

「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

1999年度の研究は、プロジェクトの中核となる音声分析変換合成法STRAIGHTの基本周波数抽出に関するアルゴリズムの精度および耐雑音性の大きな改善と、非周期的信号の情報表現に関する発明により特徴付けられる。また、これらの発明が加わることにより、昨年度の聴覚の計算理論の位置付けに新しい光が当てられたことも特筆すべき成果である。

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、アルゴリズムの頑健化であった。また、後半に向けて応用への展開に着手した。

アルゴリズム・計算論・知覚

本プロジェクトのターゲットである実時間処理に向けて、中核となるSTRAIGHTを構成するアルゴリズムの弱点である頑健性と非周期信号の表現の問題を昨年引き続き検討した。1999年度の前半においては、昨年度に発明した周波数領域での不動点に基づく基本周波数抽出方法の数理的基礎を見直すことにより、分析の精度と雑音に対する頑健性を大きく向上させる複数の方法を発明し特許化を図った。また、後半においては、これまで適切な情報表現を持たなかった非周期的信号に対し、時間領域での不動点に基づく駆動情報の抽出方法を発明し特許化を図った。本発明は、音声からの声門閉止時点の抽出のための決定的な手法として応用できるばかりではなく、機器の異常診断等、これまで聴覚が用いられてきた様々な分野の機械化に応用することができる。

ヨーロッパの音声国際会議(Eurospeech'99)の機会をとらえて、昨年ATRに招聘した英国ケンブリッジ大学CNBH所長のRoy D. Patterson博士との研究協力を、ATR人間情報通信研究所の入野主任研究員中心として進めた。以前から交流のあった両研究者が萌芽的に形成しつつあったMellin変換を中心とした聴覚末梢系の計算理論が、本プロジェクトの中核アルゴリズムであるSTRAIGHTの非周期信号への拡張の本質に関わるものであることが、現在、明らかになりつつある。上記の発明の優れた性能は、それらの発明が聴覚の計算理論の本質的な部分を反

映しているためであるとも考えられよう。

ワークショップ

「聴覚の情景分析」の提唱者であるBregman教授の訪日の機会をとらえて、熊本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、Bregman教授の専門である聴覚心理学の側面から見た「音声・音響処理システム」の問題を、この10年の「聴覚の情景分析」の展開を踏まえて議論することを狙った。本ワークショップは、Bregman教授に加え、このテーマに関連する聴覚心理の分野で精力的に研究を進めている九州芸術工科大学の中島教授、熊本大学の宇佐川助教授を講師としてして迎えて議論を交わすことにより、これまで進めてきた工学的・計算論的取り組みをより大きな枠組みの中に位置付けて考える貴重な機会となった。

外部発表・特許等

9月にハンガリーで開催されたヨーロッパの音声国際会議（Eurospeech'99）を戦略的に重要な発表の場と位置付けて、重点的に発表を行った。また、11月にオハイオ州コロンバスで開催された米国音響学会では、生物音響のセッションにおける招待講演として招かれた。この米国音響学会においては、欧米の研究者による4件を含め、STRAIGHTを用いた発表が計6件あり、先のEurospeechとも併せて国際的な場における本プロジェクトの存在感が確立してきたと考えられる。また、STRAIGHTの本質的な改良につながる3件の特許の出願を行っている。なお、本年度、ATR人間情報通信研究所の入野主任研究員と鷓木博士の論文が、日本音響学会の論文賞を受賞していることを特記する。

応用への展開への着手

本年度は、本プロジェクトの後半において重要となってくる情報表現の様々な音声技術への展開に着手した。この過程において、複数音源による音声・音響信号の処理が、未解決の重要課題として前面に出てきている。この解決のための重要な手がかりである、聴取者の能動性、両耳・位相情報の利用、群遅延情報の利用の各々について、自走ロボット、発声・聴覚相互作用測定器等の実験装置を準備し、並行して検討を進めている。

3. 主な研究成果の発表（論文発表）

Speech Communication, Vol.27, No.3-4, Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds pp.187-207, 河原英紀、勝瀬郁代、A シュベニエ

Speech Communication, Vol.27, A Method of Signal Extraction from Noisy Signal based on Auditory Scene Analysis, pp.261-279, 鷓木祐史、赤木正人

日本音響学会英文誌, Vol.20, No.6, An analysis/synthesis auditory filterbank based on an IIR implementation of the gammachirp, pp.397-406, 入野俊夫、鵜木祐史

6 Th European Conference on Speech Communication and Technology FIXED POINT ANALYSIS OF FREQUENCY TO INSTANTANEOUS FREQUENCY MAPPING FOR ACCURATE ESTIMATION OF F0 AND PERIODICITY Hideki Kawahara, Haruhiro kawahara, Alain de Cheveigne, Roy D,Patterson

6 Th European Conference on Speech Communication and Technology Speech Coding Using Mixture of Gaussians Polynomial Models Parham S. Zolfaghari, A.J.Robinson

6 Th European Conference on Speech Communication and Technology Stabilised Wavelet Mellin Transform : An Auditory Strategy for Normalising Sound-Source Size Toshio Irino,Roy D.Patterson

6 Th European Conference on Speech Communication and Technology Segregation of vowel in background noise using the model of segregating two agoustic sources based on auditory scene analysis Masashi Unoki, Masato Akagi

6 Th European Conference on Speech Communication and Technology SPEAKER CONVERSION THROUGH NON-LINEAR FREQUENCY WARPING OF STRAIGHT SPECTRUM kazuya takada, Noriyasu Maeda, Hideki Banno, Shoji Kajita, Fumitada Itakura

6 Th European Conference on Speech Communication and Technology AN OBJECTIVE DISTORTION ESTIMATOR FOR HEARING AIDS AND ITS APPLICATION TO NOISE REDUCTION Mitsunori Mizumachi, Masato Akagi