

「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

聴覚によって捉えられる外界の情報表現と同型の情報表現を求めることにより、具体的なシステムの実現を通じて音声・音響処理技術にブレークスルーをもたらす技術体系の創出を狙い、研究開発を進めている。2001年度までの研究により、本プロジェクトの中核技術である STRAIGHT で用いていた信号の周期性に基づく情報表現に加え、現象のサイズや形状を自然な形で表現している wavelet Mellin 変換に基づく情報表現、それらの橋渡しをする不動点に基づく情報表現、独立成分分析システムとしての両耳聴などを発明・発見し、聴覚における処理の工学的実現の範囲が拡大しつつある。また、プロジェクト終盤への移行に伴い、応用への展開、成果の技術移転に向けた活動を併行して進めた。

### 2. 研究実施内容

#### 2.1 基礎アルゴリズム・計算論・知覚

中核技術である STRAIGHT は、同じ名前であっても内部のアルゴリズムが全て本プロジェクトにおける発明によって置き換えられたため、発足時のシステムとは別物となっている。本年度は、こうして改定されて来た現状の STRAIGHT の品質の再評価を行うとともに、鮮明になって来た問題点の解明を進めた。再評価の結果は、表面的な波形を破壊してしまう STRAIGHT の方が波形の復元に基づく従来の方法よりも常に高い品質の再合成音声を生成できることを明らかにしている。これは、本プロジェクトの目標である、聴覚にとって本質的な表現の領域での処理が STRAIGHT によって実現されていることを示唆するものである。STRAIGHT に残された問題の一つは、非周期性から周期信号に至る中間領域の音響信号の表現と、再合成のためのパラメタの安定な抽出方法である。この問題に対し、一方では工学的な手段による対策を用意することとし、組織的な知覚実験によりその対策に必要な基礎データを蓄積した。また、それと平行して、聴覚によるイベントの検出のモデル化に基づく本質的な解決策の検討を知覚ならびに信号処理アルゴリズムの両面から進めた。また、本プロジェクトで開発した高精度のアルゴリズムを用いることで初めて見つけられた子音周辺での音声の基本周波数軌跡の規則的な偏倚に関して報告するとともに、それらの原因と音響処理における意味を明らかにした。

本プロジェクトの入野らによる wavelet Mellin 変換装置としての聴覚末梢系の最適性を示した一

連の仕事の進展は、周期性に基づく STRAIGHT の背景である Fourier 変換とどのようにして両立し得るのかというもう一つの課題をますます目立つものとした。wavelet Mellin 変換の核関数である gammachirp は、聴覚末梢系における実際の処理のモデルとしての妥当性をますます明らかにしつつある。また、gammachirp の計算自体、僅かな非線形時変部分を付加した線形時不変システムとして高い精度で効率良く実現できることを具体的に明らかにし、工学的な応用可能性を大きく広げた。ここで、wavelet Mellin 変換による聴覚イメージから、非線形写像の学習を通じて STRAIGHT の合成部を利用して音声を再合成する方法を発明したことは、この矛盾を統合する一つの重要なステップである。もう一つの統合の可能性は、STRAIGHT 関連の発明である周波数領域と時間領域でのアルゴリズムにおいて用いられた不動点を介するところにあるが、最終年度に詳細はゆずる。なお、不動点に対する別の解釈に基づくことで、日常環境のような雑音の多い状況においても精度の劣化が少ない基本周波数の抽出方法を開発したことを付記する。

## 2.2 応用への展開

STRAIGHT の素直な情報表現を生かし、簡単な滑らかなパラメタ変換によって音声の話者を入れ替えて高品質な合成音声を作成する技術を開発した。また、STRAIGHT を用いることで、音声が表出する感情についても、高い品質を保ったまま変換し再合成できる可能性を実証した。音声知覚についても、これまで合成音声でしか確認されていなかった母音連鎖の知覚における補償や予測現象の存在を、STRAIGHT による実音声に近い加工音声で確認することが可能となった。これらの応用の実例の蓄積は、本プロジェクト内にとどまらず、内外の様々な研究グループでも進められている。これは、『生態学的に妥当な状況下での人間の音声知覚機構の解明』という、成果の応用の重要な目標として設定していた課題に、本プロジェクトが実際に大きく貢献して行くサイクルが本格的に回りだしたことを意味するものである。

## 3. 研究実施体制

### (1) 基礎アルゴリズムグループ

- ① 河原英紀(和歌山大学、教授)
- ② 基礎アルゴリズム、感性情報変換、アルゴリズム実装

### (2) 計算論グループ

- ① 入野俊夫(NTT コミュニケーション科学基礎研究所、主任研究員)
- ② 初期聴覚系計算理論

### (3) 知覚グループ

- ① 河原英紀(国際電気通信基礎研究所、招聘研究員)
- ② 基礎アルゴリズム、聴覚心理特性、聴覚心理実験の企画・遂行

### (4) 符号化グループ

- ① 板倉文忠(名古屋大学、教授)
- ② 符号化、特徴抽出、符号化応用、聴覚符号化、アルゴリズムの最適化、群遅延特性符号化、ささやき声の音響分析

- (5) 認識グループ
  - ① 中川聖一(豊橋技術科学大学、教授)
  - ② 音声認識応用、音声認識用特長パラメータ
- (6) 合成規則グループ
  - ① 峯松信明(東京大学、助教授)
  - ② 音声の規則性の定式化
- (7) 変換・音場グループ
  - ① 鹿野清宏(奈良先端科学技術大学院大学、教授)
  - ② 音声変換、音情報処理、旋律制御、音場認識・制御、音声変換
- (8) 音源分離グループ
  - ① 赤木正人(北陸先端科学技術大学院大学、教授)
  - ② 聴覚モデル、情報表現、位相情報による分離、方向知覚機能モデル、基本周波数抽出・音源分離
- (9) オブジェクト記述グループ
  - ① 嵯峨山茂樹(東京大学、教授)
  - ② 聴覚的オブジェクトの記述と再合成

#### 4. 研究成果の発表

##### (1) 論文発表

- 坂野秀樹(名古屋大学),陸金林,中村哲(ATR),鹿野清宏(奈良先端科学技術大学院大学),河原英紀(和歌山大学,ATR,CREST) 「時間領域平滑化遅延を用いた短時間体相の効率的表現方法」 電子情報通信学会論文誌 D-II,Vol.J84-D-II,No.4,pp621-628,2001/4
- Hideki Banno, Kazuya Takeda, Fumitada Itakura (Nagoya Univ.) 「A Study on Perceptual Distance Measure for Phase Spectrum of Stimuli」 IEEE Signal Processing Society, ICASSP, International Conference on Acoustics, Speech, and Signal Processing, Vol.VofVI,pp3297-3300, 2001/5/7-11
- 猿渡洋(奈良先端科学技術大学院大学), 栗田悟史, 武田一哉(名古屋大学) 「Blind Source Separation Combining Frequency-Domain ICA and Beamforming」 IEEE Signal Processing Society, ICASSP, International Conference on Acoustics, Speech, and Signal Processing, pp464-467, 2001/5/10
- Toshio Irino(NTT communication Science Labs./CREST), Roy D. Patterson (CNBH, Univ. Cambridge) 「A compressive gammachirp auditory filter for both physiological data」 Journal of Acoustical Society of America, Vol.109 No.5 pt.1,May 2001
- 立蔵洋介, 猿渡洋, 鹿野清宏(奈良先端科学技術大学院大学) 「Compensation of Temperature Fluctuation in Inverse System of Sound Field Reproduction by Linear Warping Processing」 17th International Congress on Acoustics (ICA) , Rome, Italy, 2001/9/3

- Hideki Kawahara(Wakayama Univ./CREST/ATR), Parham Zolfaghari (CIAIR/Nagoya univ./CREST) 「Systematic F0 Glitches around Nasal-Vowel Transitions」 7th European Conference on Speech Communication and Technology (Eurospeech 2001), Aalborg,DENMARK, 2001/9/7
  - Yuichi Ishimoto,Masashi Unoki,Masato Akagi (School of Information Science, japan Advanced Institute of Science and Technology) 「A Fundamental Frequency Estimation Method for Noisy Speech Based on Instantaneous Amplitude and Frequency」 7th European Conference on Speech Communication and Technology (Eurospeech 2001), Aalborg,DENMARK, 2001/9/7
  - Hiroshi Saruwatari, Toshiya Kawamura, Kiyohiro Shikano(Nara Institute of Science and Technology) 「Blind Source Separation for Speech Based on Fast-Convergence Algorithm with ICA and Beamforming」 7th European Conference on Speech Communication and Technology (Eurospeech 2001), Aalborg,DENMARK, 2001/9/7
  - 西野隆典(名城大学), 梶田将司, 武田一哉, 板倉文忠(名古屋大学) 「水平方向及び仰角方向に関する頭部伝達関数の補間」 日本音響学会誌, Acoustical Science and Technology, 57 卷 11 号,pp685-692, 2001
  - Hideki Banno, Kazuya Takeda, Fumitada Itakura (Nagoya Univ.) 「The effect of group delay spectrum on timbre」 Acoustical Science and Technology,23 卷 1 号,pp1-9,2002
  - 河原英紀(和歌山大学, ATR 人間情報科学研究所, CREST), 片寄晴弘(和歌山大学) 「高品質音声分析変換合成システム STRAIGHT を用いたスキヤット生成研究の提案」 情報処理学会論文誌 第 43 卷,第 2 号,pp208-218,2002/2
- (2) 特許出願  
国内なし、外国なし