卒業論文要旨

乳児音声検出システムの開発とその応用

（知能情報システム学）　山本翔太

1. 緒言

近年、人と機械とのコミュニケーションへの音声認識・音声合成の応用が期待されている。しかし、音声認識はいまだ発展途上であり、さまざまな方法が提案されている。さらに、幼児音声の音響的特徴に関する研究は少なく、これらの値を推定したという報告も少ない[1]。この分野の研究が少ないのは、幼児の泣き声（以下「乳児音声」と表記）は言語獲得前の非言語であり、しかも意思表示（感情）の内容はあくまでも第三者が推定する形になり、また、発声の途中で感情が変化するなど、的確に捉えるのが難しいからであると考えられる。本研究では、Julius[2]による音声認識と基本周波数の特徴を用いて、連続音声から乳児音声区間を検出するシステムを開発した。そして、本手法を用いて切り出した音声に対し、既報の研究[3]で用いた FFT による周波数解析により見出した特徴量をもとに、主成分分析を用いて感情特徴のパターン認識を行った。

2. 開発環境

・OS:Windows XP　　・使用プログラム言語:Microsoft Visual C++ 6.0
・PC:DELL OPTIPLEX GX260(CPU:INTEL Pentium 4　2.4 GHz, メモリ:512 MB)

3. 処理概要

連続音声に含まれる乳児音声区間を検出するために、以下の 2 つを特徴量に用いた。

3－1.　Julius により得られる単語信頼度

高性能な汎用大語彙連続音声認識エンジンである Julius を用い、入力された連続音声を単語区間に分け、区間毎の認識結果を出力する。その際、認識結果の各単語について示される単語信頼度（0～1 の値で表され、1 に近いほど、競合候補に比べて尤度の差が大きいことを示す）が、成人音声区間に比べ乳児音声区間の方が、値が低い傾向があることが本研究で観察された。

3－2.　基本周波数

乳児音声は、成人音声とは異なる音響的特徴を持っており、基本周波数 F_0 が 400Hz、第 1 フォルマント周波数 1000Hz、第 2 フォルマント周波数 3000Hz、第 3 フォルマント周波数 5000Hz となっている（成人では 500，1500，2500Hz）。また、成人音声に比べ F_0 のとる範囲が広く、高域まで強いエネルギーを持つ有声音の特性をもっている[1]。本研究で、乳児音声において基本周波数が短時間で大きく変化することが観察された。

4. 処理フロー

はじめに、連続音声（前後に数百 msec 程度の無音区間を含む）に対し、Julius を用いて単語認識を行った。次に、Julius により有音声区間として認識された区間に対し、先頭無音区間（1s）の振幅の平均と標準偏差を用いて閾値を決定し、閾値未満の振幅をもつ区間を無音区間、閾値以上の振幅をもつ区間を有音声区間とした。

有音声区間に対し、次の 2 つの条件のいずれかを満たす区間を乳児音声区間とした。

条件 1：Julius による単語信頼度が 0.15 未満
条件 2：基本周波数の値が 0.1s で 150Hz 以上変化する

5. 乳児音声検出システムの評価実験

言語獲得前の月齢 1.5 ヶ月の乳児（男児）における音声データを収集し、それら 57 の連続音声（1 連続音声につき 10～12 発声を含む）と比較対象として成人音声を 7 連続音声（1 連続音声につき 14～22 発声を含む）と、保育園で収集した乳児と成人音声が混在している音声データである混合①を 3 連続音声（1 連続音声につき 17～21 を含む）と混合②を 5 連続音声（1 連続音声につき 8～21 発声を含む）とを用いて本システムの評価実験を行った。また、Julius により検出された音声区間に対し、音声波形の目視と聴音確認を行い、乳児音声であるか否かの判定を行った。音声は、サンプリング周波数 16000Hz、wave 形式によるモノラルで録音した。実験結果を表 1 に示す。

表 1　本システム評価実験結果

	乳児	成人	混合①	混合②
Julius による検出区間数	1172	333	186	239
乳児音声区間数	569	0	60	126
本システムによる検出区間数	407	66	82	97
（検出された乳児音声区間数）	(395)	(0)	(39)	(94)
検出率	69.4%	——	65%	66.2%

検出率: 本システムにより検出された区間の内の乳児音声区間数　/　乳児音声区間数

6. 応用実験

本システムを用いて切り出した 407 区間（57 連続音声）について、32 次元 FFT と主成分分析を組み合わせ、音声区間を不快、空腹、眠気の 3 感情に分類した。その際、切り出した音声データの前半部 28 の連続音声を学習データ（感情既知）とし、後半部 29 の連続音声を認識データ（感情未知）とした。

まず、32 次元 FFT により各区間 30 個の特徴量を得た。次に、得られた特徴量を元に学習データに対し主成分分析を行い、音声モデルを構築した。認識データに対しては、学習データを元に作成した主成分空間に投影後、ユークリッド距離による最近傍法を用いて認識を行った。また、認識する際、時系列的に連続音声に並べなおし、一連の中で一番多い感情をその連続音声全体の感情とした。実験の結果、29 個の連続音声の内、18 連続音声（62.1%）の感情を認識することができた。

7. 結言

乳児音声の特徴を元に Julius と基本周波数を用いて乳児音声検出システムを提案した。さらに、切り出された乳児音声に対して FFT と主成分分析を組み合わせて感情を認識し、その有効性を示した。

本研究では、Julius の単語信頼度と基本周波数の変化を特徴量として用いたが、成人が非言語を発した区間（乳児をあやしている場合など）や乳児音声と成人音声が混在する区間においては誤検出した。今後は、精度向上のために、2 つの特徴量以外に乳児音声に固有の特徴量を見出し、本システムに加える必要がある。事例の積み重ねと手法の改善により乳児音声検出率の向上が期待される。

参考文献

[1] 中谷智広，天野成昭，入野俊夫，「幼児音声の基本周波数および有声区間の推定法」，
音響学会秋季研究発表会, 1-P-11, vol.1, pp.393-394, 2002.

[2] 大語彙連続音声認識システム Julius Web ページ，
http://julius.sourceforge.jp/

[3] 櫛田康,田伏正佳,吉冨康成，「乳児の音声における感情特徴のパターン認識」，
ヒューマンインタフェースシンポジュウム 2008 論文集, pp.25-28, 2008