

二国間交流事業 共同研究報告書

平成23年3月31日

独立行政法人日本学術振興会理事長 殿

共同研究代表者所属・部局 奈良先端科学技術大学院大学・情報科学研究科

(ふりがな)

職・氏名 助教 戸田 智基

1. 事業名 相手国 (フランス) との共同研究 振興会対応機関 (仏外務省)

2. 研究課題名 カシス ～コミュニケーション補助のための静かな音声インタフェース～

3. 全採用期間

平成21年4月1日～平成23年3月31日 (2年0ヶ月)

4. 経費総額

(1) 本事業により執行した研究経費総額 1,800,000円

初年度経費1,000,000円、 2年度経費800,000円

1. (2) 本事業経費以外の国内における研究経費総額 100,000円

5. 研究組織

(1) 日本側参加者（代表者は除く）

氏名 <small>(ふりがな)</small>	所属・職名	研究協力テーマ
かわはら ひでき 河原 英紀	和歌山大学・教授	音声分析合成技術の改善
しかの きよひろ 鹿野 清宏	奈良先端科学技術大学院大学・教授	非可聴つづやきマイクの研究
なかむら けいご 中村 圭吾	奈良先端科学技術大学院大学・博士 後期課程学生（2010年度からは日本 学術振興会特別研究員－PD）	サイレント音声変換技術の構築
おおたに やまと 大谷 大和	奈良先端科学技術大学院大学・博士 後期課程学生（2009年度まで）	声質変換基礎技術の構築
どい ひろのり 土井 啓成	奈良先端科学技術大学院大学・博士 後期課程学生	サイレント音声変換技術の構築
きさき きゆうすけ 木佐木 雄介	奈良先端科学技術大学院大学・博士 後期課程学生	テキスト音声合成技術の構築

(2) 相手国側研究代表者

所属・職名・氏名

GIPSA-lab, UMR 5216 STIC・DR CNRS・BAILLY Gerard

(3) 相手国参加者（代表者は除く）

氏名	所属・職名（国名）	研究協力テーマ
BADIN Pierre	GIPSA-lab, Grenoble・DR CNRS	サイレント音声データ収録技術の研究
ELISEI Frederic	GIPSA-lab, Grenoble・IR CNRS	サイレント音声データ収録技術の研究
JUTTEN Christian	GIPSA-lab, Grenoble・PR UJF	サイレント音声データ収録技術の研究
LOEVENBRUCK Helene	GIPSA-lab, Grenoble・CR CNRS	マルチモーダル音声変換技術の構築
VILAIN Coriandre	GIPSA-lab, Grenoble・IR Stendhal	サイレント音声データ収録技術の研究
TRAN Viet-Anh	GIPSA-lab, Grenoble・PhD student	マルチモーダル音声変換技術の構築
YOUSSEF Atef Ben	GIPSA-lab, Grenoble・PhD student	マルチモーダル音声変換技術の構築
CHOLLET Gerard	ENST, Paris・DR CNRS	音声分析合成技術の改善
WEI Jianguo	ENST, Paris・Postdoc CNRS	サイレント音声データ収録技術の研究
DENBY Bruce	ESPCI, Paris・PR UPMC	サイレント音声インタフェースの研究
DREYFUS Gerard	ESPCI, Paris・PR UPMC	サイレント音声データ収録技術の研究
HUEBER Thomas	ESPCI, Paris・PhD student	マルチモーダル音声変換技術の構築

6. 研究実績概要（全期間を通じた研究の目的・研究計画の実施状況・成果等の概要を簡潔に記載してください。）

我々の音声コミュニケーションにおいて、静かな環境下では発声することで周囲に迷惑をかける、周囲に人がいる際に秘匿性の高い内容を発声することができない、騒音下では音声が伝えにくい、身体的障害により発声機能の一部を失うと音声を発声すること自体が困難となる、など様々な問題が存在する。これらの問題を解決するために、音声コミュニケーションにおいて、周囲に聞こえないほど小さな声を収録することができる非可聴つぶやきマイクロフォン、発声時に生じる筋電信号、発声時の顔画像、調音運動を捉える磁気センサ、口腔内の超音波画像、などといったサイレント音声インタフェースの利用を可能とする基礎技術の構築に取り組んだ。日本チームが持つ最先端の音声特徴量変換技術、音声分析合成技術、非可聴つぶやきマイクロフォンによる收音技術と、フランスチームが持つ最先端の顔画像収録技術、超音波画像収録技術、磁気センサによる調音運動データ収録技術、マルチモーダル信号処理技術を組み合わせることで、多様なサイレント音声インタフェースの利用技術の構築を目指した。

実験用データベースを構築するために、フランスチームにおいて、各種サイレント音声データの収録を行った。日本チームにおいても、博士後期課程メンバーが Grenoble の GIPSA-lab に滞在し、サイレント音声データ収録を行った。これらのデータを整理して、日本語及びフランス語からなる非可聴つぶやき音声データや顔画像データが含まれるマルチモーダルデータベース、磁気センサデータベース、超音波画像データベースなどを構築した。

サイレント音声をを用いた円滑な音声コミュニケーションを実現するために、サイレント音声データから通常音声への変換技術の構築に取り組んだ。日本チームとフランスチームの博士後期課程メンバーによる1～2週間程度の短期滞在を実施することで、日本チームが持つ最先端の音声特徴量変換技術と音声分析合成技術の共有を行った。異なる研究機関で分担して、非可聴つぶやき音声、磁気センサによる調音運動データ、超音波画像データといった各種サイレント音声データから通常音声への変換法の構築に取り組むことで、効率的に研究を進めた。また、マルチモーダル音声データの利用として、通常音声と同期した顔画像合成技術の構築にも取り組んだ。構築した技術を実験的に評価することで、その高い有効性を示した。両国チームの代表者と他数名により、これらの研究成果の一部を共同で学術論文としてまとめ、学術論文誌にて発表した。

人対人の音声コミュニケーションにおいては、リアルタイム処理が重要となるため、リアルタイム音声特徴量変換処理の構築にも取り組んだ。高い変換精度を保ちながらリアルタイム変換処理を実現するために、短遅延で動作可能な変換アルゴリズムを構築し、その有効性を確認した。また、CPU パワーが限られた状況下での利用も考慮して、計算量を削減する手法を構築し、その有効性を示した。さらに、サイレント音声データを収録する上で、収録機器の設置状況などの収録条件により、得られるサイレント音声データの音響特性が大きく変化する問題に着目し、その解決策を検討した。音声特徴量変換モデルを収録されたサイレント音声信号に対して自動的に適応することで、収録条件の変化に起因する音響特性変化を補正する手法を提案し、その高い有効性を示した。また、サイレント音声インタフェースの一つの応用例として、喉頭摘出者に対する発声補助に着目し、非可聴つぶやきマイクロフォンによる収録技術と非可聴つぶやき音声から通常音声への変換技術を適用することで、新たな代替発声法の構築に取り組んだ。その結果、従来の代替発声法と比較し、提案技術を用いることで、より自然な音声の発声が得られることが分かった。

博士後期課程学生による相手国での短期滞在や、両チームの代表者を含むメンバー数名によるミーティングなどを実施し、お互いの研究動向の把握や技術交流を深めることで、上記研究を円滑に進めた。