

機関番号	研究種目番号	審査区分番号	細目番号	分割番号	整理番号
12601	12	-	1011	B	0001

平成21年度 (2009年度) 若手研究 (S) 研究計画調書

平成 20 年 11 月 4 日
2 版

新規

研究種目	若手研究(S)							
分野	総合領域							
分科	情報学							
細目	生体生命情報学							
細目表 キーワード	脳型情報処理							
細目表以外の キーワード	メディア工学・普遍表象・自閉症							
研究代表者 氏名	(フリガナ)	ミネマツ ノブアキ						
	(漢字等)	峯松 信明						
年齢 (H21.4.1現在)	42 歳 (S . 41年12月生まれ)							
所属研究機関	東京大学							
部 局	工学(系)研究科(研究院)							
職	准教授							
学 位	博士(工学)							
現在の専門	情報工学					フォート	35%	
研究課題名	構造不変性に基づくメディア普遍の運動表象とそれに基づく脳型メディア情報処理							
研究経費 (千円未満の 端数は切り 捨てる)		研究経費 (千円)	使用内訳(千円)					
	年度		設備備品費	消耗品費	旅費	謝金等	その他	
	平成21年度	17,620	6,600	1,600	2,300	6,320	800	
	平成22年度	19,620	8,600	1,600	2,300	6,320	800	
	平成23年度	18,820	5,100	1,600	2,300	9,020	800	
	平成24年度	22,420	6,000	1,600	2,300	11,720	800	
	平成25年度	21,420	5,000	1,600	2,300	11,720	800	
総計	99,900	31,300	8,000	11,500	45,100	4,000		
開示希望の有無	審査結果の開示を希望する							

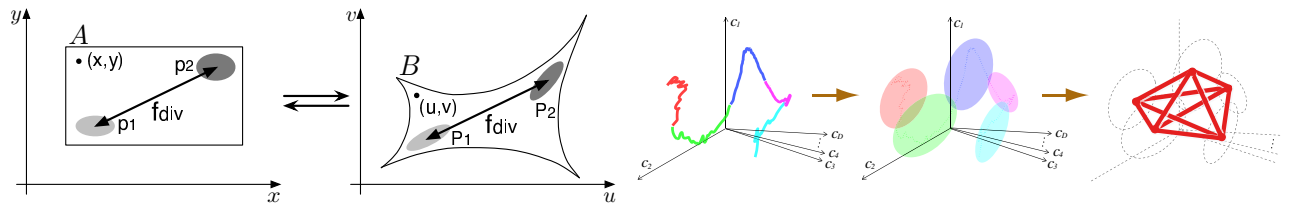
研究目的

本欄には、研究の全体構想及びその中で本研究の具体的な目的について、冒頭にその要旨を記述した上で、適宜文献を引用しつつ記述し、特に次の点については、焦点を絞り、具体的かつ明確に記述してください。（記述に当たっては、「科学研究費補助金（基盤研究等）における審査及び評価に関する規程」（公募要領 52～99 頁参照）を参考にしてください。）

- ① 研究の学術的背景（本研究に関連する国内・国外の研究動向及び位置づけ、応募者のこれまでの研究成果を踏まえ着想に至った経緯、これまでの研究成果を進展させる場合にはその内容等）
- ② 研究期間内に何をどこまで明らかにしようとするのか
- ③ 当該分野における本研究の学術的な特色・独創的な点及び予想される結果と意義
- ④ 科学研究費補助金の研究代表者として受入予定の研究課題について、本研究と関連する場合には、到達目標等の相違点、また、関連のない場合には、研究内容等の相違点（該当者は必ず記述してください。）

① 本研究は「音声の音響的多様性と知覚的不変性（異なる話者の同一言語内容の音声は、音響的には大きく異なることがあるが、両者に、同一性を容易に知覚できる）」という音声科学に古くから存在する疑問に対し、「構造的不变性」という数学的にその不变性が保証された枠組みを用いて、一つの説明理論を構築したことに端を発する。この構造的不变性は色知覚やメロディー知覚の不变性（恒常性）を含有する考え方である。既に単語音声認識において、従来の技術では認識率が1.4%と極めて低くなるタスク（4千人以上の成人音声を用いて音響モデルを構築し、小人音声を認識する）に対し、提案手法では、数名の成人音声を用いて90%以上の精度を出すなど、驚異的な技術革新を実現した。また、従来困難であった子供音声の高精度な情報処理を可能とし、小学生向けの英語発音評価技術を構築し、数年後に開始される小学校英語教育に備えるなど、実用アプリケーションも構築した。

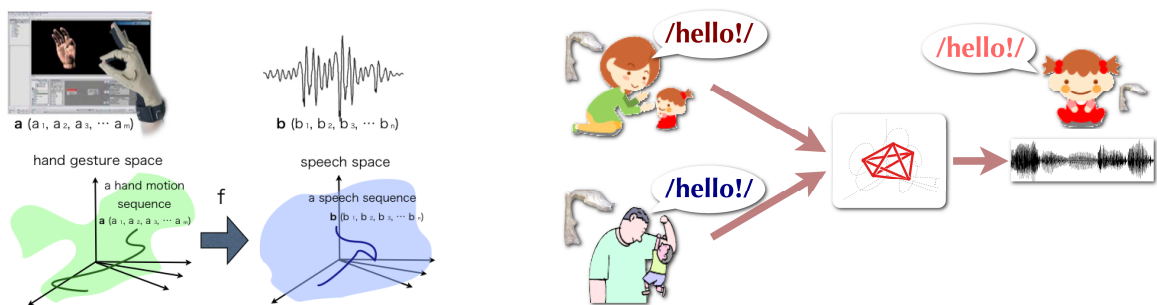
さて「構造的不变性」とは、観測対象に対して、如何なる連続的かつ可逆の線形・非線形写像（変換）を施しても、一切不変の量（情報）が、観測対象の「運動」の中にあることを数学的に証明した定理である。



図：「完全変換不変量としての f-divergence」と「f-div. を用いた運動の不变表象」

上図左は、複数の事象群（個々の事象は確率密度分布として表象される）が存在する空間を如何に変形しても、事象間の f-divergence は一切不変であることを示している。f-div. とは、 $f_{div}(p_1(x), p_2(x)) = \int p_2(x) g(\frac{p_1(x)}{p_2(x)}) dx$ で表現される分布間距離の一般形であり、KL-div. やバタチャリヤ距離などはその一例である。我々は f-div. の変換不変性に対する十分性及び必要性を数学的に証明した（f-div. であれば不変であり、不変量は f-div. でなければならない）。上図右は、ある位相空間における観測対象の「運動」を f-div. のみで表象する方法を示している。「運動」を一旦分布系列へ変換し、その後、全ての分布間距離を f-div. で求め、「運動」を距離行列を用いて表象する。こうすることで、観測対象物の「運動」は、完全に不変の枠組みで表象されることになる。この枠組みを用いて、上記した通り、年齢や性別といった音声の音響的特性を容易に変形（変換）する要因に対して超頑健な音声認識の枠組みを提案し、緊急性を要する実アプリケーションを既に構築している。

上記では「変換＝話者性の変換」と捉えていたが、「変換＝メディア変換」として考えることも可能である。例えば、我々は構音障害者支援を目的として、手の動きをそのまま音の動きに変換して音声生成を行なう技術を開発している。健常者は舌や唇などの構音器官の運動を介して音の運動を生成する訳だが、構音障害者にはこれが難しい。そこで、これを手の運動を介して行なう。この場合、手の運動空間と音の運動空間とが写像関数で、直接結ばれ、異なるメディアの運動が同期されることになる。この時、両者の「運動」を構造的に表象すれば、手の動きと音の動きは全く区別できなくなる。即ち、メディアの違いを超えて「運動」が普遍的に表象される。各メディアは個々の物理空間に存在しているが、そこから「運動」のみを取り出し、それを抽象的な数学空間で議論する。物理（具象）から数学（抽象）への昇華である。換言すれば、位相幾何学（トポロジー）における新しい位相不変量（f-div.）を、異メディア間写像（変換）を対象として扱うことに相当する（下図左参照）。



図：「異なるメディア間の空間写像」と「音声の構造的不变表象からの音声生成」

研究目的(つづき)

また我々は、ある話者の音声から構造抽出を行ない(声道サイズなどの身体情報を削除する)、これに別話者の身体情報を戻すことで、構造からの音声生成系を試作した(上図右参照)。別話者が幼児であれば、これは幼児の音声模倣に相当する。幼児の音声模倣は九官鳥のそれとは異なり、「声の模倣」は行なわない。音韻意識が薄弱な幼児は、親の声を音韻列に変換し、各音韻を音へ変換することもできない。発達心理学は「語の全体的な音形を獲得する」と説明するが、構造→音声生成系は、幼児の音声模倣(言語獲得)モデルとして評価されている。

本研究は「**メディア普遍の運動表象に基づくメディア情報処理の基盤技術の構築**」を目的とするが、その時に、ヒトの脳の連合野・前頭前野における情報処理と関連付けながら検討する。周知のように、視覚、聴覚、触覚などに対応する形で感覚野は存在し、各情報が連合野で「統合」され、前頭野、運動野を通して環境に働きかける。

＜環境＞ → [感覚野] ↔ [連合野] ↔ [前頭前野] ↔ [連合野] ↔ [運動野] → ＜環境＞

ここで、**異メディア情報の統合は如何になされるのか**、という問題がある。当然のことながら、光と音を3[カンデラ]+4[デシベル]=7という異なる単位間の計算で結びつけることは無意味である。異なる物理量を統合するには、**情報の抽象化、物理単位の消失**を図ることが必要であり、これを実装する枠組みとして、メディア普遍の運動表象を考える。例えば「サルは写真で提示された積み木を、袋の中から手で探すことが出来ない」「同様の作業は、一部の自閉症者も困難である」などの知見がある。両者の共通項は、[感覚野] ↔ [連合野・前頭前野間]の神経回路の接続がヒト(健常者)と比べて貧弱なため、メディア情報の抽象化が困難であり、異なるメディア(物理量)を超えた柔軟な情報処理が困難となる点である。脳型情報処理の工学的実装は多くの場合、人工ニューラルネットによって検討されて来たが、ニューロン単体の工学モデルを集めて使うことと、脳全体のシステム理解とが乖離していることは、昨今の脳科学、人工知能の研究例が示す通りである。**様々な物理単位を持つ、具体的な物理刺激に対してどのような抽象化(数学化)を実装することが、肥大化した連合野・前頭前野を持つヒトの柔軟な情報処理に迫れるのか、ヒトの脳全体をシステムとして捉えることで可能となるメディア情報処理パラダイムを提案することが本研究の到達点である。**例えば、メディア普遍の運動表象を担うニューロンが存在したとすれば、それは異メディアの運動に対して(例えば、自らの身体運動と、他者による同様の身体運動を見た時の網膜上の視覚運動)等しく発火するはずである。申請者は、ミラーニューロンの機能をこのように捉えている。なお、申請課題の直接的な比較研究例、及び位置付けについては、研究計画・方法の欄において述べる。

② 本研究では「鉄腕アトムを作ろう」と意気込んでいる訳ではない。生態心理学、進化心理学、脳科学、意識研究、認知障害(自閉症)研究など、様々な分野で得られた知見を俯瞰し、各種メディア工学で培われて来た技術を導入し、従来試みられたことがない**複数のメディアを普遍的に扱うことのできる、連合野・前頭前野型メディア情報処理の実装系を構築**することを目的とする。この実装系は、「動き」に着眼したメディア普遍の情報表象に支えられており、また、その表象は、強固(かつ新しい)数学的な基盤の上に構築される。本研究は、新しいメディア情報処理の理論研究と応用研究を並列して進めることになる。数学的な理論(基盤)を常に意識しつつ、その理論の、実世界での検証を常に行ない、実用価値のある新しいメディア情報処理系を構築する。現時点では、1)話者不変表象を用いた超頑健な音声認識系と、2)手の動きと音の動きを空間写像で結びつけた、手の動きからの音声生成系が具体的な実装系となっているが、これら以外にも、様々なメディアの「動き」を捉え、具体的な実装系を通して検討していく。ロボットを作りながら、連合野・前頭前野型のメディア普遍情報処理を考える。

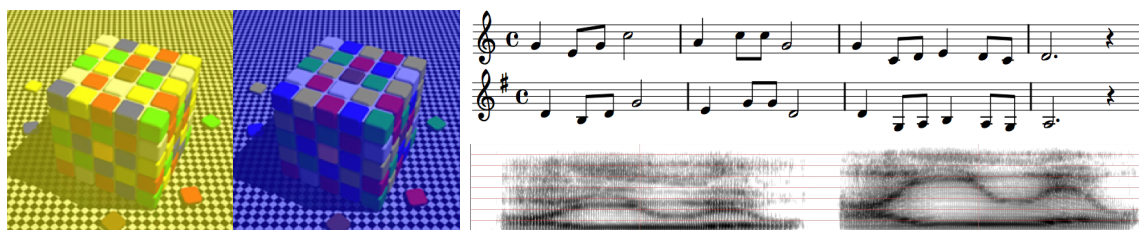
③ 従来のメディア工学では、身体、視覚、聴覚は、各々ロボット工学、コンピュータビジョン、音声声響工学といった分野で、縦割りの議論されており、これらのメディアの違いを抽象化によって消失させ、メディア普遍の情報処理系を構築する試みは、少なくとも申請者は知らない。**脳と対比して考えれば、感覚野のみ存在し、連合野・前頭前野は存在しない状態に相当する。**本研究におけるメディア普遍の情報表象は「**不変かつ普遍な構造**」の上に構築される。この構造的不変性は色知覚や音高知覚の不変性を含有する枠組みであるが、色知覚の不変性は昆虫でも観測される進化的に古い能力であるのに対し、音高知覚の不変性(音高の相対音感、音高の動きの中に不変的情報を感覚する能力)はマカク猿でも観測されない進化的に極めて新しい能力である。申請者が実装した話者の違いに超頑健な音声認識系は、**音色**の相対音感を有する(音色の動きに不変的情報を感覚できる)機械に相当する。ヒトの言語能力とは、その個体の多種多様な経験(即ち記憶)を検索し、整理整頓し、書き換え、生成(予測)する能力(思考能力)、及び、他者の記憶までもを検索し、整理整頓し、書き換え、生成する能力(コミュニケーション能力)である。この場合、**記憶本来のメディア(物理量)に依存しない、メディア普遍の記憶操作手段を我々は有している**ことになる。申請者は約20年に渡り音声コミュニケーションを、その物理的な側面から研究して来たが、何故「音」という物理的には極めて不安定な現象の上に、これだけ頑健かつ効率よい記憶操作系が構築されたのかが、全くの「謎」であった。しかし、不変構造という数学的な武器を携えることで超頑健な音声認識系を構築し、「謎」の一部は氷塊した。本研究は、**構造の話者不変性ではなく、メディア普遍性という側面にも着眼し、何故「音」を通して自己及び他者の記憶を操作できるのか、という問いに対する物理的、数学的な解答を得ることを最終的な目的の一つに掲げている。**これは言語の起源、心の起源、意識の起源、に対する物理的、数学的な一解答を提示することであり、本研究の高い学術性を示すものである。

④ 来年度、基盤(B)「音声の分節的・韻律的特徴を包含する発音の構造的表象に基づく外国語学習支援」及び、特定領域研究「多様な目的に適した形態素解析システム用電子化辞書の開発」にて援助を得ることになっている。これらは、外国語学習及び音声合成のテキスト処理に関する研究であり、申請課題とは研究目的が大きく異なる。

研究計画・方法

本欄には、研究目的を達成するための具体的な研究計画・方法について、冒頭にその要旨を記述した上で、平成21年度の計画と平成22年度以降の計画に分けて、適宜文献を引用しつつ、焦点を絞り、具体的かつ明確に記述してください。ここでは、研究が当初計画どおりに進まない時の対応など、多方面からの検討状況について述べるとともに、研究代表者が自ら組織を率いて研究計画を遂行するための研究体制について、研究代表者及び必要に応じて研究協力者（若手の研究者、大学院生、海外共同研究者、研究を補助する者等（氏名、員数を記入することも可））の具体的な役割（図表を用いる等）についても述べてください。

個体が受け取る身体刺激、視覚刺激、聴覚刺激は、様々な変形を被る。例えば、ある犬を見た場合、角度を変えて見直せば網膜像は異なる。しかし視点を変える前後で、我々は二つのイメージに同一性を容易に感覚できる。ある花を日中見る。そして夕焼け空の下で見る。当然、「色み」は異なる。しかし同一性を感覚できる。男性話者のハミングと、同一メロディーの女性話者のハミングは基本周波数（音高）が異なるにも拘らず、同一性を感覚できる。男性話者の発声と同一言語内容の女性の発声とでは、音色が大きく異なるが、同一性を感覚できる。本研究では、これらの不変的知覚（**静的バイアスの除去戦略**）を普遍的な枠組みで捉えている（下図参照）。



図：異なる色バイアス、音高バイアス、音色バイアスを含有する刺激（我々は容易に同一性を感覚する）

一部の自閉症者は上記の同一性を感覚することが困難である。彼らは**刺激の局所的、具体的な詳細を記憶する一方で、変形を通して変わらない、共通した「何か」を感覚することが困難となる**。当然、彼らの知覚世界は健常者のそれと大きく異なり、自閉症者の多くは自らを「異星人」と語る。昆虫ですら持ち合わせている情報処理が困難であれば「地球という生態系・進化系とは異なる枠組みの上に情報処理系が構築されてしまった」と言及したとしても無理は無い。しかし語れる自閉症者はまだ良い。多くの自閉症者にとって音声言語は非常に困難なツールである。母親の声は理解できるが、父親の声は認識できない自閉症児もいる。九官鳥のように父親の声を模倣しようとする自閉症児もいる（当然、音声言語は獲得できない）。自閉症者かつ動物学者の T. Grandin 教授は、自閉症者の情報処理と動物のそれは高い類似性を有することを主張している。「**刺激を“そのまま”記憶する。異なる感覚（メディア）を抽象化して統合する能力が欠落している**」ことを主張している。なお、認知障害研究の中には、チンパンジーを自閉性の極めて高い個体として利用している例もあることを言及しておく。

近年、脳の情報処理パラダイムを計算機の上に実装する研究プロジェクトが発足している。Redwood Neuroscience Institute を設立した J. Hawkins 博士が提唱している、Memory Prediction Theory (MPT) に基づいて各種メディア情報を実装する枠組みである。彼は、自身の図書 (*On Intelligence*) の中で、**1) 各種変形・変換に不変な情報表象の存在、2) 異種メディアにおいて共通して存在する情報処理アルゴリズムの存在（即ち、各感覚野皮質で行なわれている情報処理アルゴリズムの等価性）**、を強く主張している。また、彼の提唱する情報処理パラダイムに基づいて、幼児の言語能力の獲得を計算機上に実装する一大研究プロジェクトもヨーロッパで発足している (ACORNS プロジェクト / Acquisition of Communication and Recognition Skills)。MPT と本研究の違いは明確である。Hawkins は最終的に「**予測を行なう機能の実現**」を第一義的に目標として掲げ、その実装系を検討している。申請者が提案する情報処理パラダイムは、予測を行なう機能の実装を第一義的には考えていない。予測機能は派生的に生まれる能力である、と考えている。申請研究では、**1) 各種変形・変換不変な情報表象、及び、2) 異種メディアを「動き」に着眼して抽象化し、メディアの違いを超えた操作を可能とする普遍的な情報表象、を提案し、それを具体的な実装系を通して検討することが目的である**。この不変かつ普遍的な情報表象が確立すれば、メディアの違いを超えた情報（記憶）の検索、整理整頓、書き換え、生成機能は、自然に派生することになる。また、この不変かつ普遍的な表象が、対象刺激の時間的な「動き」を抽出したものである事実を考えれば、未来を見据えた情報処理となるのは当然の帰結である。「動き」表象に基づく異メディア間の柔軟な情報処理が実現できれば、観察者はそれを「予測」と解釈することになる。更に、MPT の枠組みでは、**彼らの主張する情報処理パラダイムが、何故、1) 各種変形・変換に不変となるのか、何故、2) 各種メディア情報を、メディアの違いを超えて普遍的に扱うことができるのか？**についての数学的、物理的説明に乏しいことを指摘したい。この研究戦略の違いは、申請者がメディア工学（音声・音響工学）を背景に脳型情報処理を模索し、Hawkins は人工知能を背景に脳型情報処理を模索している点に帰着される。物理現象（自然現象）を直接的に扱って来た者として、数学的・物理学的根拠に乏しい方法論は説得力に乏しく、その妥当性も懐疑的にならざるを得ない。

- T. Grandin 「動物感覚 ～自閉症者が読み解くアニマル・マインド～」日本放送出版協会、2006
- ニキリンコ「スルーできない脳 ～自閉は情報の便秘～」生活書院、2008（著者は自閉症者である）
- J. Hawkins and S. Blakeslee, *On intelligence*, Henry Holt, 2004
- ACORNS プロジェクト (<http://www.acorns-project.org/>)

研究計画・方法（つづき）

本研究では、ロボット工学、コンピュータビジョン、音響工学など、異なる分野で構築されてきたメディア処理系を、**メディア普遍の運動表象**という観点から纏め上げ、脳型（特に連合野・前頭前野型）情報処理を構築することが目的である。この場合、各メディア工学の専門家が集結し、身体、光、音といった個々のメディア特有の現象を**敢えて捨て**、抽象的な情報表象の下に、個々のメディア処理系を再構築する作業が必要になる。各メディア工学は、各メディア特有の処理系を構築することが多く（各研究者はそのメディアに興味を持つのが普通であり、メディア特有の情報処理系を好む）、それを**敢えて捨てた上で**、抽象的な情報処理系の構築に賛同する研究者を募ることは困難な作業である。幸いなことに申請者は2007年度より、コンピュータビジョンの専門家（若手研究者）である喬宇氏をポストドクとして迎え入れ（学振の外国員特別研究員。学振に対する申請課題は「**構造不変の定理に基づく聴覚・視覚・身体メディアの普遍的表象に関する基礎研究**」である。本年度一杯で学振の雇用は終了する。）、申請者が構築した音声の不変構造論を精緻化すると共に、他メディアへの応用について基礎的検討・議論を繰り返して来た。手運動と音声運動の空間写像生成はその一環である。本申請は5年に渡る長期研究を予定しているが、喬宇氏を今後も雇用し、また、3年目からロボット工学の若手研究者も別途募り、聴覚、視覚、身体メディアを真に統合することを考えている。また、ポストドクの研究者以外にも、本学博士課程の齋藤大輔氏、國越晶氏、修士課程の鈴木雅之氏と協力して進める予定である。

数年来の研究において、我々は構造表象の音響工学応用について深い理解を得ることができた。f-div. は如何なる写像でも不変性を示すため、「**不変性が強すぎる**」問題が生じる。即ち、話者の違いを消すあまりに、単語の違いまで消えてしまう（異なる単語が同一と評価される）。この場合必要なのは、1) **話者の違いがどのような変換関数群としてモデル化できるのかの検討**、及び、2) **対象とする変換関数群のみに不変性が成立する音響照合方式の構築**であるが、これをマルチストリーム構造化という技術を導入することで解決した。これとは別に、パラメータ次元数が極端に増加する問題もあったが、LDAやPCAを適切に導入することで、識別力の高い低次元空間を求めることに成功した。以上の経緯を踏まえて、21年度、及びそれ移行の研究計画を記す。

■平成21年度の研究計画■

与えられた刺激群に内在する多様性を表現する写像関数群の設計（喬，峯松）

「強すぎる不変性」問題は、対象とする多様性を数学的にモデル化し、その写像群のみに不変性が成立する枠組みを求めることで解決できる。音声の場合、上記したように、マルチストリーム構造化（位相空間を次元分割し、個々の部分空間にて構造照合を行なう）を提案した。しかし、音声以外のメディアへの構造表象応用を考えた場合、与えられた刺激群をボトムアップに観測し、そこに内在する「多様性を表現する写像関数群」を推定・設計する**一般的な枠組み**を構築する必要がある。ここでは、ガウス混合分布を用いた空間写像推定を拡張し、確率的な回帰写像の混合分布を考え、非線形性も扱える写像関数群設計を検討する。更には、これら写像群のみに不変性を有する照合方式についても検討する。主に音声メディアを用いた検証実験を行なう。

参考文献 = 1,2,3,4,5,7,8,9,12,13,14,17,20,22,23,27,28,29（研究業績番号）

共鳴音と非共鳴音とを分割した構造的音声認識の検討（峯松，鈴木）

構造的音声認識はまず、母音連結単語（「あうえおい」など）といった人工的な語彙セットを使って検討し、次に、通常の音素バランス単語を用いた孤立単語音声認識を検討した。既に音素バランス単語使用時も、従来技術ではおよそ不可能な頑健性を実現しているが、モデル学習時と評価時に何ら mismatches が無い場合、認識率の低下が若干観測されている（母音単語の場合は認識率は低下しない）。共鳴音と非共鳴音とでは話者性による音の変形の様子が異なるため、これらを異なる2ストリームとして扱い、個別に構造構築・照合する方法を検討する。また、既に我々は、構造表象に基づく音声認識アルゴリズムと非常に類似した議論が、音声知覚研究の分野において展開されていることを確認している。音声知覚研究は、「音声のどの側面に語彙情報が符号化されているのか」の解明の一つの目標としているが、「音ではなく、音の動きの中に符号化されている」との説明理論を提唱している研究グループが複数ある。彼らの実験事実との擦り合わせなどについても検討する予定である。

参考文献 = 1,3,5,7,8,9,12,14,17,22（研究業績番号）

抽象的な動き表象である構造表象から実音声を構成する、構造音声合成の実装（峯松，齋藤）

構造抽出は「具象→抽象」のプロセスである。その逆プロセスである構造からの具体的なメディア運動を生成する技術を音声メディアを題材として検討する。研究目的の欄にも示した「構造からの音声合成」である。これは既述した様に「幼児の音声模倣プロセスの計算機上の実装」として捉えることができる。幼児は音声模倣を通して言語を獲得するが、親の声をそのまま真似ようとはしない。音韻意識が希薄なため、発声を平仮名化し、個々の平仮名を音声化するというプロセスは困難である。発達心理学では「語全体の音形、語のゲシュタルトを真似る」と説明される。語ゲシュタルトは当然話者不変な音声表象である必要があるが、語ゲシュタルトの物理実装として音声の構造表象を捉える訳である。既に、簡単な実装系は構築済みであり、ここでは、その精緻化（子音や韻律的特徴の考慮）と他メディアでの実装（音声から抽出された構造からの手運動の生成など）を検討する。

参考文献 = 1,2,4,6,7,8,17,18,19,21,22（研究業績番号）

手運動と声運動との空間写像の設計と、それに基づくメディア変換とメディア普遍表象（峯松，喬，國越）

本研究は、理論構築とその実験的検証を同時進行的に行なう。ここでは、構音障害者支援の名の下、異メディア間の運動空間写像を設計する。構音障害者支援という意味では、異メディア間の写像が実装できればよいが、

研究計画・方法（つづき）

ここでは、情報の抽象化（手運動→構造）、具象化（構造→手運動）についても検討する。構造からの音声合成の技術がそのまま利用できる。既述した通り、旧式のデータグローブを用いた初期検討は既に完了している。初期検討を通して、手から声への変換に適したセンサーを準備する必要性を強く感じており（より高い空間分解能が必要）、モーションキャプチャーなど、種々のセンサーを購入、検討することを考えている。更には、簡単な手運動認識も検討する。これは、構造に基づく音声認識の手運動版である。これが実現できれば、構造というメディア普遍の（全く同一の）情報表象を用いて、音声認識と手運動認識の両方が一切プログラム・データを変更することなく実装可能となる。メディアを問わない連合野・前頭前野型情報処理の一つの実装系である。

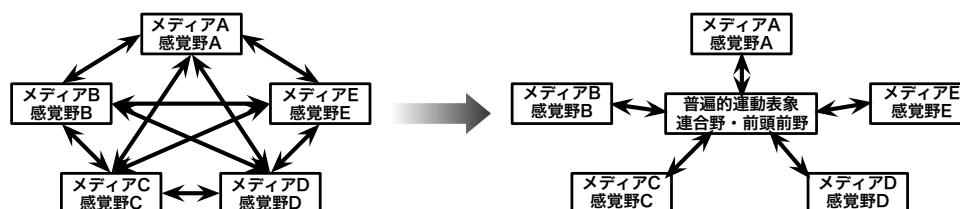
参考文献= 1,5,6,7,13,18（研究業績番号）

■平成 22 年度以降の研究計画■

21 年度の 4 つの研究テーマはそのまま 22, 23 年度前半も継続し、その精緻化を図る。例えば、多様性を表現する写像関数群の設計は、聴覚メディア、身体メディア、更には、視覚メディアを含め、様々なメディアにおける実験的検討を繰り返し、その妥当性を検証する。構造的音声認識は、孤立単語音声認識から、連続単語音声認識へとタスク難度を上げ、また、従来の方法論との融合も検討する。更には、話者性の違いのみならず、背景雑音や反響音歪みなどへの頑健性についても検討する。構造音声合成に関しては、幼児の音声模倣プロセスをどこまでシミュレートできているのか、その評価を発達心理学者と共同で行なう（既に、共同研究は開始している）。言語獲得の一側面のモデル化が可能となるので、例えば外国語学習者がその外国語を完全に習得できた時の、その学習者による対象言語の発音を生成するシステムの構築なども検討したい。手運動と声運動間の写像関数の設計に関しては、例えば、具象→抽象、抽象→具象の実装が完成すれば、新たな単語を手運動で教えなくても、音声で教えれば、自動的にその手運動を学習するようなシステムが構築可能となる。より一般的に言えば、メディアの違いを超えた普遍的な表象に基づいて各メディア情報を処理するシステムの構築が可能となる。

なお、構造表象に基づく普遍的メディア情報処理であるが、既に音声工学の分野では音声認識、音声合成、音声応用の分野で実績を積んでおり、音声工学における理論構築及び応用研究においては、進捗状況が、当初の予定と大きく異なることは無いものと考えている。問題があるとすれば、5ヶ年計画の後半、異メディアとの積極的な統合を検討する中で、想定外の問題に遭遇する可能性は否めない。特に、ロボット工学との融合を図る場合、峯松や喬はロボット工学の専門では無いため、新たな若手研究者を雇用する予定である。

5ヶ年計画の前半を通し、構造理論の精緻化、及び、構造理論の音声工学応用に関して膨大な知見が得られ、他メディア応用への準備が整うものと考えている。23 年度後半より、ロボット工学の若手研究員をポスドクとして受け入れ（人選は 22 或は、23 年度に行なう）、聴覚、視覚、身体の 3 つのメディア情報処理系を「不変かつ普遍的な構造表象」の元で再構築する。この場合、システムの実装を通して種々の問題を洗い出し、その解決を図る。目標となるのは、例えば、「メディア A の運動 ↔ メディア B の運動」及び「メディア B の運動 ↔ メディア C の運動」を学習させると、メディア A の新規な運動パターンに対して、自ずとメディア C の運動パターンが生成・認識されるシステムの構築である。メディアを超えた情報表象を保有し、それに基づいて、具体的な入力刺激を認識し、かつ、具体的な身体運動を生成できるシステムの構築である。自閉症者の中には優れたピアノ演奏が可能者がいる。しかし、そのメロディーを（ピアノよりも操作が容易な）縦笛で演奏することは困難である。メロディーの動き（A）と鍵盤上の手の動き（B）の対応はとれても、メロディーの動き（A）を縦笛上の手の動き（C）に転化することが難しい。彼らの場合、異なるメディア（鍵盤上の手と縦笛上の手は異なるメディアと考える）の数を N とすれば、任意の 2 メディア変換を訓練する場合、 $N C_2$ 種類だけのメディア変換（転化）訓練が必要となる。これが「**感覚野で生きている自閉症者**」（T. Grandin 教授の言葉）の情報処理となる。本研究では、メディアに捕われない抽象的な運動表象と、個々のメディアとのインタフェースの構築を目指す（下図参照）。



図：個々のメディア変換（転化）から、普遍的な運動表象と各種メディア間のインタフェース構築へ

更には、多種多様なメディア情報を記憶した際に、それを検索、整理整頓、書き換え、生成する作業に対して、音メディア（特に音声メディア）がどのような優位性を持ち得るのか、についても検討する（最終年度）。研究目的の欄にも示した様に、ヒトの情報処理において、「**音声という不安定な物理メディアが、何故これほど高効率な記憶検索・操作ツールになり得たのか？**」という問いに対する一つの答えを模索する。「**何故、空気振動は記憶を、意味を生成できるのか？**」に対する物理的解答を模索したい。この検討に対して、被験者実験を検討している。光トポグラフィー（NIRS）や経頭蓋磁気刺激法（TMS）などの計測機器を用い、音声からの構造生成が困難な状況にヒトを陥れた場合に、脳がどのように反応するのか、について検討する。申請者は、心理学者、脳科学者とも綿密に連絡をとりあっており、これらの実験計画についても十分な援助を受けられる立場にある。

今回の研究計画を実施するに当たっての準備状況等

本欄には、次の点について、焦点を絞り、具体的かつ明確に記述してください。

- ① 本研究を実施するために使用する研究施設・設備・研究資料等、現在の研究環境の状況
- ② 必要に応じ組織に参画させる予定の研究協力者（若手の研究者等）の状況（募集方法や選考基準、具体的に決まっている場合にはその者の研究履歴・学歴等）
- ③ 本研究の研究成果を社会・国民に発信する方法等

① 本研究遂行のための研究環境について、「物」的環境と「知」的環境の準備状況について記す。「物」的環境とは各種メディア情報を取得するための装置（各種センサーなど）、及び、それらを情報処理するための計算機環境を指す。「知」的環境とは、種々の知識獲得のための図書などの情報源確保の環境を指す。

過去2年間の研究遂行によって、音声分析及び音声認識装置としての計算機環境を、一通り整備することが出来た。しかし現時点で所有している計算機群では、「刺激の多様性を表現する写像関数群の設計」の遂行に多大な時間を必要とすることが分かっている。また、5ヶ年計画の前半で遂行を考えている、視覚情報の融合に関しては膨大な記憶媒体（HDD）が必要となる。これらを考慮し、計算機環境のより一層の充実を図る。計算機の設置場所についても、サーバー室（空調完備の部屋）などの整備が必要になる可能性が高い。こちらの方も、計算機環境向上のための経費として割当て、随時サーバー室の拡張を行なう。一方、各種メディア情報を取得するセンサー類であるが、音情報に関しては既に現有設備で十分であるが、特に身体運動に関するセンサーが現有設備では非常に貧弱である。現在使用しているデータグローブも旧式のものであり、より広範囲な身体運動のセンシングも含め、様々な身体運動データの取得装置（モーションキャプチャ）を、本申請課題を通して揃える予定である。最終年度での遂行を考えている各種脳測定装置であるが、これについても、本申請課題を通して購入する必要がある。学内の研究所に設置されているものは、そちらの設備を使うことも可能であるが、頻繁に使用することになる機器については是非、自分の研究室に常備しておきたい。

「知」的環境であるが、本研究の一つの特色は、工学という狭い視野に捕われず、脳科学、言語学、認知科学、言語障害学、進化心理学、進化生物学、生態心理学など、様々な研究者との情報交換を通して、提案手法を「科学的に妥当な」方法論へと纏め上げるプロセスにあると考える。その意味において、異分野の研究者との交流は非常に大切な側面である。関連する図書や研究資料などは積極的に購入、勉強会を開くなどして、各分野の知識獲得に務めている。また、国内外を問わず、face-to-faceでの議論の場を求め、人脈の確保についても積極的に行なってきた（研究略歴の欄参照）。彼らとの議論は、「研究業績」の欄を見て頂くと分かるように、赤ちゃん研究、言語障害研究など、従来申請者があまり参加してこなかった研究活動の場（各分野の国際会議など）において研究発表するなど、十分に結実されている。このような活動は今後も積極的に続ける予定である。このような異分野交流的・学融合的活動無しに、本申請課題の遂行は不可能である。

② 本研究は、基本的に峯松（研究室）の単独研究であるが、種々のメディア情報処理を統合するというプロジェクトの性質上、異分野の若手研究者の協力が必須である。峯松自身は音声・音響工学を専門としており、メディア普遍情報の聴覚情報処理を担当する。視覚情報処理の担当としてはポスドクの喬宇氏を考えている。彼は学振の援助を受け、1.5年に渡り峯松研で研究を進めてきた（研究課題は既述した通り。なお、2007年3月に電気通信大学にて博士（工学）を取得している）。彼の専門はコンピュータ・ビジョンであるが、峯松の提唱する普遍・不変構造の枠組みに興味を示し、音声認識の頑健性の向上について基礎的検討を行なっている。また、構造表象に基づく視覚情報処理についても理論的な考察を進めている。学振からの経済支援は本年度一杯で終了するため、次年度以降はこの若手（S）の財源を使って彼を雇用する。音とビジョンを核にして構造的なメディア普遍情報処理を検討し、3年目後半以降からロボット工学の若手研究者をグループに迎え、身体メディアまでを含む枠組みに拡張する予定である。ポスドク研究員の選考はプロジェクト3年目前半に行なう。その時点での、聴覚・視覚メディアの統合状況を示した上で、身体情報にどう統合できるのか、に関して提案を募り、それを考慮して選定する。また、峯松研究室の博士学生、修士学生にも随時研究協力を依頼する予定である。

③ 峯松が提唱している新しい音声情報処理体系は、既にマスコミによって一部報道されている。「研究目的」の欄にも記したように、提案する音声情報処理体系によって初めて頑健な処理が可能となった音声アプリケーションがある。子供を対象とした外国語発音教育支援である。既に小・中学校にて検証しており、この技術は、英語教育が小学校授業に導入される数年後には、多くの国民によって使われる可能性が高い。このように具体的な実用システムを通して新しい技術体系、新しい情報処理の枠組み、そしてその高い実用性を広く国民に発信することができる。また、本研究で目指すメディア普遍の情報処理体系が確立され、例えば、ヒューマノイドロボット研究へ応用されれば、「見た目が人間らしい」ロボットではなく、「中身が人間らしい」ロボットとして位置づけられることとなり、様々なメディアを通して本研究の成果を公表することが可能となる。

研究目的・計画・方法の欄で、自閉症者の情報処理と健常者のそれとの対比を記述した。提案する情報処理系が健常者の連合野・前頭前野の情報処理モデルとして妥当であれば、これらを一歩破損することで構成される計算機上の処理系は、認知障害者や自閉症者の情報処理モデルを与える。このような観点から障害者教育を支援することも可能となる。人間固有の柔らかい情報処理が計算機上で実装できれば、工学応用のみならず、様々な応用の場面が広がることは想像に難くない。既に、申請者は自閉症者のセラピーの現場などにも足を運び、また、自閉症児を持つ家庭／機関を訪問するなど、各種情報の取得を積極的に行なっている。研究略歴に示す様に、訪問先には当然海外も含まれている。訪問先での招待講演なども数多い。

研究業績

本欄には、これまでに発表した論文、著書（教科書、学会抄録、講演要旨は除く。）、産業財産権、招待講演のうち、主要なものを選定し、現在から順に発表年次を過去にさかのぼり、通し番号を付して記入してください。

なお、学術誌へ投稿中の論文を記入する場合は、掲載が決定しているものに限ります。

発表論文名・著書名 等

（例えば発表論文の場合、論文名、著者名、掲載誌名、査読の有無、巻、最初と最後の頁、発表年(西暦)について記入してください。)(以上の各項目が記載されていれば、項目の順序を入れ替えても可。著者名が多数にわたる場合は、主な著者を数名記入し以下を省略(省略する場合、その員数と、掲載されている順番を○番目と記入)しても可。なお、研究代表者には下線を付してください。)

1. 峯松信明, “音声言語運用が要求する認知的能力と音声言語工学が構築した計算論的能力”, 音声言語シンポジウム 10 周年記念特別招待講演 (2008, 招待講演)
2. N. Minematsu, T. Nishimura, “Consideration of infants’ vocal imitation through modeling speech as timbre-based melody,” Lecture Notes of Artificial Intelligence, 4914, 26–39, Springer (2008, 査読有)
3. 村上隆夫, 峯松信明, 廣瀬啓吉, “音声の構造的表象に基づく日本語孤立母音系列を対する音声認識”, 電子情報通信学会論文誌, J91-A, 2, 181–191 (2008, 査読有)
4. 峯松信明, “「あ」という声を聞いて母音「あ」と同定する能力は音声言語運用に必要なか?”, 日本語学, 4, 187–197, 明治書院 (2008, 査読有)
5. N. Minematsu, “Speech as music –Yet another framework for speech recognition,” ESF-JSPS Frontier Science Conference Series for Young Researchers – Experimental Cognitive Robotics – (2008, 招待講演)
6. D. Saito, S. Asakawa, N. Minematsu, K. Hirose, “Structure to speech – speech generation based on infant-like vocal imitation –,” Proc. INTERSPEECH, 1837–1840 (2008, 査読有)
7. Y. Qiao, N. Minematsu, “f-divergence is a generalized invariant measure between distributions,” Proc. INTERSPEECH, 1349–1352 (2008, 査読有)
8. N. Minematsu, T. Nishimura, D. Saito, S. Asakawa, Y. Qiao, “Holistic and prosodic representation of the segmental aspect of speech,” Proc. Int. Conf. Speech Prosody, 169–172 (2008, 査読有)
9. S. Asakawa, N. Minematsu, K. Hirose, “Multi-stream parameterization for structural speech recognition,” Proc. Int. Conf. Acoustics, Speech, and Signal Processing, 4097–4100 (2008, 査読有)
10. A. Rebordao, M. K. I. Molla, K. Hirose, N. Minematsu, “An adaptive speech denoising system based on ICA and voice activity detection,” Proc. Int. Workshop on Nonlinear Circuits and Signal Processing, 228–231 (2008, 査読有)
11. Q. Sun, K. Hirose, N. Minematsu, “Improved tone component prediction of tone nucleus for F0 contour generation of Mandarin speech,” Proc. Int. Workshop on Nonlinear Circuits and Signal Processing, 112–115 (2008, 査読有)
12. 朝川智, 喬宇, 峯松信明, 広瀬啓吉, “判別分析と構造構造表象を用いた話者の多様性に超頑健な音声認識”, 日本音響学会講演集, 2-P-3, 113–116 (2008, 査読無)
13. 國越晶, 喬宇, 峯松信明, 広瀬啓吉, “空間写像に基づく手の動きを入力とした音声生成系”, 日本音響学会講演集, 1-Q-23, 375–376 (2008, 査読無)
14. S. Asakawa, N. Minematsu, K. Hirose, “Automatic recognition of connected vowels only using speaker-invariant representation of speech dynamics,” Proc. INTERSPEECH, 890–893 (2007, 査読有)
15. 朝川智, 峯松信明, 広瀬啓吉, “音声の構造的表象に基づく英語学習者発音の音響的分析”, 電子情報通信学会論文誌, J90-D, 5, 1249–1262 (2007, 査読有)
16. 峯松信明, “小特集「言語障害を通して再考する音声言語情報処理」にあたって”, 日本音響学会誌, 63, 7, 363–364 (2007, 査読有)
17. 峯松信明, 西村多寿子, 朝川智, 櫻庭京子, 斎藤大輔, “要素論から全体論へ ～全体から入る音声情報処理への招待～”, 情報処理学会音声言語情報処理研究会, 2007-SLP-67-14, 75–80 (2007, 査読無)
18. 斎藤大輔, 朝川智, 峯松信明, 広瀬啓吉, “音声の構造的表象を入力とした音声合成に対する基礎的検討”, 日本音響学会講演集, 1-P-2, 399–402 (2007, 査読無)
19. N. Minematsu, K. Sakuraba, T. Nishimura, “Development of a spoken word recognizer without phonemic awareness – Is this machine a Dyslexia simulator? –,” Proc. The 2nd Riken and Oxford Joint International Symposium – Reading and Dyslexia in Different Languages –, 49–51 (2007, 査読有)
20. N. Minematsu, “From fragile substances to robust contrasts – Yet another framework for speech recognition –,” Seminar talk at Microsoft Research Asia and Institute of Linguistics, Chinese Academy of Social Sciences (2007, 招待講演)

研究機関名 | 東京大学

研究代表者氏名 | 峯松 信明

研究業績(つづき)

21. N. Minematsu, T. Nishimura, “Universal and invariant representation of speech,” Proc. Int. Conf. Infant Study, CD-ROM (2006, 査読有)
22. N. Minematsu, T. Nishimura, K. Nishinari, K. Sakuraba, “Theorem of the invariant structure and its derivation of speech Gestalt,” Proc. Int. Workshop on Speech Recognition and Intrinsic Variations, 47-52 (2006, 査読有)
23. N. Minematsu, T. Nishimura, T. Murakami, K. Hirose, “Speech recognition only with supra-segmental features – hearing speech as music –,” Proc. Int. Conf. Speech Prosody, 589-594 (2006, 査読有)
24. 渡辺美知子, 広瀬啓吉, 伝康晴, 峯松信明, “音声聴取時のフィラーの働き ～「エート」による後続句の複雑さ予測～”, 日本音響学会誌, 62, 5, 370-378 (2006)
25. Md. Khademul Islam Molla, K. Hirose, N. Minematsu, “Separation of speech and interfering audio signal from single mixture by subspace decomposition,” Journal of Signal Processing, 9, 6, 487-495 (2005, 査読有)
26. N. Minematsu, “Structural representation of individual learners,” ASSTA Research Workshop of Assessing Spoken Language Proficiency (2005, 招待講演)
27. T. Murakami, K. Maruyama, N. Minematsu, K. Hirose, “Japanese vowel recognition based on structural representation of speech,” Proc. INTERSPEECH, 1261-1264 (2005, 査読有)
28. N. Minematsu, “Mathematical evidence of the acoustic universal structure in speech,” Proc. Int. Conf. Acoustics, Speech, and Signal Processing, 889-892 (2005, 査読有)
29. N. Minematsu, “Yet another acoustic representation of speech sounds,” Proc. Int. Conf. Acoustics, Speech, and Signal Processing, 585-588 (2004, 査読有)
30. 峯松信明, 他 6 名, “英語 CALL 構築を目的とした日本人及び米国人による読み上げ英語音声データベースの構築”, 日本教育工学会論文誌, 27, 3, 259-272 (2004, 査読有)
31. 峯松信明, 倉田岳人, 広瀬啓吉, “米語における音素体系及び語彙体系に着眼した日本人英語の発声と聴取に関するコーパス統計分析”, 音声研究, 7, 3, 77-91 (2003, 査読有)
32. 峯松信明, 仁科喜久子, 中川聖一, “外国語学習用読み上げ音声データベース”, 日本音響学会誌, 59, 6, 345-350 (2003, 査読有)
33. 成澤修一, 峯松信明, 広瀬啓吉, 藤崎博也, “音声の基本周波数パターン生成過程モデルのパラメータ自動抽出法”, 情報処理学会論文誌, 43, 7, 2155-2169 (2002, 査読有)
34. 広瀬啓吉, 峯松信明, 森谷高明, “単語の関連性を利用した音声認識用言語モデルのドメイン適用”, 情報処理学会論文誌, 43, 7, 2065-2074 (2002, 査読有)
35. 峯松信明, 広瀬啓吉, 関口真理子, “話者認識技術を利用した主観的高齢話者の同定とそれに基づく主観的年代の推定”, 情報処理学会論文誌, 43, 7, 2186-2196 (2002, 査読有)
36. 小橋川哲, 峯松信明, 広瀬啓吉, ドナ・エリクソン, “文強勢検出に基づく英語文リズム学習支援”, 日本音響学会講演集, 2-5-22, 117-118 (2002, 査読無)
37. 河原達也, 他 10 名, 峯松は⑤番目に掲載, “日本語ディクテーション基本ソフトウェア (99 年度版)”, 日本音響学会誌, 57, 3, 210-214 (2001, 査読有)
38. 峯松信明, 中川聖一, “PSOLA 分析合成に基づく F0 変換音声の品質向上に関する実験的検討”, 電子情報通信学会論文誌, J83-D-II, 7, 1590-1599 (2000, 査読有)
39. 中川聖一, 花井健豪, 山本一公, 峯松信明, “HMM に基づく音声認識のための音節モデルと triphone モデルの比較”, 電子情報通信学会論文誌, J83-D-II, 6, 1412-1411 (2000, 査読有)
40. 河原達也, 他 10 名, 峯松は⑤番目に掲載, “日本語ディクテーション基本ソフトウェア (98 年度版)”, 日本音響学会誌, 56, 4, 255-259 (2000, 査読有)
41. 峯松信明, 藤澤友紀子, 中川聖一, “英単語発音上の癖の自動推定・視覚化とそれに基づく発音能力の韻律的評定”, 電子情報通信学会論文誌, J83-D-II, 2, 486-499 (2000, 査読有)
42. 峯松信明, 中川聖一, “音声単語の同定におけるアクセント核の働きに関する知覚的検討”, 日本音響学会誌, 56, 1, 39-46 (2000, 査読有)
43. 峯松信明, 藤澤友紀子, 中川聖一, “HMM を用いた英単語音声からの強勢音節の自動検出とそれに基づく発音能力の韻律的評定”, 電子情報通信学会論文誌, J82-D-II, 11, 1865-1876 (1999, 査読有)
44. 峯松信明, 中川聖一, “F0 変化に伴うスペクトル変動に対する分析とモデル化”, 日本音響学会誌, 55, 3, 165-174 (1999, 査読有)
45. N. Minematsu, K. Hirose, “Role of prosodic features in the human process of perceiving spoken words and sentences in Japanese,” J. Acoust. Soc. Japan(E), 16, 5, 311-320 (1995, 査読有)

研究略歴

〔本欄には、最初に研究代表者の研究履歴、学歴、特筆すべき事項（研究成果、受賞歴等）等を現在から順に年度をさかのぼって記入してください。次に線で区切って、これまで行ってきた研究内容を研究業績欄に記入した論文等を引用しつつ記述してください。なお、本応募研究課題の立案に密接に関連したこれまでの研究の内容は「研究目的」の欄にも記述してください。〕

2008年 電子情報通信学会 ISS 活動功労賞を受賞。
 2008年 人工知能学会優秀論文賞を受賞（研究業績番号 2）。
 2008年 信号処理学会優秀論文賞を受賞（研究業績番号 10）。
 2008年 信号処理学会優秀論文賞を受賞（研究業績番号 11）。
 2007年 信号処理学会最優秀論文賞を受賞（研究業績番号 25）。
 2007年 日本音響学会全国大会ポスター論文賞を受賞（研究業績番号 18）。
 2002～2003年 スウェーデン・ストックホルムにある王立工科大学に一年間在外研究員として滞在。
 2001年 日本音響学会全国大会ポスター論文賞を受賞（研究業績番号 36）。
 2000年 東京大学大学院・工学系研究科に助教授として転任。
 1995年 豊橋技術科学大学情報工学系に助手として赴任。
 1995年 東京大学大学院博士課程電子工学専攻を修了。博士（工学）を取得。
 1992年 東京大学大学院修士課程電子工学専攻を修了。工学修士を取得。
 1990年 東京大学工学部電気工学科を卒業。工学学士を取得。

申請者は、学部時代より約 20 年に渡り、音声コミュニケーションの研究を行なっている。言語学（音声学 & 音韻論）、外国語教育学、言語障害学、認知科学、脳科学、音楽学、発達心理学、進化心理学、生態心理学、進化生物学など様々な分野に造詣があり、非常に幅広い観点から人間を観察し、人間の行なう音声情報処理の計算機実装を試みて来た。2002 年から一年間、王立工科大学（スウェーデン）に滞在したが、これが研究者としての大きな転機となった。「計算機が行なう音声情報処理」と「人間が行なう音声情報処理」との根本的な差異について徹底的に熟考し、全く新しい音声情報処理（**話者不変の音声構造に基づく音声処理**）を提案し、今日に至っている。音声認識の分野では従来の枠組みでは性能がほぼ 0% となるような劣悪な環境下でも 90% 以上の認識精度を出し、また、発音教育支援の分野では、小学生の音声でも精度高く評定できる実用的技術を提供している。革新的な技術を提供し続ける中で、**従来の音声認識技術が（音声言語の運用が困難な）重度自閉症者の音情報処理と極めて類似していることを指摘し、また、音声合成技術が動物的（即ち九官鳥的）音声模倣の実装系であり、例えば、幼児の音声模倣プロセスとは全く異なることなどを指摘している。**端的に言えば、**従来の音声科学・音声工学は、音声によって伝搬される言語メッセージは音声の「音としての実体」が伝えるもの、という大前提の上で議論して来た。申請者はこれを完全否定し、「音の差異（コントラスト）が構成する音声構造」が言語メッセージを担うことを主張した。**言語学や認知科学の中には、類似した主張は見受けられるが、それを物理学・数学を用いて実用可能な情報処理体系として実装したのは、申請者が世界で初めてである。特に「可逆かつ連続的な如何なる写像にも不変なる f-divergence に基づく情報表象」を導いているが、これは「不変なる情報は f-divergence でなければならない」という必要性までも導出していることを強調しておきたい。更に、言語学者からは「ソシール言語哲学の物理的実装」と評されており、申請者の主張が言語学的に高い妥当性を有することも強調しておきたい。本申請は、構造の話者不変性から、メディア普遍性へと視点を移し、非音声メディアに対する応用を検討し、ヒトのみが持つ連合野・前頭前野のメディア普遍の柔軟な情報処理を計算機上に実装する試みである。

上記した主張は当初、音声科学・工学研究者からの理解を得ることが難しかったが、近年、申請者の主張に真摯に耳を傾け、賛同する研究者が急増して来た。ここ数年、招待講演・依頼講演を、各学会や各種研究所にて行なう機会が急増したのは、その現れである。MIT メディアラボ、University College of London、ナイメーヘン大学（オランダ）、シェフィールド大学（英）、九州大学、University of Science and Technology of China、東北大学 21 世紀 COE 「言語・脳・認知」公開講演会、及び同 COE シンポジウム、東大 21 世紀 COE 「言語から読み解くゲノム」公開講演会、オーストラリア音響学会、米音響学会、ESF-JSPS Frontier Science Conference Series for Young Researchers (Experimental Cognitive Robotics)、The Neurosciences Institute (San Diego)、Max Planck Institute (Brain Science, Leipzig)、中国社会科学院（言語学、音声工学）、電子情報通信学会、視聴覚情報研究会 (AVIRG)、幼児のコモンセンス研究会講演会、音声言語シンポジウム、日本 IBM 基礎研究所、Microsoft アジア研究センター、など多数の招待・依頼講演を行ない、多様な分野の研究者と議論を重ねて来た。特に音声言語シンポジウムは日本最大の音声言語工学の研究発表会であるが、10 周年を記念した特別招待講演として発表するに至っている（研究業績番号 1）。これらは、申請者が提唱する新しい音声情報処理に対する並々ならぬ関心の高さを示している。本申請は、その期待を実現するために必要な経済的支援、物質的支援を申請するものである。

申請者の研究者としての特異性を考える場合「網羅する領域の広さ」を特筆しておきたい。上記招待講演の開催場所、機関や、研究業績を見て頂くと分かるが、脳科学、言語障害、認知科学、霊長類研究など、分野横断的に発表の場を持つだけでなく、積極的に、国内外の研究者を訪問し、議論の場を持っている。そして、人間を非常に幅広い観点から眺めた上で、計算機上に構築すべきアルゴリズムを絞り込む戦略をとっている。

次ページに、研究業績欄を参照しつつ、申請者の行なって来た研究内容を年代の古いものから示す。

研究機関名 東京大学

研究代表者氏名 峯松 信明

研究略歴(つづき)

申請者の音声研究は、**音声知覚研究**(人間の音声情報処理に対するモデル構築)から始まった。卒業論文のタイトルは「人間における音声言語処理過程のモデル化」であり、博士学位論文のタイトルは「音声を媒体とした情報の受容に関する基礎的研究」であった。修士時代より、音声分析、音声合成、更には音声認識研究にも従事するようになる。研究業績欄には申請者の発表論文の抜粋が示されているが、音声知覚に関する論文としては、45, 42, 35, 24, 23, 22, 19, 17, 4 などがある。心的辞書への検索過程に関する実験的検討が多い。

音声分析に関する研究としては、音声の韻律的特徴である、基本周波数パターンモデル化や、基本周波数とスペクトルの相関分析、更には、分析合成技術の高精度化を目的とした、PSOLA 分析合成技術の高精度化、最近では、音源分離に関しても研究を重ねて来た。いずれにせよ、音声に含まれる種々の情報は、どのように分解できるのか、という問いに答える試みである。特に音源分離に関する論文は論文賞を獲得することができた。論文番号としては、44, 38, 35, 33, 29, 28, 25, 10, 7 などがある。

音声知覚研究を通して得た知識、或は、自ら実験的に実証してきた種々の事実の応用先として、計算機を用いた**音声認識研究**において多くの成果を上げている。論文番号としては40, 39, 37, 34, 29, 28, 27, 23, 22, 17, 15, 14, 12, 9, 8, 5, 3, 1 などがある。例えば40, 37は現在日本で最も広く使われている日本語ディクテーション基本ソフトウェアの前身となるソフトウェアキットの開発に携わって執筆した論文である。本国の音声認識研究に大きく貢献している。また、39, 34などは音声認識の要素技術の高精度化に相当する。その後、2002~2003年時の海外在外研究において、研究者としての大きな転機を迎え、従来の音声認識手法とは全く異なる、「音声構造に基づく音声認識」を提案するに至っている。実は、この新しい音声認識系の構築は、卒論執筆時以来常に抱いて来た問い(音声物理の多様性と音声知覚の不変性)に対して、初めて自身納得の行く解を導くことに成功し、その解の妥当性を、工学的実装を通して世に問うているものである。音声構造に関する論文は29, 28, 27, 26, 23, 22, 21, 20, 19, 18, 17, 15, 14, 12, 9, 8, 7, 6, 5, 4, 3, 2, 1 などがある。人間の音声知覚過程と機械の音声認識過程との差を埋めることを目的として提唱した方法論であるが、外国語発音矯正技術や、音声合成技術、更には、幼児の言語獲得のモデル化、自閉症や失読症者の情報処理モデルの構築など、様々な分野への応用に成功している。

例えば、21, 19, 18, 16, 4, 2, 1は**言語獲得、言語障害**に焦点を絞って執筆した論文である。国内外にて、これらの分野に対しても情報発信している。特に2は「幼児の音声模倣」という行為に対して音声構造に基づく説明理論を構築した論文であるが、これは論文賞を獲得している。また、18はその説明理論に基づく音声合成系(幼児の音声模倣プロセスの計算機上での実装)の報告であるが、これも賞を獲得している。なお、16は、言語障害学の専門家・臨床家に自閉症、失読症、及び、失語症について工学者向けの説明を依頼し、それを通して現在の音声言語情報処理を再考した日本音響学会誌の特集記事である。このような啓蒙活動を強く押し進めている。

申請者の研究略歴を述べる上で、**外国語発音学習支援**という分野は欠かせない。申請者自ら、日本人学習者を相手に英語発音を指導していた経験を持ち、それを音声工学の技術と統合する形で種々の要素技術、及び、システム開発を手がけて来た。論文番号では43, 41, 36, 32, 31, 30, 26, 15が相当する。音声構造論提唱後は、音声構造に基づいた発音矯正技術の構築や、システム開発を手がけている。特に音声構造は、話者の体格差に起因する音声の音響差異を消失できるため、小学生の音声でも問題無く処理できる。数年後に小学校に導入される英語教育を支援すべく、各種システム開発を行っており、音声構造論のその実用性の高さを裏付けている。なお30は、科学研究費特定領域研究「メディア教育」の計画班にて構築した、日本人学生による英語読み上げ音声データベースに関する論文である。日本の英語教育史上最大の、学習者による読み上げ音声データベースであり、その後の発音教育及び発音教育支援技術の進展に大きな貢献を果たした。なお、音声コミュニケーションに関する世界最大の国際学会であるISCA(International Speech Communication Association)が主催する、外国語教育支援のためのWorkshop(SLaTE, Speech and Language Technology in Education)は2年に一度行なわれているが、2011年は、申請者がGeneral Chairを務め、日本で開催することになっている。

音声構造論は、当初、話者の体格差に起因する音声の音響差異を消失することを目的として提案された。話者Aの声を話者Bの声に変換する話者変換の技術は、音声の話者間写像として実装される。話者差に起因する音響差異に影響を受けない、音声の話者不変の物理表象は、話者変換において用いられる写像に対して不変な物理量を求めることで規定できる。その後、この不変量(f-divergence)が線形、非線形を問わず全ての可逆かつ連続的な変換に対して不変であることが分かり、更には、可逆かつ連続的な変換全てに不変な量を定義しようとすれば、それはf-divergenceにならざるを得ないことも数学的に証明することに成功した。これらの経緯を踏まえ、話者不変の枠組みを、メディア普遍の枠組みへと拡張し、それを、脳の連合野・前頭前野の情報処理として捉える枠組みを提唱するに至った(論文番号13, 7, 1)。メディア普遍の枠組みへと拡張するに際し、申請者がこれまで網羅・調査して来た幅広い人間科学研究(言語学、認知科学、言語障害、脳科学、進化心理学、生態心理学、人類学など)を通して獲得した種々の知識が、当初考えていた以上に、見事に融合され、本申請課題に結実している。ヒトとサル(動物)の脳の情報処理は何が違うのか、自閉症者と健常者の脳の情報処理は何が違うのか、健常なヒトにのみ音声言語獲得を可能ならしめている認知的能力とは何か、ヒトのみが有する極めて柔軟な情報処理は、どのような数学的・物理的・情報論的枠組みの上で議論すべきなのか。本研究課題は、申請者がこれまで行なって来た研究を、更なる高みへと推進すべく、申請するものである。

これまでに受けた研究費とその成果等

本欄には、研究代表者がこれまでに受けた研究費（科学研究費補助金、所属研究機関より措置された研究費、府省・地方公共団体・研究助成法人・民間企業等からの研究費等。なお、現在受けている研究費も含む。）による研究成果のうち、本研究の立案に生かされているものを選定し、科学研究費補助金とそれ以外の研究費に分けて、次の点に留意し記述してください。

- ① それぞれの研究費毎に、研究種目名（科学研究費補助金以外の研究費については資金制度名）、期間（年度）、研究課題名、研究代表者又は研究分担者の別、研究経費（直接経費）を記入の上、研究成果及び中間・事後評価（当該研究費の配分機関が行うものに限る。）結果を簡潔に記述してください。（平成20年度の科学研究費補助金の研究進捗評価結果がある場合には、若手S-13「研究計画と研究進捗評価を受けた研究課題の関連性」欄に記述してください。）
- ② 科学研究費補助金とそれ以外の研究費は線を引いて区別して記述してください。

● **基盤研究(B)(一般)2008～2010年度「音声の分節的・韻律的特徴を包含する発音の構造的表象に基づく外国語教育・学習支援」研究代表者、合計 14,700 千円**

音声構造論に基づく外国語発音教育支援環境の精緻化を目指して始まったプロジェクトである。従来の音声構造に基づく発音評定では、単語音声を対象としており、ここでは、文音声を対象とした要素技術の構築と、プロトタイプシステムの構築を検討している。現在、文音声における構造解析結果が発音評定というタスクにおいて、どの程度の精度を出すのか、またどのような問題が生じるのか、について検討を行っており、随時、問題解決に向けた検討を開始する予定である。なお、若手(S)の申請が受理された場合、本プロジェクトは規則により途中で終了することになる。

● **萌芽研究 2007～2008年度「言語情報とパラ言語情報を統合した音声の構造的表象の提案とその音声合成への応用」研究代表者、合計 3,300 千円**

音声構造の音声認識系への応用とは別に、音声の抽象的表象である音声構造を再度、音声の具象的な空間に戻す作業、即ち、構造からの音声生成系について検討している。現時点では母音連続発声に対して、完全に発話者の身体情報（声道の長さなど）を削除した形で（構造として）表象し、それを出発点として、所望の体格（声道形状）を与える形で構造を具体的な音へと変換するシステムの構築に成功した。これは、語ゲシュタルトを獲得すると言われる幼児の音声模倣プロセスとして捉えることができ、論文賞を受賞するなど高く評価されている。現在、子音への着眼、韻律的特徴への着眼などを行なっている。

● **特定領域研究「情報爆発」2006～2008年度（公募研究）「構造不変の定理に基づく音声アフォーダンスの提案とそれに立脚した音声認識系の構築」研究代表者、合計 11,900 千円**

申請者が「構造不変の定理」と呼ぶ数学定理が導出する音声の構造的表象（音声アフォーダンス）による音声認識系の構築に着手した。理論的進展としては、本表象が可逆かつ連続的な如何なる変換に対しても不変であること、更に、このような不変性の高い表象は、必ず f-divergence を用いざるを得ないことを数学的に証明することができた。更には、幼児の言語獲得（音声模倣）を説明する一つの数理モデルを与えられること、自閉症、失読症が示す一部の症例の説明モデルとして妥当であることを示すことができた。工学的進展としては、連続母音系列については、完全に話者性を削除した形で音声認識系を構築することに成功し、子音を含む連続発声に対しても、極めて高い話者頑健性を示すことができた。

● **基盤研究(B)(一般)2005～2007年度「英語教育の小学校導入を考慮した発音の包括的記述とその伝達能力に立脚した発音支援」研究代表者、合計 15,400 千円**

音声から性別や年齢の情報をそぎ落とすことができる、音声構造論に基づいた外国語発音評定技術の精緻化とプロトタイプシステムの構築を行なった。ここで、話者性を事前にそぎ落とすことができる点に着眼し、様々な要素技術を構築した。1) 個々の学習者の発音変化の様子をカルテとして記録する技術、2) 大勢の学習者を、その年齢や性別は無視して、純粹に発音構造の偏り、歪みに着眼して分類する技術、3) ある学習者と所望の先生の発音とを比較し、どこから修正することが、その学習者にとって最も効率の良い発音矯正となるのかを推定する技術、4) 一般的に退屈になりがちな発音訓練の意欲向上を目的として、学習者の向上心を煽るユーザ・インタフェースの構築、などである。母音学習システムをプロトタイプシステムとして構築し、700名以上の学習者に対してカルテを発行し、彼らの発音診断を行なうことができた。

● **萌芽研究 2003～2005年度「発音の明瞭性に着眼した発音構造の包括的記述とその語学学習への応用に関する研究」研究代表者、合計 3,100 千円**

学習者の発音状況を物理的に記述する際に、発声者の性別や年齢といった非言語情報が見える物理表象ではその技術的安定性は保証されない。本研究では言語学の一部である構造音韻論に着眼し、音声に不可避的に混入する非言語情報を表現する次元を末梢する形で、構造音韻論を物理学の上に実装することに成功した。この研究成果により、学習者の発音をその母国語依存性のみに着眼する形で物理的に表象することが可能となり、202名の発音を幾つかのカテゴリに分類することを可能としている。

● **特定領域研究「情報学」2001～2005年度（公募研究）「人間による音声情報処理過程の分析とそれを応用した音声対話インタフェースの構築」研究代表者、合計 27,200 千円**

人間は五感を用いながら、様々な情報を環境から得ている。ここでは聴覚に絞り、また対象を音声に絞った上で、非言語的情報（話者、年齢、性別等）や、パラ言語的情報（感情、意図等）の抽出過程を計算機上に実装した。知覚的年齢の推定に際しては、大規模音声データベースに対して聴取実験より年齢ラベリ

研究機関名 東京大学

研究代表者氏名 峯松 信明

これまでに受けた研究費とその成果等（つづき）

ングを施し、それに基づいて各年代の音響モデルを作成し、知覚的年齢推定を実装した。また、パラ言語情報の推定に対しては、音声アフォーダンスの基盤ともなる音声の構造的表象を用いて分析を行ない、より高精度な話者意図の推定を実現した。音声に不可避的に混入する非言語的な特徴（年齢、性別、マイク特性など）を一切含まない音声の物理表象への着眼は、本公募研究（情報学）においてその発想を得た。

- **特別領域研究「韻律と音声情報処理」2001～2003年度（計画班）「音声認識・理解における韻律情報の利用」（代表：尾関和彦）研究分担者，合計8,000千円**

韻律の特徴に焦点を絞った音声情報処理の高度化に関する研究プロジェクトである。申請者は本プロジェクトにおいて、音声認識における韻律利用に関する研究を遂行した。ここでは申請者が先行研究として行なった単語音声知覚における韻律の役割に着眼し、句頭のアクセント核の存在によって仮説探索を制御する方法などを提案し、その有効性を実験的に示した。デコーディングの処理は第一、第二パスに分かれるが、第一パスの情報を元に、第二パスの語頭処理について、核の有無に基づいた候補予備選択を実装した。

- **特別領域研究「メディア教育」2000～2002年度（計画班）「音声言語処理技術と学習者モデルを用いた語学学習システムの研究」（代表：中川聖一）研究分担者，合計7,000千円**

日本人学生による英語読み上げ音声データベースの設計・構築（英語教育史上最大データベース）と共に、韻律（特にストレス）に関する学習者モデリング技術の構築を行なった。前者においては、「自らが正しいと考える発音ができるまでその収録を繰り返すこと」を原則とし、その結果「学習者にとって正しい英語音声」のデータベースが構築された。これはある意味、英語発音教育の現在の「性能」を示すデータベースである。ストレスに関する学習者モデルとしては、学生がストレス生成において陥ってしまっている「癖」を定量的に記述する方法を提案し、それに基づくシステム開発を行なった。また、システム開発のみならず、ストレス検出技術を用いて、日本人英語音声に対するストレスラベリングを行なった。

- **NTT データ「音声からの話者特性の抽出に関する研究」2001～2002年度，研究代表者，合計2,400千円**

音声の音響情報からの種々の話者情報・特性の抽出について検討した。知覚的年齢を主な対象としてその基礎技術の構築を行なった。ここでの検討は、後に、特定領域研究「情報学」の公募研究に繋がり、知覚的インタフェースの構築を行なった。

- **科学技術振興事業団「音声の基本周波数とスペクトル構造間の依存性の研究」2000～2002年度，研究代表者，合計6,300千円**

音声の韻律的特徴と分節的特徴は、従来独立要因として扱われることが多い。しかし、音声技術の高精度化を実現する場合、両者の相互依存性も定量的にモデル化すべきとの声が高い。ここでは、両者の相関を音韻性、話者性に着眼して定量的に分析し、音韻依存性、話者依存性について評価を行なった。

- **情報処理振興事業協会「擬人化音声対話エージェント基本ソフトウェアの開発」（代表：嵯峨山茂樹）2000～2004年，研究分担者，合計4,500千円**

擬人化エージェント基本ソフトウェアの開発に際して、音声合成部の韻律制御（アクセント辞典及びアクセント結合規則の構築）を担当した。従来より広く使われている規則の問題点を洗い出し、聴取実験を通して規則の改良を試み、より高精度なアクセント結合のモデル化が可能となった。

- **IPA 独創的情報技術育成事業「日本語ディクテーション基本ソフトウェアの開発」（代表：鹿野清宏）1997～2000年度，研究分担者，合計5,200千円**

大語彙音声認識技術の社会的普及を目的として「音響モデル」「言語モデル」「デコーダ」「形態素解析器」の各々を無償で利用出来るプラットフォームを実現した。現在の国内の音声認識研究の多くが、また、海外の幾つかの研究サイトが、本プロジェクトの成果（及び、後継コンソーシアムの成果）を研究プラットフォームとして利用している。申請者は本プロジェクトの中で、主に、音響モデルの構築作業に携わった。

研究計画と研究進捗評価を受けた研究課題の関連性

- ・本欄には、本応募の研究代表者が、平成20年度に、「特別推進研究」、「基盤研究(S)」又は「学術創成研究費」の研究代表者として、研究進捗評価を受けた場合に記述してください。
- ・本欄には、研究計画と研究進捗評価を受けた研究課題の関連性(どのような関係にあるのか、研究進捗評価を受けた研究を具体的にどのように発展させるのか等)について記述してください。

人権の保護及び法令等の遵守への対応（公募要領9頁参照）

本欄には、研究計画を遂行するにあたって、相手方の同意・協力を必要とする研究、個人情報の取り扱いの配慮を必要とする研究、生命倫理・安全対策に対する取組を必要とする研究など法令等に基づく手続きが必要な研究が含まれている場合に、どのような対策と措置を講じるのか記述してください。

例えば、個人情報を伴うアンケート調査・インタビュー調査、患者から提供を受けた試料の使用、ヒト遺伝子解析研究、組換えDNA実験、動物実験など、研究機関内外の倫理委員会等における承認手続きが必要となる調査・研究・実験などが対象となります。

なお、該当しない場合には、その旨記述してください。

該当しない。

研究経費の妥当性・必要性

本欄には、「研究計画・方法」欄で述べた研究規模、研究体制等を踏まえ、次頁以降に記入する研究経費の妥当性・必要性・積算根拠について記述してください。また、研究計画のいずれかの年度において、各費目（設備備品費、旅費、謝金等）が全体の研究経費の90%を超える場合及びその他の費目で、特に大きな割合を占める経費がある場合には、当該経費の必要性（内訳等）を記述してください。

平成21年度の研究計画は、1) 多様性を表現する写像関数群の推定方法の確立、2) 共鳴音と非共鳴音とを分割した上での構造化の検討、3) 構造を具体的なメディア空間に定位する構造音声合成の実装、4) 手運動と音運動を題材として実装する異メディア変換と両メディア普遍の構造表象の検討、である。購入予定物品は基本的に、これらの研究項目を推進するために不可欠な物品である。特に1)の検討に際しては当初考えていたよりも遥かに高い計算機性能（計算サーバ）が必要であることが分かって来た。更に、構造表象の応用先を音声に留めず、視覚や身体運動にまで拡張しながら検討を進める場合、膨大なデータ量を蓄積する必要があり、データサーバの性能向上も必須である。データ収集に際しては、音メディアの収集は既存の設備で賄えるが、それ以外のデータ収集に関しては、視覚メディア、身体メディアの両者において、機器を新たに購入する必要がある。手運動と音声運動との空間写像に関しては、既に旧式のデータグローブを使った実装を行なったが、より分解能の高いセンサーが埋め込まれたグローブの使用により、より細かな運動モデル、制御が可能になる。また、手に限らず身体全身の動きをモニターするためにはモーションキャプチャなどの機器も必要となる。24年度以降は、構造表象に基づくメディア情報処理の脳科学的妥当性を検証するために、TMSの購入を考えている。

「準備状況」の欄に記述した様に、本申請課題では、異なるメディア情報処理を統合する必要があるため、視覚メディア、身体メディアの若手研究者を雇用する。視覚メディアについては、既に学振のポスドクとして1.5年に渡り峯松研究室を支援してくれた喬宇氏を採用する。学振相当の金額を支給することを考え、年棒として540万円を考えている。また、23年度以降、二人目のポスドク（ロボット工学）を雇用することを考えており、彼にも年棒として540万円を考えている。

備品費や人件費以外にも、各種消耗品、被験者謝金、更には、情報収集・研究打合せ・成果発表などの国内外の出張旅費など、必要不可欠な経費のみを申請している。なお、既述している通り、本申請課題を遂行するに当たって、他分野の研究者との交流が必要不可欠であり、情報収集のための旅費として必要十分な額を申請している。これらを考慮すると、次頁以降に示す経費は、極めて健全で妥当である。

主な現有設備 (記入に当たっては、若手研究(S) 研究計画調書作成・記入要領を参照してください。)						
研究機関	設備名	仕様(形式・性能)	専用・共同利用の別	設置年度	備考	
東京大学	音声分析サーバ	DELL PowerEdge1950・ク アドコアインテル Xeon プ ロセッサ(×1)	専用	2006	上記2台では処理力不足 今後も利用予定	
東京大学	音声分析サーバ	DELL PowerEdge1950・ク アドコアインテル Xeon プ ロセッサ(×1)	専用	2007		
東京大工	音声収録装置	DELL Inspiron PC (×10)	専用	2006		
設備備品費の明細 (記入に当たっては、若手研究(S) 研究計画調書作成・記入要領を参照してください。)(金額単位:千円)						
年度	品名・仕様	数量	単価	金額	主として使用する研究者 及び設置機関名	購入予定 時期
21	計算サーバ・DELL Pow- erEdge1950	2	1,000	2,000	東京大学	6月
	高分解能データグローブ・ 5DT社 Data Glove 14 Ultra	2	1,000	2,000	東京大学	7月
	動画収録装置・DELL Inspiron PC	3	200	600	東京大学	8月
	動画収録装置・SONY ビデオ カメラ DXC-D55WSL	1	2,000	2,000	東京大学	8月
	計			6,600		
22	身体データ収録装置・Vicom 社 Motion Capture System T160	1	8,000	8,000	東京大学	8月
	身体データ収録装置・DELL Inspiron PC	3	200	600	東京大学	8月
	計			8,600		
23	各種メディアデータサーバ・ DELL PowerEdge6950	3	1,500	4,500	東京大学	6月
	メディア普遍表象デモンスト レーション装置・DELL Insp- iron PC	3	200	600	東京大学	6月
	計			5,100		
24	研究用コミュニケーションロ ボット・Vstone社 Robovie-R	1	5,000	5,000	東京大学	6月
	ロボット制御装置・DELL In- spiron PC	2	200	400	東京大学	6月
	評価実験装置・DELL Inspiron PC	3	200	600	東京大学	10月
	計			6,000		
25	経頭蓋磁気刺激装置・ミュキ 技研マグスティム TMS	1	4,000	4,000	東京大学	6月
	評価実験装置・DELL Inspiron PC	5	200	1,000	東京大学	6月
	計			5,000		
研究機関名	東京大学			研究代表者氏名	峯松 信明	

消耗品費等の明細

(記入に当たっては、若手研究(S)研究計画調書作成・記入要領を参照してください。)

(金額単位：千円)

年度	消耗品費		旅費		謝金等		その他	
	事項	金額	事項	金額	事項	金額	事項	金額
平成21年度	聴覚・視覚・身体メディアの各種データベース	300	(国内)研究打合せ	300	学術支援研究員(1人×12月×450千円)	5,400	通信費	100
	信号処理ソフトウェア	500	成果発表	500	専門知識の提供(3人×12月×20千円)	720	研究打合せ会場費	200
	プリンタ・プロッタトナー	300	(外国)研究打合せ	500	音声収録用謝金	200	論文投稿料	500
	Flushメモリなど各種記録メディア	500	成果発表	1,000				
	計	1,600	計	2,300	計	6,320	計	800
平成22年度	RAID用HDD	500	(国内)研究打合せ	300	学術支援研究員(1人×12月×450千円)	5,400	通信費	100
	各種ソフトウェア	300	成果発表	500	専門知識の提供(3人×12月×20千円)	720	研究打合せ会場費	200
	ライセンス更新	300	(外国)研究打合せ	500	身体運動記録用謝金	200	論文投稿料	500
	プリンタ・プロッタトナー	300	成果発表	1,000				
	Flushメモリなど各種記録メディア	500						
計	1,600	計	2,300	計	6,320	計	800	
平成23年度	RAID用HDD	500	(国内)研究打合せ	300	学術支援研究員1(1人×12月×450千円)	5,400	通信費	100
	各種ソフトウェア	300	成果発表	500	学術支援研究員2(1人×6月×450千円)	2,700	研究打合せ会場費	200
	ライセンス更新	300	(外国)研究打合せ	500	専門知識の提供(3人×12月×20千円)	720	論文投稿料	500
	プリンタ・プロッタトナー	300	成果発表	1,000	評価実験用謝金	200		
	Flushメモリなど各種記録メディア	500						
計	1,600	計	2,300	計	9,020	計	800	
平成24年度	高速SDD	500	(国内)研究打合せ	300	学術支援研究員1(1人×12月×450千円)	5,400	通信費	100
	各種ソフトウェア	300	成果発表	500	学術支援研究員2(1人×12月×450千円)	5,400	研究打合せ会場費	200
	ライセンス更新	300	(外国)研究打合せ	500	専門知識の提供(3人×12月×20千円)	720	論文投稿料	500
	プリンタ・プロッタトナー	300	成果発表	1,000	評価実験用謝金	200		
	Flushメモリなど各種記録メディア	500						
計	1,600	計	2,300	計	11,720	計	800	
平成25年度	高速SDD	500	(国内)研究打合せ	300	学術支援研究員1(1人×12月×450千円)	5,400	通信費	100
	各種ソフトウェア	300	成果発表	500	学術支援研究員2(1人×12月×450千円)	5,400	研究打合せ会場費	200
	ライセンス更新	300	(外国)研究打合せ	500	専門知識の提供(3人×12月×20千円)	720	論文投稿料	500
	プリンタ・プロッタトナー	300	成果発表	1,000	評価実験用謝金	200		
	Flushメモリなど各種記録メディア	500						
計	1,600	計	2,300	計	11,720	計	800	

研究費の応募・受入等の状況・エフォート

本欄は、合議審査において、「研究資金の不合理な重複や過度の集中にならず、研究課題が十分に遂行し得るかどうか」を判断する際に参照するところであり、研究代表者の応募時点における、(1) 応募中の研究費、(2) 受入予定の研究費、(3) その他の活動、について、次の点に留意し記入してください。なお、複数の研究費を記入する場合は、線を引いて区別して記入してください。

- ① 「エフォート」欄には、年間の全仕事時間を100%とした場合、そのうち当該研究の実施等に必要となる時間の配分率(%)を記入してください。
- ② 「応募中の研究費」欄の先頭には、本応募研究課題を記入してください。
- ③ 科学研究費補助金の「特定領域研究」及び「新学術領域研究」の領域提案型にあつては、「計画研究」、「公募研究」の別を記入してください。
- ④ 所属研究機関内で競争的に配分される研究費についても記入してください。

(1) 応募中の研究費

資金制度・研究費名・研究期間 (配分機関等名)	研究課題名 (研究代表者氏名)	役割 (代表・分担の別)	平成 21 年度 研究経費 (期間全体の額) (千円)	エフォート (%)	研究内容の相違点及び他の研究費に加えて本応募研究課題に応募する理由
【本応募研究課題】 若手研究 (S) (H21~H25)	構造不変性に基づくメディア普遍の運動表象とそれに基づく脳型メディア情報処理	代表	17,620 (99,900)	35	
特定領域研究 (公募班) (H21 ~ H22) 東京大学	構造不変の定理に基づく音声アフォーダンスの提案とそれに基づく音声認識系の構築	代表	4,700 (9,300)	20	構造表象に基づく音声認識系の構築に特化しており、メディア普遍の情報表象の構築はその目的ではない。
挑戦的萌芽研究 (H21 ~ H22) 東京大学	幼児の音声模倣と音声ゲシュタルトに着眼した構造表象に基づく音声生成系の構築	代表	2,650 (5,000)	—	音声合成 (生成) に主眼を置いた構造表象の応用研究であり、メディア普遍の情報表象の構築はその目的ではない。なお、若手 (S) が採択された場合、萌芽研究は辞退することになる。
基盤研究 (B) (H21 ~ H23) 東京大学	生成過程モデルに基づく表現力豊かな多言語音声合成とそれによる音声自動翻訳 (広瀬啓吉)	分担	600 (1,800)	5	複数言語の音声合成・音声翻訳に主眼を置いた研究であり、本申請研究との関係は無い。
基盤研究 (B) (H21 ~ H23) 東京国際大学	音読とシャドーイングに音声情報処理技術を活用した外国語能力評価システムの開発 (山内豊)	分担	3,000 (6,000)	5	シャドーイングという外国語学習に特化した発音・聴取訓練方法の自動化に関する研究であり、本申請課題との関係は無い。
研究機関名	東京大学		研究代表者氏名	峯松 信明	

研究費の応募・受入等の状況・エフォート（つづき）					
(2) 受入予定の研究費					
資金制度・研究費名・研究期間（配分機関等名）	研究課題名（研究代表者氏名）	役割（代表・分担の別）	平成21年度研究経費（期間全体の額）(千円)	エフォート(%)	研究内容の相違点及び他の研究費に加えて本応募研究課題に応募する理由
基盤研究(B) (H20～H22) 東京大学	音声の分節的・韻律的特徴を包含する発音の構造的表象に基づく外国語教育・学習支援	代表	4,800 (14,700)	－	構造表象に基づく外国語発音学習支援環境の構築であり、申請課題とは目的が大きく異なる。なお、若手(S)が採択された場合、基盤研究(B)は辞退することになる。
特定領域研究(計画班) (H18～H22) 東京大学	多様な目的に適した形態素解析システム用電子化辞書の開発（伝康晴）	分担	2,500 (13,300)	10	音声合成を目的としているが、テキスト処理に関する研究なので、申請課題と直接的な関係は無い。
(3) その他の活動 〔上記の応募中及び受入予定の研究費による研究活動以外の職務として行う研究活動や教育活動等のエフォートを記入してください。〕				25	
合計 (上記(1)、(2)、(3)のエフォートの合計)				100 (%)	