社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

複合ウェーブレットモデルに基づく音声の分析合成

槐　　武也† 　松本　恭輔†† 　酒向　慎司† 　嵯峨山茂樹†

† 東京大学大学院情報理工学系研究科
‡ 東京大学工学部
〒 113–8656 東京都文京区本郷 7–3–1
E-mail: {saikachi, k-matsumoto, sako, sagayama}@hil.t.u-tokyo.ac.jp

あらまし　本稿では、パラメトリックな音声分析合成モデルとして、複合ウェーブレットモデル (Composite Wavelet Model、以下 CWM) 法を提案し、その有効性について議論する。従来の巡回型フィルタによる音声合成では、その時間特性が音声品質低下の一要因である可能性があり、提案法ではこれを改善することが期待できる。提案法では音声のスペクトル包絡を混合ガウス関数モデル (GMM) で近似することで少数のパラメータによって表現する。合成時にはこの GMM の逆フーリエ変換である複合 Gabor ウェーブレットを基本波形として、これをピッチ周期ごとに重ね合わせて有声音を合成する。検証のため、提案法により音声を分析合成し、時間特性が改善されていることを確認した。
キーワード　音声分析合成, 混合ガウス関数モデル, 複合 Gabor 関数

Speech Analysis and Synthesis based on Composite Wavelet Model

Takeya SAIKACHI†, Kyosuke MATSUMOTO††, Shinji SAKO†, and Shigeki SAGAYAMA†

†Graduate School of Information Science and Technology, The University of Tokyo
‡School of Engineering, The University of Tokyo
7–3–1 Hongo, Bunkyo-ku, Tokyo, 113–8656 Japan
E-mail: {saikachi, k-matsumoto, sako, sagayama}@hil.t.u-tokyo.ac.jp

Abstract　In this paper, we propose and discuss Composite Wavelet Model (CWM) for speech analysis and synthesis. Conventional speech synthesis using recursive filter involves a problem of temporal characteristics which may cause degradation of synthetic speech quality. In our new method, speech is analyzed by approximating the spectral envelope with a Gaussian Mixture Model (GMM). In the synthesis stage, the basic waveform is chosen to be composite Gabor function which is inverse Fourier transform of Gaussian mixture and is pitch-synchronously overlapped and added. For preliminary evaluation of the proposed method, we analyzed and synthesized speech and showed that it improved the temporal characteristics.
Key words　Speech Analysis/Synthesis, Gaussian Mixture Model, Composite Gabor Wavelet

1. はじめに

本稿では、新しいパラメトリックな音声分析合成モデルである複合ウェーブレットモデル (Composite Wavelet Model、以下 CWM) 法を提案する。提案法は Gabor wavelet をピッチ周期で重ね合わせる方法であり、高品質な合成音声を実現し、かつ加工性に優れることが期待される。今回は与えられた音声を低次元のパラメータに分析した上で音声の再合成を試みる音声分析合成実験を行い、結果を報告する。

音声合成の研究は従来から盛んに行われてきた。PSOLA 方式 [2] や波形接続型の音声合成手法 (たとえば CHATR [3]) は、与えられたテキストを聴き取りやすい高品質な音声で読み上げるといった目的において一定の成功を収めた。

しかしコンピュータにおける音声情報処理が進展するに伴い、ただ読み上げるだけにとどまらない、より高度な要求がなされるようになった。例えば Galatea [1] などの擬人化音声対話エージェントシステムでは、人間らしい対話を実現するために会話音声や感情音声、すなわち状況に合わせた様々な意志や感情が含まれた音声を合成することが求められる。これらの要因による音声の変化は極めて多様であるため、それらをカバーするデータを用意するのは記憶容量や労力が多く必要となり、良い解決と言えない。音声変換等の手法を適用するにしても、さ

まざまな制限が予想される。

音声の特徴がパラメトリックなモデルで表現された音声合成法であれば、そのパラメータの操作によって合成音の柔軟な操作が可能であると考えられる。LPC [4] や PARCOR [5]、LSP [6]、ケプストラム合成 [7] などの巡回フィルタ型音声合成はパラメトリックな音声の分析合成法の代表であり、加工性の高さが期待される。これらの手法は主に音声の符号化や分析・合成法として広く利用されているが、例えば HMM による音声合成手法 [9] によって統計的にパラメータを生成すれば、テキスト音声合成に応用することが可能になる。しかし、これらのフィルタ型音声合成をもちいたテキスト音声合成は、一般に波形接続型音声合成に比べ合成音声の品質が低いという問題がある。

パラメトリックな分析合成法で、巡回フィルタによらない合成音声波形生成の手法としては、複合正弦波モデル (CSM) による手法 [8] がある。本稿で提案する音声合成法も、そのアイディアを基にしている。

我々は、従来の巡回フィルタ型音声合成には時間特性に関する問題が内在していると考え、それが音声品質の問題の一因となっている可能性を検討する。そして、その問題を解決する可能性のある新手法を提案し実験的に検証する。

なお、STRAIGHT 法や正弦波合成法などはノンパラメトリックな音声分析合成法と分類し、ここでは論じない。但し、Zolfaghari らによる GMM を用いた音声スペクトル解析法 [10] は、本稿の音声分析法として利用する。

以下、2 章ではこれまでの音声合成方法とその問題点を述べ、3 章で提案法である CWM 法について述べる。そして 4 章で提案法の性能の評価を行い、考察を加える。

2. 従来の音声合成法の問題点

2.1 波形接続方式とフィルタ方式

対話調の合成音声など様々な要求に適用可能な、高品質かつ多様なスタイルの音声を生成できる音声合成の方法を考える。

PSOLA 方式や波形接続型の音声合成手法は、十分なバラエティの音声素片のデータがあれば高品質な合成音声が期待できるが、データベースに含まれない条件の音声を合成したり、話者適応するような音声の特徴を操作する加工性は高くない。合成したい音声のスタイルに応じた音声データを補うことによって対処するとしても、様々な発話スタイルに対応したデータを収集することは困難が予想される。このため、感情音声や対話音声などを生成するには効率が悪いと考えられる。

これに対して、パラメトリックな音声合成手法の代表例であるフィルタ型の音声合成では、スペクトル包絡と微細構造を (近似的に) 分離して扱う。そのため、F_0 は任意に変化させられ、フィルタ特性を比較的少数のパラメータで制御して音声スペクトルを生成するため、加工性が高いと期待されている。フィルタ特性を与えるパラメータとして LPC、PARCOR、LSP やケプストラムなどが提案されており、それぞれ比較的品質が高い音声分析合成方式が確立されている。しかし、これらの方法ではフィルタパラメータと音声の声質やスタイルの間の関係が一意には定まらないため、音声の性質を自在に制御することは容易ではない。

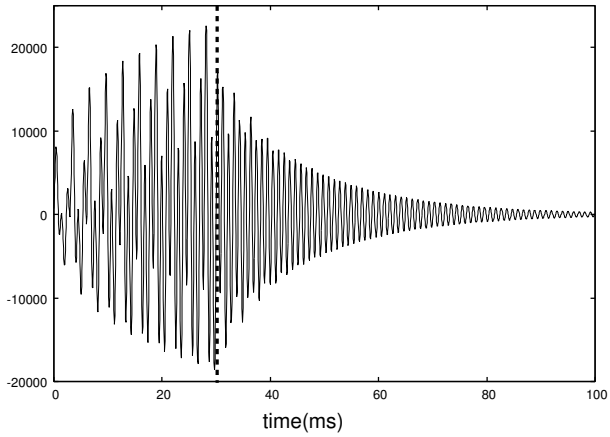


図 1　LPC フィルタ出力の時間特性の例
Fig. 1　An example of temporal characteristics of LPC filter

古典的なフォルマント合成方式も広義のフィルタ方式であるが、フォルマント周波数を直接制御パラメータとする。フィルタ係数とは異なり、フォルマント周波数と声質等の関連については音声学で盛んに研究されており、加工に際してその知識を適用することができる可能性がある。しかし、現状ではフォルマント音声合成に適した優れた分析方法が存在しない。

2.2 フィルタ方式音声合成の高 Q 値問題

フィルタ型の音声合成は、音声分析合成系として使われる場合はかなり高い品質を示す。しかし、分析時とは異なる F_0 で駆動した場合など、一般に波形接続型音声合成に比べ音声品質が低い。その一因として、われわれはフィルタの利得特性と時間特性に注目する。

全極型フィルタによる音声分析合成方式 (LPC 系) における有声音の分析合成について考察しよう。一般に音声スペクトル包絡の山と谷の間には数十 dB に達する大きなレベル差 (スペクトルダイナミックレンジ) があることが多く、これを少数のパラメータを用いたモデルで表現するために、十数次のような比較的次数が低い全極型フィルタを用いる。全極型フィルタは多重共振系であるが、このような理由によっておのおのの極の共振特性の Q 値は、実際の声道の特性よりも大きな値をとる傾向がある。

このような周波数特性のフィルタの時間特性は、共振周波数の信号成分に対して Q 値にほぼ比例した利得が生じるとともに、Q 値にほぼ比例した時定数で出力振幅が立上り、減衰する。アクセント (ピッチ) を制御して音声を合成するような場合を考えると、分析時と異なる F_0 で全極型フィルタを駆動し、たまたま駆動音源信号の倍音成分が高 Q 値の共振周波数に一致した場合などには、出力振幅の立ち上りにも減衰 (立ち下がり) にも時間がかかり、その結果として合成音声の時間制御特性が悪くなる。そして、このような音が後続の音声に重畳することで、エコーが掛かっているような印象の「歯切れの悪い」音になる一因となっている可能性がある。

図 1 は、ある音声データにおいて、音素/o/に該当する区間

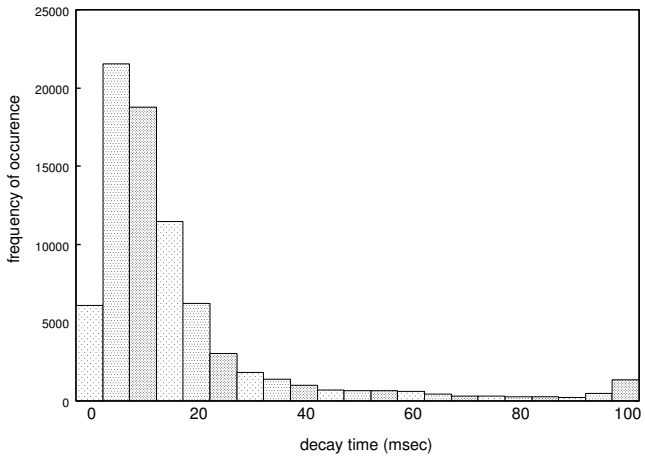


図 2　LPC フィルタ出力の時間特性の傾向
Fig. 2　Temporal characteristics of LPC filters

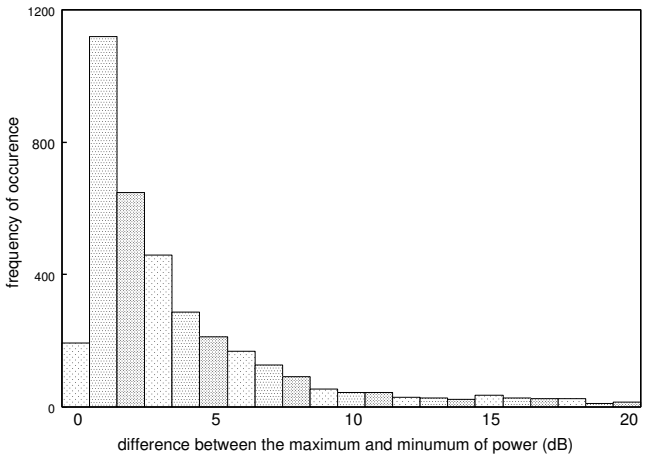


図 3　LPC フィルタ出力の利得特性の傾向
Fig. 3　Gain characteristics of LPC filters

を LPC 分析して得た全極型フィルタに、1 フレーム分の長さ (30msec) のインパルス列 (有声音駆動に相当) を入力したときの出力波形である。入力に対して出力振幅は増大を続ける (定常状態に達するまでに時間が掛かる) とともに、入力が終了した後も数十 msec にわたり出力が持続している。

また、フィルタでは出力信号の利得が Q 値に比例するため、その利得は駆動音原信号のピッチ周波数によって大きく変動する。このような現象のため、フィルタ型音声合成では合成音声のパワーを制御しにくい。

これらの問題は決して特殊な状況ではなく、LPC 系においてはしばしば起こりうる。実験的にそれを示すために、ある程度長い (1 分程度) 音声を用意し、LPC 系で分析合成を行った。

まず、時間制御特性を調べるための実験を行った。ピッチ周期を 0.8 倍から 1.2 倍まで 0.02 刻みで変更し、分析したフィルタに 30msec 間入力した。その後入力をせずに合成を続け、各フレーム、ピッチ周期で減衰時間を調べた。ただし、減衰時間は入力停止から合成音声のパワーが 30dB 低下するまでの時間と定義する。また、速い変化に追従するためパワーを 10msec 間の振幅の二乗和として定義した。図 1 においては、55msec が減衰時間である。そして、図 2 に減衰時間を 5ms 単位のヒストグラムで示した。分布が右に偏るほど、減衰時間が長くなりやすいと言える。

さらに、利得特性を調べるためにピッチ周波数を同様に変化させて音声全体の合成を行い、有声区間の各フレームのパワーを調べた。同一のフレームで、駆動音源のピッチ周波数を変えることでパワーが変化するが、その最大になる場合と最小になる場合のパワーの差を図 3 にヒストグラムで示した。やはり分布が右へ偏るほど、利得の変化が大きいと言える。

これらの結果より、LPC フィルタにおいて、時間特性の問題や利得が大きく変化する現象が確認できる。

LPC 系の分析合成では、原音声のピッチ周波数を用いれば比較的高い品質の分析合成音が得られるが、原音と異なるピッチ周波数で駆動すると品質が劣化する現象については、以上の考察がその一因として考えられる。

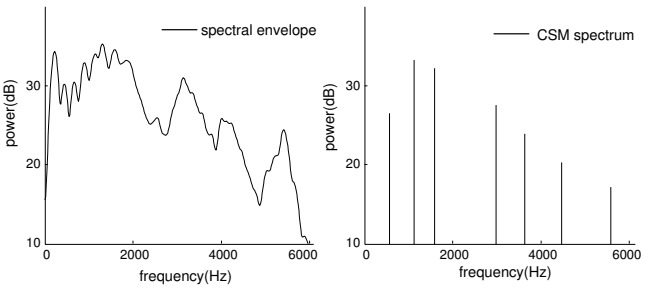


図 4　CSM 法による音声スペクトルの線スペクトル表現
Fig. 4　A spectral envelope represented by CSM with line spectra

このような問題は、有極型フィルタ (巡回型ディジタルフィルタ) の本質に根ざす問題で解消は難しい。仮にそれを改善するために Q 値を下げると、包絡の山と谷のレベル差が形成できず明瞭性の低い buzzy な印象の音が生成されてしまう。また、全極モデルでなく極零モデルを用いることはやや有望であるが、分析合成法として確立していない。対数スペクトル特性近似フィルタ (ケプストラム合成フィルタ) [7] はこの点で LPC 系より有利である可能性が有る。

2.3　複合正弦波モデル (CSM) 音声合成

CSM 法 [8] では、線スペクトルモデルに基づく音声分析法である CSM 音声分析によって、フォルマント周波数にほぼ対応する複数個の正弦波周波数 (CSM 周波数) を得る。そして、それらの周波数の正弦波の和を基本波形として、位相を基本周期ごとに 0 にリセットすることで音声を合成する。線スペクトルを広げる目的で振幅に指数関数減衰を乗じることも行われた。

これは、巡回型フィルタを用いずにパラメトリックに音声合成が行える方式なので、振幅の制御は極めて容易であるため「歯切れのよい」音声合成が期待できる。しかし、CSM 法は音声スペクトルを図 4 のように線スペクトルで近似することに相当するため、スペクトルの再現方法としては検討の余地が残っていた。

3. 複合ウェーブレットモデル (CWM)

3.1 基本波形の接続による音声合成

以上の方式における有声音の合成を、ピッチ周期のインパルス列を入力したある線形系と考えて、その線形系のインパルス応答により整理すると、PSOLA 方式あるいは波形接続型では音声波形のピッチ周期波形そのものをインパルス応答とするのに対し、全極型フィルタでは推定されたスペクトル包絡の逆 Fourier 変換が対応する。

これを基本波形の繰り返しとして解釈し比較すると、波形接続型におけるピッチ周期波形は、これを構成する基本正弦波とその多数の高調正弦波の重ねあわせととらえられるが、これら個々の振幅位相はピッチそのものに大きく依存するため、ピッチと独立した制御には適さない。

一方、CSM 合成においてはほぼフォルマント周波数に対応する正弦波断片が、全極型フィルタにおいては単振動 (二次系) のインパルス応答である指数型減衰正弦波が、それぞれ基本波形となっており、いずれもこれら基本波形の重ねあわせと解釈できる。これら基本波形に必要な性質は、音声のスペクトル包絡をよく近似するスペクトルをもつことである。この意味からは必ずしも巡回型フィルタの場合のような長い基本波形は必要ではなく、巡回型フィルタでは単に時間特性を悪化させる要因になっていると言える。

以上の考察から、パラメトリックでかつ時間特性が良い音声合成は、少なくとも有声音の合成においては、巡回型フィルタを用いず、スペクトル包絡の逆 Fourier 変換をピッチ周期で繰り返し、それに希望する振幅を乗じる方法が有望であることになる。

3.2 基本波形のモデル化

以上から、新しい音声分析合成方式を設計する。合成音声の基本波形を少数の扱いやすいパラメータによって表現することができれば、合成音声の声質や感情を操作するなどの加工がしやすくなる可能性がある。その要求条件には、以下の 2 点が挙げられる。

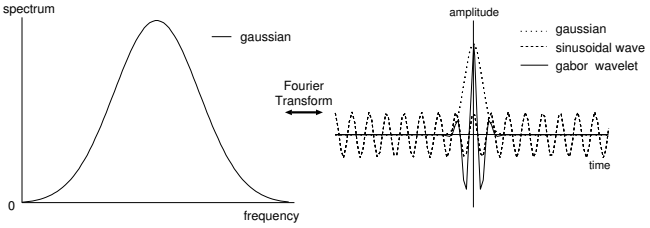
- 多様な音声を少数のパラメータで表現したい (パラメトリックな方式であること)
- 音声スペクトルの大きなダイナミックレンジを表現でき、かつ Q 値は低く抑えたい (巡回型によらない方式であること)。

である。

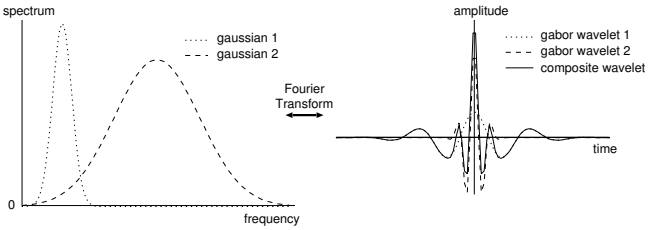
そこで、次の Fourier 変換公式に着目する。ω を周波数、t を時間、a, b, c を任意の実数とすると、

$$\mathcal{F}\left[\frac{a}{2\sqrt{b\pi}}e^{-\frac{t^2}{4b}+jct}\right] = ae^{-b(\omega-c)^2} \qquad (1)$$

が成り立つ。すなわち、周波数領域のガウス関数は、図 5(a) に示すように、時間領域ではガウス関数と正弦波の積である Gabor 関数で表される。ガウス関数は dB 尺度で見れば下に開いた放物線であり、これを共振特性と考えると Q 値を抑えつつ、かつ大きな山と谷を形成するのに都合がよい。これらの関数対は、スペクトル領域でも時間領域でも大きく拡がらない利点を持つ。これを音声のフォルマントに対応づけて考える。



(a) 単一ガウス関数 (Single gaussian) のフーリエ変換



(b) 混合ガウス関数 (Gaussian mixture) のフーリエ変換

図 5　ガウス関数の Fourier 変換対

Fig. 5　Fourier transform of Gaussian function

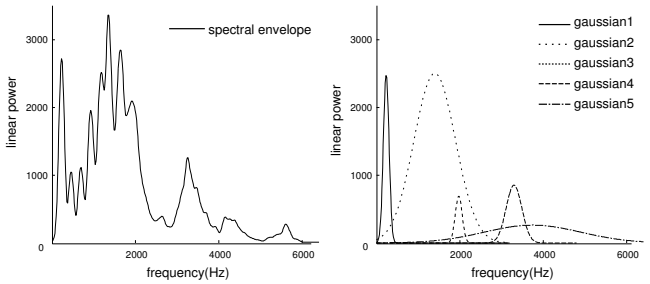


図 6　GMM による音声スペクトルの近似の例

Fig. 6　An example of spectral envelope approximated by GMM

従って、図 6 に示すように音声スペクトル包絡を混合ガウス関数モデル (GMM)(注1)で近似すれば、GMM で表されたスペクトル包絡から、基本波形は容易に生成できる。従って (振幅) スペクトル包絡を図 5(b) のようにガウス関数の重ねあわせによって近似すると、基本波形は Gabor 関数の重ね合わせである。このため、本手法を複合正弦波モデル (Composite Sinusoidal Modeling) の精神を継ぎ、正弦波の代わりに Gabor Wavelet の重ね合わせを基本波形とするという意味で、複合ウェーブレットモデル (Composite Wavelet Model) と名付ける。

3.3 GMM の EM アルゴリズムによる音声分析法

少数のガウス関数でスペクトル包絡の近似を行うと、各混合成分の平均がフォルマント周波数に、分散がフォルマントの広がりに対応することが期待でき、分析パラメータによって音声のフォルマント構造を直接操作できる可能性がある。これにより、フォルマント音声合成同様に音声学の知見を活かした声質変換の点で有利であると考えられる。また、逆に多数のガウス関数でスペクトル包絡の近似を行う場合、加工は難しくなるが

(注1)：通常は GMM は Gaussian Mixture Model の略で、混合ガウス分布密度モデルを意味し、その積分値は 1 に等しくなければならない。しかし、本稿ではスペクトル (パワースペクトルあるいは 0 位相化した振幅スペクトル) のモデルとしての混ガウス関数モデルを意味するものとする。本稿中では混乱は生じないので、同一の略語 GMM を用いる。

近似の精度がよくなり音声品質が向上することが期待できる。

Zolfaghari ら [10] は、音声スペクトルのフォルマント分析のために包絡を GMM で近似する手法を提案した。しかし、分布密度関数推定に関する EM(Expectation-Maximization) アルゴリズムがパワースペクトルのモデル化にそのまま使用できるかどうかは自明でない。その議論は亀岡ら [13] によりなされ、EM アルゴリズムと同型のアルゴリズムにより、観測したスペクトルに対するモデルスペクトルの KL 尺度 (Kullback-LeiblerL 情報量と同型の関数間の擬距離) を最小化 (あるいは極小化) することができることが示されている。筆者らはその原理に基づいて、EM アルゴリズム (に同型なアルゴリズム) に基づいて、分析フレーム単位の音声スペクトルの GMM 推定によりスペクトルパラメータを抽出する。

また、Zolfaghari らは GMM 化が包絡でなくピッチ構造に収束する場合を指摘しているので、筆者らは自己相関関数にラグ窓 [11] を掛けてフーリエ変換することにより平滑化パワースペクトルを得て用いることによりその問題を回避している。

3.4 CWM 音声分析合成手順

以上をまとめて、本稿の音声分析合成法の手順を示す。

まず分析系のアルゴリズムの一例を以下に示す。

（1） 音声信号の自己相関関数をフレーム単位に計算する。

（2） 自己相関関数にラグ窓 [11] を掛けてフーリエ変換することにより平滑化パワースペクトルを得、その平方根を取る。(これによりゼロ位相化した振幅スペクトルを得る。)

（3） EM アルゴリズムにより、これを適切な混合数 m の GMM で近似し、各ガウス関数の平均 μ_i、分散 σ_i^2、重み w_i (但し $i = 1, \cdots, m$) をフレームごとのスペクトル分析結果とする。

以上は一例であり、音声スペクトル包絡を GMM で近似する手法ならば、これに限定しない。また、有声無声判定と F_0 推定には、既存のピッチ抽出手法が利用できる。

次に、合成系の手順を示す。(有声音の場合)

（1） フレームごとのガウス関数の平均 μ_i、分散 σ_i^2、重み w_i (但し $i = 1, \cdots, m$) から、GMM のフーリエ変換に対応する Gabor 関数の重みつき和を求める。

（2） これを、ピッチ周期間隔で周期的に配置して音声合成出力とする。

また、無声音については、音声の生成モデルに基づき雑音源に基本波形を畳み込む方法やランダムな間隔のピッチを与える方法などが考えられる。

この音声合成は、non-recursive フィルタと音源パルスの畳み込みとも理解できるから、品質向上のために multi-pulse 音源と組み合わせる手法や、波高率や品質を改善するために、複数の Gabor 関数を同期させるのでなく、適度にずらして重畳する方法なども考えられる。

4. 音声合成実験

上記提案手法の有効性を確認するために、分析合成によって音声が再現されるかを確認した。また、従来法の問題点の解決に向けて、LPC 法との比較を行った。

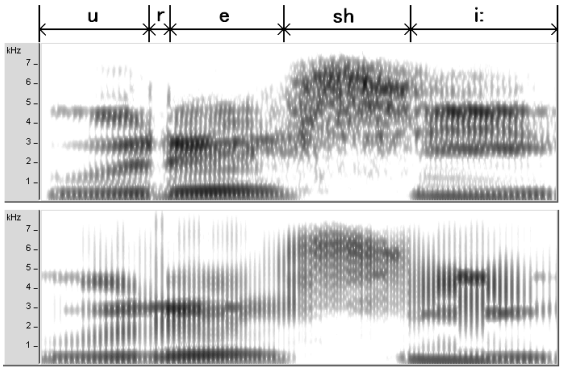


図 7 原音声 (上) および合成音声 (下) のスペクトログラム例

Fig. 7 Spectrograms of original (upper) and synthetic (lower) speech samples

4.1 実験条件

まず、3.1 節の合成法及び 3.2 節で述べた GMM による近似の動作検証のために、提案法によって音声を低次元のパラメータに分析し、パラメータから合成を行った。実験には ATR 音声データベースより 3–5 秒程度の女性話者による文音声を 5 程度選び、用いた。サンプリング周波数 16kHz、サンプルサイズ 16bit の音声に対して、ラグ窓法によるスペクトル包絡の抽出を行った。さらに、スペクトル包絡を 5 個のガウス関数の和に近似した。したがって分析パラメータは 1 フレームにつき 15 次元である。今回は、F_0 は Snack Sound Toolkit [12] 付属の F_0 抽出ツールによって抽出した。また、フレーム長 30ms、フレームシフト 10ms で分析した。

まず、ピッチ周期や分析パラメータに変更を加えず、音声を合成した。無声音については、ランダムなピッチ周期を与える方法で合成した。そして聴取による比較の他、スペクトルの比較を行った。さらに、ピッチ周期や分析パラメータの平均を 0.7 倍–1.3 倍程度に変化させ、音声を合成し、音声として破綻していないか聴取によって確認を行った。後者は、フォルマント周波数を変更したことに相当する。

提案法により時間特性が改善することを示すため、2.2 節と同様の実験を行い、時間特性と利得特性を調べた。

4.2 実験結果と考察

聴取実験によって、良好な音声が合成されることを確認したが、背景にブザー的な雑音が聴かれた。図 8 に「うれしいはずが...」の冒頭部分の原音声と提案法の合成音声のスペクトルを示す。この図から分かるように、合成音声はかなり原音声の特徴を再現できているが、基本波形をゼロ位相化しているためにエネルギーの集中が著しくなっていることがわかる。図 8 に提案法により合成される「あ」の音の一部を示す。原音声とは明らかに異なる波形を持つが、スペクトルはほぼ同じである。

ピッチ周期やフォルマント周波数を変更する試験を行ったところ、いずれの条件においても破綻することなく音声を合成することができた。

図 9 および図 10 に提案法の時間特性と利得特性を示す。図 2 および図 3 との比較より、提案法によって時間特性が改善し、かつ利得が安定したことがわかる。

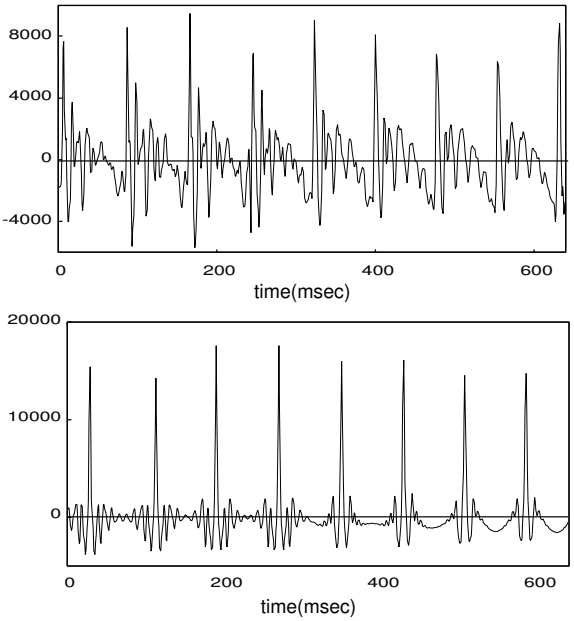


図 8　原音声 (上) および合成音声 (下) の波形例

Fig. 8　Waveforms of original (upper) and synthetic (lower) speech samples

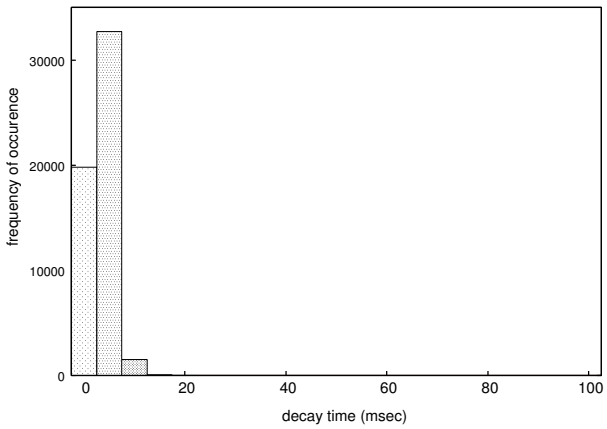


図 9　提案法の時間特性

Fig. 9　Temporal characteristics in the proposed method

5. おわりに

本稿では、従来のフィルタ型音声合成の音声品質が低下する要因として時間特性の問題があることを指摘し、それを解決する手法として複合ウェーブレットモデル (CWM) による音声の分析合成法を提案した。提案法では、音声のスペクトル包絡を GMM で近似して分析パラメータを得る。そして、この GMM の逆フーリエ変換である Gabor 関数の重ね合わせを基本波形とし、それをピッチ周期ごとに配置して有声音を合成する。ピッチ周期をランダムにすれば無声音も合成できる。

本手法の動作検証のために、音声の分析合成を行った。また、時間特性と利得の改善を実験的に確認した。

スペクトルの再現性を向上するためには、文献 [14] などの精度の高いスペクトル包絡推定方法を用いることを考えている。今後は合成音品質の改善手法の検討とともに、歌声合成、音声学的知見の適用、会話音声や感情音声の生成、HMM 音声合成

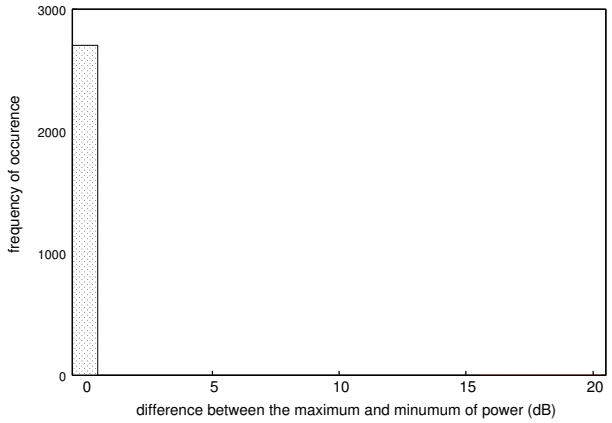


図 10　提案法の利得格差

Fig. 10　Gain difference in the proposed method

系と組み合わせた擬人化エージェントへの搭載を目標として研究を進めたい。

文　献

[1] “Galatea Project,” http://hil.t.u-tokyo.ac.jp/ galatea/index-jp.html

[2] E. Moulines, and F. Charpentier: “ Pitch-synchronous Waveform Processing Techniques for Text-to-speech Synthesis Using Diphones,” , Speech Communication, no. 9, pp. 453-467, 1985.

[3] ニック・キャンベル, アラン・ブラック: “CHATR: 自然音声波形接続型任意音声合成システム,” 信号処理学会技術報告, vol.96, no. 39, pp. 45-52, 1996.

[4] F. Itakura and S. Saito: “Analysis Synthesis Telephony Based on the Maximum Likelihood Method,” Proc. 6th Int. Congresson Acoustics, 1968.

[5] 北脇信彦, 板倉文忠, 斉藤収三: “ PARCOR 形音声分析合成系における最適符号構成,” 電子通信学会論文誌, J61-A, pp. 119-126, 1978.

[6] 管村昇, 板倉文忠: “線スペクトル対 (LSP) 音声分析合成方式による音声情報圧縮,” 電子通信学会論文誌, J64-A, pp. 599-606, 1981.

[7] 今井聖, 北村正, 竹谷博行: “2 次元ケプストラムを利用する音声分析,” 電子通信学会論文誌, J59-A, pp. 1096-1103, 1976.

[8] 嵯峨山茂樹, 板倉文忠: “複合正弦波による音声合成,” 音声研究会資料, S79-39, pp.293-300, 1979.

[9] 徳田恵一, 益子貴史, 小林隆夫, 今井聖: “動的特徴を用いた HMM からの音声パラメータ生成アルゴリズム,” 日本音響学会誌, vol.53, no.3, pp.192-200, 1997.

[10] Parham Zolfaghari, Tony Robinson, “Formant Analysis Using Mixture of Gaussans,” Proc. ICSLP 96, vol. 2, pp. 1229–1232, 1996.

[11] 嵯峨山茂樹, 古井貞熙: “ラグ窓を用いたピッチ抽出の一方法,” 電子情報通信学会全国大会予稿集, 1235, Vol. 5, p. 263, 1978.

[12] “The Snack Sound Toolkit,” http://www.speechkth.se/snack/

[13] 亀岡弘和, 西本卓也, 嵯峨山茂樹, “調波時間構造化クラスタリング (HTC) による音楽の音響特徴量同時推定,” 情報処理学会研究報告, 2005-MUS-61-12, pp. 71-78, 2005.

[14] 亀岡弘和, 小野順貴, 嵯峨山茂樹: “スペクトル包絡と調波構造の合成関数モデルによる音声分析,” 日本音響学会 2005 年秋季研究発表会講演論文集, 2-6-4, 2005.