

機関番号：13302

研究種目：基盤研究（B）

研究期間：2008～2010

課題番号：20300064

研究課題名（和文）

聴覚フィードバック下での音声知覚・生成の同時脳活動計測に関する研究

研究課題名（英文）

A study on measurement of brain activities with speech production and perception under transferred auditory feedback conditions

研究代表者

赤木 正人 (AKAGI Masato)

北陸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：20242571

研究成果の概要（和文）：

本研究では、音声生成と音声知覚の密接な関係を示す例として「聴覚フィードバック」を取り上げ、「聴覚フィードバック時に脳のどの部位がどの順番で関与しているか」について、fMRI等を用いた測定から明らかにする。測定のために、fMRIでも使用可能な聴覚フィードバック呈示装置および収録装置の開発を行い、これらの装置を使用したfMRIによる脳活動計測を行った。また、EMAを用いた舌運動測定による音声生成と知覚の関係の検討を行った。

研究成果の概要（英文）：

This study focuses on “auditory feedback” as an example of affinity between speech production and perception, and reveals which parts of the brain are activated with which turn when transferred auditory feedback is presented, through brain activity measurements with fMRI. To measure the brain activities, we develop presentation and collection devices for the transferred auditory feedback that is able to be used with fMRI, and then we measure the brain activities by fMRI using these devices. Moreover, we examine relations between speech production and perception by tongue movement measurements using EMA.

交付決定額

(金額単位：円)

| | 直接経費 | 間接経費 | 合計 |
|--------|------------|-----------|------------|
| 2008年度 | 7,600,000 | 2,280,000 | 9,880,000 |
| 2009年度 | 3,600,000 | 1,080,000 | 4,680,000 |
| 2010年度 | 3,400,000 | 1,020,000 | 4,420,000 |
| 年度 | | | |
| 年度 | | | |
| 総計 | 14,600,000 | 4,380,000 | 18,980,000 |

研究分野：

総合領域

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：

音声情報処理，音声生成，音声知覚，聴覚変形フィードバック，脳イメージング，fMRI，MEG，ホルマント

1. 研究開始当初の背景

- (1) 音声知覚・生成は、音声によるマン・マシンコミュニケーションの根幹を成すものである。また、マン・マシンコミュニ

ケーションにおいても、ヒトの音声生成・知覚機構を基礎として、これを工学的に実現した音声合成・認識が重要な役割を果たそうとしている。本来、音声知

覚・生成は、音声コミュニケーションにおいて表裏一体を成すものであり、コミュニケーションを円滑に保つためには双方が一体となって働く必要がある。また、人が多様な音韻を生成するためには、多様な音韻を聞き分ける能力が不可欠である。このため人間は、音声生成過程とその逆過程の音声知覚の繰り返しにより、母国語を含む言語音声知覚機構の発達を遂げてきた。このような発達には音声の生成と知覚とを密接に結ぶ情報交換の通路”ことばの鎖”が大きく関与している。ことばの鎖について、話者内部（脳内）での音声生成と音声知覚との相互交信を説明するため、Libermanらは1960年代より音声知覚の運動理論（Motor theory of speech perception）を提唱してきた。しかし、話者内部（脳内）において音声生成・知覚がどのように関連し存在しているのかについてはいまだ明らかにされていない。

- (2) 我々のグループでは、音声生成と音声知覚の密接な関係を示す一例として「聴覚フィードバック」を取り上げ、知覚・生成の相互作用の解明を図ることを目的として、様々な生理指標の測定を試みてきた。「聴覚フィードバック」とは、話者が発した声を自分の耳で聞き取りながら発話機構を制御するための音声によるフィードバックのことであり、これを用いて話者は正確な発話を行っている。「聴覚フィードバック」が発話に重要な役割を演じているのならば、フィードバックされる音声を人為的にリアルタイムで変形し耳に呈示した場合、その変形によって音声生成に何らかの影響があるはずである。我々は、リアルタイムで変形されたフィードバック音声の変形量と生理的指標の関係および生成された音声との関係を調べることにより、音声知覚とタスクプランニングも含めた音声生成、およびこれらの相互作用について、検討を行ってきた。その研究の結果、今までに次の知見を得ている。
- (3) 被験者のフィードバック音の変形への反応は、変形の方向と反対方向であり、発話において変形に対する補正がリアルタイムで行われていることが確認できた。
 - ① ホルマント周波数分析から、第1、第2ホルマントにおいて明確な補償動作が観測された。これは、スペクトルに関する聴覚フィードバックにおいて短時間での補正反応を捕らえた初めての結果である。
 - ② 筋電（EMG）および舌運動（EMA）の分析結果から、摂動に対してこれを

補償するような筋肉および舌の動きが観測された。

- ③ 補償動作は、変形開始から約150 msで始まり、290 msで最大値に到達した。

2. 研究の目的

本研究では、音声生成と音声知覚の密接な関係を示す一例として「聴覚フィードバック」を取り上げ、聴覚フィードバックの実験パラダイムの中で、脳磁計（MEG）およびfMRIを用いて、「聴覚フィードバック時に脳のどの部位がどの順番で関与しているか」を明らかにすることを目指す。このために

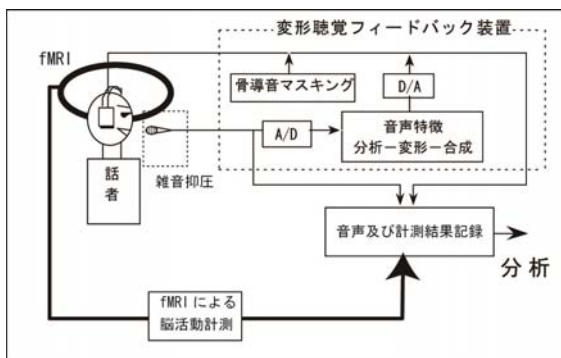
- (1) MEGおよびfMRIでの使用に耐えうる刺激音呈示・音声収録装置の開発を行う
（理由）申請者が開発したリアルタイムでのスペクトル変形呈示システムは、手作りの装置であり、これを用いて変形聴覚フィードバックの実験を行ってきた。スペクトル変形・呈示アルゴリズムについては実績があるが、MEGおよびfMRIによる計測に使用するためには、新たな工夫が必要である。なぜならば、MEGでは脳内で発生する微弱磁気を計測するため磁気測定に影響を与えない呈示方法が必要となる。また、fMRIでは撮像中に励磁騒音が発生するため、騒音下での発話音収録・刺激音呈示が必要となるからである。送話チューブの音響特性改善、気導のみならず骨導による刺激呈示、光マイクロホンアレイ等による発話音収録などを考慮して、装置を開発する。
- (2) 刺激音呈示装置を用いて聴覚フィードバック刺激を呈示しMEGおよびfMRIによる脳内活動の計測を行う
（理由）fMRIを用いて聴覚フィードバック呈示時の脳内活動部位の特定を行うことにより、知覚と生成にかかわる領域およびこれらを介在する領域の特定を行うことが可能となる。さらに、MEGにより刺激呈示に同期したリアルタイム計測を行うことにより、脳内のどの部位がどの順番で働いているかが明らかとなる。
- (3) fMRI等の測定に加えて、他の生理指標（舌運動など）、物理指標（生成された音声の音響特徴など）の測定も行う
（理由）様々な指標を統合して聴覚フィードバック時の模様を測定することにより、生成・知覚の全体像の把握が可能となる。

3. 研究の方法

- (1) 本研究では、変形聴覚フィードバックの実験パラダイムの中で、時間分解能・空間分解能の高い脳活動計測、すなわち、

MEG および fMRI を用いて、聴覚フィードバック時に脳のどの部位がどの順番で活性化し、知覚・生成に関与しているか、を明らかにすることを試みる。

- (2) このために、まず、MEG および fMRI での使用に耐えうる刺激音呈示装置の開発を行った。我々がこれまでの研究で用いてきた変形聴覚フィードバック呈示システムは、高騒音が発生する fMRI では騒音が音声呈示・収録に影響を与える可能性がある。また、より繊細な磁気測定を必要とする MEG に対して影響を与える可能性がある。このため、磁気測定に影響を与えない呈示手法および高騒音対策に対して新たな工夫が必要である。
- (3) 次に、新たに設計した刺激音呈示装置を用いて、下図に示すように聴覚フィードバック刺激を呈示し fMRI による脳内活動の計測を行った。同時に、発話された音声を録音し、これらのデータについて分析を行った。fMRI を用いて聴覚フィードバック呈示時の脳内活動部位の特定を行うことにより、知覚と生成にかかわる領域およびこれらを介在する領域を知ることが可能とする。また、音声の分析から潜時を推定し担っている処理量を明らかにする。なお、MEG および fMRI は国際電気通信基礎技術研究所 (ATR) の装置を借用した。



実験パラダイム (fMRI の場合)

- (4) さらに、聴覚フィードバックが介在した場合の発話動態を計測するために、EMA を用いた復唱時の舌運動測定を行った。

4. 研究成果

- (1) 変形聴覚フィードバック呈示装置の製作
 - ① アルゴリズムの見直しおよび呈示装置の作成
リアルタイムでの変形が可能な音声分析、変形、合成方式のアルゴリズムを再度検討し、変形できるホルマントの数を増やした。これにより、

柔軟な実験設定が可能となった。また、このプログラムをもとに、DSP により可搬型の変形聴覚フィードバック呈示装置を作成した。これにより、fMRI が設置されている場所へ変形聴覚フィードバック呈示装置を移動させ、刺激呈示が可能となった。

② fMRI 内での音声収録

fMRI 内での音声収録のために、E-C 理論に基づいた雑音源方向の推定法および音声強調手法 TS-BASE を提案し、その評価を行った。雑音源方向の推定では、角度で数度以内の誤差を達成した。また real-time TS-BASE を作成した。これにより、fMRI の撮像時騒音の影響を軽減した音声の収録が可能となった。

(2) fMRI を使用した変形聴覚フィードバック呈示時の脳活動計測

被験者 6 名に対して、構築したシステムを用いて、被験者が発話した音声/e/を/i/もしくは/a/に変形し、その音声に耳にフィードバックされている際の脳活動を fMRI により計測した。計測データを分析した結果、聴覚フィードバックにより発話に影響を受けている場合に、顕著な差が見られなかったため、さらに音声呈示装置の改良を行い、新たに被験者 8 名に対して、改良したシステムを用いて、被験者が発話した音声/e/を/i/もしくは/a/に変形し、その音声に耳にフィードバックされている際の脳活動を fMRI により計測した。現在、測定した結果を解析中である。この結果をもとに、脳活動部位の特定および MEG を用いた時系列計測を行う予定である。

(3) 発話音声分析による潜時計測

fMRI 内での変形聴覚フィードバック呈示時の発話音声を解析し、フィードバック音声の変形がどのように発話音声に影響するかを解析した。ホルマントの変形に対して、約 140 - 230 ms で発話音声に変化が現れることがわかった。

(4) EMA を用いた復唱時の舌運動測定

母語の母音カテゴリにない英語母音を目標音として復唱による学習過程に着目し、調音と音響の計測により、音声生成と知覚の関係を調べた。その結果、新しい母音カテゴリの形成は母語の母音カテゴリに依存することが分かった。また、調音運動の視覚情報の呈示は学習過程にある音声と調音の一对多の関係を抑えることができることも分かった。これらの結果をもとに、学習目標母音と復唱音声の知覚により調音を変化させる母音学習過程の制御モデルを考案した。

また、生理学的発話モデルにより、生成機構のモデル化を行った。

(5) 総括

当初予定していた MEG による計測までには至らなかったが、次の成果が得られた。

- ① F1, F2, F3 ホルマント周波数を任意にリアルタイムで変形可能な聴覚フィードバック呈示装置の開発
- ② fMRI 内でも使用可能な聴覚フィードバック呈示装置用音声収録装置の開発
- ③ これらの装置を使用した fMRI による脳活動計測
- ④ fMRI 内での変形聴覚フィードバック呈示時の発話音声解析により潜時が約 140 - 230 ms
- ⑤ EMA を用いた舌運動測定による音声生成と知覚の関係の検討

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 2 件)

- [1] Junfeng Li, Shuichi Sakamoto, Satoshi Hongo, Masato Akagi and Yōiti Suzuki, Two-Stage Binaural Speech Enhancement with Wiener Filter for High-Quality Speech Communication, Speech Communication, 53, 677-689, 2011, 査読有
- [2] Hamada, Y., Kitamura, T., and Akagi, M., A study of brain activities elicited by synthesized emotional voices controlled with prosodic features, Journal of Signal Processing, 14, 4, 265-268, 2010, 査読有

[学会発表] (計 4 1 件)

国際会議

- [1] K. Fujii, Q. Fang, and J. Dang, Investigation of Auditory-Guided Speech Production for during Learning a New Speech Sound, RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP' 11), 2011/3/1-3, Tianjin, China
- [2] X. Wu, Q. Li, J. Wei, and J. Dang, Simulation of Speech Babbling Based on a Physiological Articulatory Model, RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP' 11), 2011/3/1-3, Tianjin, China
- [3] Shih, T, Suemitsu, A., and Akagi, M., Influences of transformed auditory feedback with first three formant frequencies, RISP International

Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP' 11), 2011/3/1-3, Tianjin, China

- [4] X. WU, Q. FANG and J. DANG, Investigation of Muscle Activation in Speech Production Based on an Articulatory Model, International Symposium on Chinese Spoken Language Processing (ISCSLP 2010), 2010/11/30-12/2, Tainan, Taiwan
- [5] Chau, D. T., Li, J., and Akagi, M., A DOA estimation algorithm based on equalization-cancellation theory, INTERSPEECH2010, 2010/09/30, Makuhari, Japan
- [6] Hamada, Y., Kitamura, T., and Akagi, M., A study on brain activities elicited by emotional voices with various F0 contours, International Congress on Acoustics, ICA 2010, 2010/08/24, Sydney, Australia
- [7] Shunsuke Kidani, Ryota Miyauchi, and Masashi Unoki, Effects of the presence of cue tone in signal detection varies with relationships between cue tone and signal frequencies, International Congress on Acoustics, ICA 2010, 2010/08/24, Sydney, Australia
- [8] K. Fujii, J. Wei, A. Suemitsu, J. Dang, The relationship between speech production and perception in the process of learning vowels, The 7th International Conference on Cognitive Science, 2010/8/17-20, Beijing, China
- [9] Jianwu Dang, Investigation of vowel structure and its learning process, The Ninth Phonetic Conference of China (PCC 2010) (招待講演), 2010/5/26, Tianjin, China
- [10] Hamada, Y., Kitamura, T., and Akagi, M., A study on brain activities elicited by synthesized emotional voices controlled with prosodic features, RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP' 10), 2010/03/04, Hawaii, USA
- [11] Li, J. Sasaki, Y., Akagi, M. and Yan, Y., Experimental evaluations of TS-BASE/WF in reverberant conditions, RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP' 10), 2010/03/04, Hawaii, USA
- [12] Fujii, F., Wei, J., Suemitsu, A., and Dang, J., Interaction between speech

production and perception on learning process during imitation, RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP' 10), 2010/03/04, Hawaii, USA

- [13] Naoki Kuroda, Junfeng Li, Yukio Iwaya, Masashi Unoki, and Masato Akagi, Effects of spatial cues on detectability of alarm signals in noisy environments, IWPASH2009, 2009/11/11, Zaou, Japan
- [14] Li, J., Sakamoto, S., Hongo, S., Akagi, M., and Suzuki, Y., Two-stage binaural speech enhancement with Wiener filter based on equalization-cancellation model, WASPAA2009, 2009/10/20, New Paltz, NY, USA
- [15] Li, J., Sakamoto, S., Hongo, S., Akagi, M., and Suzuki, Y., Advancement of two-stage binaural speech enhancement (TS-BASE) for high-quality speech communication, WESPAC2009, 2009/9/21, Beijing, China
- [16] Shunsuke Kidani and Masashi Unoki, The effect of presence of cue-tone on tuning of auditory filter derived from simultaneous masking, 15th International Symposium on Hearing (ISH2009), 2009/6/1, Salamanca, Spain
- [17] Nakamura, T., Kitamura, T., and Akagi, M., A study on nonlinguistic feature in singing and speaking voices by brain activity measurement, RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP' 9), 2009/03/01, Honolulu, USA
- [18] Kuroda, N., Li, J., Iwaya, Y., Unoki, M., Akagi, M., Effects from Spatial Cues on Detectability of Alarm Signals in Car Environments, RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP' 9), 2009/03/01, Honolulu, USA
- [19] Li, J., Sakamoto, S., Hongo, S., Akagi, M., Suzuki, Y., Improved two-stage binaural speech enhancement based on accurate interference estimation for hearing aids, IHCON2008, 2008/08/16, Lake Tahoe, USA

国内研究会・口頭発表

- [20] 木谷俊介, 宮内良太, 鶴木祐史, 同時マスキングにおける cue 音呈示の効果 - 検出効率による検討 -, 日本音響学会 2011 年度春季研究発表会, 2011/3/9-11,

早稲田大学 (東京)

- [21] X. WU, Q. FANG and J. DANG, Inverse estimation of motor command based on a 3D physiological articulatory model, 日本音響学会 2011 年度春季研究発表会, 2011/3/9-11, 早稲田大学 (東京)
- [22] 濱田康弘, 北村達也, 赤木正人, 基本周波数包絡が異なる感情音声聴取時の脳活動測定, 日本音響学会聴覚研究会, 2010/10/16, 加賀白山荘 (石川)
- [23] 木谷俊介, 宮内良太, 鶴木祐史, cue 音呈示によって生じる周波数選択性の変化に関する検討, 日本音響学会聴覚研究会, 2010/10/16, 加賀白山荘 (石川)
- [24] Shih, T., Suemitsu, A., and Akagi, M., Influences of real-time auditory feedback on formant perturbations, 日本音響学会聴覚研究会, 2010/10/16, 加賀白山荘 (石川)
- [25] 藤井一哉, 末光厚夫, 党建武, 母音学習過程における音声知覚と調音変化の関係について, 日本音響学会聴覚研究会, 2010/10/16, 加賀白山荘 (石川)
- [26] Chau, D. T., Li, J., and Akagi, M., A DOA estimation algorithm based on equalization-cancellation theory, 電子情報通信学会応用音響研究会, 2010/06/11, 北海道医療大学 (札幌・北海道)
- [27] 藤井一哉, 魏建国, 末光厚夫, 党建武, 復唱による学習過程における音声生成と知覚の相互作用に関する考察, 日本音響学会平成 22 年春季研究発表会, 2010/3/10, 電通大 (東京)
- [28] 吉川裕介, 徳田功, 王永鑫, 党建武, 変調の概念を用いた感情音声における感情成分の特徴解析, 日本音響学会平成 22 年春季研究発表会, 2010/3/10, 電通大 (東京)
- [29] 濱田康弘, 北村達也, 赤木正人, fMRI による感動詞「ええ」呈示時の脳活動測定, 日本音響学会平成 22 年春季研究発表会, 2010/3/8, 電通大 (東京)
- [30] Wu, X. and Dang, J., Investigation of speech motor control using a 3D physiological articulatory model, 日本音響学会平成 22 年春季研究発表会, 2010/3/8, 電通大 (東京)
- [31] 木谷俊介, 宮内良太, 鶴木祐史, 同時マスキングにおける cue 音呈示の効果 - プローブ周波数の違いが効果に与える影響 -, 日本音響学会平成 22 年春季研究発表会, 2010/3/8, 電通大 (東京)
- [32] 木谷俊介, 宮内良太, 鶴木祐史, 同時マスキングにおける cue 音呈示の効果 1 kHz プローブ音検知に対する cue 音の存在とその周波数配置に関して, 日本音響

学会聴覚研究会, 2009/12/4, 熊本大学 (熊本)

- [33] Li, J., Sakamoto, S., Hongo, S., Akagi, M., and Suzuki, Y., Subjective evaluation of TS-BASE/WF for speech enhancement and sound localization, 日本音響学会平成21年秋季研究発表会, 2009/9/16, 日本大学 (郡山)
- [34] 末光, 朴, 党, 赤木, 櫻庭, 峯松, 広瀬, 異文化間における3感情音声認知と6感情音声認知の比較検討, 日本音響学会21年度春季研究発表会, 2009/03/19, 東京工業大学 (東京)
- [35] 木谷俊介, 鶴木祐史, 同時マスキングにおけるcue音呈示の効果 —1 kHz プローブに対する雑音マスクの周波数配置に関する検討—, 日本音響学会21年度春季研究発表会, 2009/03/18, 東京工業大学 (東京)
- [36] 中村, 北村, 赤木, 脳活動測定による歌声と話声に関する非言語特徴の検証, 日本音響学会21年度春季研究発表会, 2009/03/18, 東京工業大学 (東京)
- [37] 黒田, 李, 岩谷, 鶴木, 赤木, 方向性的手掛かりが雑音環境下での報知音の検知能力に及ぼす影響, 電子情報通信学会音声研究会, 2009/03/06, 東京工科大学 (東京)
- [38] 木谷俊介, 鶴木祐史, 同時マスキングから推定された聴覚特性の同調特性におけるcue音呈示の効果, 電子情報通信学会福祉工学研究会, 2009/2/21, 愛媛大学 (愛媛)
- [39] 中村, 北村, 赤木, fMRIを用いた歌声と話声における脳活動の差異の検討, 音響学会聴覚研究会, 2008/10/17, 神戸セミナーハウス (神戸)
- [40] 木谷俊介, 鶴木祐史, 同時マスキングにおけるcue音呈示の効果に関する検討, 日本音響学会20年度秋季研究発表会, 2008/9/10, 九州大学 (福岡)
- [41] 木谷俊介, 鶴木祐史, 同時マスキングにおけるcue音呈示の効果 —1 kHz プローブに対するマスキング閾値の変化について—, 電子情報通信学会音声研究会, 2008/6/28, 北海道医療大学 (札幌)

[図書] (計1件)
Yasuki Murakami and Masashi Unoki, “Nonlinear responses of a nonlinear cochlear model with the function of an outer hair cell model,” Concepts and Challenges in the Biophysics of Hearing, N. P. Cooper and D. T. Kemp Eds, pp. 343-349, World Scientific, London, 2009.

6. 研究組織

(1)研究代表者

赤木 正人 (AKAGI Masato)
北陸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: 20242571

(2)研究分担者
党 建武 (TOU Takeshi)
北陸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: 80334796

鶴木 祐史 (UNOKI Masashi)
北陸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: 00343187

李 軍鋒 (LI Junfeng)
北陸先端科学技術大学院大学・情報科学研究科・助教
研究者番号: 50431466
辞退 平成22年9月13日

末光 厚夫 (SUEMITSU Atsuo)
北陸先端科学技術大学院大学・情報科学研究科・助教
研究者番号: 20422199

宮内 良太 (MIYAUCHI Ryota)
北陸先端科学技術大学院大学・情報科学研究科・助教
研究者番号: 30455852

北村 達也 (KITAMURA Tatsuya)
甲南大学・理工学部・准教授
研究者番号: 60293594

(3)連携研究者
西本 博則 (NISHIMOTO Hironori)
株式会社国際電気通信基礎技術研究所・認知情報科学研究所・研究員
研究者番号: 50463716