

## 自然対話中の発話対における単独提示と連続提示の感情受容の比較

松澤 直之<sup>†</sup> 政倉 祐子<sup>††</sup> 大野 澄雄<sup>††</sup>

<sup>†</sup> 東京工科大学大学院 バイオ・情報メディア研究科 <sup>††</sup> 東京工科大学 コンピュータサイエンス学部

### 1 はじめに

感情認識において、話者の感情状態を感情音声から推定するために様々な手法が研究されている。既存研究 [?] では演技音声からの韻律的特徴を用いて感情の遷移を抽出し、話者の感情状態を推定する手法が提案された。しかし、演技音声による話者ラベルでは日常の感情を正確に網羅していない可能性がある。また、韻律的特徴以外に声質などのスペクトル特徴量を用いることで、より詳細に感情を認識できることが既存研究 [?] により明らかとなった。

本研究では話者の感情状態の推定手法の1つとして自然対話中の感情の遷移に着目し、さらに発話を単一で聞くか、連続で聞くかによって遷移時の傾向にどのような特性があるか音響特徴量から分析を行った。

### 2 評価実験

#### 2.1 単独提示と連続提示

我々は普段会話をしているときに単独発話で相手の感情を推定している場合もあるが、多くの場合は発話の前後からどう変化したのかを自然に読みとってコミュニケーションを行っている。この前後発話の変化をより詳細に調べるために何らかの出来事を挟んだ発話を単独発話で聞く場合と、出来事の前後発話を連続で聞いた場合の感情の評価傾向を音声に対して6感情から判別した評価者の回答の割合と音響特徴量を用いて分析を行った。また、本報告ではこれらを実験における単独提示、連続提示と定義する。

#### 2.2 感情音声コーパス

本研究で用いた音声は男子大学生4人を話者としたゲームプレイ中の自然対話である。音声の形式はPCM フォーマット、サンプリング周波数16kHz、量子化ビット16bitである。評価実験に用いた音声は発話対中の200ms以上の無音区間を境界とする発話を切り出し、単独発話と前後発話を結合した連続発話でそれぞれ580、290発話用意し以後の分析に用いた。評価実験で用いた各発話の構成を図1に示す。

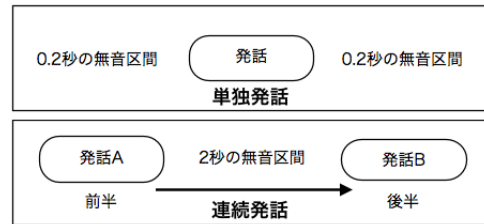


図1: 評価実験に用いた発話の構成

### 3 評価方法

本研究ではEkman(1992)の定義した基本感情である“喜び”、“驚き”、“怒り”、“恐れ”、“嫌悪”、“悲しみ”の6感情を取り扱う。ラベリングは男子学生8名に単独提示と連続提示の2種類の聴取評価実験を行い、6感情の中から音声を聞き感情の判別回答を行った。

### 4 音響特徴量の抽出

今回扱った特徴量を以下の表に示す。特徴量の抽出は音声分析ソフトpraatを用いて行い、発話単位の音声から抽出を行った。表1中の90%値、10%値ははずれ値を除去するためである。

表1: 扱った音響特徴量

特徴量	説明
$P_{max}$	パワーの発話内最大値
$P_{med}$	パワーの発話内中央値
$P_{min}$	パワーの発話内最小値
$P_{mean}$	パワーの発話内平均値
$P_{range}$	パワーの発話内ダイナミックレンジ
$P_{sd}$	パワーの発話内標準偏差
$F0_{max}$	F0の発話内最大値
$F0_{med}$	F0の発話内中央値
$F0_{min}$	F0の発話内最小値
$F0_{mean}$	F0の発話内平均値
$F0_{range}$	F0の発話内ダイナミックレンジ
$F0_{sd}$	F0の発話内標準偏差
$MFCC_{max}$	第1次メル周波数ケプストラム係数の発話内最大値
$MFCC_{min}$	第1次メル周波数ケプストラム係数の発話内最小値
$MFCC_{mean}$	第1次メル周波数ケプストラム係数の発話内平均値
$MFCC_{sd}$	第1次メル周波数ケプストラム係数の発話内標準偏差

### 5 分析結果

評価実験により6感情を判別した回答結果は音声ごとに回答の割合を算出し、音響特徴量を分析する際の尺度とした。

#### 5.1 エントロピーによる曖昧な感情の分類

各音声の感情として、回答の最も多い感情に属するとみなした。ただし、ばらつきの傾向を反映させるためエントロピーを用いることとした。エントロピーの

Comparison of acceptance characteristics of utterance pairs during a natural dialogue presented between isolatedly and continuously

<sup>†</sup> Naoyuki MATSUZAWA (Graduate School of Bionics Computer and Media Sciences, Tokyo University of Technology)

<sup>††</sup> Yuko MASAKURA and Sumio OHNO (School of Computer Science, Tokyo University of Technology)

算出方法には後述の式 (1) を用いた。最も曖昧な感情と判別された音声は6感情であるため  $\log_2 6 \doteq 2.6$  となり、多くの評価者がその感情だと判別した場合はエントロピーは0に近くなる。

$$E = - \sum_{k=1}^6 P_i \log_2 P_i \quad (1)$$

$E$  はエントロピー、 $P_i$  は各感情の回答の割合を表す。

### 5.2 エントロピーの値による音響特徴量の比較

表2はエントロピーの値から音響特徴量の抽出を3段階に分け、音響特徴量に変化が見られるか調べた。表2の太字は比較的特徴量の差が顕著なものを示している。単独提示では発話単体の感情の強弱がはっきりしているが、連続提示では各感情に明確な特徴量差が少ない。

### 5.3 感情遷移時における発話前後の特徴量比較

それぞれの提示で聴者の評価傾向が異なるため、2つの提示では感情遷移の傾向も異なる。図2はPowerのレンジが感情の遷移前後(平静から各感情)で変化するかによって前発話の影響を受けているかどうかを調べた。前の発話を聞かない単独提示の場合では前発話との差は小さく、連続提示では特徴量の最大値と最小値が単独提示よりも大きくなる傾向にある。

## 6 まとめ

本研究では6種類の感情の遷移をより詳細に抽出するために、単独提示と連続提示で感情受容の傾向と感情の遷移時にどのような特性があるかを声質を含めた音響特徴量から比較を行った。

その結果、表2の太字のように単独提示では感情の強弱が特徴量に現れやすく、連続提示では図2のように評価者が前発話を聞いて比較することで、前後発話の特徴量の変化を認識しやすい。これは感情の程度や同一感情内の変化など、より詳しい感情の変化を見る為に有効であると考えられる。しかし、連続提示による評価は単独に比べ困難であるため評価方法の改善や、より明瞭な特徴量差の検出など検討が必要である。

今後は音声データ数を増やし、さらに詳細な感情の程度と、より多くの音響特徴量から感情の遷移を抽出するために評価方法の改善などを行っていく予定である。

### 参考文献

- [1] 多田和彦, 矢野良和, 道木慎二, 大熊繁, “感情遷移における急激な韻律的特徴変化の検出による感情遷移判別方法”, 日本知能情報学会誌, Vol22, No.1, pp90-101, 2010.
- [2] Atassi H, Esposito A, Smekal Z, “Analysis of High-level Features for Vocal Emotion Recognition”, Telecommunications and Signal Processing(TSP), 34th International Conference, pp361-366, 2011

表 2: 各エントロピーの感情の特徴量平均

		単独提示			連続提示		
		$\geq 1.8$	$\geq 1.0$	$\geq 0$	$\geq 1.8$	$\geq 1.0$	$\geq 0$
平静	エントロピー	$\geq 1.8$	$\geq 1.0$	$\geq 0$	$\geq 1.8$	$\geq 1.0$	$\geq 0$
	$P_{max}$	79.7	82.1	82.3	83.1	82.4	82.4
	$F0_{max}$	<b>13.4</b>	<b>8.3</b>	8.3	11.9	12.0	12.2
	$MFCC_{max}$	<b>357</b>	<b>365</b>	<b>366</b>	356	349	356
	$LPCC_{max}$	2.51	2.62	2.58	2.65	2.54	2.56
驚き	エントロピー	$\geq 1.8$	$\geq 1.0$	$\geq 0$	$\geq 1.8$	$\geq 1.0$	$\geq 0$
	$P_{max}$	83.6	84.0	84.6	82.7	82.6	83.1
	$F0_{max}$	<b>14.1</b>	<b>15.9</b>	<b>17.1</b>	<b>9.40</b>	<b>11.3</b>	<b>12.4</b>
	$MFCC_{max}$	<b>361</b>	<b>354</b>	<b>342</b>	352	350	360
	$LPCC_{max}$	2.65	2.53	2.52	2.66	2.53	2.54
喜び	エントロピー	$\geq 1.8$	$\geq 1.0$	$\geq 0$	$\geq 1.8$	$\geq 1.0$	$\geq 0$
	$P_{max}$	79.9	82.7	82.9	81.7	82.9	83.6
	$F0_{max}$	<b>7.83</b>	<b>13.68</b>	12.30	12.5	12.1	15.1
	$MFCC_{max}$	<b>364</b>	<b>348</b>	<b>345</b>	<b>370</b>	<b>349</b>	<b>335</b>
	$LPCC_{max}$	2.55	2.57	2.57	2.64	2.50	2.55
怒り	エントロピー	$\geq 1.8$	$\geq 1.0$	$\geq 0$	$\geq 1.8$	$\geq 1.0$	$\geq 0$
	$P_{max}$	<b>84.9</b>	<b>82.1</b>	<b>81.7</b>	83.3	82.3	83.0
	$F0_{max}$	11.7	9.42	11.1	<b>16.1</b>	<b>10.3</b>	<b>8.40</b>
	$MFCC_{max}$	<b>377</b>	<b>345</b>	346	<b>336</b>	<b>343</b>	<b>385</b>
	$LPCC_{max}$	2.59	2.50	2.70	2.46	2.61	2.51
悲しみ	エントロピー	$\geq 1.8$	$\geq 1.0$	$\geq 0$	$\geq 1.8$	$\geq 1.0$	$\geq 0$
	$P_{max}$	82.8	82.3	83.0	83.3	82.9	82.4
	$F0_{max}$	<b>19.7</b>	<b>11.8</b>	<b>10.8</b>	<b>11.2</b>	<b>10.1</b>	<b>9.8</b>
	$MFCC_{max}$	358.9	356.3	367.1	<b>372</b>	<b>362</b>	<b>356</b>
	$LPCC_{max}$	<b>2.53</b>	<b>2.50</b>	<b>2.44</b>	2.61	2.67	2.57
恐れ	エントロピー	$\geq 1.8$	$\geq 1.0$	$\geq 0$	$\geq 1.8$	$\geq 1.0$	$\geq 0$
	$P_{max}$	83.4	83.2	83.9	82.7	84.2	83.4
	$F0_{max}$	20.1	18.3	19.6	<b>5.19</b>	<b>15.3</b>	<b>15.8</b>
	$MFCC_{max}$	328	337	333	366	356	367
	$LPCC_{max}$	2.42	2.56	2.54	2.47	2.51	2.54

※単位は Power は dB、F0 は 100Hz semitones を用いた

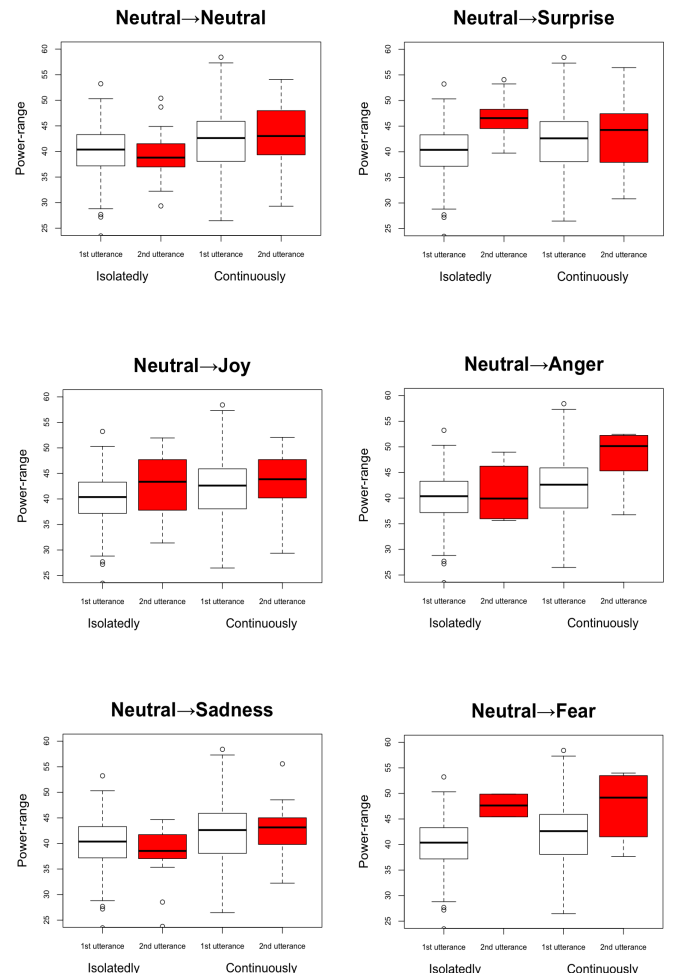


図 2: 平静から他感情遷移時の  $P_{range}$  の分布