

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成 25 年 6 月 6 日現在

機関番号：12601

研究種目：挑戦的萌芽研究

研究期間：2011 ～ 2012

課題番号：23650133

研究課題名（和文）ゲシュタルト知覚モデルに基づくパラ言語情報処理系の構築と発達障害者の行動理解

研究課題名（英文）Development of para-linguistic information processing systems based on Gestalt perception and understanding of behaviors of individuals with developmental disorder

研究代表者

峯松 信明（MINEMATSU NOBUAKI）

東京大学・大学院工学系研究科・教授

研究者番号：90273333

研究成果の概要（和文）：

- 1) 音声構造の動的特徴を定義することで音声構造による単語音声認識率の精度向上を実現した。
- 2) 従来の音声認識系におけるリランキング処理に構造特徴を導入することで、精度向上を実現した。
- 3) 幼児の言語獲得モデルとして音声構造を導入し、その妥当性を示すことができた。
- 4) サルとヒトとの音情報処理の差異を考察することで、音声言語発生の必要条件について論じた。
- 5) 自閉症児童の経時的セラピーを通して、彼らの感情コントロールや感情推測の能力の向上に貢献することができた。

研究成果の概要（英文）：

- 1) Dynamic features of speech structure are introduced and used for improving word recognition performance.
- 2) Speech structure is used in a re-ranking process in LVCSR systems and it is found to improve the performance.
- 3) Modeling of infants' language acquisition process is investigated by using speech structure.
- 4) A theory is proposed to explain the birth of spoken language, where cognitive differences between human and ape are well considered in terms of sound processing.
- 5) A longitudinal study is carried out to teach autistic children how to interact with others socially adequately. Our method of therapy is found to be effective in improving their social skills.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
交付決定額	2,800,000	840,000	3,640,000

研究分野：複合領域

科研費の分科・細目：情報学・認知科学

キーワード：認知工学

1. 研究開始当初の背景

音声コミュニケーションにおいて、意図や感情など（パラ言語情報）の適切な認知と生成は不可欠な機能である。自閉症の一種であるアスペルガー症候群は、音声コミュニケーション障害を持つことが多いが、主には、パラ言語情報処理における障害である。その理由として、与えられた刺激の要素的な側面ではなく、全体的な側面（ゲシ

タルト)を捉える困難が指摘されている。本研究では、代表者が提唱している構造的な情報表象をゲシュタルトの数学的な実装系と位置づけ、この表象に基づく音声認識系の精度向上を検討すると共に、経時的な自閉症者セラピーを通して、ゲシュタルトモデルの適用、モデルを通じた彼らの行動理解について検討する。

2. 研究の目的

本研究のより詳細な目的は以下のようになる。1) 音声の構造的表象は時間軸を非常に粗く離散化することが必要となり、このままでは動的特徴の定義そのものが困難である。これを解決する手法を提案する。更には、現在の音声認識系との効果的な融合を図る。2) (発達障害の中でも特に) 自閉症児童を対象としたコミュニケーションスキル (特に感情コントロール) セラピーを経時的に行なう。この場合、観測者として両親にも参加して戴き、両親とのカンファレンスを通して自閉症児童の変化を記録する。その様子をゲシュタルトモデルを通して分析し、彼らの行動様式の理解に努める。

3. 研究の方法

本研究開始当初の目的は上記の通りであるが、研究を推し進めるに際して、上記以外についても、a) 自閉症のみならず健常児を対象とした言語獲得の様子のモデル化とそのシミュレータの構築についても検討する、b) 動物と人間の音情報処理の差異に関する各種論文調査を行なうと共に、言語の起源 (言語が生まれるための必要条件) についても考察し、論文として世に問う、などの検討も行なうこととした。自閉症者に見られる音声言語獲得の困難さは、逆に健常者が何故音声言語を獲得できるのか? を問うこととに繋がる。その意味において、言語の起源について論じることは有益なことであると判断した。

音声の構造的表象に関する技術精緻化に関しては、以下のような方法を採用した。

1) 入力音声を一端トラジェクトリ HMM を用いてモデル化し、その結果を用いて音声構造を抽出する。こうすることで、非常に細かい時間離散化が可能となり、動的特徴の定義がより容易となる。

2) 音声構造は離れた音声事象間の距離のみを特徴量として抽出するため、フレーム時系列として音声を表象する従来技術とは非常に相性が悪い。その結果、デコーディング処理の中に構造的特徴を導入することが難しい。ここでは、デコーディング処理結果を複数の仮説として保持し、これらの仮説を再評価刷る時に構造的特徴を用いる識別的リランキング法を検討する。

健常者を対象とした音声言語獲得のモデル化とそのシミュレーションに関しては以下の方法を採用した。

3) 音声の構造的表象は話者の違いを捨象した音声モデルと位置づけられ、話者の違いに「鈍感に」音声を学習する、健常者の音声言語獲得モデルとして利用可能である (自閉症

者の場合、話者の違いに敏感になる様子が報告されている)。ここでは構造表象を用いるだけではなく、リズムへの敏感さを考慮した音節内構造と音節間構造を定義し、それらを認識系に導入する効果を見る。

更に、言語起源に関する考察に関しては下記の方法を採用した。

4) 人間と動物 (チンパンジーなど) との音の情報処理の差異を論文調査していくと、音ストリームの中に存在する相対的な不変項を抽出することが困難である様子が観測される。例えば移調前後のメロディーの同一性が認識できないなどは良い例である。同時に、彼らの発声器官の物理的制約を考慮すると、このような音の動きを多様に生成することが困難であることも示される (構音器官の制御が人間に比べて非常に単純)。これらを考慮し、音声言語を有するための必要条件について、理論としてまとめた。

自閉症者のセラピー及び行動観察については以下の通りである。

5) 獨協医科大学にて自閉症児童を対象にコミュニケーションスキルセラピー、特に、感情コントロールセラピーを開催している櫻庭京子博士に協力を依頼し、セラピーによる効果を経時的に記録することとした。この場合、ビデオカメラなどを使い、彼らの成長の様子を記録すると共に、特に、音声言語の獲得に困難を示す自閉症児童に関しては、別途、聴取実験などを行なうことで、彼らの音情報処理について分析対象とすることとした。このセラピーの効果を見るために、自閉症児童の両親にも協力してもらい、セラピー後にカンファレンスを行ない、彼らの家庭での変化、セラピーを通じた変化を口述してもらい、それについても資料として記録することとした。

得られた結果を、ゲシュタルト知覚理論と照らし合わせ、技術的に構築している音声構造技術との接点についても検討する。

4. 研究成果

1) トラジェクトリ HMM を通して時間的分解能を高めた上で音声構造を定義し、それに基づいて動的特徴 (Δ 特徴) を定義し、これを用いた構造的音声認識系を構築した。タスクは孤立単語の音声認識である。実験の結果、構造的な動的特徴をリランキング処理に使うことで、単語音声認識の精度向上に成功し、動的な構造的特徴の有効性を示すことができた。なお、本成果は国際ワークショップにて学生論文賞を受賞した。

2) 構造的特徴は時間的に離れた二事象間の

距離のみを特徴量として採択しており、音声をフレーム時系列として表象して認識する従来の音声認識技術との相性は悪い。そこで、柔軟な特徴が使える、ランキング手法に着眼し、構造的特徴を識別的ランキング手法に導入し、その精度向上を検討した。その結果、大語彙連続音声認識をタスクとした場合においても、構造的特徴による精度向上が確認でき、提案手法の妥当性を示すことができた。なお、この成果は音響学会の粟屋賞を受賞している。

3) 単語音声を対象として構造的表象を応用すれば、単語は音素の並びとして表象されるのではなく、あくまでも、音声事象群が成す全体的な「まとまり」として表象される。これは単語知覚のゲシュタルトモデルと言える。一方、幼児の音声習得においてリズムへの敏感さは広く報告されている。ここでは、リズムへの着眼を、音節核抽出と、音節核を用いた(部分)構造の定義として実装した。音節核を中心とした部分構造と、離れた音節核と核の音節対を対象にした部分構造とを定義し、これらを、単語全体を構造として表象する従来方法に組み入れることで単語認識精度の向上を実現した。

4) 調不変のメロディー表象として相対音感がある。任意の二音の音高差(音程)のみに着眼する表象であるが、音声の構造的表象は音色の相対音感として解釈できる。この場合、音色を動的に変化させるのは、調音器官(舌)の柔軟な制御によるところが大きい。一般にチンパンジーなどの霊長類ではこのような柔軟な舌運動ができず、音色に変化をつけることが難しい。つまり、喋ることが困難である。また、チンパンジーなどの霊長類は相対音感を持っていないので、移調前後のメロディーの同一性認知が難しい。このように、音ストリームの動きを十分に生成できず、また、刺激として与えられた音ストリームの動きに対して、相対的な不変量の抽出にも困難がある。このような観点からチンパンジーと人との音情報処理能力差を論じ、音声言語が生まれる必要条件として論じた。

5) 自閉症児童を対象にした経時的セラピーによって、彼らの感情コントロールや、感情を読み取る能力に徐々にではあるが確実に変化が見られた。両親による日常生活のモニタリングにおいても、彼らの判断基準に変化が見られ、より適切な他者とのインタラクション、より社会的なインタラクションを行なうことが可能となった。また、重度自閉症児童に関しては、別途、複数話者による音声を使った聴取実験などを通して彼らの音に対する反応を試験的に観測するなどの実験も

行なった。セラピーとしては十分な効果はあったが、1) - 2) で検討した音声構造抽出技術、それに基づく認識技術との強い接点を見いだすには至らなかった。

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 1 件)

S. Shimizu, M. Suzuki, N. Minematsu, and K. Hirose, "An experimental study on dynamic features of speech structure," Journal of Research Institute of Signal Processing, vol.16, no.4, pp.319-322 (2012-7)

[学会発表] (計 12 件)

1 櫻庭京子, 峯松信明, 作田亮一, 今泉敏, 堤剛, "感情のコントロールを目的とした SST グループに参加した母親の養育態度の変化", 日本発達心理学会第 24 回退会発表論文集, P1-032, p.109 (2013-3)

2 尾崎洋輔, 峯松信明, 広瀬啓吉, Donna Erickson, "波形包絡を用いた音節核の自動抽出とそれを用いた構造的表象による単語獲得プロセスのモデル化の書記検討", 電子情報通信学会音声研究会資料, SP2012-94, pp.113-118 (2012-12)

3 M. Suzuki, G. Kurata, M. Nishimura, N. Minematsu, "Discriminative reranking for LVCSR leveraging invariant structure," Proc. INTERSPEECH (2012-9)

4 尾崎洋輔, 峯松信明, 広瀬啓吉, エリクソン・ドナ, "波形包絡を用いた音節核の自動抽出とそれを用いた構造的表象による単語獲得プロセスのモデル化", 日本音響学会秋季講演論文集, 2-Q-a4, pp.519-522 (2012-9)

5 鈴木雅之, 倉田岳人, 西村雅史, 峯松信明, 広瀬啓吉, "音声の構造的表象を用いた大語彙音声認識の識別的ランキング", 日本音響学会秋季講演論文集, 2-1-4, pp.63-66 (2012-9)

6 峯松信明, "ことばって一体何だろう? ~ 音声を物理的に眺める一人の研究者が考えること~", 第 56 回日私小連全国教育夏季研究会外国語部会 (2012-8)

7 S. Shimizu, M. Suzuki, N. Minematsu, and K. Hirose, "An experimental study on

dynamic features of speech structure,” Proc. RISP International Workshop on Nonlinear Circuits, Communication and Signal Processing (NCSP), pp.166-169 (2012-3)

8 峯松信明, “音声の音響的多様性と認知的容易性 ～言葉は音の何が伝えるのか?～”, 第五回千葉工業大学音響情報フォーラム「音声の情報性」(2011-11)

9 櫻庭京子, 堤剛, 峯松信明, 田山二郎, 工藤翔永, 今泉敏, “性同一性障害者(MtF)の音声に対する話者認識技術を用いた女声度の自動推定とその臨床”, 日本音声言語医学会学術講演会, 101 (2011-10)

10 清水信哉, 鈴木雅之, 峯松信明, 広瀬啓吉, “音声の構造的特徴の動的特徴に関する実験的検討”, 日本音響学会秋季講演論文集, 1-10-2, pp.3-6 (2011-9)

11 鈴木雅之, 倉田岳人, 西村雅史, 峯松信明, “音声の構造的表象を用いた連続数字音声認識”, 日本音響学会秋季講演論文集, 1-10-14, pp.41-44 (2011-9)

12 M. Suzuki, G. Kurata, M. Nishimura, N. Minematsu, “Continuous digits recognition leveraging invariant structure,” Proc. INTERSPEECH, pp.993-996 (2011-8)

[図書] (計 1 件)

N. Minematsu, “A modulation-demodulation model for speech communication and its emergence,” In The Evolution of Language, edited by T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill, and J. R. Hurford, pp.234-241, World Scientific (2012-3)

[産業財産権]

○出願状況 (計 0 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
出願年月日 :
国内外の別 :

○取得状況 (計 0 件)

名称 :
発明者 :

権利者 :
種類 :
番号 :
取得年月日 :
国内外の別 :

[その他]
ホームページ等

6. 研究組織
- (1) 研究代表者
峯松 信明 (MINEMATSU NOBUAKI)
東京大学・大学院工学系研究科・教授
研究者番号 : 90273333
- (2) 研究分担者
- (3) 連携研究者
熊谷晋一郎 (KUMAGAYA SHINICHIRO)
東京大学先端科学技術研究センター・講師
研究者番号 : 00574659