和歌山大学システム工学部　教授

河原　英紀

「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

本プロジェクトでは、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た。『聴覚脳を創る』を標語として進めてきた本プロジェクトは、聴覚の本質的な理解に基づく『初期聴覚系の計算理論』や『高品質音声変換合成システム STRAIGHT』を生み出してきた。プロジェクトの初期段階から、これらの成果および派生した様々な要素技術は、プロジェクトの存在自体が持つメッセージとともに、音声・音響処理の様々な応用分野にも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という歴史的使命を果たしたと言えよう。また、本プロジェクトを通じて脳科学の観点から聴覚研究の次代を担う何人かの若手研究者が育ったことを特記しておきたい。

１．１　基本構想

聴覚・音声の研究には長い歴史がある。前世紀の初期までには、聴覚では周波数分析が行われていることが知られていた。真空管の発明からそれほど遠くない 1939 年のニューヨーク万博において、既に、真空管を用いたアナログ回路によって音声を駆動源の情報と声道共振の情報とに分離する VOCODER の原理に基づいて、オペレータの『演奏』による電気的合成音声が披露されている。聴覚器官の機構と機能についても、ノーベル賞に結びついた 1920 年代からの Bekesy の一連の研究により、1940 年代の半ばまでには内耳基底膜の運動の進行波モデルが明らかにされた。また、急速に進展する電話システムの普及と相互接続性の確保のために、音声知覚と物理特性との関連が精力的に調べられたのも同じ頃である。技術的資源が限られていた当時においては、「いかに少ない物理的リソースで、最低限の音声によるコミュニケーションを保証するか」といういわば機械の論理に立った問題意識は正当なものであった。しかし、この問題意識が VOCODER の潜在的可能性の一面に過ぎない情報圧縮の側面だけを強調し、それ以降の聴覚・音声の研究を偏ったものとして来たことも事実である。この偏りが、1980 年代から前世紀末にかけての情報処理能力の超指数関数的向上、脳活動のミクロならびにマクロな観測手段の発展につり合う方法論の不在を招いたとも極論できよう。

本プロジェクトは、この閉塞状況を打開するため、1939 年の VOCODER の原点に戻り、もう一つの可能性であった「聴覚的に意味のある情報に分解する」という側面を追求することから出発した。研究の基本理念として、カナダの聴覚心理学者である Bregman が 1990 年の著書において提唱した「聴覚の情景分析（Auditory Scene Analysis）」のメッセージ「聴覚の研究を生態学的な視点から再構築せよ」を受けとめ、具体的な着手点として、生態学的に重要な『周期性を有する音』に注目し、計算理論への展開を意識しながら、聴覚と同型の情報表現と処理の研究を進めることとした。

１．２　実施の概要

プロジェクトの実施に当たっては、聴覚の情報表現に基づく音声分析変換合成システムSTRAIGHTを研究の核とするとともにテストベッドとして利用することで、情報表現、アルゴリズムの開発と評価を緊密に連携させることに留意した。この戦略は、本プロジェクトの特色であり終盤での計算理論の構築において、真価を発揮することとなった。研究組織としては、それぞれの拠点の特色を生かし、STRAIGHTおよび聴覚の計算理論につながるモデルおよびアルゴリズムを開発するグループ、関連する周辺技術の開発および基礎データの蓄積を行うグループ、成果展開と代替技術の開発を行うグループの三種類のグループからなる体制をとり、研究開発を進めた。

１．３　研究成果の概要

本プロジェクトの最大の成果は、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、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、1939年に発明されたVOCODERの基本構造を踏襲しながらも、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の共通の基礎資料として、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が行われ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した方法である。このようなSTRAIGHTの存在は、聴覚情報処理の本質の追求を回避して音声波形の復元を高品質化の指針としてきた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認識されるようになった結果、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、多くの技術開発がスタートされ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もなりつつある。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、これら応用のすそ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提案した「組織的ダウングレーディング」という方法論を実施するための技術的基盤を提供する。

本プロジェクトのもう一つの特記すべき成果は、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問題を解いているとする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。別の見方をするなら、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言うこともできる。

１．４　各研究グループの概要

本プロジェクトを構成する研究グループは、主要メンバーの異動とプロジェクトの進展による新課題の開始等に伴い、何度かの再構成を経ている。ここでは、先に上げた研究グループの大分類に照らして、各グループにおける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。また、ワークショップの企画をはじめ、プロジェクト全体の統括を行った。和歌山大学、ATR、から構成されるグループである。

(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。奈良先端科学技術大学院大学、名古屋大学大学院、豊橋技術科学大学、東京大学大学院峯松研究室から構成されるグループである。

(3) 音源分離・感性情報変換・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する。中間評価で既存の音声コミュニティーへの指向性が強過ぎると指摘されたことを受けて、プロジ

ェクトの終盤に向けて強化された部分である。研究拠点としては、北陸先端科学技術大学院大学の赤木研究室と嵯峨山研究室（後に東京大学大学院に異動）、再掲形で和歌山大学から構成されるグループである。

２．研究構想

本プロジェクト開始時に目指した直接的な目標は、聴覚の機能についての生態学的理解に基づいて研究代表者によって発明された新しい音声分析変換合成方法 STRAIGHT（Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrogram）の実時間処理システムの実現であった。また、間接的な目標として、生態学的に妥当な聴覚刺激に対する聴覚の機能を明らかにするための新しい強力なパラダイムとツールを提供することで、工学のみならず、聴覚生理・心理の領域に大きな波及効果をもたらすことを狙った。しかし、最も大きな目標（強い願い）は、当初の計画書に明示的な目標としては記載していない『聴覚の計算理論』の構築にあった。この大きな目標に照らせば、STRAIGHT は計算論レベルでの聴覚脳の実現の端緒であり、計算論を試験するためのテストベッドである。実際、プロジェクト後期では、『聴覚の計算理論』の構築が現実的な目標として前面に出ている。

これらの目標の実現に向けて、当初の研究計画では、STRAIGHT を軸として要素技術の研究開発を進めるとともに、成果を関連する技術分野に展開するための適用方法の開発を併行させ、プロジェクトの終盤では、工学のみならず聴覚生理・心理の領域にインパクトを与えるための応用技術の開発を進めることとした。プロジェクト発足時の研究グループの構成は、この研究計画を反映させて、第一の STRAIGHT を軸とするグループとして、基礎アルゴリズム・計算論・知覚、第二の適用方法開発のグループとして、符号化・認識・音声変換・音場制御、第三の応用技術を担当するグループとして、音源分離・感性情報変換を設定し、それぞれの拠点として、第一のグループに、和歌山大学（河原）、ATR（入野、津崎）、第二のグループに、奈良先端科学技術大学院大学（鹿野）、名古屋大学大学院（板倉）、豊橋技術科学大学（中川）、第三のグループに、北陸先端科学技術大学院大学（赤木）、和歌山大学（片寄）、を設定した。なお、括弧中の人名は、拠点のキーパーソンである。

これらの構想と研究計画は、研究の進展に伴う新発見と新しい可能性の出現、グループのキーパーソンおよび実働メンバーの異動、中間評価結果等に基づいて見直しが行われた。目標設定に関しては、『聴覚の計算理論』への指向性と聴覚生理・心理へのインパクトの強化を図ることとした。また、プロトタイプとして実現した実時間 STRAIGHT の評価に基づき、STRAIGHT に関する直接の目標を加工性と品質の向上に向けることとした。研究計画を実施する組織に関しては、これらの変更を反映して、第一の研究グループに、NTT コミュニケーション科学基礎研究所（入野）、第二の研究グループに、東京大学大学院（峯松）、第三の研究グループに、北陸先端科学技術大学院大学（嵯峨山）（後の異動に伴い東京大学大学院に変更）を新たに設定した。

２．１　研究グループの役割分担

ここでは、研究構想での分類に基づいて、各研究グループの役割分担を説明する。

（１）グループ大分類：基礎アルゴリズム・計算論・知覚

基礎アルゴリズム・計算論グループ

和歌山大学システム工学部河原研究室：代表者（河原英紀）

研究代表者の研究室。プロジェクトの中核技術である STRAIGHT を構成する基礎アルゴリズムの開発を担当する。プロジェクト終盤の異動により、NTT コミュニケーション研究所の代表であった入野が加わることで、聴覚の計算論の研究も併せて担当することとなった。

知覚グループ

国際電気通信基礎技術研究所：代表者（河原英紀）

STRAIGHT の要素技術の開発、計算理論の構築に必要となる聴覚心理実験、音声知覚実験を企画・遂行する。

（２）グループ大分類：符号化・変換・音場制御・認識・合成

変換・音場制御グループ

奈良先端科学技術大学院大学情報科学研究科鹿野研究室：代表者（鹿野清宏）

STRAIGHT を用いた音声の変換とそのための技術開発を担当。また、STRAIGHT の限界を広げるための音場の解析と制御方法の研究を担当する。

符号化グループ

名古屋大学大学院工学研究科板倉研究室：代表者（板倉文忠）

音響信号の統計的解析と符号化およびアルゴリズムの実時間化を担当する。

認識グループ

豊橋技術科学大学情報工学系中川研究室：代表者（中川聖一）

STRAIGHT により求められるパラメタを既存の音声認識エンジンに応用するための方法の研究と評価を担当する。

合成規則グループ

東京大学大学院情報理工学研究科峯松研究室：代表者（峯松信明）

STRAIGHT による変換音声品質の向上と規則合成を目的とした、基本周波数とスペクトル間の拘束条件の解析を担当する。

（３）グループ大分類：音源分離・感性情報変換・オブジェクト記述

音源分離グループ

北陸先端科学技術大学院大学情報科学研究科赤木研究室：代表者（赤木正人）

単耳・両耳情報に基づく音源分離および神経回路における聴覚情報処理機構のモデル化

を担当する。

感性情報変換グループ（再掲）

和歌山大学システム工学部河原研究室：代表者（河原英紀）

STRAIGHT を応用した音に含まれる感性情報の変換と制御を担当する。

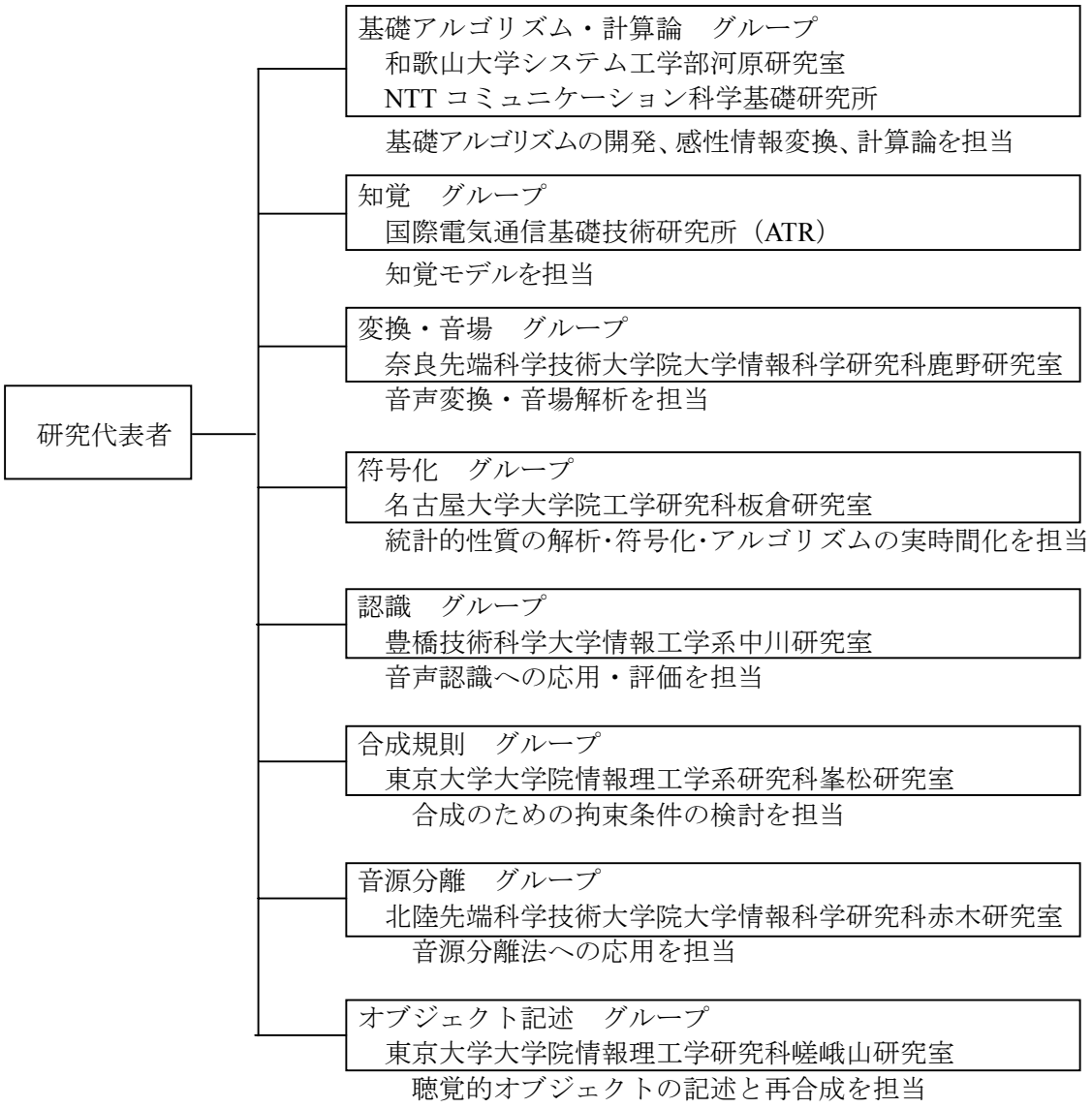
オブジェクト記述グループ

東京大学大学院情報理工学研究科嵯峨山研究室：代表者（嵯峨山茂樹）

音楽を具体的な題材として、時間構造を有する聴覚的オブジェクトの記述と抽出を担当する。

３．研究実施体制

(1) 体制



４．研究期間中の主な活動

(1) ワークショップ・シンポジウム等

年月日	名称	場所	参加人数	概要
平成 10 年 12 月 17 日 ～ 平成 10 年 12 月 18 日	CREST workshop on Stable representations for periodic sounds	名古屋大学	45 名	信号の周期性は、聴覚的には非常に重要であるにもかかわらず、これまでの音響／音声処理技術においては、有効な処理モデルが欠けていた。本ワークショップでは、聴覚と同様に周期信号を安定に表現する方法の最新の知見に基づきプロジェクトの方向付け等の議論を行った。
平成 11 年 11 月 22 日	Auditory Scene Analysis in Speech Signal Processing	熊本大学	56 名	「生態学的妥当性に基づいて聴覚情報処理を見直す」という本プロジェクトの理念の支柱となった「Auditory Scene Analysis（聴覚の情景分析）」の提唱者である Bregman 教授を招き、教授の心理学的視点と本プロジェクトの工学的／計算論的指向性の擦合わせを行った。
平成 12 年 11 月 22 日	CREST Workshop on Stable Representation of Periodic Sounds	名古屋大学	28 名	本プロジェクトで発見／定式化された聴覚情報の周期性に関する二種類の情報表現と、周期性ならびに時間情報の処理に関する最新の聴覚心理／生理学的知見の突き合わせを行い、構築すべき聴覚の計算理論の方向付けを行った。
平成 13 年 11 月 19 日	Crest Auditory Brain Project Meeting for Auditory Object Representation and Analysis	東京大学	42 名	聴覚におけるオブジェクトの表現と処理がどのように行われているか、それらの処理を工学的に実現するために本プロジェクトの成果がどのように応用／展開されるかについて、周期性だけではなく空間情報やより大きなスケールでの時間的構造を含めて議論を行った。
平成 14 年 7 月 8 日 ～ 平成 14 年 7 月 9 日	CREST Workshop on Computational Models of Auditory Processing	ATR	55 名	本プロジェクトの成果の集大成と、今後の成果展開、より広い枠組みでの聴覚の計算論の構築に向けて、内外の聴覚処理モデル研究者を招き議論を行った。

５．主な研究成果

(1) 論文発表（国内　11 件、海外　５件）

1. "Hideki Kawahara (ATR/Wakayama Univ./CREST), Ikuyo Masuda-Katsuse（九州システム情

報研）, Alain de Cheveigne (CNRS/University’Paris７)", Restructuring speech representations using a pitchadaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of repetitive structure in sounds, SPEECH COMMUNICATION, "27 (3-4), 1999, pp187-202", 1998-1999EURASIP 最優秀論文賞

2. "Masashi Unoki, Masato Akagi (JAIST)", A Method of Signal Extraction from Noisy Signal based on Auditory Scene Analysis, SPEECH COMMUNICATION," 27 (3-4), 1999, pp261-279",
3. "Alain de Cheveigne (CNRS), Hideki Kawahara (Wakayama Univ./ATR/CREST)", Missing-Data Model of Vowel indentication, Journal of Acoustical of America," Vol.105, No.6, 1999, pp3479-3508",
4. "Toshio Irino (NTT Communication Science Labs.), Roy D. Patterson (CNBH/Univ. of Cambridge)", A compressive gammachirp auditory filter for both physiological and psychophysical data, Journal of Acoustical of America," Vol.109, No.5, 2001, pp2008-2022",
5. "Toshio Irino (ATR), Roy D. Patterson (CNBH/Univ. of Cambridge)", Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilized wavelet-Mellin transform, SPEECH COMMUNICATION," 36 (3-4) 2002, pp181-203",
6. 河原英紀（和歌山大学/ATR/CREST）、自然性の極めて高い音声分析変換合成法、音声研究、"第２巻第２号（1998 年８月）、pp28-36",
7. "水町光徳、赤木正人（北陸先端大）"、マイクロフォン対を用いたスペクトルサブトラクションによる雑音除去法、電子情報通信学会論文誌（A）、"Vol.J82-A, No.4, 1999, pp 503-512",
8. "Toshio Irino (ATR), Masashi Unoki (ATR/JAIST)",An analysis/synthesis auditory filterbank based on an IIR implementation of the gammachirp, 日本音響学会論文誌（英文誌） J. Acoust. Soc. Jpn. (E), "20 (6) 1999, pp397-406", 第 40 回日本音響学会佐藤論文賞
9. "鵜木祐史、赤木正人（北陸先端大）"、聴覚の情景解析に基づいた雑音下の調波複合音の一抽出法、電子情報通信学会論文誌 (A)、"Vol.J82-A, No.10, 1999, pp1497-1507",
10. "Mitsunori Mizumachi (ATR), Masato Akagi (JAIST)",The auditory-oriented spectral distortion for evaluating speech signals distorted by additive noises, The Journal of the Acoustical Society of Japan (E)," 21 (5) 2000, pp251-258",
11. "阿竹義徳（奈良先端大）、入野俊夫（ATR/CREST）、河原英紀（和歌山大学/CREST/ATR）、陸　金林、中村　哲、鹿野清宏（奈良先端大）"、調波成分の瞬時周波数を用いた基本周波数推定方法、電子情報通信学会論文誌　D-II," Vol.J83-D-II, No.11, 2000, pp2077-2086",
12. "戸田智基（奈良先端大）、坂野秀樹、梶田将司、武田一哉、板倉文忠（名古屋大学）、鹿野清宏（奈良先端大）"、側抑制性重み付けを用いた雑音環境下における STRAIGHT 分析合成系の品質改善、電子情報通信学会論文誌　D-II ,"Vol.83-D-II, No.11, 2000, pp2180-2189",
13. "坂野秀樹、陸　金林、中村　哲、鹿野清宏（奈良先端大）、河原英紀（和歌山大学）"、時間領域平滑化群遅延による位相制御を用いた音質制御方式、電子情報通信学会論文誌　D-II, "Vol.83-D-II, No.11, 2000, pp2276-2282",
14. "坂野秀樹、陸　金林、中村　哲、鹿野清宏（奈良先端大）、河原英紀（和歌山大学）"、時間領域平滑化群遅延を用いた短時間位相の効率的表現、電子情報通信学会論文誌 D-II、"Vol.84-D-II, No.4, 2001, pp621-628",

15. "Masashi Unoki (ATR/JAIST/CREST), Toshio Irino (ATR/CREST),Roy D. Patterson (CNBH/Univ. of Cambridge)", Improvement of an IIR asymmetric compensation gammachirp filter,日本音響学会論文誌（英文誌）Acoust. Sci. & Tech.,"22 (6) 2001, pp426-430",
16. "河原英紀（和歌山大学/ATR/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 を用いたスキャット生成研究の提案、情報処理学会論文誌、"Vol.43, No.2, 2002, pp208-218"

(2) 特許出願（国内　３件、海外　１件）

①国内

１．特許番号：特許 3251555
　　発明者　：河原英紀
　　発明名称：信号分析装置
　　出願日　：平成 10 年 12 月 10 日

２．公開番号：特開 2001-249674
　　発明者　：河原英紀
　　発明名称：駆動信号分析装置
　　出願日　：平成 12 年３月６日

３．公開番号：特開 2001-249676
　　発明者　：赤木正人
　　発明名称：雑音が付加された周期波形の基本周期あるいは基本周波数の抽出方法
　　出願日　：平成 12 年３月６日

②外国

１．PCT 国際出願・指定国（米国）特許願　第 09/786, 642 号
　　発明者　：河原英紀、入野俊夫
　　発明名称：「METHOD OF EXTRACTING SOUND-SOURCE INFORMATION」
　　（音源情報の抽出方法）
　　出願日　：2001 年３月７日

(3) 受賞等

①受賞

１．第 40 回日本音響学会佐藤論文賞
　　Toshio Irino, Masashi Unoki 「An analysis/synthesis auditory filter bank based on an IIR implementation of the ganmmachirp」 J. Acoust. Soc. Japan (E), 20(6) 1999, pp397-406

２．1998-1999 EURASIP 最優秀論文賞
　　Hideki Kawahara, Ikuyo Masuda-Katsuse, Alain de Cheveigne’ 「Restructuring speech representations using a pitchadaptive time-frequency smoothing and an

instantaneous-frequency-based F0 extraction : Possible role of repetitive structure in sounds」
Speech Communication 27, 3-4, 1999, pp187-207