

| | | |
|---------|--|--------|
| 氏名(本籍) | 小野 ^の 茂 ^{しげる} | (神奈川県) |
| 学位の種類 | 博士(情報科学) | |
| 学位記番号 | 情博第12号 | |
| 学位授与年月日 | 平成8年3月26日 | |
| 学位授与の要件 | 学位規則第4条第1項該当 | |
| 研究科, 専攻 | 東北大学大学院情報科学研究科(博士課程)情報基礎科学専攻 | |
| 学位論文題目 | 低ビットレート音声符号化方式に関する研究 | |
| 論文審査委員 | (主査) 東北大学教授 丸岡 章 東北大学教授 曾根 敏夫 東北大学教授 牧野 正三 | |

論文内容要旨

第1章 序 論

本章は序論であり、本研究の目的と位置付け、本研究が解決する課題を明確にすると共に本研究の内容及び本論文の構成を説明することを目的としている。

音声符号化は音声波形をできるだけ少ない情報量で忠実に表現することを目的とする技術であり、これまで電話端局装置をはじめとする経済的な通信サービスを実現するための音声多重化装置、高い周波数利用効率を必要とする移動体通信、限られた記憶容量の中に多くの通話内容を蓄えたい音声録音やボイスメールシステムなどで中心的な役割を果たしてきた。近年、移動体通信網を含むデジタル通信網の発達や DSP (Digital Signal Processor) を筆頭とする高度なデジタル信号処理を支える LSI 技術の進歩により、従来より高品質な低ビットレート音声符号化方式への期待が高まると共に、情報通信のマルチメディア化の流れを受け、その適用範囲が拡大されている。音声符号化技術に課せられる基本要件は低ビットレート化と高品質化の両立であるが、このような実システムへの適用においては、それらの追求に際して処理量並びに記憶容量の制約を満足することが要求される。

線形予測符号化は、音声生成のモデルは音声の生成理論との整合性が高く、線形予測分析という効率的なモデルパラメータの推定法が確立しているため、現在の音声符号化方式の中で非常に重要な位置を占めている。線形予測符号化において低ビットレートで高品質な再生音声を得るためには、線形予測係数で形成される線形予測合成フィルタと線形予測合成フィルタの駆動音源として動作する予測残差の高エネルギー符号化が不可欠である。線形予測合成フィルタに関しては、PARCOR 及び LSP という線形予測係数に比べて量子化感度が格段に低いパラメータが発明され、高エネルギー符号化が可能となった。しかし、一方の予測残差については、有効なパラメータ表現が難しく、予測残差を表現する有効な音源モデルの構築と効率的なモデルパラメータの推定法の開発が現状の主要課題になっている。

本研究は、線形予測符号化における予測残差の効率的な符号化を実現するための課題を解決し、線形予測符号化の高品質化と低ビットレート化を達成することを目的とする。

現在提案されている有効な音源モデルにマルチパルス音源とコード音源とがある。マルチパルス音源モデルは、音源を複数のパルスで表すもので、モデルパラメータを制御することで、広いビットレートに適用できる。しかし、モデルとしての自由度が高いためパラメータの決定が難しく、低ビットレートで高品質な再生音声を得るためには、音源パルスパラメータの探索と量子化の最適化が不可欠である。一方、コード音源は、予測残差にベクトル量子化を適用したもので、予め定められた複数の予測残差のパターンで音源をモデル化する。ベクトル量子化の能力を高めることで入力音声

波形に近い再生音声波形が得られる可能性がある。しかし、ベクトル量子化の特性を高めるには膨大な演算量と記憶容量が必要になる。したがって、コード音源の適用にあたっては、ベクトル量子化特性の向上と共に演算量と記憶容量の低減化が必須課題となっている。尚、マルチパルス音源を用いる音声符号化方式はマルチパルス音声符号化方式 (MPLP : Multipulse Excited Linear Prediction), コード音源を用いる音声符号化方式は CELP (Code Excited Linear Prediction) と呼ばれている。

本研究では、マルチパルス音源とコード音源に関する諸課題の解決を図ると共に、それぞれのモデルに基づいた低ビットレート音声符号化方式を提案している。

本論文の構成は以下のものである。第2章では、マルチパルス音源モデルの高品質化にとって必要な音源パルス探索並びにパルス量子化問題に対する最適解を提案する。第3章では、マルチパルス音声符号化の低ビットレート化の例として、一ピッチ区間に限定したマルチパルス音源モデルに適応セグメンテーションとピッチ予測の技術を適用した 2.4 kb/s の音声符号化方式を提案する。そして主観評価により、本方式が従来の LPC ボコーダーより格段に自然性の高い合成音声を生産することができることを示す。第4章では、コード音源の適用に際して問題となるベクトル量子化に要する演算量と記憶容量を低減化するための構造化ベクトル量子化について検討し、従来より少ない演算量で高い量子化能力を持つ新たなベクトル量子化法を提案する。第5章では、前章で提案した構造化ベクトル量子化法に基づいた 3.6kb/s の音声符号化方式について報告する。主観評価により、本符号化方式が 6.7kb/s VSELP (Vector Sum Excited Linear Prediction) 方式とほぼ同等の品質を提供ができることを示す。

第2章 マルチパルス音声符号化方式における高品質化に関する諸検討

マルチパルス音源に基づくマルチパルス音声符号化方式の高品質化と低ビットレート化のためには、できるだけ少ないパルス数で入力音声波形をできるだけ正確に近似する再生音声波形を生産するパラメータを決定する必要がある。本章では、マルチパルス音声符号化方式の高品質化と低ビットレート化を実現するために従来方式が抱える諸問題点を明らかにし、それらに対する解を与えている。解決した問題は以下の通りである。

・音源パルス探索の最適化

マルチパルス音声符号化において音源パルス探索問題が難しい理由は、線形予測合成フィルタのインパルス応答で形成される基底インパルス応答系列が直交系をなしていないこと、及び探索問題が音源パルスの位置パラメータに関して非線形になっていることである。

本章では、音源パルス探索問題に対して、基底インパルス応答系列を Schmit の直交化により逐次直交系列に変換しながら音源パルス探索を行なう方法を提案している。本方法は、逐次探索という枠組では最適な音源パルス探索アルゴリズムになっている。

フレーム長を160サンプル (20ms) として、フレーム当たりのパルス数を変化させてセグメンタル SNR を測定した結果、本方式は、従来方式が32パルスで達成する特性を21パルスで、20パルスで達成する特性を16パルスで実現している。また、セグメンタル SNR の改善度は、16パルスで1.3dB、32パルスで4.1dBである。

・均一ビット割り当てにおけるパルス振幅量子化歪みの低減化

線形予測合成フィルタのインパルス応答で形成される基底インパルス応答系列は互いに直交してない。よって、基底インパルス応答の係数であるパルス振幅の大きさだけに着目する量子化構成では、再生音声波形の領域から見た最適化を図ることができないばかりか、インパルス応答間の相関によりパルス振幅の量子化誤差が増幅されるという問題が生ずる。また、このようなパルス振幅の値だけに着目する構成では、量子化誤差を重み付けフィルタ等を介して量子化過程にフィードバックする機構がないため、聴覚のマスクング特性を利用した量子化誤差の制御ができないという欠点がある。

本章では、基底インパルス応答系列を正規直交系列に変換し、パルス振幅直接量子化するのではなく、パルス振幅を前記正規直交系列上の係数に変換して量子化するという方法を提案している。

本方式は、再生音声波形領域から見た最適量子化を可能にすると共に、インパルス応答間の相関によりパルス振幅の量子化誤差が増幅されるという問題を解決する。また、重みフィルタの掛かった再生音声波形領域で量子化するため、聴覚のマスクング特性を利用した量子化誤差の制御も可能になる。

実験によれば、前節で提案した音源パルス探索法と本節で提案した直交係数上の量子化を組合せることで、従来方式に対して、4ビット/パルスの場合に、16パルスで1.1dB、32パルスで2.3dB、5ビット/パルスの場合に、16パルスで1.1dB、32パルスで2.9dBの改善が得られている。

・音源パルスへの適応ビット割り当て

マルチパルス音源における量子化の最適化を図るには、本来、線形予測残差への適応的なビット割り当てを行う必要がある。

本章では、音源パルスの適応ビット割り当て問題を解決するために、インパルス応答系列の共分散行列の固有ベクトルを成分に持つユニタリー変換を線形予測残差に適用する方式を提案している。本方式は、線形予測残差の直交変換符号化と同等であり、マルチパルス音声符号化における音源パルスの最適探索並びに最適量子化を同時に解決する枠組となっている。

実験により、前節で提案した逐次直交化による音源パルス探索と直交係数上の量子化を適用したものに比べて、1 bit/sample で約0.2dB、1.2bit/sample で約1.1dB、1.5bit/sample で約1.7dBの改善が得られ、音源パルス探索と量子化の最適問題を解決した本手法が全体的に高い特性を示すことを確認している。

・基底インパルス応答の最適化

マルチパルス音声符号化で用いられる線形予測合成フィルタは、入力音声波形を直接線形予測分析することによって得られる。マルチパルス音声符号化方式は、再生音声波形を位相の異なる合成フィルタのインパルス応答の線形結合で表す。したがって、本符号化方式の特性を更に改善するためには、音源パルスモデルに適したインパルス応答を形成することが望まれる。

本章では、音源パルス系列に適した合成フィルタの推定として、入力音声波形から音源パルスの影響を除去したサンプルを用いて合成フィルタの係数推定を行なう方法を提案している。具体的には、入力音声波形の内、音源パルスを含む区間を除去して線形予測分析を行なう。

実験では、本方式により、パルス数に依らず有声区間のセグメンタル SNR を一様に1.5dB程度改善できることが示されている。

第3章 適応セグメンテーションによる2.4kb/s ピッチ予測マルチパルス音声符号化方式

従来2.4kb/s以下のビットレートではLPCボコーダーが適用されてきた。LPCボコーダーの音源モデルは、有声音ではピッチ間隔で立つ等振幅のパルス列、無声音では白色雑音という単純なものになっているため生成される合成音声は明瞭性はあるものの自然性や話者性に乏しいになっている。

本研究では、マルチパルス音源モデルをベースに高品質な2.4kb/s音声符号化を提案している。本方式の特徴は以下の通りである。

- ・音声波形のような時変な情報源を自由度の少ないモデルパラメータで高品質に符号化することを可能にするため、符号化パラメータの更新速度を入力音声波形の動特性に適應させて制御させる適応セグメンテーションを採用
- ・セグメント境界における周期構造の接続を滑らかにするため、セグメント内におけるパラメータの一次変動を許容したピッチ予測合成フィルタモデルを導入
- ・音源モデルに要するビット数を削減するため、セグメント内の音源を一ピッチ区間に限定したマルチパルス音源モデルを導入

主観評価実験により、本方式の合成音声はLPCボコーダー型の合成音声に比べて格段に自然性が高く高品質であることが示されている。

第4章 構造化ベクトル量子化

ベクトル量子化を適用するコード音源モデルにおいては、コードベクトルの探索と格納に要する膨大な演算量と記憶容量の低減化が主要課題となっている。

ベクトル量子化に要する演算量並びに記憶容量の削減については、ベクトル量子化の構造化を中心に種々の検討がなされている CELP 方式における代表的な構造化ベクトル量子化にベクター・サム (Vector Sum) ベクトル量子化、共

共役構造 (Conjugate) ベクトル量子化, そして代数コード (Algebraic codes) ベクトル量子化がある。ベクター・サムベクトル量子化はコードベクトルを量子化ビット数分の基底ベクトルの加減算で表現することに特徴がある。この構造では、非構造のベクトル量子化法を使った場合と比較して、演算量とメモリ量が大幅に削減できるが量子化特性も大きく劣化するという問題がある。共役構造ベクトル量子化はコードベクトルを2種類のベクトル量子化器に属するコードベクトル間の加減算で表現するもので、記憶容量はある程度削減できるが演算量を削減するためには予備選択が必須である。このため、量子化特性及び演算量の削減の程度は予備選択の巧拙に大きく依存するという問題がある。特に、演算量を削減するために強力な予備選択を施すと特性が急速に劣化する傾向にある。一方、代数コードベクトル量子化は、コードベクトルを代数的構造を持った0と±1の組合せによって表現するものである。演算量並びに記憶容量は大幅に低減できるが、量子化特性は実音声資料に基づいて設計したベクトル量子化器に比べて大幅に劣化するという問題がある。このように、CELP方式におけるベクトル量子化においては、実音声資料に基づいて設計される構成であって、予備選択の巧拙に関わらず演算量が削減でき且つ優れた量子化特性を持つものが望まれる。

本章では、本課題を解決する一手法として、木構造ラティスと基本ラティス構造を組み合わせた新しい構造化ベクトル量子化を提案している。本構造は、基本ラティス構造の生成ベクトル数を操作することで演算量、記憶容量、並びに量子化特性が自由に制御できるという特徴を併せ持つ。また本章では、本構造化に適した設計アルゴリズム並びに基本ラティス構造の性質を利用した効率的な探索アルゴリズムについても論じている。

正規白色雑音を用いた実験では、本提案方式が、従来のベクタ・サムベクトル量子化に比べて、半分以下の処理量でそれを上回る量子化特性を示すことが確認されている。また、本構造の基本ラティスにおける整数係数を、コードブック探索後に予め設計された実数係数で置き換えることで、量子化歪みの大幅な低減が可能であることも示されている。

第4章 構造化ベクトル量子化に基づく3.6kb/s音声符号化方式

本章では、CELP方式に基づく3.6kbps音声符号化方式について報告している。本章では、前記構造化ベクトル量子化の音源コードブックへの適用の他に、本符号化方式の低ビットレート化と高品質化のために検討した諸結果についても併せて報告している。本符号化方式の特徴は以下の通りである。

- 適応コードブックのラグを差分符号化：フレーム全体に渡って最適化するラグトラッキング法を導入。従来の60%のビットレートでセグメンタルSNRの劣化を0.1dB以下に抑えている。
- モード制御：当該フレームの入力音声の特性に応じて、適応コードブックのオン/オフとゲインコードブックの切替えを行なう構成を提案。切替えを行わない場合より、約0.34dBのセグメンタルSNRの改善が可能となっている。
- 音源コードブックに木構造ラティスに基づく構造化を適用。従来のベクタ・サムベクトル量子化より、6割以下の演算量で0.57dB高いセグメンタルSNRが実現されている。
- 音源コードブックにコムフィルタを導入。構造化ベクトル量子化と組み合わせることで、低演算量で高品質化が図られている。コムフィルタの導入によりセグメンタルSNRが0.84dB向上している。

本方式のMOS値を評価した結果、6.7kb/sのVSELP方式とほぼ同等の主観特性を持つことが明らかとなっている。

第6章 結論

本章は本論文の結論であり、本研究によって得られた主要な成果を総括している。

審査結果の要旨

移動体通信等の分野では、音声波形を少ない情報量で忠実に符号化することが必須となる。しかし、低ビットレート化と高品質化という相矛盾する要求を同時に満足する方式は、これまでに確立されているとは言い難い。本論文は、線形予測音声符号化における予測残差の効率的な符号化方式を提案し、その有用性を示したもので、全編6章よりなる。

第1章は序論である。第2章では、マルチパルス音声符号化方式に関して、音源パルス探索とパルス振幅の量子化を効率よく実行する方式を提案している。この方式では、音源パルスの探索は、線形予測合成フィルタのインパルス応答系列を Gram-Schmit の直交化により正規直交系列に変換したものに基づいて実行し、残差信号の各量子化レベルに対するビット系列の割当ての問題を、行列の固有値を求める問題に帰着させ、符号化の効率化を図っている。これは優れた成果である。

第3章では、ピッチ予測マルチパルス音声符号化方式に基づき、音声波形の動特性に応じた長さ可変のセグメント毎にパラメータの更新を行う符号化方式を提案している。また、主観評価実験を行い、本方式が従来の LPC ボコーダよりも優れているとの評価結果を得ている。

第4章では、ベクトル量子化のための演算量と記憶容量の低減化を図るため、木構造ラティスと基本ラティスから構成される新しい構造化ベクトル量子化を提案している。この方式は、情報源のもつ確率的な偏りを有効に利用して量子化できるという利点に加えて、基本ラティス構造のサイズを変化させることにより、演算量、記憶容量、および量子化特性を制御できるという優れた特徴をもつ方式である。

第5章では、第4章で提案した構造化ベクトル量子化に基づき、音声符号化器のシミュレータを構成して主観評価実験を行い、この方式に基づく 3.6kb/s の符号化器が、従来の 6.7kb/s のベクター・サムベクトル量子化に基づくものと同程度と評価されることを明らかにしている。

第6章は結論である。

以上要するに本論文は、線形予測音声符号化の代表的な音源モデルであるマルチパルス音源とコード音源に対して、低ビットレートで高品質の符号化方式を提案するとともに、シミュレーションによりその有用性を示したもので、音声情報科学の発展に寄与するところが少なくない。

よって、本論文は博士（情報科学）の学位論文として合格と認める。