

「脳を創る」
平成9年度採択研究代表者

河原 英紀

(和歌山大学システム工学部 教授)

「聴覚の情景分析に基づく音声・音響処理システム」

1. 研究実施の概要

2000年度の研究は、プロジェクトの中核となる音声分析合成法STRAIGHTを母体とした様々な応用への展開と、聴覚の計算理論に結びつく情報表現の枠組みの展望の獲得により特徴付けられる。また、STRAIGHT自体は、Speech Communication誌の1998,1999年度の最優秀論文賞を受賞し音声知覚や高品質音声合成研究の一つのスタンダードの地歩を固めるに至った。脳における音声言語機能の研究に強力なツールを提供するという本プロジェクトの目標の一つが達成されようとしていることも特筆されるべき成果である。

2. 研究実施内容

概要に挙げた成果に直接関連するものを中心に説明する。当該年度の重点は、応用への展開と計算理論への模索であった。

応用への展開

本プロジェクトの中核技術である音声分析変換合成法STRAIGHTの構成とその聴覚の情景分析研究における位置付けを提案したSpeech Communication誌掲載の論文が、1998, 1999両年度を通じた最優秀論文としてEURASIP (欧州信号処理学会連合) によって表彰された。これは、本プロジェクトが国際的にもユニークな価値のあるものであるとして認められたことを意味すると同時に、STRAIGHTを言語音声知覚研究のためのツールとして様々な応用分野に展開するための強力な足場を確保したことをも意味する。本年度は、当プロジェクトの外部においてもSTRAIGHTを用いた研究の成果が出始めており、本プロジェクトの当初の目標の一つであった『聴覚の生理学的側面からの研究用の強力なツールの提供』も、これらの成果の延長上に達成可能性が見え始めている。また、音声認識、話者変換への応用も着実にSTRAIGHTに基づくアイデアの有用性を実証しつつある。さらに、聴覚と発声系の知見とモデルを総合的に検証することのできる応用課題として、歌唱の様式であるスキャット生成システムの構想を提案した。

アルゴリズム・計算論・知覚

本年度は、昨年度に触れたMellin変換に基づく聴覚末梢系での情報表現の研究を展開するとともに、昨年度に特許化を図ったSTRAIGHTの要素となる新アルゴリズム群の評価と最適化を進めた。その結果、提案したアルゴリズムの性能は従来の同様なアルゴリズムを大きく凌駕するものであり、これまで見出すことの出来なかった新しい音声信号の性質の発見が可能となった。この発見に関しては2001年度に詳しく検討する予定である。計算論に関しては、STRAIGHTとMellin変換の共通項である音源の周期性とシステムのサイズに関するパラメタの分離が聴覚の情報表現の本質であることを明らかにした。

知覚に関しては、STRAIGHTの残された課題である非周期成分の表現と知覚に関する検討を進めた。この検討の結果、STRAIGHTによる処理音声は、自然音声と区別がつかない高い品質を実現できる場合があるものの、特有の劣化が認められる場合も未だに残されていることが明かとなった。これらの要因を詳しく解析することにより、STRAIGHTのツールとしての完成を目指すことは、プロジェクト終盤の重要な課題である。

外部発表・特許等

10月に北京で開催された音声言語処理会議（ICSLP'2000）を、本年度の戦略的に重要な会議と位置付けて重点的に発表を行った。先に挙げたEURASIPでの受賞、プロジェクト外でのSTRAIGHTを応用した研究の発表、幾つかの企業からの照会等、継続的に進めて来た本プロジェクトの成果の展開の努力が、具体的な結果を生み出しつつある。

ワークショップ

聴覚の生理と心理的研究の重要な拠点である英国Cambridg大学生理学部のCNBH(Center for Neural Basis of Hearing)の代表であるRoy Patterson博士の日本滞在の機会をとらえて、名古屋大学において小規模なワークショップを開催した。今回のワークショップでは、STRAIGHTにおける情報表現、Mellin変換を中心とした聴覚情報表現、群遅延特性の知覚に関する最先端の話題を集中的に議論するため、CNBHのUppenkamp博士と、電総研の後藤真孝博士を招いて行われた。この会議での議論は、それらの話題が聴覚末梢系の計算理論に収斂する可能性を示唆している。この方向に関しては、2001年度も引き続き検討して行くことが重要である。

その他の成果

方向情報の導入によりICAの性能と安定性を向上させる方法の提案、基本周波数抽出における耐雑音性の大きな改善、音響からのリズム構造の抽出、スペクトル包絡と基本周波数の依存関係の検討、STRAIGHTを補完する音声符号化方式の提案等、STRAIGHTをより一般的な聴覚刺激の処理に適用するための周辺技術の

検討を進展させ、プロジェクト終盤への体制整備を進めた。

3. 主な研究成果の発表（論文発表）

Toshio Irino, Roy D Patterson Auditory Modelling of extracting size and shape information of sound source : Stabilised Wavelet Mellin Transform Speech Communication(Eurospeech99) 1999/ 8 /20

Roy D Patterson, Stefan Uppenkamp, Toshio Irino Mellin Images of vowel sound and the phonological distinctiveness of multi-formant vowels British Society of vowel sound and the phonological distinctiveness of multi-formant vowels 1999/ 9 /22

入野俊夫 音源の形状情報と寸法情報を分離する聴覚でのイメージング 日本音響学会誌 2000/ 2 /20

Roy D Patterson 抄訳：入野俊夫 聴覚イメージ：複雑な音が聴覚システムでいかに表現されるか(抄訳) 日本音響学会誌 2000/ 2 /20

Masashi Unoki, Toshio Irino, Roy D Patterson Improvement of an IIR asymmetric compensation gammachirp filter 日本音響学会論文誌（英文誌）

Masato Akagi, Hironori Kitakaze Preception of synthesized singing voices with fine fluctuations in their fundamental frequency contours 6th International Conference on Spoken Language Processing 2000/10/17

Hiroshi Saruwatari, Satoshi Kurita, Kazuya Takeda, Fumitada Itakura, Kiyohiro Shikano Blind Source Separation Based on Subband ICA and Beamforming 6th International Conference on Spoken Language Processing 2000/10/17

Yoshinori Atake, Toshio Irino, Hideki Kawahara Robust fundamental frequency estimation using instantaneous frequencies of harmonic components 6th International Conference on Spoken Language Processing 2000/10/18

Seiichi Nakagawa, Hiromitsu Nisizaki A system for retrieving broadcast speech documents using voice input keywords and similarity between words 6th International Conference on Spoken Language Processing 2000/10

Kazumasa Yamamoto, Seiichi Nakagawa Relationship among Speaking Style, Inter-Phoneme's Distance and Speech Recognition Performance 6th International Conference on Spoken Language Processing 2000/10

Hideki Kawahara, Yoshinori Atake, Parham Zolfaghari Accurate Vocal Event Detection Method Based on a Fixed-point Analysis of Mapping From to Weighted Average Group Delay 6th International Conference on Spoken Language Processing Vol.4, pp.664-667 2000/10

Mamoru Iwaki A vocal tract area ratio estimation from spectra ; parameter

extracted by STRAIGHT 6th International Conference on Spoken Language
Processing 2000/10