

乳児の音声における感情特徴のパターン認識

(知能情報システム学) 櫛田 康

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]。それは幼児の泣き声（以下音声とする）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 FFT による周波数解析とクラスタリングを組み合わせることで見出した幼児音声の特徴量を用いて、感情特徴のパターン認識を行った。

2. 乳児の発達段階による感情表現・言語獲得について

子どもに見られる感情の発達は「喜ぶ」「やさしい」、「怒る」、「嫌がる」「恥ずかしい」「悲しい」「やきもち」「怖がる」の順になっている[2]。乳児の言語獲得は、学習を中心とする知覚と生成の二つの側面があり[3]、母語の影響をつよく受け生得的なものである。聴覚は胎生7ヶ月前後で完成し、胎児段階から音声を聴き、学習している。生後200日ごろから母語への選択的注意が起こることが指摘されている[3]。乳児は生後6ヶ月を過ぎたころに規範喃語期を迎える。喃語発声初期は、発声器官が未熟で子音の音質はあまり良くないが、成長に伴って明瞭な子音があらわれ、音節の組み合わせも、月齢が上がるにつれて複雑になる。喃語後期には、初語が出現し、10ヶ月児で発声頻度の分布が母語と一致することが報告されている[4]。日本語を母語とする乳児は、子音の出現頻度が少なく、10ヶ月前後からは、音韻、韻律双方のレベルで、母語の音響特徴が乳児の生成する音声に現れてくる。母語の影響は、生後1年を過ぎたころ表れ、1歳半前後には50語程度の単語を表出し、これ以降語彙の獲得速度が急激に速まり、語彙爆発の時期を迎える。2歳前には2つの単語を組み合わせた二語文が出現し複雑な語彙構造を持つ文を表せるようになる。

乳児では、平均的な発声時間は二語文があらわれるようになった後には、直線的に増加していく[5]。発話時間は出生直後は1秒以下の短い発声ほとんどだが、喃語が出現する生後6ヶ月前後で長い発声が増え、規範的な喃語がよく出る最盛期の8ヶ月には、その出現頻度がピークを迎える。やがて1歳半を過ぎ二語文の時期になると、文の伸びに伴って、再び1秒以上の長い発声が増えていく。実際は、発達過程と呼応した回帰的な性質を持つ発声時間長の発達が存在する[6]。幼児の語彙爆発は15-20ヶ月の間に始まり、語彙爆発の期間は1年余である。また、二語文、三語文のあらわれる時期がそれぞれ、100語、300語を表現できる時点にほぼ一致し、三語文の出現と、疑問文や一部の助詞の出現との間には相関がある。このように語彙の獲得と文の獲得の間に密接な相互作用があることがわかっている。

3. 乳児音声の特徴と音声収録環境

乳児音声は、成人とは異なる以下の音響的特徴を持つ。基本周波数 F_0 が 400 Hz、第1フォルマント周波数 1000 Hz、第2フォルマント周波数 3000 Hz、第3フォルマント周波数 5000 Hz。また、成人より F_0 のとる範囲が広く（成人では 500, 1500, 2500Hz）、 F_0 の不連続度が2倍もしくは1/2倍に突然遷移し、高域まで強いエネルギーを持つ有声音の特性をもつ[1]。

音声収録には、実験環境で発話を録音する方法と、日常的な発話を録音する方法がある。実験環境下では、発声に影響を与える様々な要因をコントロールできるが、日常的な発話では、自然な音声発達の過程を観察できる利点がある。多数の同月齢乳児の発声を録音し、乳児間の共通点を見出す横断的観察は、データ取得が短期間でできるが、音声発達の個人差が非常に大きく、多数の発声を比較する必要が

ある。特定の個人を長期にわたって録音する縦断的観察では、長期的な負担を強いることになるが、発達を継続的に詳細に検討できる点で優れている。本研究では、横断的に日常的環境下での乳児音声を収集することとした。

4. 周波数解析とクラスタリング

4.1 離散フーリエ解析 DFT

画像や音声、映像などのメディアデータの処理で用いられる方法の1つとして離散的フーリエ変換と逆変換があげられる。離散的フーリエ変換では、入力波形をいくつかのグループに分けて計算する。その計算順序を工夫することにより計算量を大幅に減少させたアルゴリズムが FFT であり、グループの数を N とした時の演算の回数は、通常変換は N の 2 乗に比例するが、FFT では $N \log N$ に比例する。 N を整数とし T をサンプリング周期とすると、周期 NT の関数 $x^*(t)$ は、デルタ関数 $\delta(t)$ を用いて、次のように表すことができる。

$$x^*(t) = \sum_{n=0}^{N-1} x(nT) \delta(t - nT)$$

1 周期分では $x^*(t)$ は離散信号 $x_0, x_1, x_2, \dots, x_{N-1}$ の関数である。この関数 1 周期分について、複素フーリエ級数展開し、小さな正数を ϵ とすると、以下の式が求められる。

$$\begin{aligned} c_k &= \frac{1}{NT} \int_{-\epsilon}^{NT-\epsilon} x^*(t) e^{-j2\pi k f_0 t} dt \\ &= \frac{1}{N} \sum_{n=0}^{N-1} x(nT) e^{-j\frac{2\pi k n}{N}} \end{aligned}$$

ただしデルタ関数 $\delta(t)$ の面積は T である。複素フーリエ級数展開の係数 c_k ($k=0, 1, 2, \dots, N-1$) について、 c_{k+N} を求めると

$$\begin{aligned} c_{k+N} &= \frac{1}{N} \sum_{n=0}^{N-1} x(nT) e^{-j\frac{2\pi k n}{N}} e^{-j2\pi n} \\ &= c_k \end{aligned}$$

となり、複素フーリエ級数展開の係数 c_k ($k=0, 1, 2, \dots, N-1$) は N の周期性がある。一方、離散周期信号は、 c_k を用いて次のように表わされる。

$$x(nT) = \sum_{k=0}^{N-1} c_k e^{j\frac{2\pi k n}{N}}$$

以上より、離散フーリエ変換に関する以下の式が導かれる [7]。

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi k n}{N}} \quad x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{j\frac{2\pi k n}{N}}$$

4.2 最近傍法とウォード法

音声がい通っているかどうかの判定にユークリッド距離を用いる。 n 個のサンプルについて、 p 個の変数 $X_{i1}, X_{i2}, \dots, X_{ip}$ ($i = 1, 2, \dots, n$) があり、初期状態として、 n 個のクラスターがある。(各クラスターは 1 サンプルずつを含むと考える)。

Step1 クラスター間のユークリッド平方距離 d_{ij}^2 を計算する。

$$d_{ij}^2 = \sum_{k=1}^p (X_{ik} - X_{jk})^2, \quad (i, j = 1, 2, \dots, n)$$

Step2 ユークリッド平方距離の最も小さいクラスターを併合して、1つのクラスターとする。クラスターaとクラスターbが併合されてクラスターcが作られるとする。 d_{ab} 、 d_{xa} 、 d_{xb} を、クラスターaとクラスターbが併合される前の各クラスター間の距離としたとき、併合後のクラスターcとクラスターx ($x \neq a, x \neq b$)との距離は(1)、(2)式で表される。

$$d_{xc} = \alpha_a d_{xa} + \alpha_b d_{xb} + \beta d_{ab} + \gamma |d_{xa} - d_{xb}| \quad (1)$$

$$d_{xc}^2 = \alpha_a d_{xa}^2 + \alpha_b d_{xb}^2 + \beta d_{ab}^2 + \gamma |d_{xa}^2 - d_{xb}^2| \quad (2)$$

Step3 2個のクラスターが1個のクラスターにまとめられたので、総クラスター数が1個減る。クラスター数が1になるまでStep2を繰り返す。

表1. クラスター分析の各種法で距離の再定義において使用されるパラメータ

	α_a	α_b	β	γ	使用される式
最短距離法	0.5	0.5	0	-0.5	(1)
ウォード法	$(n_x+n_a)/(n_x+n_c)$	$(n_x+n_b)/(n_x+n_c)$	$n_x/(n_x+n_c)$	0	(2)

n_a , n_b , n_c , n_x は、クラスターaに含まれるデータの個数

(1)式、(2)式で併合後のユークリッド距離を計算するときの定数 α_a , α_b , β , γ を変えることよって、各クラスター分析が行える。ウォード法が最も分類感度が高い。最短距離法では分類感度は低く、鎖状のクラスターを作る傾向がある[8]。

5. 実験

5.1 実験条件

言語獲得前の0~1歳の乳児(男児)における発声データを横断的に収集し感情特徴のパターン認識を行った。乳児の音声を1発声毎に切り出し各発声ごとに周波数解析を行い、無発声時の雑音成分を減算したものをクラスター分析後、各クラスターとのユークリッド距離により各感情毎での発声の学習モデルの構築を行った。録音は宇治市被験者A宅、京都市北区被験者B宅にて行った。

5.2 評価方法

事前の調査で眠気感情での音声パワーがやや弱いことがわかっていたので、各音声をFFT処理後、眠気感情を表す音声を基準として、その標準偏差を基に各感情をふくむ音声を正規化し、音声パワーが平均値+標準偏差の値以上になるものを学習用と認識用にわけ(値以下の音声はすべて眠気感情と認識)、学習用をクラスター分析後に三感情毎(不快・空腹・眠気)の音声モデルを構築し、各認識用音声を認識した。クラスター分析にはウォード法を用い、各クラスターの重心を測定後、残りの認識用データは最近傍法を用いてユークリッド距離による認識を行った。認識には以下の3つの方法を用いた。

◇手法1:感情既知の音声を8グループにわけクラスターに含まれる各感情(不快・空腹)の最も多いものをそのクラスターの感情と見なし、感情未知の音声(不快・空腹)感情を認識し、時系列的に連続した音声でみたときに一番多い感情をその連続発声全体の感情と認識した。

◇手法2:手法1のクラスター分析で各グループに含まれる感情(不快・空腹・眠気)の比率をそのままクラスターが持つ比率とし、感情未知な音声は最近接するクラスターの比率をその音声の認識結果の比率とし、時系列的に連続した音声の各感情比率の和を求め、数値が一番高いものその連続発声全体の感情と認識した。

◇手法3:クラスター分析の結果、不快・空腹・眠気が混在するグループでは1クラスター1感情になるようにクラスターを分解し、各音声は最近接する感情を時系列的にみたとき一番多い感情をその音声全体の認識結果とした。

5.3 実験結果と考察

録音した音声（月齢3か月のA252発声, 月齢2か月のB195発声）をクラスター分析したところ表2のように感情のあらわれ方に個人差があることがわかった。

表2. クラスター分析結果

Aの場合									Bの場合								
Cluster No.	1	2	3	4	5	6	7	8	Cluster No.	1	2	3	4	5	6	7	8
不快	○	○	○	○	○	○	○		不快	○	○						
空腹	○	○						○	空腹	○	○	○	○	○	○	○	○
眠気	○								眠気				○				

被験者毎に学習用に用いる感情既知の音声についてクラスター分析を行い、感情未知の音声を認識して連続した音声全体の感情を認識した。連続した乳児の10発声を1グループとし、それが不快・空腹・眠気の三感情のいずれであるかは、録音時の状況の記録（親の記録による）によりあらかじめ決定しておく、上記の三手法における認識率を求めた。

いずれの手法でも学習に用いた音声の認識については不快・眠気では60～100%と高い平均認識率を示した。また今回の実験では手法3の平均認識率が総じて一番高かった。感情表出パターンについては、事前に精緻にクラスター分析等の手法で分類しておくことが認識率の向上につながるものと考えられる。

表3. 乳児における各手法での認識率の平均 (%)

	手法1		手法2		手法3	
	学習用	認識用	学習用	認識用	学習用	認識用
不快	60.0	77.8	70.0	77.8	90.0	77.8
空腹	54.5	50.0	36.3	91.7	81.8	81.8
眠気	84.6	76.9	76.9	84.6	84.6	100

6. 結言

FFTによる周波数解析とクラスタリングを組み合わせることにより見出した乳児音声の特徴量を用いて感情特徴のパターン認識を行う手法を提案し、その有効性を示した。今後の事例の積み重ねと手法の改善により認識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] 古市久子, 川村晴子, 増原喜代, 小林光子, 山根尚美, 「幼児の感情表現について研究日常保育において感情表現が見えるとき」, 日本保育学会大会発表論文抄録, pp. 530-531. 1999
- [3] 林安紀子, 出口利定, 桐谷滋, 「選好振り向き法における4～11ヶ月齢児の音声刺激に対する反応」, 音声言語医学, vol.37, pp.317-323, 1996.
- [4] de Boysson-Bardies B, vihman MM. Adaptation to language: Evidence from babbling and first words in four languages, Language, vol.67, pp.297-319, 1991.
- [5] 天野政昭, 近藤公久, 「親子の発声時間と発声速度の月齢変化」, 日本音響学会聴覚研究会資料, vol.32, H-2002-08, pp.57-62, 2002.
- [6] 麦谷綾子, 「乳幼児の音声言語獲得」, 電子情報通信学会信学技報 TL200414, 2004.
- [7] 三谷政昭, 「信号解析のための数学 ラプラス変換, z変換, DFT, フーリエ級数, フーリエ変換」, 森北出版 1998
- [8] 宮本定明, 「クラスター分析入門 ファジィクラスタリングの理論と応用」, 森北出版, 1999.