様式 C-19

科学研究費補助金研究成果報告書

平成２３年　４月１日現在

機関番号：１４７０１
研究種目：基盤研究(A)
研究期間：2007～2010
課題番号：19200017
研究課題名（和文）　聴覚・音声機能の支援・拡張技術に関する総合的研究

研究課題名（英文）　Integrated research on augmentation and assisting technologies for auditory and speech functions

研究代表者
河原　英紀（KAWAHARA HIDEKI）
和歌山大学・システム工学部・教授
研究者番号：40294300

研究成果の概要（和文）：世界最先端の音声分析変換合成方式と聴覚末梢系における信号処理の数理モデル、統計的音声変換技術、音源分離技術を総合的に連携させることにより、聴覚障害者および喉頭摘出者などの障害者支援技術を開発した。また、人間の聴覚との整合性の良い音声処理技術の新しい基本アルゴリズムの開発により、音声コミュニケーションの研究と様々な操作による拡張のために広く利用することのできる研究基盤を確立した。

研究成果の概要（英文）：Assisting technologies for hearing impaired andesophageal speech and augmentation technologies for voice conversion and manipulation were introduced as the accomplishment of this project, which are constructed by integrating our advanced speech analysis, modification and synthesis framework, computational model of auditory periphery, statistical voice conversion method, and independent component analysis. Throughout this project, these constituent technologies were also advanced further and resulted in establishment of strong basis for speech communication research and applications.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００７年度	9,200,000	2,760,000	11,960,000
２００８年度	9,100,000	2,730,000	11,830,000
２００９年度	8,100,000	2,430,000	10,530,000
２０１０年度	11,000,000	3,300,000	14,300,000
年度			
総　計	37,400,000	11,220,000	48,620,000

研究代表者の研究分野：音声情報処理
科研費の分科・細目：情報学、知覚情報処理・知能ロボティクス
キーワード：聴覚障害、喉頭摘出、音声分析、音声変換、音声合成、音源分離、聴覚モデル

１．研究開始当初の背景

研究代表者が 1997 年に発明した音声分析変換合成技術である STRAIGHT は、音声知覚・音声合成の強力な研究ツールとして内外の研究機関に普及し始めており、分担者の入野が提案した聴覚末梢系の数理モデルである gammachirp も、強固な数理的基盤に立脚するとともに既知の聴覚心理の知見とも良く整合するものとしての評価が固まりつつある状況にあった。また、STRAIGHT の応用技術として、統計的音声変換技術への STRAIGHT の導入方法が分担者の戸田により確立されるとともに、実際の環境で有効に動作する音源分離技術も分担者の猿渡らにより確立しつつある状況にあった。さらに、自らも聴覚障害者である分担者の奥乃は、計算論的聴覚情

景分析に基づいて、ロボットにおける聴覚を本質的な部分から再構築することで、人間における聴覚の役割に、新たな光を投げかけていた。本研究課題は、このような世界最先端の研究を推進している研究者が、音声・聴覚機能の支援・拡張を軸としてチームを組むことから生まれたものである。

２．研究の目的

聴覚および音声の理解に基づいて、背景で紹介した技術を総合することにより、それらの機能に生じた障害の克服を支援し、あるいは拡張するための手段を開発することを目的とする。具体的な応用としては、実環境において混在する音の中から目的とする音を分離し、適切に提示することの可能な補聴器の開発、音声変換技術の応用として、喉頭摘出などにより声を失った方々のための会話支援技術の開発、声の品質の改善など能力を拡張・支援する技術の開発などが目標となる。それらの目標を明確に意識することにより、用いられている基盤技術を高度化し、かつ、より現実的な環境に対応できるものとすることを併せて狙う。

３．研究の方法

以下の４項目を軸として研究を進めた。

(1)要素技術：STRAIGHT
(2)要素技術：音声変換
(3)要素技術：音源分離
(4)成果の統合と評価

まず、要素技術に関しては、(1)は河原が中心となって、森勢および和歌山大のグループが推進し、(2)は鹿野と戸田が中心となって推進し、(3)は、猿渡が中心となって推進した。また、(4)は、分野横断的であり河原が全体の統括と調整を行い、奥乃が評価とロボット聴覚の観点からの総合的検討を進めた。

４．研究成果

研究の方法で説明した項目毎に成果を整理する。

(1)要素技術：STRAIGHT

STRAIGHT は、音声分析、音声変換、音声合成それぞれのアルゴリズムから構成されている。STRAIGHT は、本課題の中核であるだけではなく、音声知覚・音声合成研究のための基盤ツールとして、内外の多数の研究機関で使用されており、既に大きなインパクトを与える状況にあった。本課題の大きな成果の一つは、この STRAIGHT が新しい定式化により理論からアルゴリズム、実装の全てのレベルで再構築されたことにある。その結果、分析合成の速度ならびに精度が向上するとともに、アルゴリズムとしての見通しの良さと柔軟性が大きく改善されることで、さらに大きく発展する可能性が拓かれることとなった。

①TANDEM-STRAIGHT

従来の方法では、音声のスペクトルを分析する際に音源の周期性が時間および周波数方向の干渉として影響することを回避することができなかった。本課題の成果である TANDEM-STRAIGHT は、これらの問題を本質的に解決する方法である。具体的には、基本周期の半分だけの間隔を隔てた時刻において求めたパワースペクトルの平均として、分析時刻と波形の相対位置に依存しない表現が得られるという発見と、標本化定理の新しい観点であるコンシステント標本化定理に基づいて定式化されている。実装は、対数スペクトルのフーリエ変換として定義されるケプストラム上での演算として整理されたアルゴリズムに基づいており、効率が良く、しかも音声スペクトル近似精度の高いシステムとして実現されている。

TANDEM-STRAIGHT による分析の例を図 1 と図 2 に示す。奥行き方向に時間、左右方向に周波数、上下方向にスペクトルのレベルとした 3 次元表示と、底面に 2 次元の擬似カラー表示を併用している。210ms を境として、通常のパワースペクトル、TANDEM によるスペクトル、STRAIGHT によるスペクトルが切り替えられて表示されている。

この周期性の影響を完全に排除したスペクトル表現には、音声の柔軟な加工を容易にするとともに、理解し易いため、これまでの音声科学の蓄積を有効に生かすことができるという利点がある。

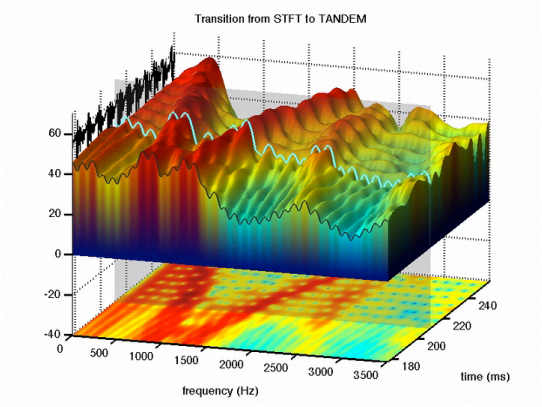

図 0 通常の分析（後部）と TANDEM（前部）

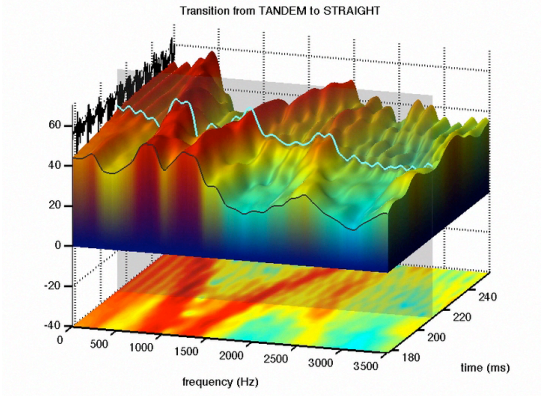

図 0 TANDEM（後部）と STRAIGHT（前部）

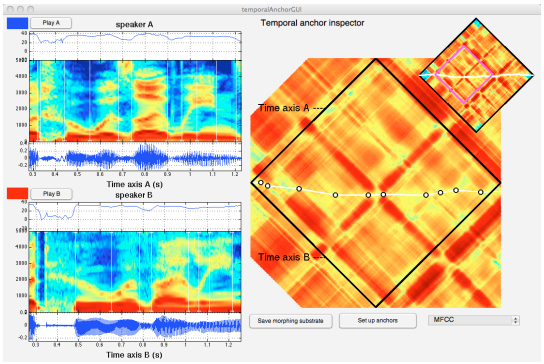

図 3 モーフィング参照点設定用 GUI

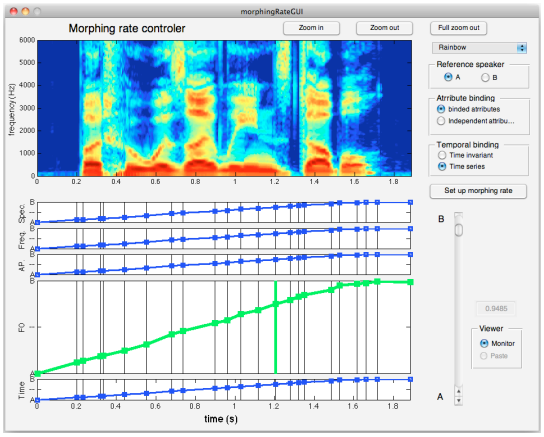

図 4 時変多属性モーフィング率設定用 GUI

②時変多属性モーフィング
TANDEM-STRAIGHT により求められる表現の透明性と、効率の良いアルゴリズムは、柔軟でかつ高品質な音声のモーフィングを可能とした。値のモーフィングにおいては、値そのものではなく対数変換されたパラメタ上で処理を行うことにより、補外されたパラメタを用いて合成する際にも破綻を生じないことを保証した。また、時間軸および周波数軸をモーフィングする場合も、軸の対応関係（変換関数）そのものではなく変換関数の導関数の対数変換したものを処理の対象とすることで、変換関数の単調性を保証した。この新しい定式化により、複数の音声の a)スペクトル、b)基本周波数、c)非周期成分、d)時間軸変換関数、e)周波数軸変換関数の五つの属性を、各時刻においてそれぞれ独立の割合でモーフィングすることを可能とした。この時変多属性モーフィングのアルゴリズムとして、コンテンツのポストプロダクションのようなオフライン処理用のものと、ライブでの使用のようなリアルタイム処理用ものの二種類を定式化した。
こうして拡張された時変多属性モーフィングを簡単に使うことができるように、TANDEM-STRAIGHT の処理と併せて、音声分析合成およびモーフィング用のグラフィカルユーザインタフェース（GUI）を開発した。

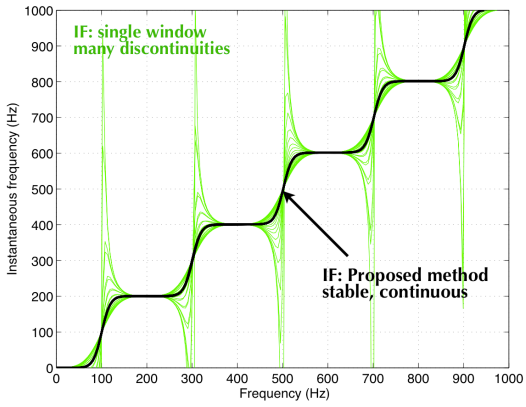

図 5 時間的変動の無い瞬時周波数計算法。緑は従来法、黒は新しい計算法による。

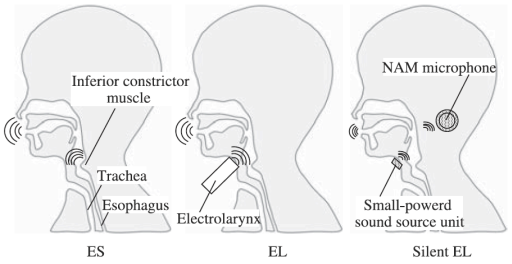

図 6 無喉頭音声の抽出法：ES 食道発声、EL 電気的人工喉頭、Silent EL 微弱振動子による人工喉頭音声

③拡張
なお、本課題の終了直前に、関連する大きな成果が得られた。①で説明した TANDEM と同様な基本的アイデアに基づくことにより、分析対象の波形と分析位置の相対関係に依存しない、瞬時周波数の計算方法が明らかとなったのである。瞬時周波数は信号の基本的な属性であるため、本計算法は、音声に限らず一般的な信号に対して応用できる基盤技術と位置づけられる。本計算法を要素技術である STRAIGHT に組込むことは、今後、別の新たなプロジェクトにおいて検討すべき課題である。
(2)要素技術：音声変換
STRAIGHT の情報表現を入力信号の分析の際に用いることにより、統計的声質変換による音声品質を大きく向上させることができる。こうして開発された音声変換技術に基づいて、極めて小さな囁き声の一種である無音声発話（肉伝導微弱音声発話：NAM）の収録と処理方法を応用することで、喉頭摘出者の音声を聞き易い明瞭な音声に変換する技術を確立した。
喉頭摘出者が発声する際には、図 6 に示すように、a)食道発声、b)電気式人工喉頭、c)肉伝達微弱音声を用いることができる。a)の食道発声は習得が困難であり、b)の人工喉頭は習得が容易であっても人工喉頭からの機械

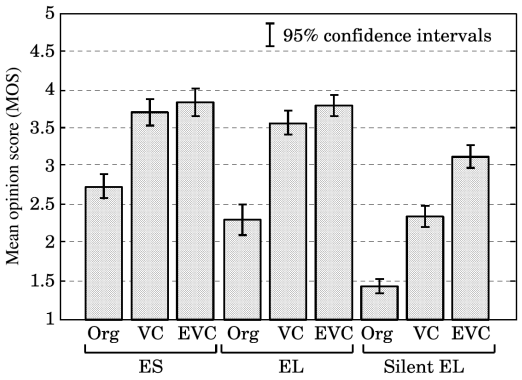

図 7　無喉頭音声の品質評価結果。Org 元の音声、VC 変換音声、EVC 一対多固有音声変換

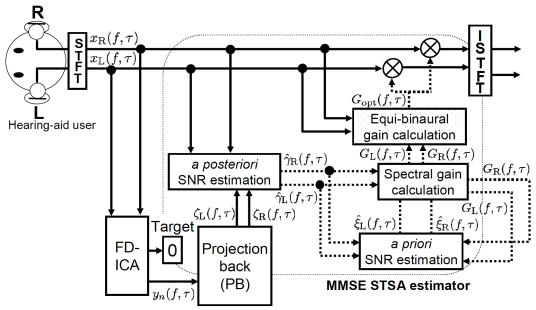

図 8　補聴器への応用を目的とした ICA と MMSE STSA に基づく両耳処理方式

的な響きとピッチの自由な制御が困難であるため、音声によるコミュニケーションの際に障害となる。c)は NAM の技術を応用したものであり、微弱振動子を用いることにより上記の機械的な響きが聴こえてしまうという問題を解決したものである。これらの音声の品質向上に音声変換技術を適用するための検討を進めた。具体的には、一対多固有声変換に基づく無喉頭音声強調法を各種無喉頭音声に対して適用し，自然性や話者性を大幅に改善できることを示した。結果を図 7 に示す。本方式は、変換時における処理が軽いため、日常的に携帯する機器に実装するに適した技術となっている。

(3)要素技術：音源分離

ディジタル技術により補聴器の動的振幅圧縮が可能となった現在、聴覚障害者を支援するための技術において残されている解決すべき最大の問題は、周囲の環境音と目的とする会話音声の分離である。本課題では、独立成分分析（ICA）を拡張するとともに、両耳効果を利用することにより、この問題を解決する技術を開発した。開発した技術は、ICA による雑音推定に基づいて目的とする音の平均二乗誤差最小化短時間振幅スペクトル推定（MMSE STSA）を行なう方法に基づいている。この方法は、非定常雑音推定に高い能力を有する ICA と、目的音声の抽出に高い性

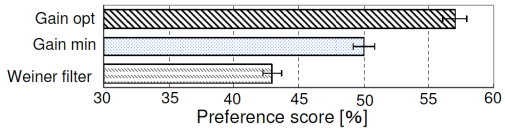

図 9 処理音声の品質評価結果。Gain opt 両耳同期処理有、Gain min 同期処理無、Weiner filter 従来法

能を有する MMSE STSA とを組み合わせた手法である。ここでは、さらに、補聴器への応用の際に有用な両耳効果の利点を生かすことができるよう、雑音抑圧処理を両耳で同期させる手法を開発した。開発した方式の処理のブロック図を図 8 に示す。図 9 の処理音声の品質評価結果は、MMSE STSA と両耳の同期処理のそれぞれが効果的であることを示している。本方式も、日常的に携帯する機器に実装する上で適したアルゴリズムにより構成されている。

(4)成果の統合と評価

開発の各段階において、これらの要素技術各々の方式パラメタによる品質および性能評価値への影響を調べるため、計算機シミュレーションおよび被験者を用いた主観評価実験が多数遂行された。また、それらの評価実験に用いるための音声収録および、収録音声へのラベリングなどの作業が行われた。また、ロボット聴覚への応用を通じて、要素技術の統合と、実環境での適合性の検討が進められた。これらの結果およびラベル等は、当面、和歌山大学およびそれぞれの機関において、成果の応用展開の際の基礎データとして蓄積されており、必要に応じて提供できるように整理されている。

(5)アウトリーチその他

多数の招待講演において本課題の成果を適宜紹介するとともに、大学等での非営利の教育研究の利用については、TANDEM-STRAIGHT に基づくツール類を提供し、成果の社会還元を進めて来た。これらにより、STRAIGHT の基本論文の引用数は、GoogleScholar では 499 となっており、音声研究の一つのデファクトスタンダードとなっている。また、入野が米国音響学会のフェローに選出されるなど、gammachirp も、聴覚末梢系の数理モデルとしての地位が確立するに至っている。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 17 件）

①赤桐隼人，森勢将雅，入野俊夫，河原英紀，スペクトルピークを強調した F0 適応型スペクトル包絡抽出法の最適化と評価，電子情報通信学会 論文誌 A，Vol.J94-A,

No.8, 2011.（査読有）
② H. Doi, K. Nakamura, T. Toda , H. Saruwatari, K. Shikano , Esophageal speech enhancement based on statistical voice conversion with Gaussian mixture models, IEICE Transactions on Information and Systems, Vol. E93-D, No. 9, pp. 2472-2482, Sep. 2010.（査読有）
③森勢将雅、河原英紀、西浦敬信、基本波検出に基づく高 SNR の音声を対象とした高速な F0 推定法、電子情報通信学会論文誌 D, Vol.J93 -D, No.2, pp.109 -117 (2010).（査読有）
④森勢将雅、高橋徹、入野俊夫、河原英紀、分析時刻に依存しない周期信号のパワースペクトル推定法を用いた音声分析、電子情報通信学会、Vol.J92-A, No.3, pp. 163-171, Mar. 2009.（査読有）
⑤河原英紀，音声モーフィングの背景と可能性，音声言語医学，Vol.50, No.2, pp.185-190 (2009),（査読無：招待論文）

〔学会発表〕（計 148 件）

①河原英紀、ロイ・パターソン、森勢将雅、坂野秀樹、津崎実、高橋徹、西村竜一、入野俊夫、成分位相の制御により声の肌触りを変える、インタラクション 2011、日本科学未来館、2011 年 3 月 11 日、(査読有)
②呉将延，猿渡洋，鹿野清宏，細井裕司，ICA による雑音推定に基づいた平均二乗誤差最小化短時間振幅スペクトル推定法の両耳補聴器への応用，日本音響学会 2011 年春季研究発表会，早稲田大学、2011 年 3 月 10 日，(査読無)
③土井啓成，中村圭吾，戸田智基，猿渡洋，鹿野清宏，統計的声質変換に基づく無喉頭音声強調法の評価，日本音響学会早稲田大学、2011 年春季研究発表会，2011 年 3 月 9 日，(査読無)
④E. Okamoto, T. Irino, H. Kawahara, Evaluation of voice morphing using vocal tract length normalization based on auditory filterbank, NCSP'11, Tianjin SaiXiang Hotel, Tianjin, China, 2011 年 3 月 2 日，(査読有)
⑤河原英紀、森勢将雅、入野俊夫、周期信号における時間的変動の影響を受けない位相関連情報の表現について、電子情報通信学会技術研究報告 SP、音声、Vol.110, No.297, pp.47-51, 愛知県立大，2010 年 11 月 18 日，(査読無)
⑥N. Yamakawa, T. Kitahara, T. Takahashi, K. Komatani, T. Ogata, H. G. Okuno, Effects of modelling within - and between-frame temporal variations in power spectra on non -verbal sound recognition, Interspeech2010, Makuhari Messe, Chiba, Japan, 2010 年 9 月 29 日，(査読有)
⑦H. Kawahara, Exploration of the other aspect of Vocoder revisited, A-Z STRAIGHT, TANDEM-STRAIGHT and morphing, SSW7, Kyoto, Japan, 2010 年 9 月 23 日，（査読無：招待講演）
⑧H. Kawahara, R. Nisimura, T. Irino, M. Morise, T. Takahashi , H. Banno, High-quality and light -weight voice transformation enabling extrapolation without perceptual and objective breakdown, ICASSP2010, Dallas Texas, USA, 2010 年 3 月 18 日，(査読有)
⑨ H. Doi, K. Nakamura, T. Toda , H. Saruwatari, K. Shikano , Statistical approach to enhancing esophageal speech based on Gaussian mixture models , ICASSP2010, Dallas Texas, USA, 2010 年 3 月 18 日，(査読有)
⑩ R. Okamoto, Y. Takahashi, H. Saruwatari, K. Shikano , MMSE STSA Estimator with nonstationary noise estimation based on ICA for high-quality speech enhancement , ICASSP2010, Dallas Texas, USA, 2010 年 3 月 17 日，(査読有)
⑪T. Takahashi, K Nakadai, K. Komatani, T. Ogata, H. G. Okuno, Missing-feature-theory-based robust simultaneous speech recognition system with non-clean speech acoustic model, IROS-2009, St. Louis, USA, 2009 年 10 月 13 日，(査読有)
⑫H. Kawahara, T. Takahashi, M. Morise, H. Bannno, Development of exploratory research tools based on TANDEM-STRAIGHT, APSIPA2009, Sapporo, Japan, 2009 年 10 月 5 日，(査読有)
⑬K. Nakamura, T. Toda, H. Saruwatari, K. Shikano, Electrolaryngeal speech enhancement based on statistical voice conversion, Interspeech2009, Brighton, UK, 2009 年 9 月 8 日，(査読有)
⑭H. Kawahara, R. Nisimura, T. Irino, M. Morise, T. Takahashi , H. Banno, Temporally variable multi -aspect auditory morphing enabling extrapolation without objective and perceptual breakdown , ICASSP2009, Taipei, Taiwan, 2009 年 4 月 23 日，(査読有)
⑮Y. Kubota, K. Komatani, T. Ogata, H. G. Okuno, Design and implementation of 3D auditory scene visualizer towards auditory awareness with face tracking, ISM08, Berkeley, CA, USA, 2008 年 12 月 16 日、(査読有)

⑯H. Kawahara, M. Morise, H. Banno, T. Takahashi, R. Nisimura , T. Irino , Spectral envelope recovery beyond the Nyquist limit for high-quality manipulation of speech sounds, Interspeech2008, Brisbane, Australia, 2008年9月24日, (査読有)

⑰K. Nakamura, T. Toda, Y. Nakajima, H. Saruwatari, K. Shikano, Evaluation of speaking-aid system with voice conversion for laryngectomees toward its use in practical environments, Interspeech2008, Brisbane, Australia, 2008年9月24日, (査読有)

⑱H. Kawahara, M. Morise, T. Takahashi, R. Nisimura , T. Irino , H . Banno, TANDEM-STRAIGHT : A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation, ICASSP2008, Las Vegas, USA, 2008年4月1日, (査読有)

〔図書〕(計3件)

①Hideki Kawahara, Masanori Morise, Toru Takahashi, Ryuichi Nisimura , Hideki Banno, Toshio Irino ,: STRAIGHT, a framework fo r speech analysis, modification and synthesis, in Computer processing of Asian spoken languages, (eds. Shuichi Itahashi and Chiu-yu Tseng), Consideration Books, Los Angeles, pp.235-243, March 2010. (総ページ数644)

〔その他〕
ホームページ等
http://www.wakayama-u.ac.jp/~kawahara/
この他、TANDEM-STRAIGHT 配付用にアクセス制限をかけた、アドレス非公開のページを設置している。
アウトリーチ活動：
①2011年3月7日に上智大学においてワークショップを開催。一部を一般に公開。
②2009年11月11日に関西TLOにおいて、本課題の技術についての講演会を実施。
③2009年10月9日に日本音響学会聴覚研究会ビギナーズセッションで本課題の技術の講習を実施。

６．研究組織

(1)研究代表者
河原 英紀 (KAWAHARA HIDEKI)
和歌山大学・システム工学部・教授
研究者番号：40294300

(2)研究分担者
入野 俊夫 (IRINO TOSHIO)
和歌山大学・システム工学部・教授
研究者番号：20346331
西村 竜一 (NISIMURA RYUICHI)
和歌山大学・システム工学部・助教
研究者番号：00379611
奥乃 博 (OKUNO HIROSHI)
京都大学・情報学研究科・教授
研究者番号：60318201
鹿野 清宏 (SHIKANO KIYOHIRO)
奈良先端科学技術大学院大学・情報科学研究科・教授
研究者番号：00263426
猿渡 洋 (SARUWATARI HIROSHI)
奈良先端科学技術大学院大学・情報科学研究科・准教授
研究者番号：30324974
戸田 智基 (SARUWATARI HIROSHI)
奈良先端科学技術大学院大学・情報科学研究科・准教授
研究者番号：90403328
森勢 将雅 (MORISE MASAHIRO)
立命館大学・情報理工学部・助教
研究者番号：60510013
(H19→H20：連携研究者)

(3)連携研究者
高橋 徹 (TAKAHASHI TOHRU)
京都大学・情報学研究科・助教
研究者番号：30419494

(4)研究協力者
坂野 秀樹 (BANNO HIDEKI)
名城大学・理工学部・准教授
研究者番号：20335003
パターソン ロイ (PATTERSON ROY)
英国Cambridge大学・CNBH・上級研究員
キューリーポート ダイアン
(KEWLEY-PORT DIANE)
米国インディアナ大学・音声聴覚学科・名誉教授
シュヴァインベルガー シュテファン
(SCHWEINBERGER STEFAN R)
独イェナ大学・心理学部・教授
ベリン パスカル (BELIN PASCAL)
英国グラスゴー大学・心理学科・教授
ヴァティキオティス-ベイソン エリック
(VATIKIOTIS-BATESON ERIC)
カナダ ブリティッシュコロンビア大学・言語学科・教授