様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年　5 月　23　日現在

機関番号：13302
研究種目：基盤研究（Ｃ）
研究期間：2010～2012
課題番号：22500150
研究課題名（和文）　音声生成における高次機能から末梢機能までの計測とモデル化に関する研究
研究課題名（英文）　Study on the measurement and modeling from higher-order functions to peripheral functions involved in speech production
研究代表者
党　建武（TOH 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

研究成果の概要（和文）：本研究では、従来の生理学的発話機構モデルを末梢器官モデルとして精密化し、ハイレベルモータ表現からローレベルモータ表現への写像を取り入れ、ニューロン計算モデルを構築して、幼児の言語音声学習の模擬によりモデルの妥当性を確認した。母音学習実験では、なじみのない音声を知覚するとき、被験者は母語音響空間に投影することでバーチャルターゲットを形成し、そのターゲットに接近するように学習していることを示唆した。

研究成果の概要（英文）：In this study, we constructed a neuro-computational model by refining the previous physiological articulatory model as a peripheral organ model, where a mapping was established between high-level motor representation to low-level motor representation. The neuro-computational model was confirmed by simulating the speech learning in babbling stage. The results of the vowel learning experiment imply that when perceiving an unfamiliar speech sound, subjects seem to build up a virtual target by projecting the sound to their own acoustic space of the native language, and then approach the target during the learning process.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2010 年度	1,600,000	480,000	2,080,000
2011 年度	1,000,000	300,000	1,300,000
2012 年度	700,000	210,000	910,000
年度			
年度			
総　計	3,300,000	990,000	4,290,000

研究分野：音声科学、音声生成生理学的モデル
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：音声生成、生理学的モデル

１．研究開始当初の背景

人間の音声生成過程は、脳の高次制御から発話器官の生理物理的な運動まで、数多くの機能に関わっている。言語音声を獲得する際には、発話機能と知覚によるスクリーニング機能とを密接に結ぶ情報交換の通路（言葉の鎖）が脳内で形成される。この情報交換通路により人間は音声を頑健に処理できるが、通路に支障があると様々な発話障害が生じる。そのため、人間の音声生成・知覚及びその制御のメカニズムに関する研究が数多く行われてきた。

これまでの研究を大別すると、生体計測技術に基づく手法とモデル模擬による手法に分けられる。計測研究として、ＡＴＲは、ｆＭＲＩを用いて発話中枢機構の機能の計測

を試みた[1]。山形大学は、癲癇患者に対して皮質電気刺激により、言語性コミュニケーションの脳内発話機構を考察した[2]。ただし、計測技術の制約と倫理規制のため、計測研究による方法には限界がある。これに対して、生理学的ニューロン計算モデルを用いる模擬手法は、非侵襲方式による脳の高次機能の研究には有力なアプローチとなる。ニューロン計算モデルを用いる研究として、米国ボストン大学は、聴覚フィードバックを含む音声生成のニューロン計算モデルを提案し、言語野や運動野における脳の活動を模擬した[3]。ドイツのアーヘン工科大学は、さらに視覚フィードバック機能を取り入れて音声知覚のマガーク効果を模擬した[4]。ただし、上記の研究では、発話運動を実現する末梢器官には幾何学的調音モデルが用いられていたため、脳の指令に対する人間の生理学的な働きを再現するには不十分であった。**正確に脳における音声生成の制御機能を理解するには、音声生成過程の末梢器官モデルとして、人間の発話機構をより忠実に再現できる生理学的モデルが必要不可欠である。**

本研究代表者らは、解剖学的データと MRI データに基づき、すべての発話器官と関連筋肉を取り入れ、生理学的発話機構モデルを構築し、人間の発話運動の全過程を実現することに成功している[5]。**舌の生理学的モデルについては**、仏国立科学研究センターGISPA-Labとカナダ UBC 大学もモデルの構築を中心として研究を展開している[6]。ただし、発話目標から、筋の収縮、発話動作及び音声合成までの音声生成の全過程を実現できたのは我々の生理学的発話機構モデルのみとなっている。

さらに、我々は、変形聴覚フィードバックを用い聴覚に摂動を与えて、調音運動と筋電図などの計測により、音声生成と知覚との相互作用に関する脳の機能を研究してきた[7,8]。

本研究は、これまでの我々の研究成果を踏まえ、生理学的発話機構モデルを基に、脳の高次機能から末梢器官の生理物理的な機能までを忠実に実現できる、生理学的ニューロン計算モデルを構築し、言語音声の獲得や音声生成のメカニズムを模擬することを目標とする。

２．研究の目的

人間の音声生成では、発話計画や調音器官の運動制御、聴覚による発話のスクリーニングなど、脳の高次機能から生理物理運動までの過程が関わっている。**人間の音声生成機能に関連する、言語音声の獲得や発話障害、未解明問題が山積しているが、技術的・倫理的な制限のため、実験的な手法により解明することが困難である。**そこで、本研究は、生理学的発話機構モデルを基に、発話計画、発話運動制御、視覚・聴覚フィードバックなど脳の高次機構を取り入れ生理学的ニューロン計算モデルを構築し、モデルによるシミュレーションと計測データとの比較により、人間の音声生成過程および発話障害を解明することを目的とする。

３．研究の方法

(1)　人間の発話機能を忠実に実現するために、観測データに基づいて、本研究グループが開発した生理学的発話機構モデルをさらに精密化する。そこで、我々はカナダ UBC 大学研究グループによる開発したツールキットを利用して、現有の生理学的発話機構モデルを離散型から連続型のモデルに発展させる。

(2)　DIVA モデルを参考にして、高次機能のモジュール化を行う。これと同時に、精緻化した生理学的発話機構モデルを取り入れるために、人間の音声生成における「運動計画→運動コマンド→筋肉の収縮→発話器官の動作」という生理学的なプロセスの定式化を検討する。音声生成の生理学的ニューロン計算モデルの構築を行う。構築したモデルに対して、パラメータを変動させながら数値シ

---

[1] 能田,本多 (2004)“fMRI による発話中枢機構の観測”, 音声研究, 8-2, pp.28-34.

[2] 鈴木(2003) “言語性コミュニケーションの脳内機構発話の脳内機構：皮質電気刺激による検討”, 神経心理,19.01

[3] F. Guenther, et al. (2006) “Neural Modeling and Imaging of the Cortical Interactions Underlying Syllable Production,” Brain & Language, 96, pp. 280-301.

[4] J. Kröger, et al (2009) “Towards a neurocomputational model of speech production and perception,” Speech Communication, 51, pp. 793-809.

[5] Q. Fang, S. Fujita, X. Lu, J. Dang, “A model-based investigation on activation of the tongue muscles in vowel production,” Acoustics of Science and Technology, 30, 277-287 (2009)

[6] S. Buchaillard, et al (2007). “3D statistical models for tooth surface reconstruction,” Computers in Biology and Medicine, 37(10), pp. 1461-1471.

[7] Y. Koba, J. Dang, “Investigation of relations between capabilities of speech production and phonemic restoration,” International Workshop on NCSP, Australia, 335-338 (2008)

[8] J. Dang, K. Akagi, K. Honda, “Communication between Speech Production and Perception within the Brain -Observation and Simulation-,” J. Computer Science and Technology, 21(1), 95-105 (2006)

ミュレーションを行い、モデルの動作を確認する。

(3)　人間が新しい言語のカテゴリを学習するメカニズムに着目して、EMA 装置を用いて、調音器官(舌、唇または顎など)の生理データと発話音声を計測する。また、不慣れな言葉の復唱により人間の学習過程を追跡することによって､図１に示した聴覚マップ、聴覚―音声マップ、調音運動計画などの各モジュールの機能を考察する。

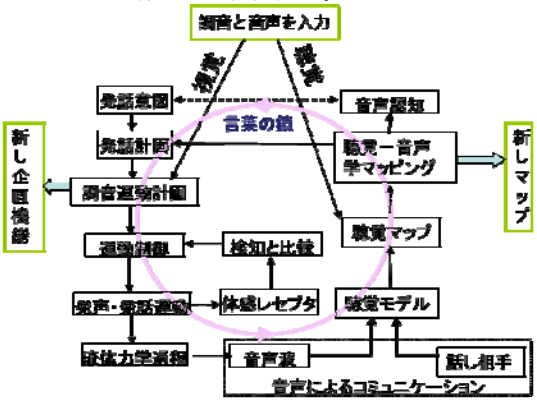

図１　音声生成と知覚の相互作用

(4)　ERP を用いて脳活動の観測を観測し、音声と画像を用いて、多感官情報整合のメカニズムを考察した。

４．研究成果

(1)　<u>生理学的発話機構モデルの精密化</u>

本グループが開発した生理学的発話機構モデルを基に、カナダUBC大学研究グループによる開発したツールキットを利用して、現有の生理学的発話機構モデルを離散型から連続型のモデルに発展させた。連続型生理学的発話機構モデルを図２に示す。モデルの計算速度は実時間の１０倍となった。従来より6倍早くなった。

発話機構モデルを精密に制御するため、新しい制御方法を検討した。まず、従来数個離散点の代わり、舌と下顎の輪郭を用いる制御法の開発を試みた。筋活動の組み合わせにより調音運動を生成し、主成分分析により調音空間を構築して、調音目標から運動指令空間への投影関係を創立した（図3を参照）。発話制御には、フィードフォーワード制御が主に熟練な発話の制御方式とされるが、フィードバック方式は調音目標から運動指令空間へのフィードフォーワードマッピングの学習には役を果たしている。本研究では、フィードバックとフィードフォーワードと取り入れ、新しい制御を構築して、モデルを用いて検証した。

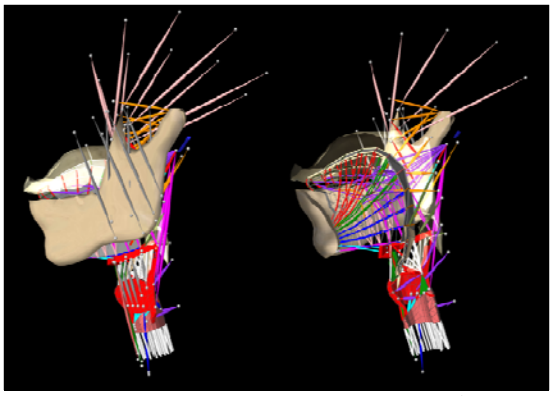

図2　連続型生理学的発話機構モデル

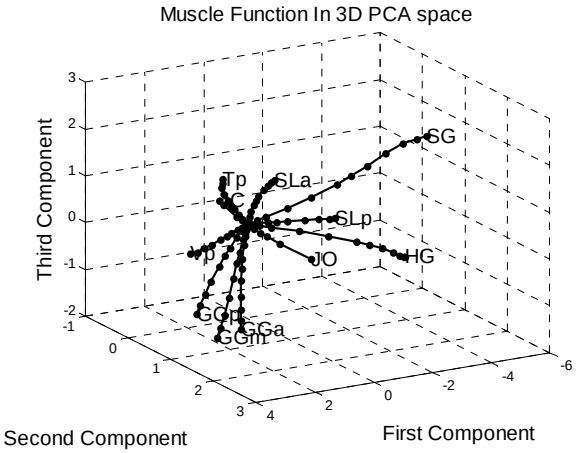

図3　各調音筋の収縮力と輪郭平衡位置の移動

(2) <u>発話過程のニューロンモデルの構築</u>

これまでニューロン計算モデルに関する研究では、発話運動を実現する末梢器官には幾何学的調音モデルが用いられていたため、脳の指令に対する人間の生理学的な働きを再現するには不十分であった。そのため、本研究では、音声生成過程の末梢器官モデルとして生理学的発話機構モデルをに取りいれ、ニューロン計算モデルを構築して、高次機能のモジュール化を行う。人間の音声生成における「運動計画→運動コマンド→筋肉の収縮→発話器官の動作」という生理学的なプロセスの定式化を検討した。提案したニューラル発話制御モデルの略図を図４に示す。さらに、ハイレベルモータ表現（制御パラメータ）とローレベルモータ表現（筋の活動パターン）への写像（Execution Map）を学習する。そのため、発話モデルの制御パラメータ（舌尖、舌背、下顎と口唇）を１６個のニューロンで表示し、調音筋を１５個のニューロンで表す。それを図５に示す。

幼児がべちゃくちゃしゃべりながら言語音声を学習すると同時に発話器官の制御方法を身につける。我々はニューロン制御モデルを用いてモータコマンドの学習を模擬した。図６に示した結果より、合理的なハイレベルモ

ータ表現とローレベルモータ表現への写像を学習できたことを示唆している。

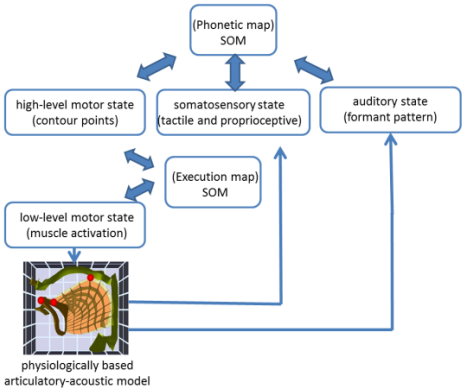

図４　提案したニューロン制御モデル

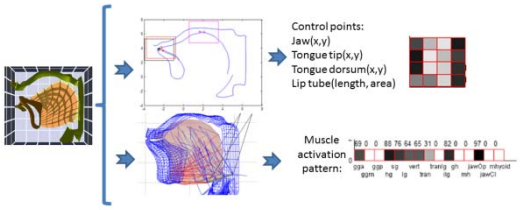

図５　制御パラメータとそのニューロン表示

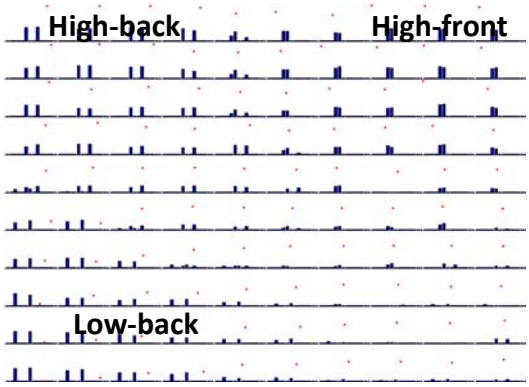

図６調音位置とモータニューロンとの関係

(3) 非母語学習過程における音声生成・知覚との関連

本研究では、復唱による母音学習過程に着目し、長期間にわたって、被験者の復唱音声と内観を分析することで、学習において利用される音響特徴量や学習過程における知覚カテゴリの変化を考察した。

日本人被験者に対して英語母音/æ/、ロシア語/ju/などなじみのない非母語母音を用い、実験のパート１では音声のみを被験者に提供して復唱させたが、パート２ではターゲット話者の調音運動を見せながら学習させた。復唱音声と調音位置を磁気センサシステム（EMA）により同時に測定した。計測した調音運動とフォルマントの変化を図７に示す。学習の初期段階で、フォルマントは徐々にターゲットのそれ（水平線）に近づいて、調音運動を見せた（縦破線で表示）時、発話運動とフォルマントと共に大きく変化された。それは、調音運動の視覚情報により発話学習を促進させることを示唆している。

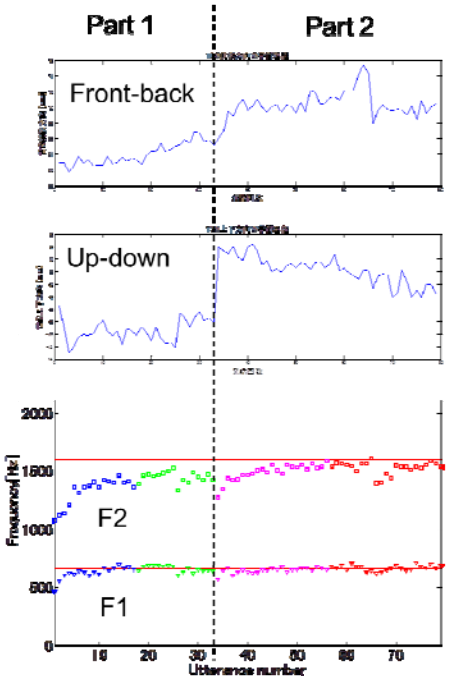

図７　/æ/の学習過程（調音前後・上限運動、F1・F2）

複数話者に対して、非母語の学習過程と母語の音響空間との関連を検討した。二人の被験者の母語音響空間を図８に示す。ターゲット音声（/æ/, /er/, /oo/, /ih/, /ju/）をその上にプロットし、復唱による学習過程には被験者１（黒マーカ）と被験者２（赤マーカ）はそれぞれ異なる初期位置からスタートし、異なる位置に収束した。しかしながら、各被験者の母語音響空間におけるターゲット音声の収束点の相対的位置は大体一致している。それは、なじみのない音声を知覚するとき、まず被験者の母語音響空間に投影し、バーチャルターゲットを形成し、それにそのバーチャルターゲットに接近するように学習していることを示唆した。

被験者の内観から学習途中には目標母音に対する知覚カテゴリの変化が生じたものと生じなかったものがあることが分かった。これは、おそらく人間の調音と同様に知覚空間にも安定した場所と不安定な場所が存在しており、音響的にも調音的にも安定した場所を探索するように学習が行われている可能性が考えられる。

(4) 脳電信号により多感官情報整合の研究

我々は、短期多感官記憶が多感官情報による認知整合への影響を研究するため、元に生態的に関連のない音声と画像を試料として、被験者にその関連を学習させた上、多感官の意味プライミング実験を行った。その結果、プライミング画像及び目標音声は多感官の記憶と一致したとき、被験者の応答時間が短く、しかも100-160ms gamma波領域での活動は不

一致である場合より著しく増強された。不一致である場合では、N400の効果を引き起こす。N400 の変化は、音声からマルティモダリティの意味ネットワークへの射影過程には意味ネットワークが短期多感官記憶の影響を受けられることを示唆していると思われる。gamma 波領域活動の増幅は、音声と画像による刺激とワーキングメモリーにある多感官記憶へのマッチング過程を反映している。N400とgamma 波の源を限定した結果は,MTGが多感官情報の意味マッチングにより促進された。本研究結果は、短期多感官記憶による多感官の情報間の関連と長期記憶は多感官意味プライミングの変調には特に差はなかった。

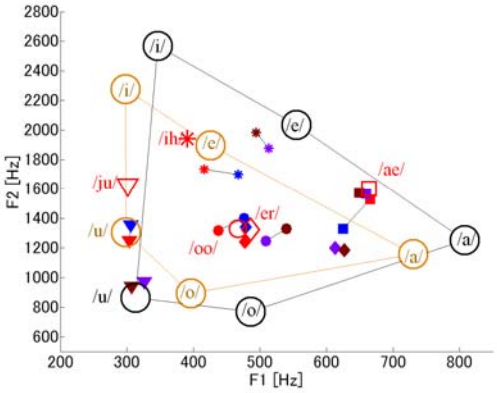

図８　非母語の学習と母語の音響空間

５．　主な発表論文等

〔雑誌論文〕（計６件）

1. B. Liu, X. Meng, G. Wu, J. Dang “Correlation between three-dimensional visual depth and N2 component: Evidence from event-related potential study”, Neuroscience, （査読有）, Vol. 237, 161-169, (2013).
2. Erickson D, Suemitsu A, Shibuya Y, Tiede M. “Metrical structure and production of English rhythm,” Phonetica, (査読有), Vol. 69, 180-190 (2012).
3. D. Ying, Y. Yan, J. Dang, and F. Soong ″Voice Activity Detection Based On An Unsupervised Learning Framework″, IEEE Trans. Audio, Speech and Language Processing, (査読有), Vol. 19, No. 8, 2624 - 2633 (2011)
4. K. Fujii, Q. Fang and J. Dang ″Investigation of Auditory-Guided Speech Production while Learning Unfamiliar Speech Sounds″ Journal of Signal Processing, (査読有), Vol. 15, No. 4, pp. 287-290 (2011)
5. X. Wu, J. Wei and J. Dang ″Study of Control Strategy Mimicking Speech Motor Learning for a Physiological Articulatory Model″ Journal of Signal Processing, (査読有), Vol. 15, No. 4, pp. 295-298 (2011)
6. A. Nishikido and J. Dang “A model-based investigation on one-to-many relationship between speech sound and articulations,” J. Acoust. Soc. Jpn, (査読有), Vol. 67, No. 1, pp. 1-12, (In Japanese) (2011)

〔学会発表〕（計 24 件）

1. 西村, 川本, 党“発話機構モデルを用いた茎突舌筋の形態・機能学的検討,”日本音響学会春季研究発表会, 東京工科大学、東京, 2013, 3, 13
2. C. Liu, J. Wei, B. Feng, J. Dang “An Anisotropic Diffusion Filter for Reducing Speckle Noise of Ultrasound Images Based on Separability”, APSIPA, Hoolywood, USA, 2012, 12, 3.
3. C. Song, J. Wei, Q. Fang, Y. Wang, J. Dang, “TONGUE SHAPE SYNTHESIS BASED ON ACTIVE SHAPE MODEL”, ISCSLP 2012, Hong Kong, 2012, 12, 6
4. C. Zhao, H. Wang, S. Hyon, J. Wei, J. Dang, “ Efficient feature extraction of speaker identification using phoneme mean F-ratio for Chinese”, ISCSLP 2012, Hong Kong, China, 2012, 12, 5
5. Y. Wang, H. Wang, J. Gao, J. Wei, J. Dang “Detailed morphological analysis of Mandarin sustained vowels,” ISCSLP, Hong Kong, China, 2012, 12, 6
6. Y. Wang, H. Wang, J. Wei, J. Dang “Mandarin vowel synthesis based on 2D and 3D vocal tract model by finite-difference time-domain method,” APSIPA, Hoolywood. USA, 2012, 12, 3
7. S. Hyon, H. Wang, J. Dang, “A research of dependencies between frequency components and speaker characteristics based on phoneme mean F-ratio contribution”, ISCSLP 2012, Hong Kong, China, 2012, 12, 7
8. 藤井, 末光, 党“母語による第二言語音韻知覚への影響に関する考察,” 日本音響学会秋季研究発表会, 信州大学, 長野, 2012, 9, 19
9. 西村, 川本, 党 “解剖学的知見に基づ

く生理学的発話機構モデルの筋配置の個人化に関する検証,” 日本音響学会秋季研究発表会, 信州大学, 長野, 2012,9,19
10. S. Hyon, H. Wang, C. Zhao, J. Wei, J. Dang “A method of speaker identification based on phoneme mean F-ratio contribution”, Interspeech, Portland, USA, 2012,9,13
11. 西村, 川本, 党 “MR 画像に基づいた変形による生理学的発話機構モデルの個人化,” 日本音響学会聴覚研究会資料,42(4), 357-362, NTT 厚木研究開発センタ, 2012,6,14.
12. A. Li, Q. Fang, Y. Jia and J. Dang. Successive Addition Boundary Tone in Chinese Disgust Intonation, NACCL24, San Francisco, USA, 2012, 6, 9
13. A. Li, Q. Fang, J. Dang. Emotional Expressiveness of Successive Addition Boundary Tone in Mandarin Chinese. Speech Prosody 2012. Shanghai, China, 2012, 5, 22
14. Y. Wang, H. Wang, J. Wei, and J. Dang ″Acoustic analysis of the vocal tract from a 3D physiological articulatory model by finite-difference time-domain method,″ The Inter. Conference on Automatic Control and Artificial Intelligence, Volume 5, 3265-3269, Xiamen, China, 2012, 3, 24-26
15. D. Ying, X. Lu, J. Li, Y. Yan, J. Dang, F. Soong “Noise Estimation Using a Constrained Sequential HMM In Log-Spectral Domain,” ICASSP 2012, Kyoto, Japan. 2012,3,27
16. 西村, 川本, 党, “形態学的情報に基づく個人化発話機構モデルの構築,”日本音響学会春季研究発表会, 1-R-11, pp.451-454, 神奈川大, 2012,3,13
17. K. Fujii, A. Suemitsu, J. Dang, “Investigation of perceptual effects during learning process via vowel imitation,” RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP’12), pp.393-396, Hawaii, USA, 2012, 3, 5
18. N. Nishimura, S. Kawamoto and J. Dang, “Morphological personalization according to human mechanism using MR images,” RISP International workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP12), pp. 497-498, Hawaii, USA, 2012, 3, 5
19. D. YING, Y. YAN, J. DANG, F. SOONG, “Noise Power Estimation Based on a Sequential Hidden Markov Model,” the 8th International Conference on Information, Comm. and Signal Processing (ICICS 2011), Singapore, 2011, 12, 15
20. 藤井, 末光, 党, “復唱による母音学習過程における音声知覚に関する考察,”聴覚研究会資料, vol.41, no.9, pp.689-694, 熊本県立大, 2011, 12, 11.
21. Li, A., Fang, Q. & Dang, J., “Emotional intonation in a tone language: experimental evidence from Chinese,” ICPhS’2011, Hong Kong, China, 2011, 8, 18.
22. X. WU, Q. FANG and J. DANG “Inverse estimation of motor command based on a 3D physiological articulatory model,” Spring Meeting of Acoustic Society of Japan, Tokyo, Japan, 2011, 3, 9.
23. D. Ying, Y. Yan, J. Dang and Soong F, Noise Power Estimation Based on a Sequential Gaussian Mixture Model, in: the 4th International Conference on Image and Signal Processing, pages 2388-2391, Shanghai, China, 2011, 10, 16
24. D. Ying, J. Li, Q. Fu, Y. Yan and J. Dang, Voice Activity Detection Based on a Sequential Gaussian Mixture Model, in: the APSIPA Annual Summit and Conference, Xi’an, China, 2011,10,20.

６．研究組織

(1) 研究代表者

党 建武 (Toh 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

(2) 研究分担者

徳田 功 (Tokuda Isao)

立命館大学理工学部・機械工学科・教授

研究者番号：00261389

（参加期間 2010－2011）

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