

Title	Audio Processing Preserving Acoustic Naturalness
Author(s)	Tachibana, Ryuki
Citation	
Issue Date	
Text Version	ETD
URL	<a href="http://hdl.handle.net/11094/215">http://hdl.handle.net/11094/215</a>
DOI	
rights	
Note	

*Osaka University Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏名	立花隆輝
博士の専攻分野の名称	博士(工学)
学位記番号	第 21579 号
学位授与年月日	平成 19 年 9 月 26 日
学位授与の要件	学位規則第 4 条第 1 項該当 工学研究科電気電子情報工学専攻
学位論文名	Audio Processing Preserving Acoustic Naturalness (人にとって自然な音響処理に関する研究)
論文審査委員	(主査) 教授 馬場口 登  (副査) 教授 小牧 省三    教授 滝根 哲哉    教授 北山 研一 教授 三瓶 政一    教授 井上 恭    教授 河崎善一郎 教授 鷺尾 隆    教授 溝口理一郎

### 論文内容の要旨

本論文は、筆者が大阪大学大学院工学研究科電気電子情報工学専攻在学中、及び日本 IBM 東京基礎研究所在籍中に研究開発を行った音響処理に関する研究成果をまとめたもので、以下の 6 章より構成した。

第 1 章は序論であり、本論文の背景となる技術分野に関して現状を述べ、本研究の目的を明らかにした。

第 2 章では、本論文で提案する音声合成と音響電子透かしという 2 種類の音響処理の自動化技術の概要について述べた。それらの技術がより大きなシステムの中でどのように位置づけられ利用されるか様々な利用状況を想定することで、技術に求められる特性を明確化した。また、本論文で実現する音響処理技術と既存研究との差異を示し、本論文の位置づけを明確化した。

第 3 章では、話者の音声のみを入力としてほぼ自動的にすべての統計モデルを全自動構築可能な音声合成システムについて述べた。統計モデルの学習に必要な各種情報を音響的処理と言語的処理を組み合わせ漸次的に求めていく手順について詳細に記述した。特に重要なアクセントの自動認識について、複数の学習コーパスやモデルを組み合わせ高精度の認識精度を実現する手法について述べた。システム全体の音質と、各要素の精度について実験によって有効性を検証した。

第 4 章では、自然な音質をほぼ保ち埋め込みが可能で、品質が劣化した後にも検出することができる音響電子透かしのアルゴリズムについて述べた。複数の擬似乱数系列を検出に併用してこのアルゴリズムの伸縮耐性をさらに向上する手法についても述べた。このアルゴリズムによって埋め込むことができるデータ量についても考察を行った。

第 5 章では、多様な状況の音響データに対して音響電子透かしの応用を拡張した。まず、圧縮された音響データに対しても埋め込み・検出が可能であることを示した。次に、アナログ音響や生演奏へと応用を拡張するための複数のシステム構成方法を述べ、それらの利点と欠点を比較した。その一つである **Sonic Watermarking** については使用シナリオや問題点などについて特に詳しく検討を行い、提案手法の有効性を音質と耐性の側面から実験によって確認した。

第 6 章は結論であり、本研究で得られた成果を総括した。

各実験の結果によって、ある範囲の入力に対して、目標としたレベルの音質や精度、耐性を達成することが可能で

あることが検証された。音声合成では想定分野のテキストの合成音声に対して優れた音質が達成されたことが主観評価により検証された。音響電子透かしにおいてはデジタル音響データに対して人間の聴覚では判別不可能なレベルの音質が達成できた。

## 論文審査の結果の要旨

音声は人間にとって最も親和性の高いコミュニケーションメディアであり、音声・音響の処理はメディア処理における中核の一つである。音声合成は、テキストデータから音声信号を生成する処理であり、いかに人間の発話に近い自然な音声を生成するかが重要な課題である。一方、音響電子透かしは音楽コンテンツ等の流通過程での著作権保護に有用な技術であり、いかに原音響信号の品質を落とすことなく透かし信号を埋め込むかが課題となっている。本論文は、人間に対する聴覚的な自然さを可能な限り保存するという指針の下で、音声合成、および音響電子透かしの各方式を取り上げ、その設計、実現、実験的考察をまとめたものである。主たる研究成果を要約すると以下の通りとなる。

(1)話者音声のみを入力として、音響・言語的統計モデルを自動学習可能な音声合成システムを実現するとともに、その要素技術である日本語アクセントの認識手法を具体化している。アクセント句境界検出、アクセント型認識実験では、各々86.2%、92.7%の正解率を得、従来手法より優れていることを検証している。また、合成音声の主観評価実験では、5段階評価で4.0以上となり、合成音声为人間の音声品質に近いことが示唆されている。

(2)時間-周波数領域での2次元擬似乱数列に従い音響信号の振幅を変化させ、電子透かし信号を埋め込む独創的な手法を考案し、ロバスト性と音質の双方の要件を満たすことを確かめている。具体的には、音質に影響がない程度の埋め込み量を対象にMP3 (MPEG1 Audio Layer3) 圧縮、エコー付加、ノイズ付加に対して十分なメッセージ検出率を達成できることを実験より明らかにしている。さらに、この手法の潜在的ペイロード (埋込容量) について理論的に解析し、音楽コンテンツに対し、3.3~13.3 bit/secであることを示し、著作権情報を埋め込むには実用上十分であることを指摘している。

(3)多様な音響データに適用可能とするための音響電子透かしの拡張手法を与えている。まず、AAC (MPEG2 Advanced Audio Coding) 圧縮を対象に、非圧縮領域で埋め込んだ電子透かしが、圧縮・非圧縮領域のいずれにおいても80%以上で検出されることを確認している。さらにアナログ音響へ拡張するために、空気中で原音響信号と透かし信号を掛け合わせるソニック透かし法を新たに考案し、音質の主観評価並びにMP3圧縮に対するロバスト性とも十分な特性が得られることを実証している。

以上のように本論文は音声合成、音響電子透かしの高機能化に多くの貴重な知見を与えており、情報通信工学、特に音響メディア工学の発展に寄与するところが大きい。よって本論文は博士論文として価値あるものと認める。