

研究課題別事後評価結果

1. 研究課題名

聴覚の情景分析に基づく音声・音響処理システム

2. 研究代表者名及び主たる研究参加者名（研究機関名・職名は研究参加期間終了時点）

研究代表者	河原 英紀	和歌山大学システム工学部 教授
主たる研究参加者	入野 俊夫	和歌山大学システム工学部 教授
	板倉 文忠	名古屋大学大学院工学研究科 教授
	中川 聖一	豊橋技術科学大学情報工学系 教授
	峯松 信明	東京大学大学院情報理工学系研究科 助教授
	鹿野 清宏	奈良先端科学技術大学院大学情報科学研究科 教授
	赤木 正人	北陸先端科学技術大学院大学情報科学研究科 教授
	嵯峨山 茂樹	東京大学大学院情報理工学系研究科 教授

3. 研究内容及び成果

本プロジェクトは、脳における聴覚情報の表現と処理の数理的な本質を明らかにすることを通じて、工学的な手段による聴覚脳の実現を目指し、聴覚の本質的な理解に基づく「初期聴覚系の計算理論」と「高品質音声変換合成システムSTRAIGHT」を生み出した。これらの成果および派生した様々な要素技術は、プロジェクトの初期段階から、音声・音響処理の様々な応用分野に大きな波及効果を与えてきた。これらの成果および波及効果は、より広い枠組みの下で本格的に「聴覚脳」研究を進めていくための確固とした基盤を築くものである。

本研究の最大の成果は、聴覚の情報表現に基づく音声分析変換合成システムSTRAIGHT である。STRAIGHT は、1939 年に発明されたVOCODER の基本構造を踏襲しながらも、本プロジェクトにおける様々な発明を取り込むことを通じて、人間の実際の話し声に匹敵する品質と自然性を有する加工音声を作成することの出来る画期的なシステムとなった。それらSTRAIGHT 改良の核となる発明は、周波数領域での不動点に基づく基本周波数の高精度抽出法、時間領域での不動点に基づく音響的イベントの高精度抽出法、時間軸の非線形伸縮に基づく音源の非周期性指標の抽出ならびに再現方法等である。これらの発明を評価するために、共通の基礎資料の基盤として、音声波形と声門の開閉状況を反映するEGG (Electro Glottograph) 信号とを同時記録し、有声/無声情報を付記した音源情報評価用データベースを構築した。このデータベースは、本プロジェクトにおける他の音源情報抽出法の発明と評価においても活用された。また、STRAIGHT に関しては、組織的な聴覚心理実験およびDRT (Dynamic Rhyme Test : 明瞭度試験の一種) により実装に関わるパラメタの最適化を行ったことも最終的な品質の改善に大きく貢献した。

STRAIGHT は、高い品質を保ったまま音声信号を聴覚的に意味のある独立な成分に分解し、それぞれの成分を変形（変換）した後に再合成することを可能にする方法である。このSTRAIGHT は、聴覚情報処理の本質の追求を回避して音声波形の復元のみを品質化の指針としてきたこれまでの通念に対する強力な反証となった。このように品質面での問題が克服可能であることが明らかにされた結果、VOCODER 型の音声処理技術が本来有している加工の柔軟性への関心が改めて喚起され、多くの関連技術開発が開花する状況を生み出した。本プロジェクトにおいても、話者変換、高能率符号化、コーパスベース音声合成を初めとする様々な応用展開が行われた。また、STRAIGHT における上記の発明は、停滞感のあった基本周波数等の音源情報抽出の分野を、様々な高精度のアルゴリズムが提案される活発な研究領域へと変化させた。さらに、分解されたそれぞれの成分が、これまでに音声知覚の分野で蓄積されてきた知見に基づく様々な操作に容易に対応付けることができる見通しの良いものだったため、STRAIGHT は、国内外の音声知覚研究のデファクトスタンダードとなりつつある。プロジェクトの最終段階で実証されたSTRAIGHT に基づく音声モーフィング技術は、応用の裾野を更に大きく広げるとともに、これまで研究が困難であった非言語・パラ言語情報の強力な研究手段として、「組織的ダウングレーディング」と呼ぶ方法論を実施するための技術的基盤を提供する。

本研究で特記すべきもう一つの成果は、聴覚初期過程が、安定化wavelet-Mellin 変換を用いて音響的信号から形状とサイズを分離して抽出するという計算理論の構築である。これは、wavelet 変換と内耳での信号処理の類似性が表面的なものではなく、時間-スケール領域での最小不確定性の要請から必然的に導かれるもの

であるとしたプロジェクト初期の理解が、本プロジェクトの過程を通じて深まり、STRAIGHT から派生した不動点に基づく音源情報抽出法と一つの計算理論として結びついたものである。別の見方をすれば、生物が周期的な音を知覚的に特別扱いするのは、時間方向の不動点を利用することでMellin 変換の原点を合理的に設定できるため、発音源の形状とサイズを安定に高精度に獲得できるという生態学的利点があるからだとすることもできる。

本プロジェクトを構成する研究サブグループについて、グループ別にその研究実施の概要を説明する。

(1) 基礎アルゴリズム・計算論・知覚グループ：

プロジェクトの中核技術であるSTRAIGHT を構成する要素技術自体を研究開発するとともに、聴覚の計算理論の構築を目指した検討を進めた。

(2) 符号化・変換・音場制御・認識・合成規則グループ：

プロジェクトの中核技術として開発されたSTRAIGHT の情報表現は、聴覚的に意味のある領域での柔軟な操作を許す代償として、従来の音声・音響処理技術で用いられてきたものと比較すると遥かに冗長なものとなっていた。また、聴覚の生態学的な機能として重要な空間情報に関する表現と処理機構を欠いていた。これらの不整合を埋めて、プロジェクトの成果を音声・音響処理技術にインパクトを与える形で展開するための技術開発を行った。

(3) 音源分離・感知情報変換・オブジェクト記述グループ：

本サブグループは、STRAIGHT ならびに計算論等の本プロジェクトの成果を生理・心理を含むより広い分野の研究者を対象として展開するための応用技術の開発と、それらの分野と共通の視線で交流することができるようにするための基盤の作成を担当した。

4. 事後評価結果

4-1. 外部発表(論文、口頭発表等)、特許、研究を通じての新たな知見の取得等の研究成果の状況

本研究は、聴覚の機能を「聴覚の情景分析」として捉え、従来の常識を超える新しい聴覚情報表現方式を生み出し、これにより音声の分析、認識、変換などを実時間で自由に行える音声・音響情報処理システムを構築することを目標として開始された。そのために、研究代表者の考案したSTRAIGHTという音声変換・合成・分析システムをその中心に据え、これに多くの工夫・改良を付け加えた。これは音声認識、音声合成などの新技術を生み出すものである。さらに、一歩進んで、内耳における音の分析モデルであるgammachirpモデルを提唱し、脳の聴覚情報処理の原理に通ずる聴覚の計算理論の構築に足がかりを築いた。こうした点で本研究は高く評価できる。

このことは、STRAIGHTが音声・音響情報処理の分野において、国際標準の一つになりつつあることから明らかであり、また海外一流国際学術誌に発表された5件の論文より見て取れる。そのひとつはこの分野の一流誌であるSpeech Communication誌の1998-1999年度最優秀論文賞を受賞した。このほか、11編を数える国内誌における論文発表、国内外の9件に及ぶ招待講演、海外76件、国内104件の学会発表がある。

STRAIGHTシステムは改良を加えられて、国内海外の非営利研究機関には、要求に応じて無償で提供されているが、これが本システムが国際標準として高い評価を得ていることにつながる。また、特許としては海外1件、国内3件が成立し、引き合いも多く、いくつかの契約が成立している。これを要するに、本研究は音響、音声の基礎技術として実用性の高い高度なものを完成させるとともに、脳の聴覚原理に迫る道を開こうとしたものであり、高く評価できる。

4-2. 成果の戦略目標・科学技術への貢献

音声情報処理は、視覚情報処理と並んで外界に接する人間の情報処理の基本である。現代の情報機械において、これらは不可欠な要素であるが、人間のように音声をこなすことはいまだに難しい課題である。本研究は、従来のパラダイムを越えた新しい視点からこれに挑み、STRAIGHTシステムを完成させた。これは、音声合成、音声変換の実用システムにそのまま有効に活用できるので、情報技術に大きな貢献をしたものといえる。また、この立場から人の脳の聴覚情報処理の原理にも迫るものである。

4 - 3. その他の特記事項 (受賞歴など)

海外の最優秀論文賞の受賞、国内での第40回日本音響学会論文賞の受賞など、本研究の高い評価を示すものである。また、特許の引き合いも多く、契約の成立したのも複数あり、技術としての完成度の高さと実用性を示しており、情報産業にインパクトを与えるものである。

[<<脳を創るトップ](#)

This page updated on September 12, 2003
Copyright(C)2003 [Japan Science and Technology Corporation](#)