$$

を作ることができる。これを話速変換と呼ぶ。

複数の周波数成分を含む信号を伸長する場合は、信号を適切に帯域分割し、それぞれの信号に話速変換を適用することが考えられる。

3. 話速変換のシミュレーション

基底膜の長さ方向のそれぞれの場所が、ある特定の周波数のときに最大の振幅で振動する。音の高周波成分は基底膜の基部近くに、低周波成分は先端付近に最大の振動を生じさせる。周波数と基底膜の最大振動する場所との関係は、対数周波数分析になる。

ウェーブレット変換は、時間的にも周波数的にも、局在した関数系での関数展開であり、スケールサイズの相似変換により基底関数を生成するため、対数周波数分析をする系であり高速演算法も知られている[2]。

3.1 位相の微分の近似と補間方法について

離散時間信号であるため位相の微分は後方差分で近似した。補間は、3次スプライン補間を採用した。3次スプライン関数は

$$\begin{aligned} s(n) = {} & M_{i-1}\frac{(a_i-n)^3}{6h_i}+M_i\frac{(n-a_{i-1})^3}{6h_i} \\ & +\left(y_{i-1}-\frac{M_{i-1}}{6}\right)\frac{a_i-n}{h_i} \\ & +\left(y_i-\frac{M_i}{6}\right)\frac{n-a_{i-1}}{h_i} \end{aligned} \tag{11}$$

ただし、

$s(a_i)=y_i$

$h_i=a_i-a_{i-1}$

M_iは定数

$i=1,\cdots,N$

のM_iが決まれば完全に決まる。M_iはa_{i-1}で$s_{i-1}(n)$と$s_i(n)$は2次導関数まで一致する。a_{i+1}で$s_i(n)$と$s_{i+1}(n)$は2次導関数まで一致する。という二つの条件から決定される。周波数特性と矩形波、のこぎり波、正弦波の補間例を図1に示す。

3.2 帯域分割をせずに話速変換をした場合

音声信号を帯域分割せずに、話速変換をした。

3.3 Daubechiesの8タップウェーブレットを使用した場合

Daubechiesの8タップウェーブレット(D8)は完全再構成が可能である。また、各スケールで等間隔のサンプリングを行い、過剰系として使用した。

D8は係数が実数である。振幅と位相を作り出すために、図2に示したように、信号をヒルベルト変換した場合と、D8の係数をヒルベルト変換する場合の2つの処理を試みた。

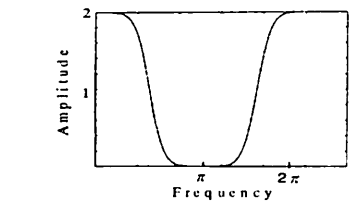
3.4 ガボール変換を用いた場合

ウェーブレット関数に中心周波数$\Omega=2\pi$(周波数分解能が低い場合)と$\Omega=4\pi$(高い場合)としたガボール関数を用いた。処理過程は図2(b)と同様である。

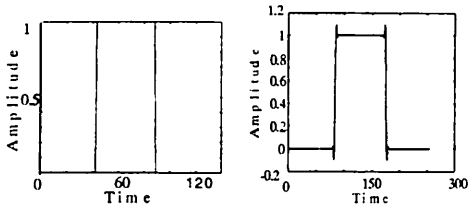
4. 比較検討

音声信号を帯域分割せずに話速変換をした場合、音声信号をヒルベルト変換しD8を用いた場合、D8の係数をヒルベルト変換した場合、ガボール関数の中心周波数Ωを2πと4πにした場合の結果を比較検討する。

すべての場合において継続時間は伸長できた。品質については、音声信号を帯域分割せずに話速変換をした場合は、音色は金属音

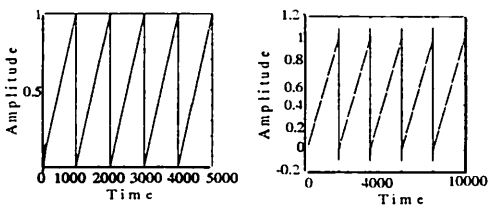


(a)補間関数の周波数特性



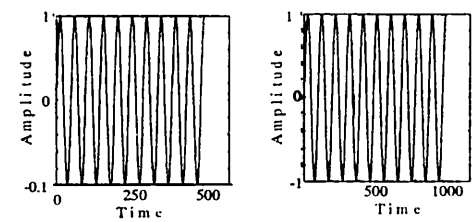
原信号　　補間後の信号

(b) 矩形波



原信号　　補間後の信号

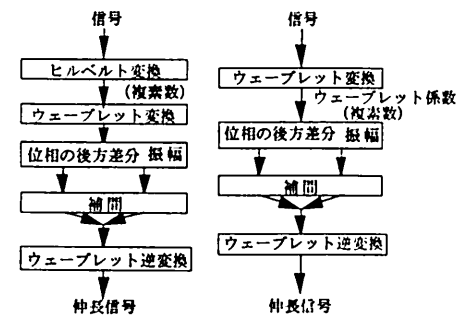
(c) のこぎり波



原信号　　補間後の信号

(d) 正弦波

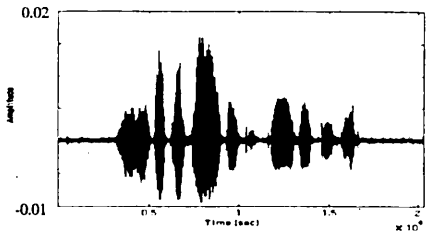
図1　補間



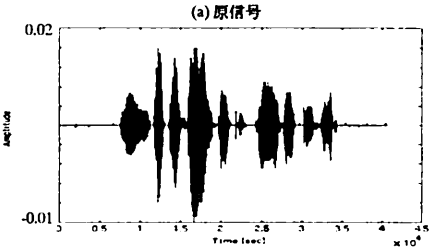
(a) 信号をヒルベルト変換した場合　　(b) D8をヒルベルト変換した場合

図2　D8を用いた場合の処理過程

的になってしまい、発音も不明瞭で聞き取ることができないほどの品質になった。D8を用いた場合は、帯域分割しなかった場合よりも多少品質が向上したが、音程が変化してしまい個人性も失われ、はっきりと聞き取ることができない。これは、分割される帯域幅が広く、ホルマントの分離が充分でないためと思われる。また、音声信号をヒルベルト変換しD8を用いた場合とD8の係数をヒルベルト変換した場合では、品質に違いは感じられなかった。ガボール関数の中心周波数Ωを2πとしたときは、D8を用いた場合よりもかなり品質は向上したが、多少の音程の変化が感じられる。Ω＝4πとした方が、ホルマントがはっきり分離できたため、最も良い品質が得られた。図3に"ウェーブレット変換"と発音した原信号とΩ＝4πとしたときの伸長信号の波形を示した。エンベロープが保持されている。



(a) 原信号



(b) 伸長信号

図3　結果波形

5．まとめ

話速変換法を定義した。複数の周波数成分を含んでいる音声信号の話速変換は、ウェーブレット変換を用い、信号帯域を幾何対称に等比分割することにより各周波数成分の信号に分離し、得られた信号それぞれに話速変換を施した。音程、音色はそのままに、継続時間だけが伸長された。

本稿では時間伸長された音声を耳で聞きその特徴を述べたが、他の話速変換の手法との比較や、ウェーブレット関数の選択などを行うために、客観的な話速変換の評価法の検討が必要である。評価法は、話速変換前と後のピッチやホルマント、有声音、無声音など準定常的な特徴量を比較するなどで実現が可能であると考える。

参考文献

(1)中村章, 清山信正, 池沢龍, 都木徹, 宮坂栄一, "リアルタイム話速変換型受聴システム," 日本音響学会誌, 50巻, 7号, pp.509-520, 1994

(2)H. Kikuchi, M. Nakashizuka, H. Watanabe, S.Watanabe and N. Tomisawa, "Fast Wavelet Transform and its Application to Detecting Detonation," *IEICE Trans. on Fundamentals*, E75-A, 8, pp.980-987, 1992