

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^{††} 中井 靖^{†††} 高田 哲^{††††}

[†] 神戸大学大学院システム情報学研究科 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

^{††} 神戸大学自然科学系先端融合研究環都市安全研究センター 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

^{†††} 川崎医療短期大学医療保育科 〒 701-0194 岡山県倉敷市松島 316

^{††††} 神戸大学大学院保健学研究科 〒 654-0142 兵庫県神戸市須磨区友が丘 7-10-2

E-mail: [†]kakyhara@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki,satoshi}@kobe-u.ac.jp, ^{†††}nakai@jc.kawasaki-m.ac.jp

あらまし 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が、この障害に特化した支援による早期療育の効果が報告されている。本研究では、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 SVM の入力として識別実験を行った。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 パーセンタイル、25-50 と 50-75 パーセンタイルの差、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、単語毎の識別、区間分割による識別、特徴分割による識別の 3 つ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。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 区間の方が大きいことが示された。

キーワード 自閉症スペクトラム、ピッチ、イントネーション、SVM

Classification of Children with Autism Spectrum and Typically Developing Children Using Pitch Features

Yasuhiro KAKIHARA[†], Tetsuya TAKIGUCHI^{††}, Yasuo ARIKI^{††}, Yasushi NAKAI^{†††}, and Satoshi TAKADA^{††††}

[†] Graduate School of System Informatics, Kobe University 1-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo, 657-8501 Japan

^{††} Research Center for Urban Safety and Security, Organization of Advanced Science and Technology, Kobe University 1-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo, 657-8501 Japan

^{†††} Department of Nursing Childcare, Kawasaki College of Allied Health Professions 316 Matsushima, Kurashiki, Okayama 701-0194 Japan

^{††††} Graduate School of Health Sciences, Kobe University 7-10-2 Tomogaoka, Suma-ku, Kobe, Hyogo, 654-0142 Japan

E-mail: [†]kakyhara@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki,satoshi}@kobe-u.ac.jp, ^{†††}nakai@jc.kawasaki-m.ac.jp

Abstract Recent investigations have demonstrated that the early support which specialized in autistic spectrum obstacle, such as Picture Exchange Communication System (PECS) Applied Behavior Analysis (ABA) Social Skills Training (SST), is effective. This paper reports the result of a classification experiment carried out using pitch features for children with autism spectrum. Pitch features consist of 24 dimensions, such as 25th, 50th, 75th percentiles, 25-50 percentile difference, 50-75 percentile difference, mean, standard deviation, kurtosis, skewness, maximum, minimum, and range.

Key words Autism spectrum disorder, Pitch, Intonation, Acoustic analysis, SVM

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、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], [12]。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。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 パーセントイル、25-50 と 50-75 パーセントイルの差、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。本研究では、単語毎の識別、区間分割による識別、特徴分割による識別の 3 つの識別実験を行った。単語毎の識別実験により、自閉症スペクトラム障害児と定型発達児の識別に有効な単語を明らかにした。また、区間分割による識別実験により、発話のどの区間が識別に貢献しているかを調べた。また、特徴分割による識別実験により、識別に有効な特徴量のセットを調べた。

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。図 1 は、ピッチ特徴算出の全体像を示す。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、25, 50, 75 パーセントイル、25-50 と 50-75 パーセントイルの差、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。ここで、ピッチ特徴量は、以下の 3 つのステップにより算出される。

- (i) 音声データからピッチ系列を抽出。
- (ii) 抽出されたピッチ系列からそのデルタ系列を計算。
- (iii) それぞれの系列に対して統計量を計算。

本論文では、(i) 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 [13] を用いた。(ii) のデルタ系列とは、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。また (iii) で扱う統計量を表 1 に示す。本論文では、これら 3 つのステップ

で算出された特徴量をピッチ特徴量と呼ぶ。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毎に正規化して、SVM の入力として用いる。

表 1 統計量

Table 1 Statistic.

Percentiles	25th Percentile, Δ 25th Percentile
	50th Percentile, Δ 50th Percentile
	75th Percentile, Δ 70th Percentile
	25-50 Percentile Difference, Δ 25-50 Percentile Diff.
	50-75 Percentile Difference, Δ 50-75 Percentile Diff.
Statistical moment	Mean, Δ Mean
	Standard Deviation, Δ Standard Deviation
	Kurtosis, Δ Kurtosis
	Skewness, Δ Skewness
Extreme	Maximum, Δ Maximum
	Minimum, Δ Minimum
	Range (Max.-Min.), Δ Range

3. データセット

本研究では、自閉症スペクトラム障害児と定型発達児のそれぞれの音声データから求めたピッチ特徴量を入力とし、SVM を用いて識別実験を行う。この音声データは、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。[14]。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。

表 2 自閉症スペクトラム障害児と定型発達児

Table 2 Children with Autism spectrum & typically developing children.

	Children with Autism Spectrum	Typically Developing Children
Num. of children	30	54
Age	3.75—10.17	3.25—3.75
Average age	7.25 \pm 1.33	7.33 \pm 1.75
Male:Female	24:6	23:31

4. 識別実験

本研究では、(A) 単語毎の識別実験、(B) 区間分割による識別実験、(C) 特徴分割による識別実験の 3 つの実験を行った。(A) 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をを用いる。(B) 区間分

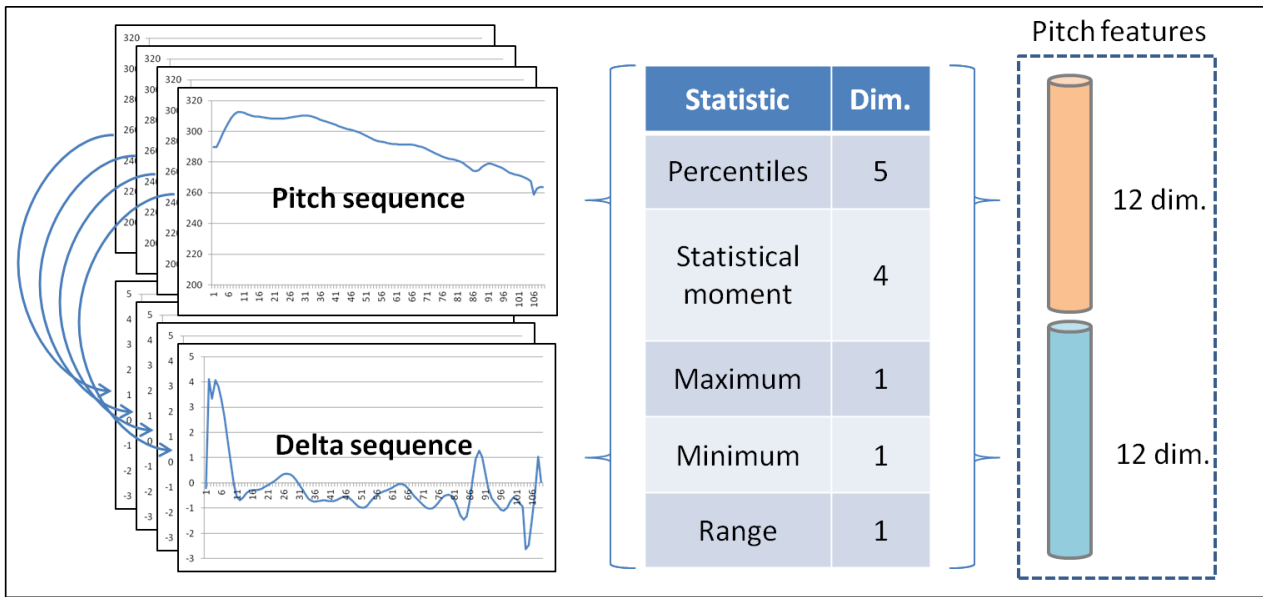


図1 ピッチ特徴量の抽出
Fig. 1 Pitch feature extraction.

割による識別実験と(C)特徴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。図2は、2つの特徴量をピックアップして、自閉症スペクトラム障害児と定型発達児の分布をプロットしたものである。

4.1 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着目し、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。特徴量としては、特徴次元毎に平均0分散1に正規化した24次元のピッチ特徴量を用いた。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。

4.2 単語毎の識別結果

表3と図3に単語毎の識別実験の結果を示す。なお表4.2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、合計発話数、識別精度[%]、適合率[%]、再現率[%]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。

表3 単語毎の識別結果

Table 3 Classification result for each word.

Word	ASD	TD	Sum	Acc. [%]	Precision [%]	Recall [%]
/ashi/	42	68	110	80.91	86.21	59.52
/enpitsu/	44	58	102	84.31	85.00	77.27
/inu/	42	67	109	75.23	89.47	40.48
/kirin/	44	68	112	68.75	71.43	34.09
/megane/	43	67	110	71.82	73.08	44.19
/ringo/	37	64	101	78.22	82.61	51.35
/sakana/	35	61	96	72.92	100.00	25.71
/semi/	40	64	104	76.92	78.57	55.00
/terebi/	42	72	114	71.05	71.43	35.71
/tsukue/	36	68	104	77.88	80.95	47.22

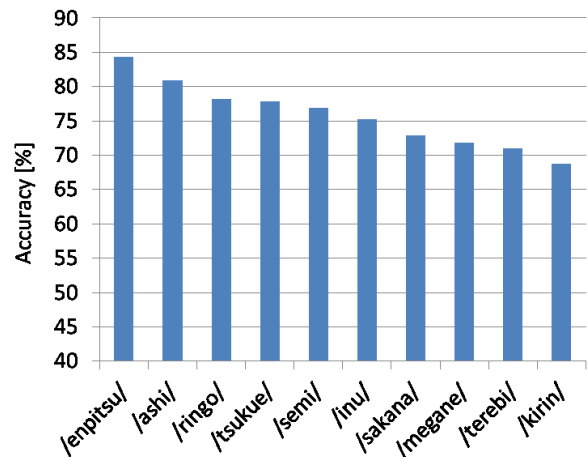


図3 単語毎の識別結果

Fig. 3 Classification result for each word.

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」と「あし」であった。これらは、識別精度が80%を超えている。一方で、「めがね、てれび、きりん」は、識別精度が低く、特に「きりん」は、70%を下回っている。これらの結果から、今回の実験では「めがね、てれび、きりん」といった単語よりも、「えんぴつ、あし」といった単語の方が、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が、要因については今後検証していく。

4.3 区間分割による識別実験

本実験では、図4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

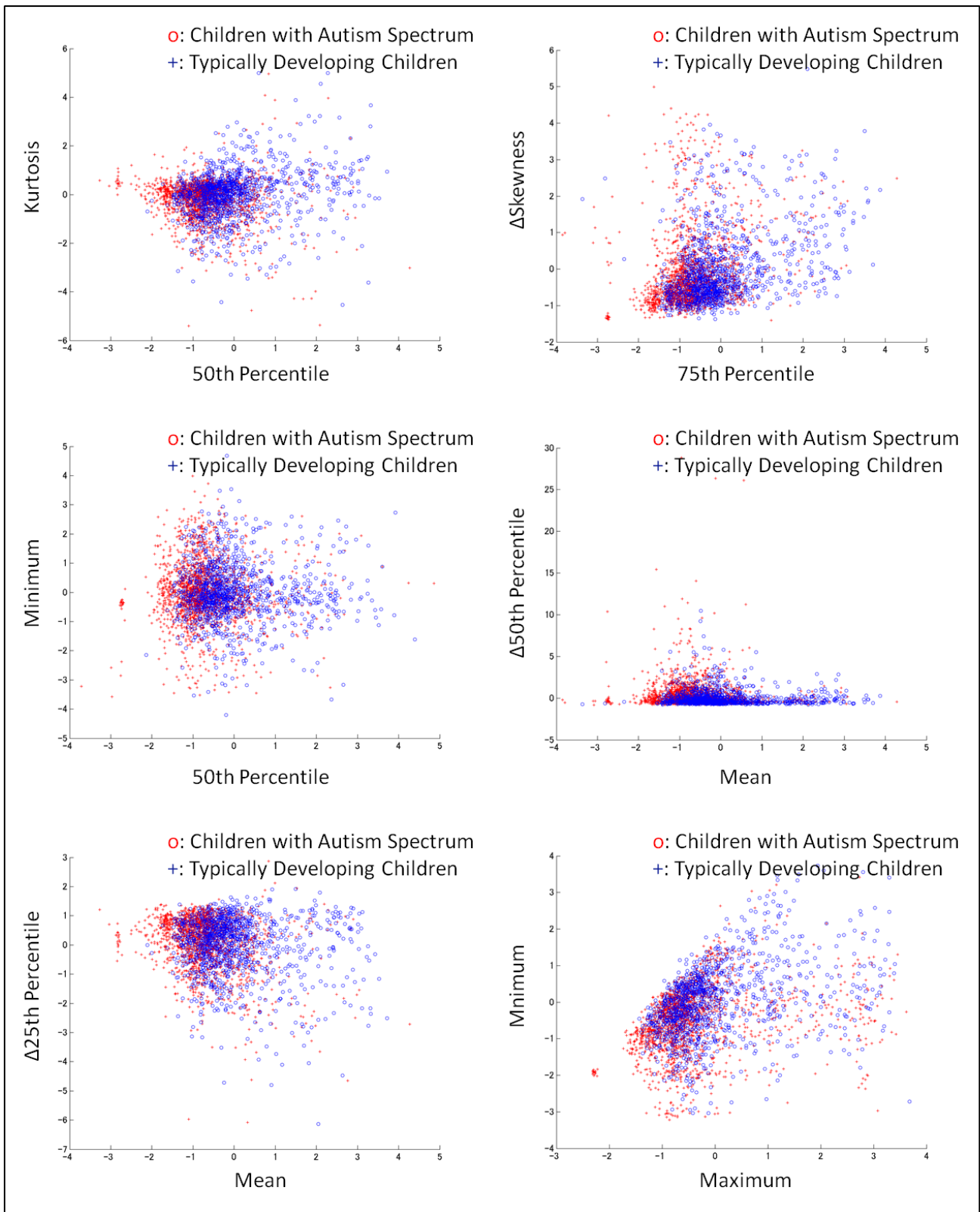


図2 自閉症スペクトラム障害児と定型発達児のデータの分布
 Fig.2 Distribution map of children with Autism spectrum and typically developing children

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かえて、7 通りの実験を行った。実験 1 では、第 1、第 2、第 3 区間から得られる計 72 次元のピッチ特徴を用いて識別を行う。実験 2 では、第 1、第 2 区間から得られる 48 次元、実験 3 では、第 1、第 3 区間から得られる 48 次元、実験 4 では、第 2、第 3 区間から得られる 48 次元特徴を用いて識別を行う。実験 5 では、第 1 区間から得られる 24 次元、実験 6 では、第 2 区間から得られる 24 次元、実験 7 では、第 3 区間から得られる 24 次元特徴を用いて識別を行う。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。

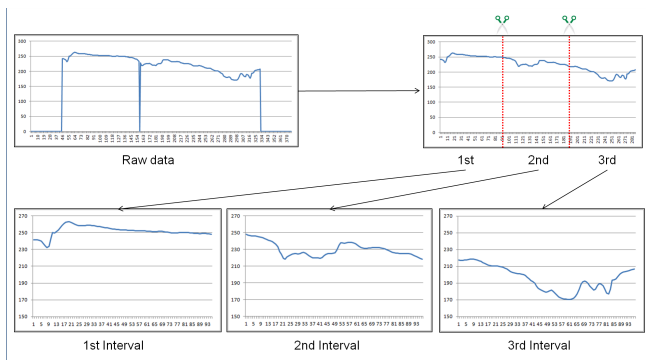


図 4 区間分割の例

Fig. 4 Example of pitch interval divided into three sections.

4.4 区間分割による識別結果

表 4 は、区間分割による識別実験の結果である。表 4 中の Case は、実験番号を表す。また、1st Interval, 2nd Interval, 3rd Interval は、どの区間の特徴を用いたかを示す。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 区間の 24 次元の特徴量と第 3 区間の 24 次元の特徴量を合わせるため、Dim=48 となっている。

表 4 区間分割による識別結果

Table 4 Classification result for three divided sections.

	1st Interval	2nd Interval	3rd Interval	Dim.	Acc. [%]
Case1	12×2	12×2	12×2	72	74.9
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

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。

4.5 特徴分割による識別実験

本実験では、24 次元のピッチ特徴量を分割して識別を行う。この実験では、前節で述べた区間分割は行わず、特徴量として、6, 8, 10, 14, 16, 18, 24 次元のピッチ特徴量を用いる。SVM の入力となるこれらのピッチ特徴量のセットは、特徴次元毎に平均 0、分散 1 となるように正規化している。ピッチ特徴量のセットは、統計量の種類を考慮し、パーセンタイルに関する特徴セット、モーメントから求められる統計量に関する特徴量セット、最大・最小・レンジに関する特徴セットとした。パーセンタイルに関する特徴セットは、ピッチ系列から求められる 5 次元の特徴と、デルタ系列から求められる 5 次元の特徴を合わせた 10 次元の特徴ベクトルからなる。また、モーメントから求められる統計量に関する特徴セットは、ピッチ系列から求められる 4 次元の特徴と、デルタ系列から求められる 4 次元の特徴を合わせた 8 次元の特徴ベクトルからなる。同様に、最大・最小・レンジに関する特徴セットは、ピッチ系列から求められる 3 次元の特徴と、デルタ系列から求められる 3 次元の特徴を合わせた 6 次元の特徴ベクトルからなる。本実験では、これらの特徴量セットの組み合わせを替えて、7 通りの識別実験を行った。

4.6 特徴分割による識別結果

特徴分割による識別結果を表 5 に示す。表 5 中の Case は、実験番号を表す。また Percentiles, Statistical moment, Extreme は、各実験でどの特徴セットを用いたかを示し、Dim. は次元数の合計を示す。

表 5 特徴分割による識別結果

Table 5 Classification result for each feature.

	Percentiles	Statistical moment	Extreme	Dim.	Acc. [%]
Case8	5×2	4×2	3×2	24	73.2
Case9	5×2	4×2	—	18	72.3
Case10	5×2	—	3×2	16	72.6
Case11	—	4×2	3×2	14	71.8
Case12	5×2	—	—	10	71.8
Case13	—	4×2	—	8	71.3
Case14	—	—	3×2	6	65.8

Case8 から Case14 の実験で最も識別率の高くなったのは、24

次元の特徴量を用いた Case8 であり、識別率は 73.2% となった。このことから、全てのピッチ特徴を用いた場合が識別率が高くなるのが分かる。特徴量を 2 セット組み合わせた Case9-11 の場合、パーセンタイルに関係する特徴セットと最大・最小・レンジに関する特徴セットを組み合わせた場合が最も識別率が高くなる。また、特徴量を 1 セットのみ用いた場合、最大・最小・レンジに関する特徴セットでの識別結果がもっとも低い識別結果となった。特徴量の次元数が異なるため、一概にどの特徴セットが識別に有効であるかを言及することは出来ないが、少なくとも、パーセンタイルに関係する特徴セット、モーメントから求められる統計量に関する特徴量セット、最大・最小・レンジに関する特徴セットを全てのピッチ特徴を用いた場合の識別率が高くなるのが分かる。

5. おわりに

本論文では、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 パーセンタイル、25-50 と 50-75 パーセンタイルの差、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、単語毎の識別実験、区間分割による識別実験、特徴分割による識別実験の 3 つを行った。

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、実験を行った。識別精度が高かった単語は、「えんぴつ」と「あし」であった。一方で、「めがね、てれび、きりん」は、識別精度が低く、特に「きりん」は、70% を下回っていた。これらの結果から、今回の実験では「めがね、てれび、きりん」といった単語よりも、「えんぴつ、あし」といった単語の方が、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が、要因については今後検証していく。

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。

特徴分割による識別実験では、最も識別率の高くなったのは、24 次元の特徴量を用いた場合であり、識別率は 73.2% となった。このことから、全てのピッチ特徴を用いた場合が識別率が高くなるのが分かった。特徴量を 2 セット組み合わせた場合、パーセンタイルに関係する特徴セットと最大・最小・レンジに関する特徴セットを組み合わせた場合が最も識別率が高くなる。

また、特徴量を 1 セットのみ用いた場合、最大・最小・レンジに関する特徴セットでの識別結果がもっとも低い識別結果となった。特徴量の次元数が異なるため、一概にどの特徴セットが識別に有効であるかを言及することは出来ないが、少なくとも、パーセンタイルに関係する特徴セット、モーメントから求められる統計量に関する特徴量セット、最大・最小・レンジに関する特徴セットを全てのピッチ特徴を用いた場合の識別率が高くなるのが分かった。

今後は、音声データ数を増やし、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。

文 献

- [1] A. Bondy and L. Frost, A Picture's Worth: PECS and Other Visual Communication Strategies in Autism. Topics in Autism, ERIC, 2002.
- [2] S. Richman, Raising a child with autism: a guide to applied behavior analysis for parents, Jessica Kingsley Publishers, 2001.
- [3] R.P. Liberman, W.J. DeRisi, and K.T. Mueser, Social skills training for psychiatric patients, Pergamon Press, 1989.
- [4] S.M. Myers and C.P. Johnson, "Management of children with autism spectrum disorders," Pediatrics, vol.120, no.5, pp.1162-1182, 2007.
- [5] A. Cruttenden, Intonation, Cambridge University Press, 1997.
- [6] D. Snow, "Intonation in the monosyllabic utterances of 1-year-olds," Infant Behavior and Development, vol.24, no.4, pp.393-407, 2001.
- [7] D. Snow, "Children's imitations of intonation contours: Are rising tones more difficult than falling tones?," Journal of Speech, Language and Hearing Research, vol.41, no.3, p.576, 1998.
- [8] L. Kanner, et al., "Autistic disturbances of affective contact," Nervous child, vol.2, no.3, pp.217-250, 1943.
- [9] L.D. Shriberg, R. Paul, J.L. McSweeney, A. Klin, D.J. Cohen, and F.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, p.1097, 2001.
- [10] H. 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] A.M. COOPER and R. Michels, "Diagnostic and statistical manual of mental disorders," American Journal of Psychiatry, vol.138, no.1, pp.128-129, 1981.
- [12] M. Garnett and A. 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.
- [13] H. Kawahara, I. Masuda-Katsuse, and A. deCheveigné, "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.
- [14] Y. Nakai, R. Takashima, T. Takiguchi, and S. Takada, "Speech intonation in children with autism spectrum disorder," Brain and Development, pp.●●-●●, 2013.