

## ピッチ特徴量を用いた自閉症スペクトラム障害児と定型発達児の識別

☆柿原康博, 滝口哲也, 有木康雄 (神戸大), △中井靖 (川崎医療短期大), △高田哲 (神戸大) \*

### 1 はじめに

近年, 自閉症スペクトラム障害の発生頻度の増加が注目されている。自閉症とは, 社会性や他者とのコミュニケーションに困難が生じる発達障害の一種である。また自閉症スペクトラム障害とは, 自閉性障害, アスペルガー障害, 特定不能の広汎性発達障害の総体である。これらの障害は多様な原因に基づいて発症するため根本的な治療は困難とされている。自閉症スペクトラム障害に特化した支援として, 絵カード交換式コミュニケーションシステム (PECS)[1], 応用行動分析 (ABA)[2], ソーシャルスキルトレーニング (SST)[3] が挙げられる。これらの支援は自閉症スペクトラム障害の特性に依拠しており, 早期発見・早期療育による効果が報告されている [4]。

これまで言語発達の研究は, 主に発話のイントネーションやリズム, アクセント等に注目して行われてきた [5]。定型発達児に関して, 抑揚のパターン (上昇調, 下降調など) は, 年齢によって差が生じると報告されている [6],[7]。また一方で, 自閉症スペクトラム障害児の発話プロソディは, 変化に乏しく単調であると報告されている [8],[9],[10], [11]。本研究では, ピッチ特徴量を入力とし, SVM を用いて自閉症スペクトラム障害児と定型発達児の識別を行う。本稿では, ピッチ特徴量として, 音声データから得られたピッチ系列とそのデルタ系列のそれぞれに対して, 12種類の統計量を計算したものをを用いて, 区間分割による識別実験と単語毎の識別実験を行った。

### 2 ピッチ特徴量の算出

本論文では, ピッチ特徴量を用いて自閉症スペクトラム障害児と定型発達児の識別を行う。ピッチ特徴量とは, ピッチ系列とそのデルタ系列のそれぞれに対して, 表 1 に示す 12 種類の統計量を計算したものである。デルタ系列とは, ある時刻のピッチとその前後の時刻のピッチの差から計算される時間的な変化量である。なお本稿では, 音声データからピッチ系列を抽出する際に, 音声信号の分析合成ツールである STRAIGHT[12] を用いた。

図 1 にピッチ特徴抽出の全体像を示す。ピッチ特徴量は, (i) ピッチ系列の抽出, (ii) デルタ系列の算出,

(iii) 統計量の計算という 3 つのステップにより求められる。ここで得られた特徴量は, 平均 0, 分散 1 となるように各次元を正規化して, SVM の入力として用いる。

Table 1 Statistic

Percentiles	25th Percentile
	50th Percentile
	75th Percentile
	25-50 Percentile difference
	50-75 Percentile difference
Statistical moment	Mean
	Standard Deviation
	Kurtosis
	Skewness
Extreme	Maximum
	Minimum
	Range (Max.-Min.)

### 3 データセット

実験で使用する音声データは, 自閉症スペクトラム障害児 30 名・定型発達児 54 名の新版構音検査絵カード (50 枚) に対する発話の中から有意味の単語のみを取り出したものである [13]。本実験では, その中でも STRAIGHT によるピッチ抽出が可能であった発話を集め, データセットとした。このデータセットは, 自閉症スペクトラム障害児の 1,143 発話と定型発達児の 1,909 発話からなる。なおこれらの音声データは, 神戸大学大学院医学研究科等医学倫理委員会の正式な認可のもと, 2010 年 4 月から 2010 年 7 月の間に, 神戸大学医学部附属病院小児科発達行動外来診療にて収録されたデータである。また表 2 は, 被験者の人数, 年齢の範囲, 平均年齢, 男女比を示している。これは, DSM-IV-TR に基づき小児神経専門医が診断したもので知的障害合併例を含んでいる。

\* Classification of Children with Autism Spectrum and Typically Developing Children Using Pitch Features.  
by Yasuhiro Kakihara, Tetsuya Takiguchi, Yasuo Ariki (Kobe univ.), Yasushi Nakai (Kawasaki College of Allied Health Professions), Satoshi Takada (Kobe univ.)

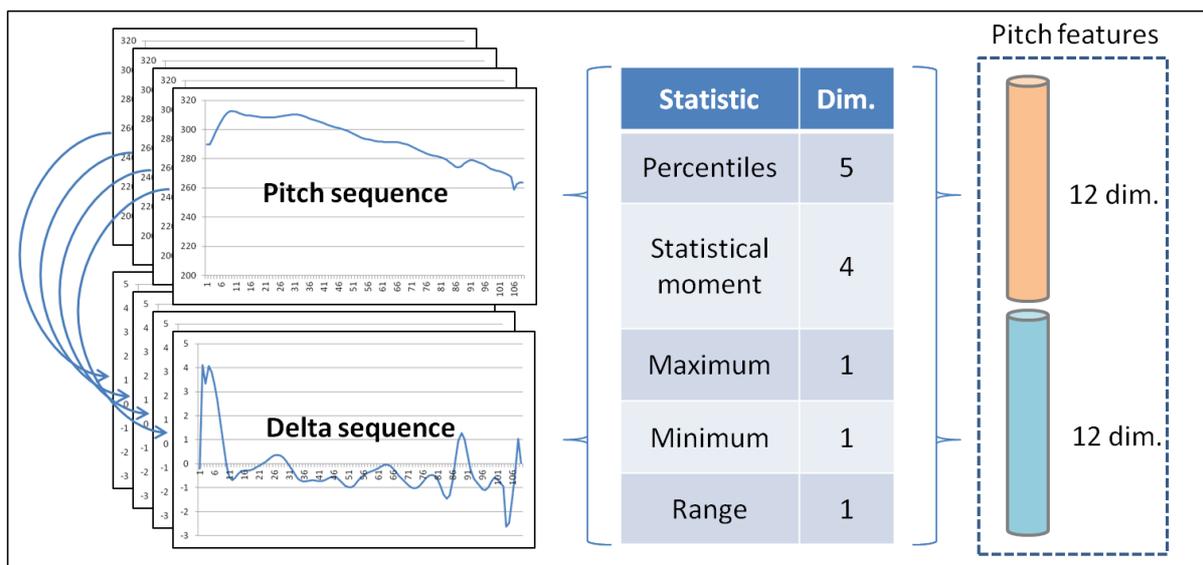


Fig. 1 Pitch feature extraction.

Table 2 Children with Autism spectrum & typically developing children

	Children with Autism Spectrum	Typically Developing Children
Num. of persons	30	54
Age	3.75—10.17	4.25—10.75
Average age	7.25±1.33	7.33±1.75
Male:Female	24:6	23:31

## 4 識別実験

本研究では、区間分割による識別実験と単語毎の識別実験を行った。区間分割による識別実験で用いるデータセットは、第3章で述べたものである。単語毎の識別実験で用いるデータセットは、第3章で述べたデータセットを単語毎に分類したものである。

### 4.1 区間分割による識別実験

本実験では、図2に示すように、1発話を3区間に分割し、それぞれの区間から24次元のピッチ特徴量を抽出し、平均0、分散1となるように各特徴次元を正規化して、SVMの入力とする。ここで、24次元のピッチ特徴量は、ピッチ系列から得られた12次元の特徴ベクトルと、デルタ系列から得られた12次元の特徴ベクトルからなる。本実験では、区間の組み合わせを変えて7通りの実験を行った。実験1では、第1～3から得られる計72次元のピッチ特徴を用いて識別を行う。実験2, 3, 4では、第1第2区間、第1第3区間、第2第3区間から得られる各48次元の特徴を用いて識別を行う。実験5, 6, 7では、

第1区間のみ、第2区間のみ、第3区間のみから得られる各24次元の特徴を用いて識別を行う。また実験8は、区間分割を行わない場合（発話の始めから終わりまでを一つの区間とみなした場合）の実験である。なお、識別実験はデータ集合を10分割し、1つをテストデータ、残りを学習データとし組み合わせ変更して、10-foldのクロスバリデーションを適用した。

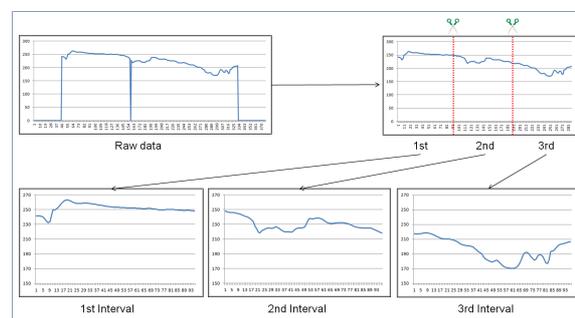


Fig. 2 Example of pitch interval divided into three sections

### 4.2 区間分割による識別結果

表3は、区間分割による識別実験の結果である。表3中のCaseは、実験番号を表し、1st, 2nd, 3rdは、特徴抽出を行った区間を示す。また、Dim.は特徴量の次元数を表す。例として、Case3の場合であれば、第1区間から得られる24次元の特徴量と第3区間から得られる24次元の特徴量を合わせて、Dim=48となっている。

Case1からCase7までの実験で最も精度が高くなったのは、各区間から特徴量を抽出したCase1で、精度は74.9%となった。このことから全区間を用いた方が識別精度が高くなると分かる。次元数毎に結果

Table 3 Classification result for three divided sections

	1st	2nd	3rd	Dim.	Acc. [%]
Case1	12×2	12×2	12×2	72	<b>74.9</b>
Case2	12×2	12×2	—	48	74.3
Case3	12×2	—	12×2	48	74.0
Case4	—	12×2	12×2	48	70.5
Case5	12×2	—	—	24	71.0
Case6	—	12×2	—	24	68.6
Case7	—	—	12×2	24	67.7
Case8	12×2			24	73.2

をみると、2区間-48次元の特徴を用いた Case2-4 では、第1、第3区間を用いた Case2 が最も識別精度が高い結果となった。Case2-4 の結果から、「各区間の識別に対する貢献度」は、第2区間や第3区間よりも第1区間の方が大きいと考えられる。次に、1区間-24次元を用いた Case5-7 では、第1区間を用いた Case5 が最も識別精度が高い結果となった。これは、Case2-4 の結果から考えられる「各区間の識別に対する貢献度」と一致する。Case6 と Case7 は、識別精度が 70% を下回っており、区間を個別にみると「各区間の識別に対する貢献度」は、第3区間よりも第2区間の方が大きいと考えられる。以上の結果から自閉症スペクトラム障害児と定型発達児の識別にピッチ特徴を用いた場合、発話の終盤よりも、序盤に識別のポイントがあると考えられる。これらの原因として、発話に対する最初の唇運動の固さや、最初に息を出すことへの緊張（発話は吸う動作ではなく、吐く動作である）等が原因ではないかと考えられる。

### 4.3 単語毎の識別実験

本実験では、自閉症スペクトラム障害児の発話が 35 以上確保できた単語に対して、単語毎に SVM による識別実験を行った。実験を行った単語は、あし、えんぴつ、いぬ、きりん、めがね、りんご、さかな、せみ、てれび、つくえの 10 単語である。なお、この単語毎の識別実験は、Leave-one-out によるクロスバリデーションを適用した。

### 4.4 単語毎の識別結果

表 4 と図 3 に単語毎の識別実験の結果を示す。なお表 4 は、左から、自閉症スペクトラム障害児の発話数、定型発達児の発話数、区間分割を行わない場合の識別精度 [%]、区間分割を行って得られた 72 次元の特徴を用いた場合の識別精度 [%] を表す。また図 3 は、識別精度の高い単語から順に並べ替えている。

本実験で用いた 10 単語の中で識別精度が高かった単語は、「えんぴつ」であり、24 次元・72 次元どちら

Table 4 Classification result for each word

Word	ASD	TD	24dim.	72dim.
/ashi/	42	68	<b>80.91</b>	78.95
/enpitsu/	44	58	84.31	<b>87.25</b>
/inu/	42	67	75.23	<b>77.06</b>
/kirin/	44	68	68.75	<b>75.89</b>
/megane/	43	67	71.82	<b>72.73</b>
/ringo/	37	64	78.22	<b>80.20</b>
/sakana/	35	61	72.92	<b>80.21</b>
/semi/	40	64	<b>76.92</b>	76.70
/terebi/	42	72	71.05	<b>79.82</b>
/tsukue/	36	68	77.88	<b>79.41</b>

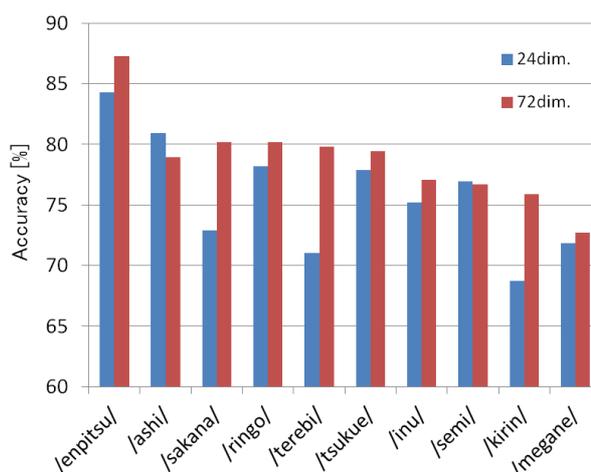


Fig. 3 Classification result for each word

の場合も識別率が 80% を超えている。一方、「めがね」については、いずれの場合も 75% を下回っている。

また、72 次元の特徴を用いることで識別率が改善した単語は、「えんぴつ、いぬ、きりん、めがね、りんご、さかな、てれび、つくえ」であり、その中でも識別率が大きく改善したのは、「さかな、てれび、きりん」であった。逆に「あし」に関しては、72 次元の特徴を用いた場合、識別率が低下した。

これらの結果から、ピッチ特徴量の最適な抽出方法は、単語によって異なると考えられる。また、今回の実験では、「えんぴつ、あし」といった単語が、自閉症スペクトラム障害児と定型発達児の発話の違いが大きいと考えられる。これらの要因については今後検証していく。

## 5 おわりに

本稿では、自閉症スペクトラム障害の早期発見を音響的な側面から目指し、ピッチ特徴量を用いて自閉症スペクトラム障害児と定型発達児の識別を行った。ピッチ特徴量とは、音声データから得られたピッチ系列とそのデルタ系列のそれぞれに対して、25、50、

75 パーセント、25-50 と 50-75 パーセントの差、平均、標準偏差、尖度、歪度、最大値、最小値、レンジという 12 の統計量を計算したものである。実験として、区間分割による識別実験、単語毎の識別実験を行った。

区間分割による識別実験では、最も精度が高くなったのは全区間を用いた場合であり、識別精度は 74.9% となった。このことから、全区間を用いた方が識別精度が高くなると分かった。また、区間の組み合わせを変えて行った実験から、「各区間の識別に対する貢献度」は、第 2 区間や第 3 区間よりも第 1 区間の方が大きいと考えられる。したがって、自閉症スペクトラム障害児と定型発達児の識別にピッチ特徴を用いた場合、発話の終盤よりも、序盤に識別のポイントがあると考えられる。これらの原因として、発話に対する最初の唇運動の固さや、最初に息を出すことへの緊張（発話は吸う動作ではなく、吐く動作である）等が原因ではないかと考えられる。

単語毎の識別実験では、データセットから自閉症スペクトラム障害児の発話数が十分確保できた 10 単語を選び、24 次元・72 次元のピッチ特徴を用いて実験を行った。識別精度が高かった単語は「えんぴつ、あし」であり、これらは、自閉症スペクトラム障害児と定型発達児の発話の違いが大きいと考えられる。また今回の実験では、ピッチ特徴量の最適な抽出方法は単語によって異なることが分かった。これらの要因については今後検証していく。

今後は、単語毎・発話者毎の発話分析を行い、SVM だけでなく Boosting 等の識別手法についても検討する。さらに、実験結果を医学的な知見と照らし合わせながら、自閉症スペクトラム障害児の発話の特性を明らかにしていきたい。

## 参考文献

- [1] Andy Bondy and Lori Frost, *A Picture's Worth: PECS and Other Visual Communication Strategies in Autism. Topics in Autism*, ERIC, 2002.
- [2] Shira Richman, *Raising a child with autism: a guide to applied behavior analysis for parents*, Jessica Kingsley Publishers, 2001.
- [3] Robert Paul Liberman, William J DeRisi, and Kim T Mueser, *Social skills training for psychiatric patients*, Pergamon Press, 1989.
- [4] Scott M. Myers and Chris Plauché Johnson, "Management of children with autism spec-

trum disorders," *Pediatrics*, vol. 120, no. 5, pp. 1162–1182, 2007.

- [5] Alan Cruttenden, *Intonation*, Cambridge University Press, 1997.
- [6] David Snow, "Intonation in the monosyllabic utterances of 1-year-olds," *Infant Behavior and Development*, vol. 24, no. 4, pp. 393–407, 2001.
- [7] David Snow, "Children's imitations of intonation contours: Are rising tones more difficult than falling tones?," *Journal of speech and hearing research*, vol. 41, no. 3, pp. 576–587, 1998.
- [8] Leo Kanner et al., "Autistic disturbances of affective contact," *Nervous child*, vol. 2, no. 3, pp. 217–250, 1943.
- [9] Lawrence D Shriberg, Rhea Paul, Jane L McSweeny, Ami Klin, Donald J Cohen, and Fred R Volkmar, "Speech and prosody characteristics of adolescents and adults with high-functioning autism and asperger syndrome," *Journal of Speech, Language and Hearing Research*, vol. 44, no. 5, pp. 1097, 2001.
- [10] Helen Tager-Flusberg, "On the nature of linguistic functioning in early infantile autism," *Journal of Autism and Developmental Disorders*, vol. 11, no. 1, pp. 45–56, 1981.
- [11] MS Garnett and AJ Attwood, "The australian scale for asperger's syndrome," *Attwood, Tony. Asperger's Syndrome: A Guide for Parents and Professionals. London: Jessica Kingsley Publishers*, pp. 45–56, 1997.
- [12] Hideki Kawahara, Ikuyo Masuda-Katsuse, and Alain de Cheveigné, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds," *Speech communication*, vol. 27, no. 3, pp. 187–207, 1999.
- [13] Yasushi Nakai, Ryoichi Takashima, Tetsuya Takiguchi, and Satoshi Takada, "Speech intonation in children with autism spectrum disorder," *Brain and Development*, 2013.