

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24年6月1日現在

機関番号：16201

研究種目：基盤研究（C）

研究期間：2009～2011

課題番号：21500517

研究課題名（和文） 聴覚障害が引き起こす音声の不明瞭要因の解析と対話型発話訓練
ロボットの構築研究課題名（英文） The analysis of unclear vocal articulations of auditory-impaired
and the construction of an interactive training system using a talking robot

研究代表者

澤田 秀之（SAWADA HIDEYUKI）

香川大学・工学部・教授

研究者番号：00308206

研究成果の概要（和文）：本研究ではまず、発話動作をおこなう器官を全て機械系によって構成した発話ロボットを開発した。本ロボットは、人間の乳幼児のように聴覚フィードバック学習によって自律的に発話動作の獲得が可能であり、物理的な発話動作と、そこから生成される音声の関係を、適応的に獲得できる。そこで、音声の不明瞭要因を抽出、分類することにより、その要因となっている口内動作を発話ロボットによって再現し、発話障がい者のための対話による訓練システムを構築した。

研究成果の概要（英文）：In this study, a vocalization training system for the auditory-impaired was constructed by applying a talking robot, which has mechanical organs like a human. With an adaptive learning strategy using an auditory feedback control, the robot autonomously learned the human vocalizations, and then reproduced the speech articulations from inputted vocal sounds. By analyzing the acquired relations between vocal articulations and the generated vocal sounds stored in the robot, unclear vocal sounds given by auditory-impaired people were classified, and then particular features were extracted, which were associated with vocal articulations reproduced by the robot. The robot was employed in the speech training for auditory-impaired, and the training results verified the effectiveness of the interactive training employing the robot.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,500,000	450,000	1,950,000
2010年度	1,200,000	360,000	1,560,000
2011年度	700,000	210,000	910,000
年度			
年度			
総計	3,400,000	1,020,000	4,420,000

研究分野：ロボット工学

科研費の分科・細目：人間医工学、リハビリテーション科学・福祉工学

キーワード：発話ロボット、発話訓練、聴覚障がい者、支援技術、音声、聴覚フィードバック、自己組織化ニューラルネットワーク、構音障害

1. 研究開始当初の背景

人間の音声は、肺、声帯、気道、舌、軟口蓋、顎、唇、鼻腔などから成る音声生成器官

の複雑な運動によって生成され、特定の音響のパターンである「言葉」として聞き手に知覚される。話者が音声によって聞き手に情報

を伝えるとき、まず音声生成のための発話動作指令が脳内で生成され、運動神経を介して音声生成器官へ送られる。この指令を基に、それぞれの音声生成器管が適応的に動作をおこない、声帯振動による音源生成から声道形状の形成、鼻腔共鳴の付加などによって音声生成される。発話された音声は同時に、骨導と気導によるフィードバックにより聴覚神経を介して発話者自身にも伝わり、常に音声生成器官が適応的にコントロールされて、安定した発話が実現されている。以上の音声生成過程と聴覚フィードバックは、発話時に話者にとって無意識のうちにおこなわれている。

このような能力は、乳幼児が生まれてから数年のあいだに言語を習得していく過程において、発声と聞き取りの試行錯誤を繰り返すことによって獲得されるものである。また音声獲得後であっても、人間は練習によって声真似をすることができる。これは、発声のために必要な器官が生まれながらに備わっている一方で、発話技術は学習によって後天的に獲得されるものであるためと考えられる。例えば、聴覚障がいや、脳性麻痺などの動作器官の障害によりフィードバック学習が正常に行なわれない場合には、音声の獲得、生成が困難となる。

申請者はこれまでの研究で、人間と同等の発話機構を全て機械的に再現した発話ロボットを構築し、聴覚フィードバック学習により、自律的に発話動作を獲得できることを示してきた。特に日本語音声の発声に着目し、日本人男性の平均値を基に各発話器官を製作して、日本語5母音、摩擦音、鼻音の発話と、声帯および空気流の制御と適応学習により、歌声の生成が可能であることを示した。発話ロボットによる発話動作の自律的な獲得には、ニューラルネットワーク(NN)を適用した聴覚フィードバック学習を提案してきた。NNにより、目的とする音声の音響特徴とモータ制御パラメータを適応的に対応づけることが可能であり、見本となる人間の音声を与えれば、学習によって同様の音声を生成するための発話動作が獲得できることが解った。

一方、食道発声患者の音声および、老化や脳性麻痺にともなう発声障がい者音声の明瞭化フィルタリングの研究では、音声の不明瞭要因の特定と、その音響的特徴の抽出をおこなってきた。またこれらの不明瞭音声を、発話ロボットによって物理的に再現できる可能性を示した。

聾学校やリハビリテーション病院における聴覚障がい者の発声訓練は、言語聴覚士(ST)が個々の障害に応じた手法でおこなっている。具体的には、鏡を使って口内の形状を確認しながらおこなうものや、触振動覚を

利用して発話を教示するといった方法がある。しかし発話時の口内形状は、口外からの観察が困難なために様々な試行錯誤を繰り返すこととなり、正しい発話の獲得には長い年月と多大な努力を要する。また、このような訓練はSTと一対一でおこなわれることが基本であるが、聴覚障がい者の数に対して言語聴覚士の数が圧倒的に不足していることも問題となっている。そこで障がい者が、専門的な知識が無くとも、手軽にかつ直感的に自学自習の発話訓練が可能な装置、システムの実現が望まれている。

2. 研究の目的

本研究ではまず、発話動作をおこなう器官をエアポンプ、人工声帯、声道共鳴部、鼻腔部など、全て機械系によって構成されるポータブルな発話ロボットを構築する。計算機による聴覚フィードバック制御によって、ロボットが自律的に発話動作を獲得し、任意の音声を生成することができる学習機構を実装する。ロボティクス、ヒューマンインタフェース開発の視点から、人間と同等の発話器官を用いて音声対話が可能な発話ロボットの開発を進めると共に、発話障がい者のための対話型発話訓練システムの構築をおこなう。また本ロボットが音声模倣学習によって獲得できる発話動作および音声を詳細に解析することにより、発話障がい者音声の不明瞭要因を解析し、更に人間の音声獲得機構の解明に繋げていく。

3. 研究の方法

人間と同様の音声生成器官を持ち、自律的に発話動作を獲得する発話ロボットを構築し、聴覚障がい者に対話的に発話動作を教示するシステムの構築を行った。本報告書では、主に発話ロボットの構築と、2つの自己組織化マップを組み合わせた dual-SOM の提案によるロボットの柔軟な自律的音声獲得ならびに、障がい者音声の再現について述べる。

(1) 発話ロボットの構築

本研究で構築した発話ロボットは、エアコンプレッサ、人工声帯、共鳴管、鼻腔、音声分析器に接続されたマイクロフォンから構成され、それぞれ人間の肺、声帯、声道、鼻腔、聴覚に対応している。図1にロボットの外観を示す。



図1 発話ロボットの外観

発声手法として、まずエアコンプレッサから圧縮空気を送り出す。送り出された空気は減圧弁、流量制御弁を経由し、人工声帯を振動させて音源を生成する。このとき流量制御弁により、生成される音声の音量と高低を変化させることができる。音源は、共鳴管としての役割を果たす声道に導かれ、声道形状を変化させることにより、様々な音響特性を音源に付加することが可能となっている。人間は軟口蓋により口腔と鼻腔を遮断することが可能であるが、本ロボットは、声道と鼻腔接続部に回転弁を持ち、鼻腔共鳴を制御することにより鼻音/m/や/n/の生成が可能である。また声道内部にシリコンゴムで成形された舌を持ち、これをモータにより上げ下げすることで、側音/l/や破裂音/t/などを生成することが可能である。音響分析器は、聴覚と脳に対応しており、発話ロボットが生成する音声を実時間で解析し、発話動作の獲得に必要な特徴量を抽出する。システムコントローラは生成音声の音響特徴量に対応するモータ制御量を算出する機能を持ち、これら2つの間の関係を聴覚フィードバック学習により学習する。

(2) dual-SOMによる発話動作の獲得

人間の発声においては、母音は定常的な声道形状を形成することにより生成され、子音は顎や舌、軟口蓋などの非定常的な動作から引き起こされる変化により生成される。よって、任意の音を発声するための声道形状を決定することができれば、ロボットがその発声器官を使用して口を動かしながら発話動作を学習することが可能となる。加えて似たような音響特徴を持つ音声は、似たような声道形状により生成されることは、我々の発話中の口内形状からもわかる。

本研究では、人の発話動作の獲得機能に着目した。人は聴覚フィードバック学習により、目標となる音声を教師信号として、自身の音声と比較しながら試行錯誤的に口内の発話動作を獲得していく。この機構を発話ロボットに実装し、人間のように柔軟に発話学習を可能とするために、2つの3次元自己組織化マップ(Self-Organizing Map: SOM)を使った自律学習アルゴリズムを提案した。

まず発話動作と生成音声の対応関係の学習のために、3次元特徴マップ上に音響特徴量を配置した3D-Phonetic_SOMを構築した。

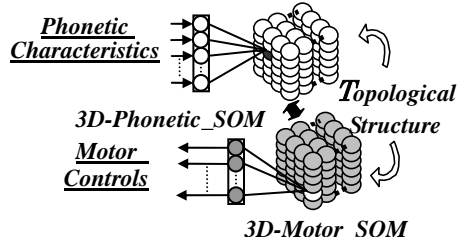


図2 dual-SOMの構造

加えて、3D-Phonetic_SOM と、モータ制御量を用いて作成される 3D-Motor_SOM との対応関係を学習する dual-SOM を構築し、声道形状と音声の対応関係を学習させた。図2に dual-SOM の構造を示す。

dual-SOM の学習としては、まず発話ロボットをランダムに動作させ、各モータの制御量とそこから生成される音声の音響特徴量の組を得る。得られた音響特徴量を 3D-Phonetic_SOM 上に配置する。マップ上では、近傍学習により互いに似たような特徴量どうしが位相関係を保持した状態で近隣に配置される。次に 3D-Phonetic_SOM の学習に使用した音響特徴量と対応関係にあるモータ制御量を用いて、3D-Motor_SOM の学習を行う。2つの SOM には、互いに同等な近隣関係をもつ位相構造が構築されている。本研究ではこのことを用いて、2つの3次元 SOM の対応関係を、勝者素子近傍の距離比により構築した。学習終了後は 3D-Phonetic_SOM 上の任意の一点を選択することで、対応する 3D-Motor_SOM 上の点が選択され、その素子に配置されているモータ制御量が出力される。なお本研究では、サンプリング周波数 8[kHz] で計算された 10次元の LPC ケプストラムを音響特徴量として用いた。

モータ制御量と各々生成される音声の特徴量 400 個の組を、それぞれ素子数 10x10x10 (合計 1000 素子) の 3次元マップ内に配置した結果、良好にパターン同士の近隣関係ならびに、モータ制御量と音響特徴量の対応関係が学習できた。これにより、dual-SOM の学習によって、発話ロボットが自律的に生成音声と発話動作の関係を構築できることがわかった。

(3) 声道断面積関数を用いた発話動作の検証

形状と音声の対応関係が、適切に構築されているかの検証を行った。学習の妥当性について検証するためには、発話ロボットの生成音声と話者の実音声を比較するとともに、ロボットと人間の声道形状について調査検討する必要がある。そこで、声道断面積関数を用いて発話者の音声から声道形状を推定し、dual-SOM により推定されたロボットの声道形状と比較した。

音声発声時の声道形状を推定する方法としては、MR 画像を用いる手法がある。発話特性解析や声道物理モデルを構築する際に応用されているが、未知の音声に対する形状が推定できない、特定話者の形状しか得られない等の問題がある。そこで、ソフトウェア的手法として、声道断面積関数を用いることとした。人間の声道を短い筒状の音響管が接続したものとして近似し、その接続面の反射係数を用いて声道内の進行波を考えると、境界面の反射係数は偏自己相関 (partial auto-correlation) 係数に等しくなる。この時、

音響管の直径は波長に比べて十分に小さいとすると、式(1)を用いて声道断面積が推定できる。

$$S_m = \frac{1 - k_m}{1 + k_m} S_{m-1} \quad m=1,2,\dots,8 \quad (1)$$

式中の k_m はPARCOR係数、 S_m は断面積をそれぞれ表しており、発話ロボットの声道の形状を制御しているモータ数と次数を同じとするために $m=8$ とし、人間の声道を8つの断面積が異なる音響管で近似した。

実音声で実験を行った結果、例えば/a/の発話は、口唇から中舌までの区間が広がり、後舌部で狭まり、喉頭付近で再度広がる傾向にある。推定された声道断面積関数も同様の傾向を示しており、1から4次の口唇から中舌に対応する区間が広がり、5次の後舌部で狭まり、喉頭付近を示す6から8次の項で再度広がりを見せた。また、/i/についても人間と同様の傾向が見られた。これらのことから、良好に話者の声道断面積が推定されることがわかった。

(4) 発話ロボットと人間の声道形状

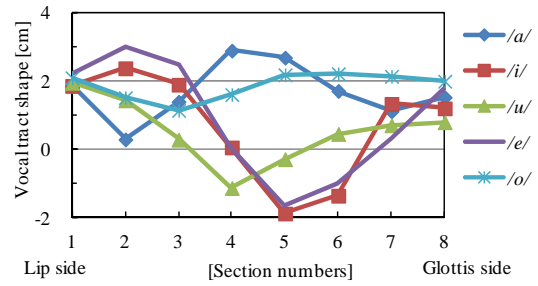
dual-SOMにより獲得された発話動作の妥当性を、人間の実音声から得られる声道断面積関数を用いて検証した。被験者1名が発話した日本語5母音をdual-SOMに入力し、発話ロボットが再現した声道形状とその時の音声を、話者の実音声と声道断面積関数により推定された声道形状と比較した。

被験者の声道断面積をロボットの声道形状と比較するため、音響管それぞれの断面積から直径を算出し、声道形状を推定した。図3に、5母音音声から推定された話者の声道形状(a)、dual-SOMにより推定されたロボットの声道形状(b)、被験者とロボットの声道形状の差(c)、そしてロボットと被験者の生成音声から得られた音響特徴量(d)を示す。なお(d)の横軸と縦軸は、それぞれ第一、第二フォルマントを示している。

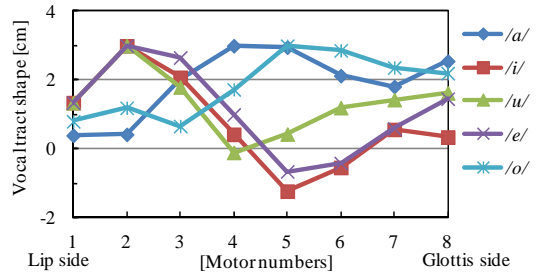
dual-SOMを用いて推定された声道形状と、声道断面積関数を用いて推定された発話者の声道形状は、全体的に似たような特徴を示すことが解った。しかしながら/u/については、他の母音と比較すると、人間とロボットとで形状に差が見られた。特に2から4番モータの値に差があることが解った。この問題は、ロボットに口をすぼめる機構がないために、/u/の発話に必要な断面積を十分に確保できずに生じた結果と考えられる。今後はロボットの声道部を改良し、口の突き出し機構を付加していく必要がある。

また生成された音声の第一、第二フォルマントは、5母音について話者のものと近い特徴を示した。聴取したロボットの生成音声は、話者のものと類似しており、dual-SOMが良好に声道形状と音声の対応関係を学習して

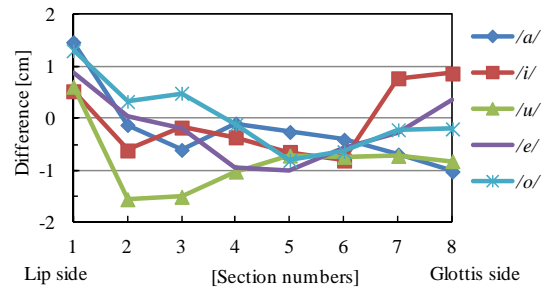
いることを確認した。



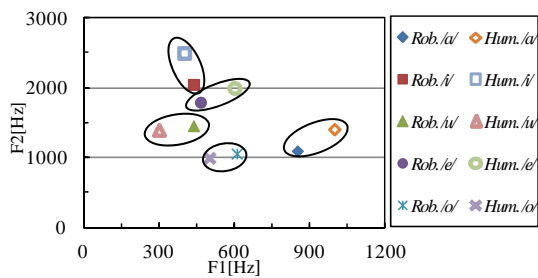
(a) 人間の推定声道形状



(b) 発話ロボットの声道形状



(c) 人間とロボットの声道形状の差

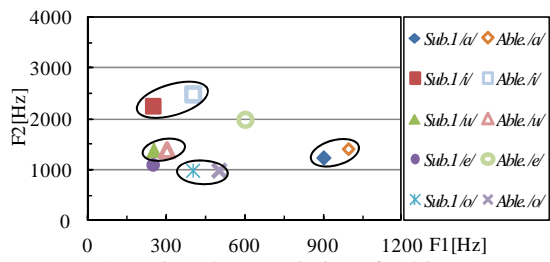


(d) 人間とロボット音声のフォルマント比較
図3 声道形状とフォルマントによる比較

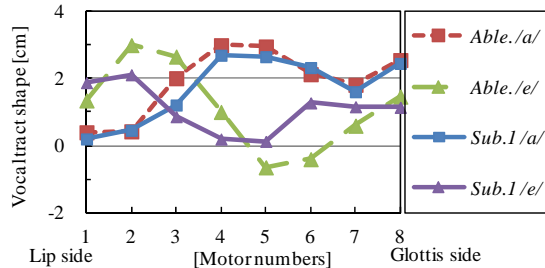
4. 研究成果

(1) 聴覚障がい者の発話動作の再現

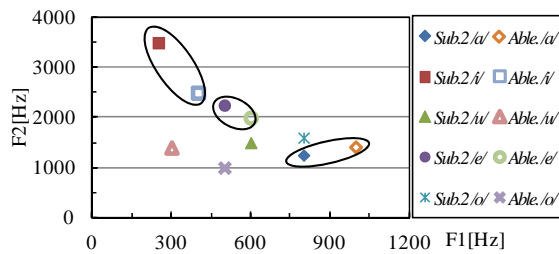
dual-SOMに聴覚障がい者の不明瞭な音声を入力し、ロボットによる発話動作の再現をおこなった。香川県立聾学校にて、様々な障害レベルの生徒から日本語5母音の発話音声を収録した。dual-SOMにより声道形状を推定し、発話ロボットを用いて発話動作を再現した。図4に、聴覚障がい者3名の音声の音響特徴により再現されたロボットの声道形状を示す。図(a-1)、(b-1)、(c-1)中のマーカは、



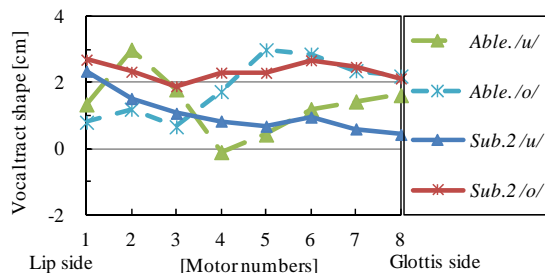
(a-1) Voice characteristics of subject 1



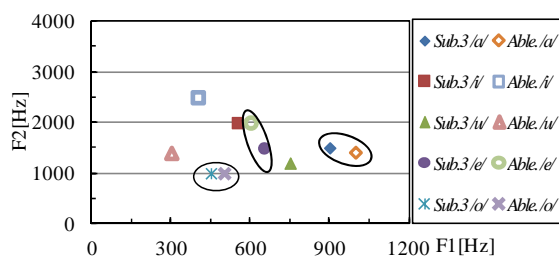
(a-2) Estimated vocal tract shapes of subject 1



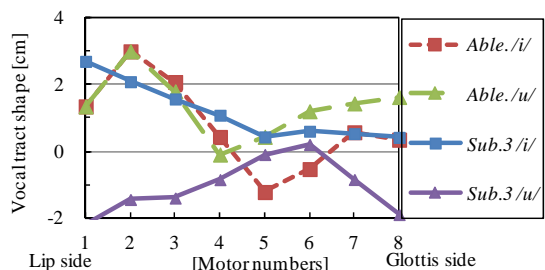
(b-1) Voice characteristics of subject 2



(b-2) Estimated vocal tract shapes of subject 2



(c-1) Voice characteristics of subject 3



(c-2) Estimated vocal tract shapes of subject 3

図4 聴覚障がい者の声道形状と音響特徴量

それぞれ健聴者(Able.)と聴覚障がい者(Sub.)が日本語5母音を発話した際の第一、第二フォルマントを示している。

被験者1の音声は、/e/を除いた母音については健聴者とほとんど同じような特徴を示した。ロボットにより再現された/a/の声道形状は、健聴者のものと似たような傾向を示しており、良好に発話がおこなえていることがわかる。一方/e/に関しては、/u/に似たような音響的特徴を示しており、再現された声道形状も/u/に近いものであった。健聴者の発話により再現された形状と比較すると、前舌から後舌に対応する2から4番モータの値が小さくなっており、これは本被験者が発話の際に舌を下げていることを示している。舌位置に注意して発話訓練を行うことで、被験者1の発話は改善されると考えられる。

被験者2の音声は、/a/, /i/, /e/については健聴者と似たような音響特徴を示したが、/u/と/o/についてはそれぞれ/e/と/a/に近い特徴を示した。また、ロボットにより再現された声道形状を確認すると、被験者2の/u/と/o/の口内形状にほとんど変化がなく、健聴者の音声により再現された形状と大きく異なっていた。/u/については、口唇部に対応する1番モータの値が大きく、前舌に対応する2番モータの値が小さい。これは被験者2が、/u/の発話において重要な口唇の狭めと前舌部の広げを十分に行えていないことを示している。また、/o/の場合1から4番モータの値が大きく、/o/の発話に必要な口唇部から後舌付近の広げが十分に出来ていないことがわかる。被験者2は、これらの点を考慮しながら発声訓練を行う必要があると言える。

被験者3は、/i/と/u/の発話音声は健聴者のものと異なっていた。特に/i/の音響特徴は/e/のものに近い傾向が見られた。再現された声道形状を確認すると、/i/については、後舌から喉頭部に対応する5番モータの値が大きく、/e/に近い形状になっていることがわかる。また/u/については、口唇から後舌に対応する箇所が大きく開いており、かつ喉頭部の狭めが不十分であった。このため、音響特徴は/a/や/o/に近い値を示した。被験者3はこれらの点に注意して発声練習を行うことにより、音声の改善が見込める。

(2) 研究の総括

本研究では、発話ロボットを用いた聴覚障がい者の発話動作の再現と、対話型発話訓練システムの開発をおこなった。人間と同様の音声生成器官を持ち、聴覚フィードバックにより自律的に発話動作を獲得することができる発話ロボットを構築し、dual-SOMの自律学習によってロボットの声道形状と生成音声を柔軟に対応づけることが可能であることを示した。ロボットによる自律学習の妥当性を検証するために、人の実音声から声道

断面積関数を推定し、dual-SOM により再現されたロボットの声道形状と比較した。その結果、dual-SOM は声道形状とそこから生成される音声の対応関係を良好に学習できていることが解った。

更に、発話ロボットを用いて、聴覚障がい者の発話動作と発話不明瞭要因の解析を行った。ロボットシステムにより訓練者は、予めロボットにより再現された健聴者の口内形状と、自身の発話音声により再現される口内形状の違いを、視覚的に確認することができる。また dual-SOM を用いることで、健聴者の音声と訓練者の音声の違いを、マップ上の距離として呈示することも可能である。今後は更に、母音だけでなく子音の発話動作を再現できるシステムを構築し、被験者が声道形状や舌の動きを直観的に理解できるように訓練手法の検討を進めていく。

5. 主な発表論文等

[雑誌論文] (計 2 件)

- [1] 木谷光来, 原達矢, 澤田秀之: 「Dual-SOM の位相構造学習に基づく発話ロボットの自律的音声獲得」, 日本機械学会論文集 C編, Vol.77, No.775, pp.1062-1070, 2011
- [2] Mitsuki Kitani, Tatsuya Hara and Hideyuki Sawada: "Voice articulatory training with a talking robot for the auditory impaired", International Journal on disability and human development, Vol.10, No.1, pp.63-67, 2011

[学会発表] (計 9 件)

- [1] Mitsuki Kitani, Tatsuya Hara, Hiroki Hanada and Hideyuki Sawada: "A Talking Robot and Its Human-like Expressive Speech Production", International Conference on Human System Interaction, pp.203-208, 2011
- [2] Hideyuki Sawada and Mitsuki Kitani: "Several Approaches to Speech and Auditory Systems for Human System Interactions", International Conference on Human System Interaction, pp.409-414, 2011
- [3] 木谷光来, 大久保和哉, 澤田秀之: 「ロボットによる聴覚障がい者の発話動作の再現」, 電子情報通信学会 HCG シンポジウム 2011, PP.196-201, 2011
- [4] **Outstanding English Presentation Award**, Mitsuki Kitani and Hideyuki Sawada: "Analysis of Robotic Speech employing the Vocal Tract Area Function", 電気関係学会 四国支部大会 講演論文集, 2011 年 9 月 23 日
- [5] **Best Paper Award**, Mitsuki Kitani, Tatsuya Hara, Hiroki Hanada and Hideyuki Sawada: "A talking robot for the vocal communication by the mimicry of human voice", International Conference on Human System Interaction, pp. 728-733, 2010

[6] Mitsuki Kitani, Tatsuya Hara, Hiroki Hanada and Hideyuki Sawada: "Robotic Vocalization Training System for the Auditory-impaired", International Conference on Disability, Virtual Reality and Associated Technologies, pp. 263-272, 2010

[7] **IEEE Best Award**, Mitsuki Kitani, Tatsuya Hara, Hiroki Hanada and Hideyuki Sawada: "A Talking Robot and the Auditory Feedback Learning for Natural Vocalization", The 7th IEEE Tokyo Young Researchers Workshop, p.25, 2010 年 11 月 20 日

[8] **Invited Talk**, Hideyuki Sawada: "A Talking Robot and the Interactive Speech Training for Hearing-Impaired", General Meeting of the Acoustical Society of America, May 18th, 2009

[9] Mitsuki Kitani, Tatsuya Hara and Hideyuki Sawada: "A Talking Robot and the Adaptive Learning of Speech Articulation Using 3D SOM", IEEE International Symposium on Biomedical Engineering (IEEE ISBME2009), CD-ROM proceedings WM1-1 #1042, 2009

[図書] (計 1 件)

[1] Mitsuki Kitani, Tatsuya Hara, Hiroki Hanada and Hideyuki Sawada, "A Talking Robot and Its Singing Performance by the Mimicry of Human Vocalization", Human-Computer Systems Interaction: Backgrounds and Applications Part 2, Advances in Intelligent and Soft Computing, Volume 99, pp.57-73, ISSN 1867-5662, 2012

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]

ホームページ等

<http://www.eng.kagawa-u.ac.jp/~sawada/>

6. 研究組織

(1) 研究代表者

澤田 秀之 (SAWADA HIDEYUKI)

香川大学・工学部・教授

研究者番号: 00308206