

時間領域平滑化群遅延を用いた短時間位相の効率的表現方法

坂野 秀樹<sup>†\*</sup>      陸 金林<sup>†</sup>      中村 哲<sup>†\*\*</sup>      鹿野 清宏<sup>†</sup>  
河原 英紀<sup>††</sup>

Efficient Representation of Short-Time Phase Based on Time-Domain Smoothed Group Delay

Hideki BANNO<sup>†\*</sup>, Jinlin LU<sup>†</sup>, Satoshi NAKAMURA<sup>†\*\*</sup>, Kiyohiro SHIKANO<sup>†</sup>,  
and Hideki KAWAHARA<sup>††</sup>

あらまし 音声の短時間位相を効率的に表現する方法を提案する。本論文では、音声の群遅延を逆フーリエ変換したものを適当な次数で打ち切った時間領域平滑化群遅延 (TSGD) を用いることにより、短時間位相を表現することを試みる。セグメンタル SNR による実験を行った結果、TSGD の 100 次以下に波形の形状に関する情報が集中していることがわかった。また、主観評価実験により、TSGD の次数を 30 次程度とした場合に、原音とほとんど区別つかない高品質な音声合成されることが示され、提案手法により短時間位相を効率的に表現できることが示された。また、提案手法を用いれば、これまで分析合成方式で問題になっていた有声無声判定の誤りによる劣化も抑えられるため、音声符号化や声質変換などの応用に有効であると考えられる。

キーワード 短時間位相, フーリエ変換, 群遅延, 時間領域平滑化群遅延

1. ま え が き

短時間位相は、振幅情報に比べると聴覚的に重要ではないと一般にいわれている。この性質を利用して分析合成系では、原音の位相情報を用いずに零位相や最小位相での合成が行われていた。また、誉田らは音声を近似的に零位相にする位相等化フィルタを用いた音声符号化方式を提案し、一定の効果があることを示した [1], [2]。しかし、“鼻づまり音,” “buzzy” などと呼ばれる零位相・最小位相合成独特の音の悪さは以前より指摘されており、高い品質を要求される応用においては、このような分析合成方式がそのまま用いられることは少ない。また、こういった分析合成系では、あるフレームが有声音であるか無声音であるかを判定する有声無声判定を行う必要がある。ところが、この有声無声判定は極めて難しく、この判定誤りが合成音の

劣化をもたらすという問題もある。そこで、これらの問題点を解決するため、いくつかの方法が提案されてきた。

符号化で最も一般的なものは、LPC 分析による残差波形を符号化して用いる方法である [3]。この LPC 残差には LPC 分析で表現しきれなかった振幅情報と原音の位相情報を含んでいるため、最小位相合成のような音質の劣化は軽減される上、有声無声判定に起因する劣化も少ない。この方式はハイブリッド符号化と呼ばれるものであり、高品質ながらビットレートを比較的小さくすることができるので符号化では広く用いられている。しかしながら、ハイブリッド符号化方式は残差を波形として扱っているため、波形符号化と同様の問題点である、ビットレートを一定の値以下に下げると品質が極端に劣化するということが知られている。更に、声質変換などの応用を考えた場合、分析合成系のように柔軟に声質を変化させるといったことは困難となる。また、同様の方法として、残差波形を用いる代わりに原音の位相を積極的に利用して音声を合成するものがある [4]~[7]。ところが、これらの手法はいずれも位相を効率的に表現しているとはいいがたい上、補間を自然に行うことは困難であると考えられる。

<sup>†</sup> 奈良先端科学技術大学院大学情報科学研究科, 生駒市  
Graduate School of Information Science, Nara Institute of  
Science and Technology, 8916-5 Takayama-cho, Ikoma-shi,  
630-0101 Japan

<sup>††</sup> 和歌山大学システム工学部, 和歌山市  
Faculty of Systems Engineering, Wakayama University, 930  
Sakaedani, Wakayama-shi, 640-8510 Japan

\* 現在, 名古屋大学工学研究科

\*\* 現在, ATR 音声言語通信研究所

また、聴覚研究の立場から、Pattersonらは、位相の知覚特性に関する知見を利用した位相情報の表現方法を提案したが[8]、現時点では、符号化や声質変換に適用できる段階にまで達していないと思われる。

本研究では、音声符号化・声質変換・音声モーフィング[9]などへの応用を目的とした短時間位相の効率的表現方法を模索している。位相の表現方法に対する要求としては、情報圧縮を行っても聴覚的には劣化が少なかったり、聴覚的に自然な補間が可能であることなどが挙げられる。本論文ではこれらの要求を満たすような短時間位相の効率的表現方法を提案する。

## 2. 群遅延

本研究では、位相を直接用いず、位相の周波数微分である群遅延を用いてパラメータ化を行う。群遅延を用いた理由として(1) DFTにより求めた位相は  $-\pi$  から  $\pi$  の範囲に制限されており、そのまま使用するのは困難であること、(2) 群遅延は、時間領域での遅延量に相当するので[10]直観的に理解しやすい上に、補間などの操作が聴感上違和感なく行える可能性があること、(3) 3. で説明する群遅延の特徴を利用すれば、位相を効率的に表現することも可能となることなどが挙げられる。

### 2.1 群遅延抽出のアルゴリズム

分析の手順を図1に示す。音声波形にはあらかじめピッチマーキングを施しておき、マーク位置を中心に約2ピッチ周期分の長さの窓を掛けて波形を切り出し、分析を行う。以下ではこの切り出した波形を単位波形と呼ぶ。ピッチマーキングについては、500 Hz程度の低域フィルタを施した音声に対してピーク抽出を行った後、ピークにおけるパワーが、フレーム内の平均パワーから求めたしきい値を超えるものをピッチマークとした。ピッチマークは、必要であれば修正を加えた。また、無声音についても、適当な間隔(今回は2.5 ms)でピッチマーキングを行い、有声音と全く同様に処理を行う。

群遅延の抽出は次のように行う。 $x(t)$ を単位波形、そのフーリエ変換を $X(\omega)$ とし、 $X(\omega)$ を以下のように表す。ここで、 $F[\cdot]$ はフーリエ変換を表す。なお、フーリエ変換を行う際には、FFTポイント数に応じた、切り出した波形に対し零詰めを行う。

$$\begin{aligned} X(\omega) &= F[x(t)] \\ &= a(\omega) + jb(\omega) \end{aligned} \quad (1)$$

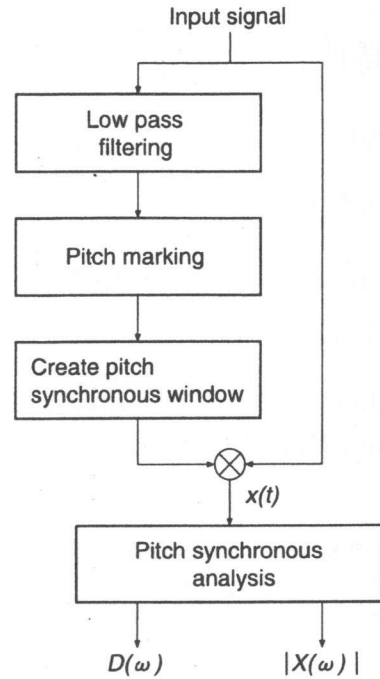


図1 分析のブロック図  
Fig. 1 Block diagram of analysis.

すると、振幅及び位相は次のように表される。

$$|X(\omega)| = \sqrt{a^2(\omega) + b^2(\omega)} \quad (2)$$

$$\theta(\omega) = \arctan \left( \frac{b(\omega)}{a(\omega)} \right) \quad (3)$$

群遅延とは、位相微分の符号を反転したものであり、次式で定義できる。

$$D(\omega) = -\frac{d\theta(\omega)}{d\omega} \quad (4)$$

ここで、

$$\tan \theta(\omega) = \frac{b(\omega)}{a(\omega)} \quad (5)$$

の両辺を微分することにより、群遅延が以下のように求まる。

$$-\frac{d\theta(\omega)}{d\omega} = \frac{-\frac{db(\omega)}{d\omega}a(\omega) + b(\omega)\frac{da(\omega)}{d\omega}}{a^2(\omega) + b^2(\omega)} \quad (6)$$

更に、フーリエ変換の周波数微分の法則  $(-jt)x(t) \leftrightarrow \frac{dX(\omega)}{d\omega}$  を利用すると、

$$\begin{aligned} \frac{dX(\omega)}{d\omega} &= \frac{da(\omega)}{d\omega} + j\frac{db(\omega)}{d\omega} \\ &= F[-jtx(t)] \end{aligned} \quad (7)$$

であるから、最終的に群遅延  $D(\omega)$  は

$$D(\omega) = \frac{-\text{Im}[F[-jtx(t)]]a(\omega) + b(\omega)\text{Re}[F[-jtx(t)]]}{a^2(\omega) + b^2(\omega)} \quad (8)$$

と表される。この式は、 $x(t)$  と  $-jtx(t)$  を用いれば、直接群遅延が求まることを示している。

### 2.2 群遅延からの音声の合成

合成の手順を図2に示す。群遅延から位相を求めるには、群遅延を0から $\omega$ まで積分すればよい。

$$\tilde{\theta}(\omega) = -\int_0^\omega D(\psi)d\psi \quad (9)$$

実際には離散時間系であるため、積分を累積和で近似して位相を導出する。

群遅延により導出した位相と振幅スペクトルを用いて複素スペクトルを得、それを逆フーリエ変換することで単位波形を再構成する。

$$\tilde{X}(\omega) = |X(\omega)|e^{j\tilde{\theta}(\omega)} \quad (10)$$

$$\tilde{x}(t) = F^{-1}[\tilde{X}(\omega)] \quad (11)$$

この単位波形をOverlap Add法[11]によって配置することにより、音声を合成する。

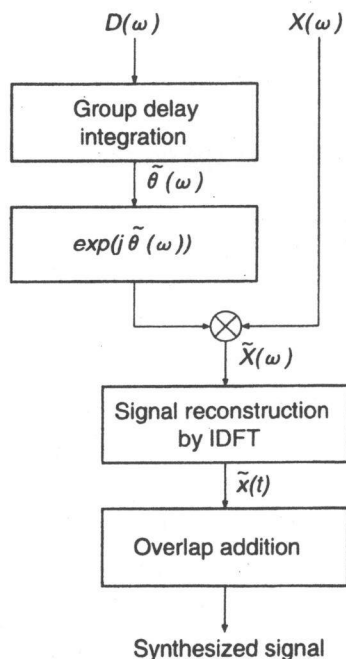


図2 合成のブロック図

Fig. 2 Block diagram of synthesis.

ただし、このままでは折返しひずみや累積和の誤差などの影響により、劣化を生ずることがあるため、次の操作を施し、誤差の発生を抑える。

[直線位相成分の付与] 実関数をフーリエ変換したときの位相は、角周波数が $2\pi$ のときに、 $2n\pi$  ( $n$ :整数)となる。抽出された位相は、式(9)の積分が累積和で近似されているため、この制約を満たさない。そのため、次のように直線位相成分の付与を行う。

$$\tilde{\theta}(\omega) = -\int_0^\omega D(\omega)d\omega + c\omega \quad (12)$$

ここで、 $c$ として $\tilde{\theta}(2\pi) = 2n\pi$ となるような定数の中で最小のものを用いる。

[FFTポイント数の増加] 位相特性の時間応答波形(振幅がフラットで位相特性のみをもつ複素スペクトルを逆フーリエ変換したものがこれに相当する)は、場合によっては極めて長い応答長をもつことがある。その場合、上述の手法により求めた群遅延は、FFTポイント数に応じた折返しひずみを含むものとなる。そこで、FFTポイント数を増やすことにより、この折返しひずみを減少させる。FFTポイント数を増やすことは、結果的に累積和の近似誤差を減少させることにもつながるため、非常に効果的である。FFTポイント数の増加による効果を調べるため、予備的な実験を行った。FFTポイント数を変化させ、原音と、群遅延を用いた合成音とのセグメンタルSNRを測定した。FFTポイント数以外の実験条件は表1と同じである。音声データとして、二つの短文章を男女各2名が発声したもの(計8文章)を用いた。実験結果を図3に示す。この図から、FFTポイント数を大きくすることで、SNRが改善されていくことがわかる。また、合成音を聴いたところ、8192ポイント程度で原音との差がほとんど感じられなかった。そのため、以下の実験では、FFTポイント数を8192ポイントとする。

以上の処理により、原音とほとんど区別つかない高品質な音声を合成することが可能となることが予備実験により確認された。

分析窓としては、ピッチマーク間隔に応じた左右非

表1 分析条件  
Table 1 Analysis conditions.

標本化周波数	8 kHz
分析窓	ガウス窓
分析窓長	可変
FFTポイント数	8192 points

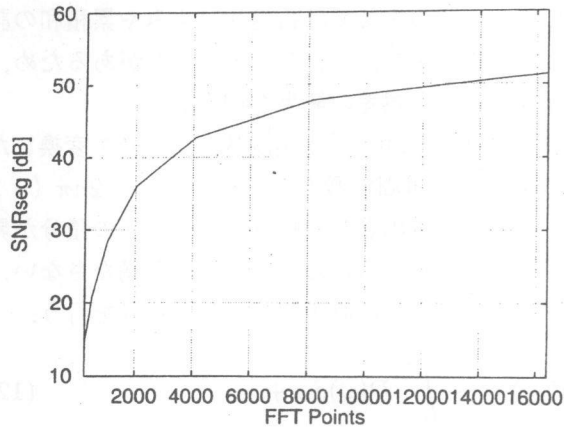


図3 FFTポイント数とセグメンタルSNRの関係  
Fig. 3 Relationship between the FFT points and the segmental SNR.

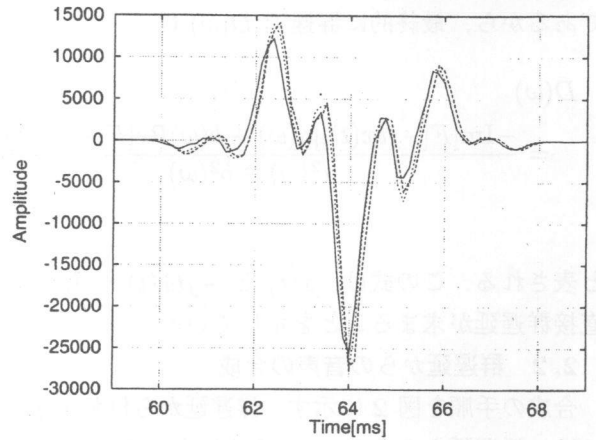


図4 単位波形の例  
Fig. 4 Examples of unit waveform.

対称の窓を用いた。ここで窓関数としてハニング窓のように足して振幅が1になるものを選べば、窓掛けによる影響で原音声に変形されることはなくなる。ところが、ハニング窓は低次のサイドローブの減衰が少ないため、ある周波数における群遅延の値が、近傍の周波数の影響によってひずむことになる。この群遅延を用いて合成を行うと、群遅延を積分する際の誤差が大きくなり、合成音が劣化することがある。そこで、分析はガウス窓を用いて行い、合成の際に、ガウス窓をハニング窓に変換するような窓関数を掛けて音声を合成する。この処理により、視察での修正を施さないピッチマークを用いた場合でも、合成音の品質が変化しないことが予備実験により確認された。また、以上の処理を行うことにより、ピッチマーキングにおける有声無声判定誤りに起因する劣化はほとんど発生しなくなる。

### 3. 群遅延の効率的表現方法

前節で説明した方法を用いて高品質な音声合成されることが確認されたが(2.2, 図3参照), 位相のみでFFTポイント数の半分(群遅延は偶関数であるため)という膨大な情報量になってしまう。ここでは、その群遅延をいかに効率的に表現するかについて議論し、効率的な群遅延の表現方法を提案する。

#### 3.1 群遅延の特徴

実際に、図4のようなある単位波形を分析したときの群遅延を図5に示す。図は時間的に連続した三つの単位波形における各々の群遅延を示したものである。この図から、次のような特徴が読み取れる。

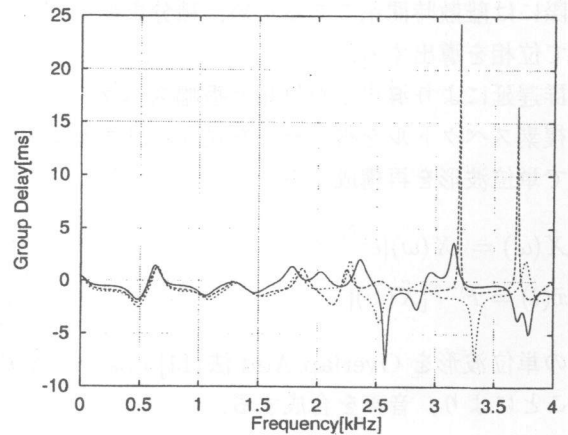


図5 群遅延の例  
Fig. 5 Examples of group delay.

(1) 時間波形の時間変化が小さい場合、群遅延の時間変化も小さい。

(2) 群遅延は、いくつかのピークを除けば、周波数方向への変化は比較的滑らかである。

(3) 群遅延の概形と時間波形の形状は高い相関をもっているが、群遅延のいくつかのピークと音声波形の形状との相関は低い。

経験的な知見によれば、三つの時間波形は極めて似ているため、群遅延の微細構造のみが異なるこれらの時間波形の違いを知覚するのは難しい。すなわち、聴覚は、群遅延の微細構造の違いをほとんど知覚できないと考えられる。しかし、聴覚は、分析合成音における零位相や最小位相による劣化は知覚することができる。例えば、零位相の場合は、すべての周波数で群遅延の値が定数となるので、この場合の合成音と、図5のよ

うな群遅延をもつ音声との違いを聴覚は知覚できることになる。これらのことから、我々は、音声の知覚においては群遅延の概形情報が特に重要であるという仮説を立てた。すなわち、群遅延の概形の情報のみを何らかの方法で抽出して用いれば群遅延の情報圧縮を音質の劣化なしに行える可能性がある、ということである。以下で具体的に群遅延の概形の情報を抽出する方法を説明する。

### 3.2 時間領域平滑化群遅延

例えば、群遅延自体に低域フィルタを施せば、群遅延は滑らかになり、概形の情報のみを抽出することができる。すなわち、群遅延をフーリエ変換したものを適当な次数で打ち切ることで、その概形のみを抽出することが可能となる。ただし、群遅延は偶関数であるからフーリエ変換と逆フーリエ変換は等化となるため、逆フーリエ変換によりこの処理を行うこととする。ここで、群遅延は周波数領域のパラメータであるため、それを逆フーリエ変換したこのパラメータは時間領域であることがわかる。よって、以下では、このパラメータを平滑化された群遅延の時間領域での表現であることから TSGD (Time-domain Smoothed Group Delay: 時間領域平滑化群遅延) と呼ぶことにする。離散時間系での群遅延を  $D(k)$  とすると、次数  $\alpha$  で打ち切る場合の  $n$  次の TSGD  $d(n)$  は、次のように表される。

$$d(n) = F^{-1}[D(k)] \quad (0 \leq n \leq \alpha) \quad (13)$$

なお、TSGD は時間領域であることから、その次数は群遅延を求める際の FFT ポイント数には依存しない。TSGD から音声を合成する場合には、以上の操作とは逆に、TSGD をフーリエ変換し群遅延を求めればよい。

この TSGD は、時間領域のパラメータであるため、入力音声のサンプリング周波数が異なる場合でも、次数を時間単位で打ち切ることにより対応が可能となる。なお、その場合は FFT ポイント数も必要に応じて増加させる必要がある。

## 4. 評価実験

### 4.1 実験 1: セグメンタル SNR による客観評価

TSGD の次数を変化させたときに、どの程度まで波形形状を再現できるかを調査するため、セグメンタル SNR による客観評価を行う。

#### 4.1.1 実験方法

TSGD の次数を変化させ、原音とのセグメンタル SNR を測定する。位相以外の情報についてはそのまま用いている。また、すべてのパラメータに対し、量子化等の処理は行っていない。

実験条件を表 1 に示す。音声データとして、二つの短文章を男女各 5 名が発声したもの (計 20 文章) を用いた。

#### 4.1.2 実験結果

TSGD の次数とセグメンタル SNR との関係を図 6 に示す。図 6 より、TSGD の次数が 100 次以上の場合には、セグメンタル SNR がほぼ一定の値をとっていることがわかる。これは、波形の形状に関する情報が、TSGD の 100 次以下の成分に集中していることを示している。つまり波形の形状に関する情報は群遅延の概形に存在するということができる。

また、男性の方が常に低いセグメンタル SNR をとっていることがわかるが、これは次の理由によると考えられる。通常、男性の方がピッチが低いため、男性の場合は分析窓が長くなる。このとき、群遅延に含まれる情報量は、男性の方が多いためと考えられる。そのため、TSGD の次数を同じ値で打ち切った場合には、男性の方がより低い値となったと思われる。また、群遅延に含まれる情報量が多いということは、位相特性の時間応答波形も長くなる可能性があり、この場合は折返しひずみも増加することになる。そのため、収束した値も男性の方が低い値となったと考えられる。性別が同じ場合には、音声データの基本周波数が高い場合に高いセグメンタル SNR をとる傾向があることもわかった。

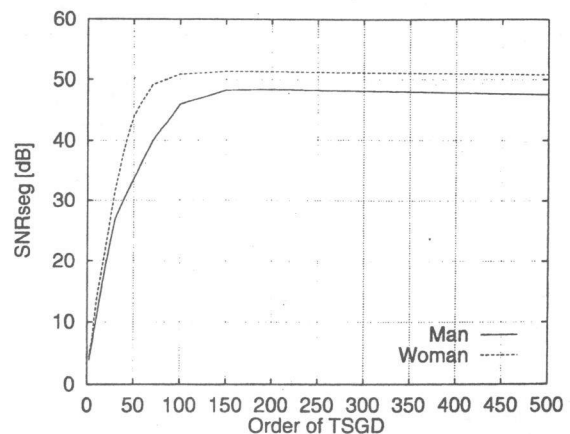


図 6 TSGD の次数とセグメンタル SNR の関係  
Fig. 6 Relationship between the order of TSGD and the segmental SNR.

た。また、TSGDの次数の増加に伴い、セグメンタルSNRが若干低下しているが、原因はわかっていない。

本実験では、波形のひずみという観点から評価を行ったが、群遅延の概形に聴覚的に重要な情報が集中するのであれば、TSGDの次数を100以下にした場合でも、高品質な音声を合成できると考えられる。そこで、次節では主観評価実験を行い、どの程度の次数であれば主観的に高品質な音声が合成できるかを調査する。

#### 4.2 実験2: オピニオン法による主観評価

聴感上TSGDがどの程度の性能をもっているかを評価するため、主観評価実験を行う。また、短時間位相の必要性を示すために、零位相・最小位相による合成音も作成し比較を行う。

##### 4.2.1 実験方法

被験者にランダムな順序で合成音を呈示し、その品質を5段階で評価してもらった。被験者には特に音色に着目して評価を行うように指示した。被験者は成人男女計12名であり、防音室内にてヘッドホン受聴を行った。音声データは二つの短文章を男女各4名が発声したもの(計16文章)を用いた。その他の実験条件については、実験1と同様である。

主観評価実験に用いた方式は、零位相音声(ZP)、最小位相音声(MP)、TSGDの次数を10次、20次、30次、100次としたもの(それぞれTSGD10、TSGD20、TSGD30、TSGD100)、群遅延による音声(GD)、原音声(ORG)の計8種類である。零位相・最小位相合成については劣化を最小限に抑えるために若干特殊な処理を行っているので、以下で詳しく説明する。

##### 4.2.2 零位相・最小位相音声の合成方法

零位相若しくは最小位相で合成した場合には、基本周波数が原音の厳密な基本周波数に対応していないなどの理由で劣化が生じる。このため、以下の手順で音声を合成し、劣化を最小限に抑える。今回は精密な基本周波数を求める方法として、河原らの提案するTEMPO法[12]を用いた。

(1) あらかじめ精密な基本周波数を求めておく。倍ピッチや半ピッチなどの極端な誤りが発生したときには修正を施す。

(2) ピッチマーキングの情報から窓掛けを行い、単位波形を切り出す。

(3) FFTを用い単位波形を零位相若しくは最小位相に変形する。

(4) 無声音の場合、雑音源を畳み込む。この場合

の位相はランダム位相となる。

(5) (1)で求めた基本周波数の値を用いて単位波形を配置する。ただし、文献[13]にある方法を用い、サンプリング間隔で決まるものより更に精密に基本周波数を制御する。

これら一連の処理を行うことによって零位相・最小位相音声とTSGDによる合成音との違いは短時間位相の差だけになる。

##### 4.2.3 実験結果

主観評価の結果を図7に示す。エラーバーは標準偏差を表している。零位相及び最小位相の評価が低いのに対し、TSGDの次数が20次以上の場合では、男女ともに高い評価を得ており、30次以上の場合には、原音に極めて近い品質で音声が合成できるといえる。実際、聴いた印象でも30次以上では原音とほとんど区別がつかず、20次では若干ざらついた感じになり、10次ではざらつきがひどくなるという具合である。

セグメンタルSNRによる客観評価の結果とは異なり、TSGDの次数を100にした場合とTSGDの次数を30にした場合の品質の差が少ないが、これは、群遅延の概形に聴覚的に重要な情報が集中することによるものと考えられる。そのため、主観的には、30次以上で極めて高品質の音声が合成できると考えられる。ただし、群遅延の概形がどれほど聴覚的に重要であるかを調べるには、より詳細な聴覚実験が必要であろう。

また、零位相や最小位相では有声音判定を行っていることも品質を劣化させる原因であると考えられる。実際の音声では、無声音から有声音へと連続的に位相

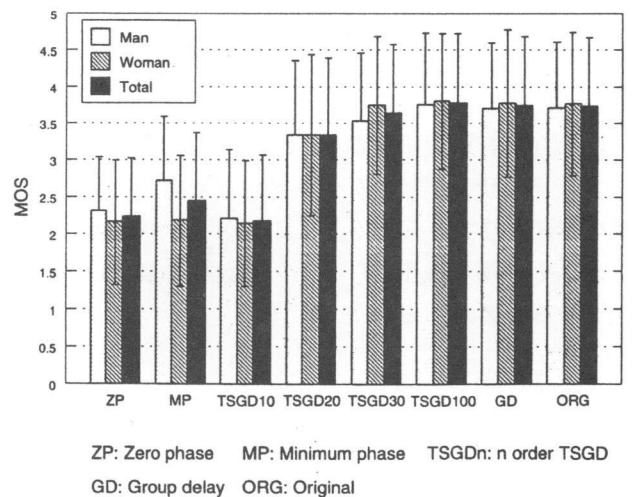


図7 オピニオン法による評価結果  
Fig. 7 Mean opinion scores.

が変化していくが、有声無声判定を行った場合、位相が無声音のランダム位相から、有声音の零位相若しくは最小位相へと急激に変化することになる。このため、有声無声判定に誤りがなかったとしても、無声音と有声音の連続点で不自然な合成音となり、劣化が発生すると考えられる。また、実際には有声無声判定は極めて難しく、実環境においては大きな問題となる。それに対し、TSGDではそのような有声無声判定を行わず、符号化や声質変換などの応用にも有効性は高いと考えられる。

## 5. む す び

音声の位相においては、群遅延の概形がより重要であるとの仮説から、群遅延の概形を表すパラメータとして時間領域平滑化群遅延 (TSGD) を提案した。セグメンタル SNR による実験では、TSGD の低次、すなわち群遅延の概形に波形形状に関する情報が集中していることがわかった。また、主観評価では、TSGD の性能は零位相・最小位相合成を凌駕しており、TSGD の次数を 30 次以上とした場合には、原音とほとんど区別のつかない極めて高品質な音声合成が合成されることが示された。

本論文で提案した手法は、LPC 残差などの位相を扱う場合にもそのまま用いることができ、更に処理時間も比較的短時間なため、符号化等への適用も容易であると思われる。今後、提案手法を用いた音声符号化・声質変換等について検討する予定である。

**謝辞** 本研究の一部は、科学技術振興事業団による戦略的基礎研究推進事業 CREST の援助を受けて行われた。

## 文 献

- [1] 誉田雅彰, 守谷健弘, “位相等化処理を用いた音声符号化,” 音響学音声研資, S84-05, 1984.
- [2] 誉田雅彰, “位相等化音源による音声分析合成,” 信学技報, SP89-124, 1989.
- [3] M.R. Schroeder, “Code-excited linear prediction (CELP): High-quality speech at very low bit rates,” Proc. ICASSP 1985, pp.937-940, 1985.
- [4] I.M. Trancoso, R.G. Gomez, and J.M. Tribolet, “A study on short-time phase and multipulse LPC,” Proc. ICASSP 1984, pp.10.3.1-10.3.4, 1984.
- [5] P. Hedelin, “Phase compensation in all-pole speech analysis,” Proc. ICASSP 1988, pp.339-342, 1988.
- [6] 池田幹男, 中村一雄, 浅田博重, “オールパスフィルタの位相特性を利用した音声符号化,” 音響学講論集, pp.305-306, 1991.
- [7] 山浦 正, 高橋真哉, “低ビットレート CELP における有

- 声品質改善の検討,” 音響学講論集, pp.263-264, 1994.
- [8] R.D. Patterson, “The sound of a sinusoid: Time-interval models,” J. Acoust. Soc. Am., vol.96, no.3, pp.1419-1428, 1994.
- [9] 坂野秀樹, 武田一哉, 鹿野清宏, 板倉文忠, “包絡と音源の独立操作による音声モーフィング,” 信学論 (A), vol.J81-A, no.2, pp.261-268, Feb. 1998.
- [10] B. Boashash, “Estimating and interpreting the instantaneous frequency of a signal,” Proc. IEEE, vol.80, no.4, pp.519-569, 1992.
- [11] L.R. Rabiner and R.W. Shafer, Digital Processing of Speech Signal, Prentice-Hall, 1978.
- [12] 河原英紀, Alain de Chaveigné, “原理的に誤りの存在しないピッチ抽出方法とその評価について,” 信学技報, SP96-96, 1997.
- [13] 河原英紀, 増田郁代, “時間周波数領域での補間を用いた音声の変換について,” 信学技報, EA96-28, 1996.  
(平成 11 年 11 月 29 日受付, 12 年 9 月 5 日再受付)



坂野 秀樹 (学生員)

平 8 名大・工・電子情報卒。平 10 奈良先端大博士前期課程了。同年、名大大学院博士後期課程入学。音声合成・声質変換に関する研究に従事。日本音響学会会員。



陸 金林 (正員)

昭 57 中国・華東工学院・電子卒。平 8 宇都宮大学大学院博士後期課程了 (工博)。平 2~5 宇都宮大学工学部助手。平 8 年 4 月より奈良先端大助手。音声合成を中心にした音声信号処理に関する研究に従事。日本音響学会会員。



中村 哲 (正員)

昭 56 京都工繊大・工芸・電子卒。昭 56~平 6 シャープ (株) 中央研究所及び情報技術研究所に勤務。昭 61~平 1 ATR 自動翻訳電話研究所に出向。平 6~12 奈良先端科学技術大学院大学情報科学研究科助教授。平 8 年 3 月~8 月 Rutgers University・CAIP Center 客員教授。平 12 年 4 月より ATR 音声言語通信研究所第一研究室長。音声情報処理、主として音声認識の研究に従事。京大博 (工学)。平 4 日本音響学会粟屋学術奨励賞受賞。IEEE, 情報処理学会, 日本音響学会, 人工知能学会各会員。



鹿野 清宏 (正員)

昭45名大・工・電気卒。昭47同大学院修士課程了。同年電電公社武蔵野電気通信研究所入所。昭59~61カーネギーメロン大客員研究員。昭61~平2 ATR自動翻訳電話研究所音声情報処理研究室長。平4 NTTヒューマンインタフェース研究所主席研究員。平6より奈良先端科学技術大学院大学情報科学研究科教授。音情報処理学講座を担当。工博。主として音声・音情報処理の研究及び研究指導に従事。昭50 本会米沢賞, 平3 IEEE SP 1990 Senior Award, 平6 日本音響学会技術開発賞各受賞。IEEE, 日本音響学会, 情報処理学会各会員。



河原 英紀 (正員)

昭47北大・工・電子卒。昭52同大学院工学研究科博士課程了。同年電電公社武蔵野電気通信研究所入所。平4 ATR人間情報通信研究所第一研究室長。平9年4月より和歌山大学システム工学部教授。戦略的基礎研究推進事業「聴覚脳プロジェクト」研究代表者。ATR人間情報通信研究所招聘研究員。工博。聴覚機能の数理的解明と工学的実現の研究に従事。平9 音響学会佐藤論文賞, 平9 電子通信協会テレコム技術論文賞各受賞。日本音響学会, IEEE, ASA, 神経回路学会, 認知科学会各会員。