

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：45309

研究種目：若手研究（B）

研究期間：2009～2011

課題番号：21700565

研究課題名（和文） プロソディの音声解析を活用した基本的言語形成期における
自閉症超早期診断ツール開発研究課題名（英文） Development of Early Diagnosis Tools for Autism during the
Formation of Basic Language Based on Speech Prosody Analysis.

研究代表者

中井 靖 (NAKAI YASUSHI)

川崎医療短期大学・医療保育科・助教

研究者番号：80462050

研究成果の概要(和文)：自閉症児と定型発達児の音声データベースを開発した。これを用いて、イントネーションの音声解析プログラムを開発し、さらに自閉症児と定型発達児の音声識別プログラムを開発した。識別精度は、認識率に加え、 F 値 = $(2 \times \text{適合率} \times \text{再現率}) / (\text{適合率} + \text{再現率})$ を指標とした。認識率は人間が.69、機器は.76、 F 値は人間が.59、機器が.73 であった。つまり、自閉症児と定型発達児の音声識別において、機器の有効性が示唆された。

研究成果の概要（英文）：This study developed the speech database of children with autism spectrum disorder and typical development. Using this, this study developed a voice of intonation analysis program and a speech identification program of children with autism spectrum disorder and typical development. The identification precision was used for accuracy and F-measure $(2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$. The accuracy of the human auditory perception and the speech identification device was .59 and .73, F-measure was .59 and .73. Thus, a higher value was obtained for the speech identification device.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	1,900,000	570,000	2,470,000
2010 年度	1,000,000	300,000	1,300,000
2011 年度	500,000	150,000	650,000
総計	3,400,000	1,020,000	4,420,000

研究分野：人間医工学

科研費の分科・細目：リハビリテーション科学・福祉工学

キーワード：自閉症・早期診断・音声データベース・プロソディ・音声解析・音声識別プログラム・識別精度・ F 値

1. 研究開始当初の背景

自閉症スペクトラム障害（Autism Spectrum Disorder；以下 ASD とする）のある子どもの話し言葉は、定型発達（Typical Development；以下 TD とする）児と比べ、音声リズムが単調であること、文章の最後が疑問文のように上がること等の独特なプロソディを持つ。プロソディとは、音声リズム、抑揚、音程、話速等の音声性質のことであり、

これは話者感情や会話文脈により流動的に変化する。そして、この独特さは語彙獲得以前にも認められることから、ASD 児は 2 歳までの基本的言語形成期において既に独特なプロソディを持つことが考えられる。すなわち、1 歳 6 ヶ月から 2 歳までの ASD 児と TD 児のプロソディの違いが解明されれば、ASD の超早期診断が可能となる。

しかし、先行研究は ASD 児に限らず、TD

児についても、プロソディを十分に解明していない。なぜなら、プロソディは客観的指標による評価が困難だからである。先行研究では、設定した文章を対象児に読ませる等の課題を用いているが、これではプロソディの特徴を適切に評価できない。また、聞き手の印象や基本周波数グラフの視覚的確認などの主観的評価が用いられてきた。

2. 研究の目的

本研究では、最新の音声解析技術を活用することで、従来のように人間の主観によってプロソディを評価するのではなく、人間では見つけることができなかった ASD 児と TD 児の違いを示す新たな特徴量空間を発見することを目的とした。

3. 研究の方法

(1) 音声データベースの作成

①参加者

2010年4月から7月の間に神戸大学医学部附属病院の小児科発達神経外来を受診し、Diagnostic and Statistical Manual of Mental Disorders 4th edition-Text Revision (DSM-IV TR)¹⁾の基準に基づいて ASD の診断を受けた3歳から10歳までの子どものうち、親から許可が得られ、知的に大きな遅れのない子ども30名が本研究に参加した。そのうち、音声データとして用いる単語が30語以上抽出できた27名(就学前6名、就学後21名: ASD群)を、普通幼稚園または普通小学校に在籍するTD児51名(就学前24名、就学後27名: TD群)とした(表1)。

表1 参加者の特徴

		ASD群	TD群
人数		27	51
生活年齢	平均±SD	7:5±1:2	7:4±1:9
	範囲	4:10-10:2	4:3-10:9
男女比	男	21	21
	女	6	30

②課題

構音検査²⁾の単語検査に準じて実施した。今回の研究では50枚の絵カードをすべて提示し終えた後、3分間の休憩を挟み、2回目を実施した。また、自閉症スクリーニング質問紙日本語版 (Autism Screening Questionnaire ; 以下 ASQ とする)³⁾への記入を参加者の親に求めた。なお、ASQ の使用には翻訳者の許諾を得た。

③手順

静かな個室において個別に単語検査を実施し、その様子を録音した。録音機材にはラベリアマイクロホン、オーディオキャプチャ、モバイルノートパソコンを使用した。録音時、ラベリアマイクロホンを参加者の襟元に固

定した。録音の設定はサンプリング周波数を48kHz、量子化を24ビットとした。

本研究では、実験者効果⁴⁾を抑制するために、さらに参加者の自然な話し言葉を録音するために以下の通り行った。すなわち、参加者が模倣の影響を受けないよう、絵カードを提示する際、検査者は口頭で「これは何？」とだけ教示し、参加者が名称を誤ったり答えなかったりした場合でも正しい名称を教示せず、次の絵カードを提示した。単語検査がすべて終了した後に、参加者に正しい名称を教示するようにした。また、ASQ については、単語検査を実施している間に参加者の親に記入を依頼した。

④音声データ

解析対象は絵カードの名称、すなわち単語とした。音声編集ソフトウェアを用いて録音データのサンプリング周波数を24kHzに変換し、雑音等を除去して参加者が話した単語を一語ずつ切り分けた。

(2) イントネーションの定量的評価

①音声解析プログラム

作成した音声データベースを用い、新たに開発した音声解析プログラムを用いてイントネーションを定量的に評価することを試みた。このプログラムではイントネーションを1/100秒ごとのピッチの連続体として表し、イントネーションの定量的指標としてピッチ幅(ピッチの最大値から最小値を引いた値)、ピッチ標準偏差(Standard Deviation ; 以下 SD とする、平均値を基準としたピッチのばらつき)、ピッチ変動係数(Coefficient of Variation ; 以下 CV とする、個々の平均値が異なることを考慮した相対的なピッチのばらつき)の3つを測定した。さらに、3つの定量的指標に基づいて測定した値を、参加者ごとにまとめて算出した場合と一語ずつ算出した場合の2通りの方法で算出した。つまり、3つの定量的指標と2つの算出方法から6つの値を導き出し、イントネーションを定量的に評価した。音声解析には数値解析ソフトウェアを使用した。

②統計解析

音声解析プログラムを通じて導き出した値を用い、ASD 児と TD 児のイントネーションが年齢に応じてどのように変化するかを比較するため、参加者種別(ASD群、TD群)と発達的变化(就学前、就学後)の2要因参加者間計画による分散分析を行った。さらに、ASD の行動特徴の重症度とイントネーションの関係について検討するため、ASQ の得点とイントネーションの相関関係を調べた。統計解析には統計解析ソフトウェアを使用した。

(3) プロソディの定量的評価

①参加者

ASD 児の言語訓練に関する臨床経験を3

年以上有する言語聴覚士 10 名が参加した。

②課題

音声データベースの 10 語の単語について音声識別を行った。単語 1 語につき、ASD 群と TD 群のデータから無作為にデータを 10 個抽出した。すなわち、10 語の単語における 100 個のデータを、音声識別の対象とした。データの配列順は 10 語の単語の順番を統一し、各単語内のデータの順番を無作為にすることで 5 通りとした。データの配列順によるバイアスを抑制するため、ひとつの配列順につき、2 名の参加者が音声識別を行った。

③手順

実験者は参加者に、実験手順書と音声データ Compact Disc (CD)、回答用紙を渡した。実験者効果を抑制するために、参加者は各自で実験手順書に基づいて、音声データ CD を再生した。ヘッドフォンを使用してデータをひとつずつ聴き、ASD 児あるいは TD 児のどちらかとして識別し、識別結果を回答用紙に記入した。実験終了後、参加者は実験者に、音声データ CD と回答用紙を返却した。

④音声識別プログラム

本研究では、新たに音声識別プログラムを開発した。音声識別プログラムの開発には数値解析ソフトウェアを使用した。音声特徴から感情を識別する研究に基づき^{5,6)}、窓幅 32 msec、フレームシフト 8 msec の条件下で分析したピッチを用い、単語ごとに算出される多変量の特徴量で評価した。特徴量はパーセンタイルの 5 次元 (25、50、75 パーセンタイル、25-50 および 50-75 のパーセンタイルの差)、統計モーメントの 4 次元 (平均、SD、尖度、歪度)、極地の 3 次元 (最大値、最小値、最大値と最小値の差) を足した 12 次元、さらに時間的変化を表現するデルタ特徴量⁷⁾についても同じ 12 次元、つまり計 24 次元とした。音声識別に用いた単語は、ピッチ値が 0 である場合や、あるピッチ値における前後のデルタ特徴量が 50Hz 以上の急激な変化がある場合を除き、分析可能なピッチ値が 60%以上あり、さらに ASD 群、TD 群のそれぞれ 10 名以上が話した単語とした。

識別には Support Vector Machine (SVM) を用いた^{8,9)}。SVM は線形の識別器であるが、今回の研究では、カーネル法を用いて特徴量を高次元空間に写像した。これにより、高次元空間において線形識別器を構成することで、写像前の空間においては非線形な識別器を構成することが可能となった。カーネル関数には RBF Kernel を用いた。評価方法は一つ抜き法を採用した。これをすべてのデータについて行い、各データの評価結果から最終的な評価値を算出する方法である。この音声識別プログラムを用いることで、ASD 児と TD 児の音声識別に有効な特徴量を、人間が選択するのではなく、機器によって自動的に

選択した。

評価基準には F 値¹⁰⁻¹¹⁾を用いた。機器が ASD と識別したうち、ASD のデータの割合を True Positive rate (TP)、TD のデータの割合を False Positive rate (FP) とし、機器が TD と識別したうち、ASD のデータの割合を False Negative rate (FN)、TD のデータの割合を True Negative rate (TN) とした。そして、機器が ASD と識別したうち、正答の割合、すなわち $(TP \times \text{ASD のデータ数}) / (TP \times \text{ASD のデータ数} + FP \times \text{TD のデータ数})$ を適合率、ASD のデータのうち、正答の割合、すなわち $(TP \times \text{ASD のデータ数}) / (TP \times \text{ASD のデータ数} + FN \times \text{ASD のデータ数})$ を再現率とし、F 値を $(2 \times \text{適合率} \times \text{再現率}) / (\text{適合率} + \text{再現率})$ によって算出した。認識率は $(TP \times \text{ASD のデータ数} + TN \times \text{TD のデータ数}) / (\text{ASD のデータ数} + \text{TD のデータ数})$ によって算出される指標である。F 値では、正答率と誤答率の両方をバランス良く評価し、ASD と TD の両方の識別精度を考慮した基準であるため、本研究に採用した。

⑤統計解析

人間の聴覚と機器による ASD 児と TD 児の識別精度を比較するため、参加者と音声識別プログラムの F 値を算出した。その際、参加者が音声識別を行った 100 個のデータを抜き出し、残りのデータを用いて機器にパターン学習をさせた。そして、参加者が音声識別を行った同じデータについて、改めて音声識別プログラムの識別精度を算出した。さらに、参加者の回答パターンの傾向を調べるため、10 語の単語における 100 個のデータと参加者の 4 種類の回答パターンについて二要因参加者間内混合計画による分散分析を行った。統計解析には統計解析ソフトウェアを使用した。

(4) 倫理的配慮

神戸大学大学院医学研究科の医学倫理委員会による承認を受けた後、調査に参加する同意を求めするために参加者と親に対して文書による説明を行い、併せて個人情報保護が確保され、調査に参加しないことによる不利益が一切生じず、参加者または親の意思によって調査を途中で中止することができることを説明した。親が同意書に署名し、文書による同意を得た後に課題を実施した。

【文献】

- 1) American Psychiatric Association (APA). Autistic Disorder. In: APA, editor. Diagnostic and Statistical Manual of Mental Disorders 4th edition-Text Revision (DSM-IV TR). American Psychiatric Publishing; Virginia, 2000, pp. 69-84.
- 2) 構音臨床研究会. 新版構音検査. 千葉テストセンター; 東京, 2010.

- 3) 大六一志、千住淳、林恵津子、東條吉邦、市川宏伸. 自閉症スクリーニング質問紙 (ASQ) 日本語版の開発. 自閉性障害のある児童生徒の教育に関する研究、7巻、2004、19-34.
- 4) Rosenthal, R. *Experimenter Effects in Behavioral Research*. Appleton-Century-Crofts; New York, 1996. 5) Kwon O., Chan K., Hao J., Lee T. *Emotion Recognition by Speech Signals*. Proc. of Eurospeech 2003; Geneva, 2003, pp. 125-128.
- 6) Schuller B., Steidl S., Batliner A. *The INTERSPEECH 2009 Emotion Challenge*. Proc. INTERSPEECH 2009; Brighton, 2009, pp. 312-315.
- 7) Furui S. *Speaker-Independent Isolated Word Recognition Using Dynamic Features of Speech Spectrum*. In: IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 34, Issue 1, 1986, pp. 52-59.
- 8) Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag New York; New York, 1995.
- 9) Bishop, Christopher M. *Pattern Recognition and Machine Learning*. Springer-Verlag New York; New York, 2006.
- 10) Van Rijsbergen, C. J. *Information Retrieval* 2nd edition. Butterworth-Heinemann; London, 1979.
- 11) Manning, Christopher D. Prabhakar Raghavan and Hinrich Schu"tze: *Introduction to Information Retrieval*. Cambridge University Press; Cambridge, 2008.

4. 研究成果

(1) 音声データベース

ASD 児 27 名から 630 語、TD 児 51 名から 1,033 語、計 1,663 語からなる音声データベースを作成した。

(2) ピッチ CV

一語ずつ算出した場合のピッチ CV において、TD 児は就学前と比べ就学後に有意に増加したが、ASD 児ではそのような変化がほとんどないことが明らかとなった (TD 群: $F(1, 74) = 11.84, p < .01$, ASD 群: $F(1, 74) = .24, n. s.$, 図 1)。

(3) ASQ

ASQ の得点と一語ずつ算出した場合のピッチ CV において、ASD 児、TD 児ともに相関がなかった (ASD 群: $r = .08$, TD 群: $r = .03$, 図 2)。ASD において話し言葉のイントネーションが単調であることは ASD の行動特徴

の重症度と関連せず、独立した要因がある可能性が示唆された。

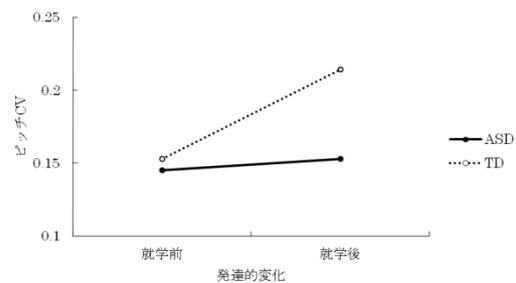


図 1 一語ずつ算出した場合のピッチ CV における発達の变化

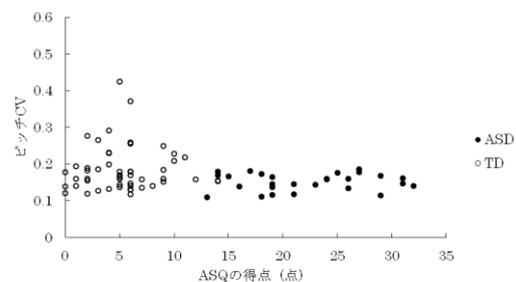


図 2 ASQ の得点と一語ずつ算出した場合のピッチ CV の関係

(4) 識別精度

参加者と音声識別プログラムについて、2 種類の評価基準 (認識率、F 値) を算出した。認識率は人間が .69、機器は .76、F 値は人間が .59、機器が .73 であり、いずれも機器の方が大きかった。つまり、自閉症児と定型発達児の音声識別において、機器の有効性が示唆された。

(5) 回答パターン

参加者と音声識別プログラムについて、4 種類の回答パターンの傾向を調べた。TP と FN の間を除いたすべての多重比較において有意差がみられ、最も多い回答パターンは TN、2 番目に多い回答パターンは TP と FP であった。回答パターンの傾向は参加者と音声識別プログラムでは異なった (図 3)。

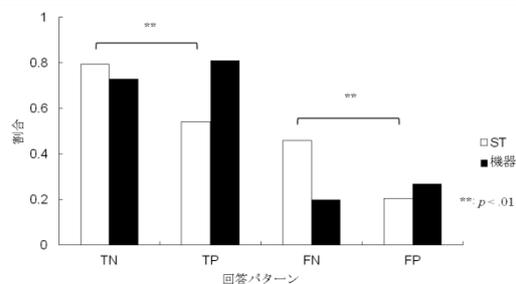


図 3 回答パターンの傾向

5. 主な発表論文等

[雑誌論文] (計 2 件)

- ① 中井靖、神垣彬子. 就学前後を一体的に捉えた発達障害のある子どもを持つ親

に対する支援モデルの構築. 小児保健研究、査読有、71 巻、2012、399-404.

- ② 中井靖. 知能検査による自閉症児と精神遅滞児の鑑別. 川崎医療短期大学紀要、査読有、30 巻、2010、80-87.

http://ci.nii.ac.jp/els/110008103445.pdf?id=ART0009629805&type=pdf&lang=jp&host=cinii&order_no=&ppv_type=0&lang_sw=&no=1338509964&cp=

[学会発表] (計 7 件)

- ① 中井靖ら (5 名中 1 番目). 言語の質的側面からの自閉症児と定型発達児の識別. 日本小児科学会、平成 24 年 4 月 20 日、福岡.
- ② 中井靖ら (5 名中 1 番目). 音声の質的側面からの自閉症児と定型発達児の識別－音声識別プログラムと言語聴覚士の識別精度の比較－. 日本小児神経学会近畿地方会、平成 23 年 10 月 22 日、大阪.
- ③ 中井靖ら (5 名中 1 番目). 音声解析を用いた自閉症児における言語の質的評価. 日本小児科学会、平成 23 年 8 月 12 日、東京.
- ④ 中井靖ら (4 名中 1 番目). 自閉症児の言語発達－音声の質的要因解析による評価－. 日本小児精神神経学会、平成 22 年 11 月 13 日、徳島.
- ⑤ 中井靖. プロソディの音声解析による自閉症児の言語発達. 日本小児保健学会、平成 22 年 9 月 17 日、新潟.
- ⑥ 中井靖ら (5 名中 1 番目). プロソディの音声解析による乳幼児期の言語発達. 日本小児科学会、平成 22 年 4 月 24 日、岩手.
- ⑦ Yasushi NAKAI. Prosodic features of individuals with autism in conversation using developmental test and cepstrum analysis. SARMAC, Jul. 27, 2009, Kyoto.

[その他]

ホームページ : <http://yas-nakai.jp/>

6. 研究組織

(1) 研究代表者

中井 靖 (NAKAI YASUSHI)

川崎医療短期大学・医療保育科・助教

研究者番号 : 80462050

以上