

論文の内容の要旨

論文題目	残差のスパース化に着目したロスレス音声コーデック
学位 申請者	峰尾 太陽

音と映像を含んだデジタルメディアコンテンツの高級化に伴い、そのデータの容量は増大を続けている。これらのデータを劣化なく圧縮する技術は、計算機やネットワーク資源を節約できるため産業からの需要が大きい。mp3, Ogg, Opusなどのロシー（非可逆）圧縮コーデックの研究開発に伴い、音声コンテンツを身近に体験できるようになった。その一方で、音声の記録、編集、高品質コンテンツ向け配信用途を主として音声データを忠実に記録することも求められている。音声データを劣化なく圧縮するシステムとしてロスレス（可逆）音声コーデックが知られており、このコーデックでは、予測符号化の枠組みで実装されることが一般的である。予測符号では、音声データに対して予測モデルを適用し、予測とのずれを残差信号として出力し、この残差信号に対してエントロピー符号を適用し、短いビット列を割り当て圧縮データとして出力する。ここで、予測が正確であれば残差がスパースとなり音声データを小さく圧縮できるという枠組みになっている。本研究では、この残差のスパース性に着目したロスレス音声コーデックを研究対象としている。

音声データを記録・再生する用途において、コーデックの圧縮率とデコード（展開）速度の改善は重要な課題である。このため、本論文では、理論的な提案と、実装時の処理速度に着目して議論を展開している。

第2章では多岐にわたる音声信号処理研究の背景を述べており、歴史的な経緯、音声信号処理の重要性、今後の発展的な話題に関する研究の方向性が網羅的に調べられており、本論文の立ち位置を確認できるようにしている。

第3章では、符号を行う上で残差がどのような分布を表しているかに着目し、再帰的Golomb-Rice(Recursive Golomb-Rice: RGR)符号の最適なパラメータ値を導出している。予測モデルを構築する際にどのような残差が分布として残るかはその符号性能に大きな影響を与える。例えばGolomb-Rice符号は残差がLaplace分布に従うときに最も性能が高くなる。このため符号をどのように設計するかは残差分布に強く依存する。本章では、RGR符号と通常のGolomb-Rice符号との比較を行い、提案しているパラメータ決定手法を伴ったRGR符号が、通常のGolomb-Rice符号より優れていることを示すことに成功した。

第4章では、予測モデルとして適応フィルタを取り上げ、これに関する議論を行っている。ロスレス音声圧縮における適応フィルタとしてSign Algorithm (SA)がよく用いられるが、一般的な適応アルゴリズムよりも収束性能が悪いことが問題点である。本章では、SAの収束性の悪さが、勾配情報の悪条件に起因するものとし、これを改善するために自然勾配(Natural Gradient)法を導入した。ここではSA法を出発点として、自然勾配法の導入した学習アルゴリズムを提案した上で、学習係数を適応的に最適化する Normalized Natural Gradient SA (NNGSA)法を提案している。また自然勾配法においてはフィッシャー情報行列の逆行列を求める操作が必要となるが、この計算量を低減させるために、音声入力に自己回帰モデルを仮定することで、自然勾配法の計算負荷を低減したモデルを計算機実装している。計算機実験ではトイプロブレムだけでなく、実際の音声信号にも適用し、NNGSAが SAよりも収束性能が高いことを示した。また、NNGSAを組み込んだコーデックNARU(Natural-gradient Auto-Regressive Unlossy audio compressor)を公開し、FLAC, TTA, WavPack などの既存コーデックよりも高い圧縮率を示すことに成功した。

第5章では、もう一つの予測符号化である線形予測符号化の拡張を提案している。提案したモデルでは線形予測符号化を多段かつ階層的に構成した予測モデルである。この構成は畳み込み層からなるResNetと同等の構成を有する。提案モデルは従来の1層のみからなる線形予測符号化よりも予測精度が高く、係数の学習は層ごとに高速に行うことができる。提案モデルを組み込んだコーデックLINNEを実装し、FLAC, TTA, WavPack, MPEG4-ALSよりも高い圧縮率を、TTA, WavPack, MPEG4-ALSよりも高速なデコード速度を示すことを実験により示した。この結果を第3章で提案しているRGR符号を適用し、自動的なパラメータ推定を行う符号化を提案している。従来のRGR符号では、音声サンプル単位でパラメータを設定するという適応をしていたが、本論文で提案する手法では一定の音声フレーム単位で適切にパラメータを設定しているため計算負荷を減らすことに貢献している。このコーデックは LINNE(Linear-predictive Neural Net Encoder)として計算機実装され公開されている。

上記の予測モデルと符号の成果を比較検討し、研究成果となるロスレス音声コーデックの構成を示す。現在LINNEを基本にした構成を提案しているが、主な課題としてエンコード速度の改善と処理遅延(レイテンシ)が挙げられる。本課題を解決することで、業界標準のMPEG4-ALSやFLACに代わる標準コーデックとして本研究成果が採用されることが期待できる。

論文審査の結果の要旨

学位申請者氏名	峰尾 太陽	
審査委員主査	庄野 逸	
委員	柳井 啓司	
委員	西野 哲朗	
委員	羽田 陽一	
委員	中鹿 亘	
委員		印*
委員		印*

(*自筆署名の場合に限り、押印省略可)

本研究論文では、ロスレス音声圧縮を対象に、音声信号予測と符号化方法に関して、残差信号のスパース性を中心にして議論を行っている。音と映像を含んだデジタルメディアコンテンツの高級化に伴い、そのデータ容量は増大を続けており、データを劣化なく圧縮する技術は、計算機やネットワーク資源を節約できるため産業からの需要が大きい。例えば音声信号を用いた高品質配信、記録、編集においてロスレス音声圧縮は、その存在意義を増している。ロスレス音声圧縮において重要な操作として、音声信号の特性に応じた予測と、予測残差に応じた符号化が挙げられる。本研究論文は、予測残差のスパース性に基づいて圧縮率と展開速度性能の向上を計算量と実装の両面から追求した論文構成となっている。

第4章では、残差をスパースにする性質を持つ適応アルゴリズムである Sign Algorithm (SA) を出発点として、これの収束性を改善する手法を提案している。トイプロブレムによる観察結果から、SAの収束性の悪さが勾配の悪条件に起因するものであることを仮説とし、SAに適応重み付き自然勾配 (Natural Gradient) 法を導入した予測モデル (Normalized Natural Gradient SA: NNGSA) を導出、提案している。計算機実験では、従来モデル (WavPack, TTA, FLAC) との十分な比較実験を行っており、NNGSA法が同等以上の性能を持つことを示している。

第 5 章では、符号化器の階層性と、それに対応した符号手法の改善についての議論を行っている。ここでの提案モデルは線形予測符号化を多段かつ階層的に構成した予測モデルであり、これは畳み込み構造とスキップ接続を用いた ResNet と呼ばれるニューラルネットワークと同様の構造を持つ。この階層的な残差予測を再帰的な Golomb-Rice 符号と組み合わせることで、高速かつ良好な圧縮率を持つ圧縮手法を提案している。こちらにも計算機実装を行った上で、従来の手法との十分な性能比較を行った上で良好な性質を持つことを示している。

以上のように、本論文は、ロッシェー音声圧縮を前提とした符号化方式を提案しており、今後の音声信号圧縮技術に関して大きく寄与することが期待できる。したがって、本論文は博士(工学)の学位論文として十分な価値を有するものであると認める。