

1 はじめに

人間間の音声コミュニケーションを観測すると、音声の音響情報から様々なパラ言語、非言語情報を抽出することで円滑なコミュニケーションを実現していることが分かる。本研究では、非言語情報である感情・意図に関してその自動抽出を検討した。この際、音響音声学に立脚した音声工学とは完全に異なる観点からの音声モデリングを行っている。「個々の音の音響特性、調音特性を記述する」音声工学に対し、本研究では音を個々に見ることはせず、音の並びや音の集合に潜む規則、関係、構造に着眼する。音声の物理現象の中に、話者収録環境に依存しない普遍構造が存在する [1] ことに着眼し、その普遍構造の大きさ・歪みが感情・意図に関連するとの仮説に基づいて実験的に検証した。例えば怒りや喜びなど positive な感情を表そうとすれば、相手に個々の音の違いを明確にはっきり伝えようとする調音努力が働き、音素間距離の広がりから構造サイズが大きくなるのではないかと、逆に悲しみなどの negative な感情では小さくなるのではないかと、という予測をたてた。本研究では特に 5 母音 /a/i/u/e/o/ のみに焦点を当て、研究を進めた。

2 音声に内在する普遍構造

本研究では、音素モデルをケプストラムベクトルによって構成されるガウス分布であると仮定して (即ち GMM で音素をモデリングすることに相当する)、音素間距離としてバタチャリヤ距離の平方根を用いた。分布 u と v 間のバタチャリヤ距離は以下の式によって与えられる。

$$BD(P_u, P_v) = -\ln \int_{-\infty}^{\infty} \sqrt{P_u(x)P_v(x)} dx$$

$$= \frac{1}{8} \mu_{uv} \left(\frac{\sum_u + \sum_v}{2} \right)^{-1} \mu_{uv}^T + \frac{1}{2} \ln \frac{|\sum_u + \sum_v|/2}{|\sum_u|^{\frac{1}{2}} |\sum_v|^{\frac{1}{2}}}$$

μ_u は u の平均ベクトル、 μ_{uv} は $u_u - u_v$ を、 $\sum_u$ は u の分散共分散行列を意味する。ここで、バタチャリヤ距離は、両分布を構成するデータに対して共通の一次変換をかけた場合でも、距離是不変である。即ち、如何なる一次変換も距離を変えない。一次変換における行列 A とベクトル b の音響音声学の意味を考慮すると、ケプストラムベクトルの足し算 b は、マイク特性、録音室の音響特性、伝送特性の違い、及び一部話者性を表現する。一方、スペクトルの非線型周波数ワーピングがケプストラムドメインでは、行列の A の掛け算になることが示されている。つまり、スペクトルワーピングによって距離行列は不変、となる。スペクトルワーピングを起こす要因の一つが、声道長の違い、即

ち人間の成長による音声の音響変化である。つまり、表列 A を掛けることによる変形は、構造の回転といった要素に分類される。一方ベクトル b を足すことによる変形は、構造のシフトとして観測される。

以上をまとめると、音素セットが成す構造を音素感の分布距離を用いて考えた場合、録音環境の違いは構造のシフトとして観測され、人間の調音器官の (特に) サイズの違い、聴覚特性による非線型ワーピングは構造の回転として観測されることになる。つまり、構造そのものは、収録環境・話者環境に一切不変である、という結論に達する。

3 音響的普遍構造に基づく音声中の感情・意図の分析

3.1 使用したデータ

プロの声優による 12 感情+12 意図の /a/i/u/e/o/ の収録をした。

感情を込めた母音：各感情とも発声は人格を統一してもらい、予め感情の種類、度合 [1] を説明した上で、感情の表現方法は声優に任せた。その後、各感情ごとに孤立的に各母音を 5 回ずつ発声してもらい (a,i,u,e,o,a,i...), その波形を切り取り分析した。

表 1. 感情パターン

	感情			
	明瞭	激怒	大喜び	深い悲しみ
大 (度合)	平静	怒り	喜び	悲しみ
小	不明瞭	押し殺した怒り	押し殺した喜び	押し殺した悲しみ

意図を込めた母音：予めこちらの意図を説明し、以下のようなシチュエーションのもと、発話意図を込めて発声してもらった。

(1) ごまかす感じに、(2) ため息のように、(3) 恐怖におののいた感じで、(4) ささやくように、(5) 驚きのあまり声が洩れてしまった感じで、(6) 声が震えてしまうような悲しみで、(7) 耳と遠いおばあちゃんに聞こえるように、(8) やる気のない感じで、(9) 心の思うままに大きな声で、(10) 言いたくないような嫌な気分、(11) 恥ずかしそうに、(12) 自慢気に、

以下では例として (1),(2),(4),(7) の結果を使用する。

3.2 分析条件

得られた孤立母音のデータを表 2 のような条件のもと、分析した。

表 2. 分析条件

データ	12 パターンで発声した母音 /a/i/u/e/o/ 5 回
サンプリング	16bit/16kHz
窓	シフト長 1 ms, ブラックマン窓長 25 ms
パラメータ	改良ケプストラム (12,24 次元) + 動的特徴
GMM	状態数 1 (全角分散共分散行列)

* Experimental examination about presumption of the utterance intention based on the acoustic universal structure in speech.

パラメータの次元数, 動的特徴を変えて実験を行った結果, 特徴パラメータを変えたことによる変化はあまり見られなかった. これは孤立母音を分析する際のシフト長が短かすぎて Δ 成分はほとんど変化しないためだと考えられる. よって, 以下では 24 次元の改良ケプストラムのみを用いた分析結果を示す.

3.3 実験の結果と考察

3.3.1 構造木のサイズ

各母音をバタチャリヤ距離を用いることにより, 母音間距離の近いものから結んでいき, 最終的な木の高さを比較することでそれぞれのパターンの発話の正確さ, 明瞭さといったものが表われてくると予測される. 図に各感情パターン (普通の度合) + 4 意図の 5 母音/a/i/u/e/o/の作成された木の構造を示す. なおデータは 5 回の発声のうちの適当なものを使用し, 表 3,4 には 各パターンの 5 回の発声の木のサイズの平均値を示す.

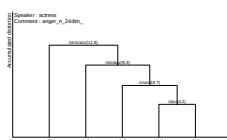


図 1. 怒り

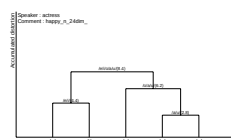


図 2. 喜び

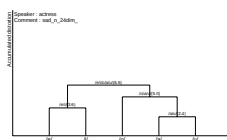


図 3. 悲しみ

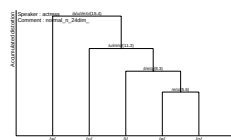


図 4. 平静



図 5. ごまかし

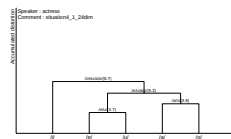


図 6. 囁やき

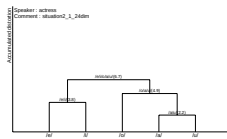


図 7. ため息

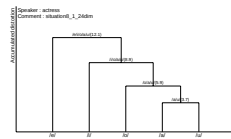


図 8. 明確

表 3. 木のサイズ (感情)

	平静	怒り	喜び	悲しみ
大	20.2	11.1	7.0	6.6
(度合)	15.6	11.2	8.9	6.8
小	10.6	10.6	6.1	6.4

表 4. 木のサイズ (意図)

明確	ごまかし	囁やき	ため息
12.7	8.8	7.0	6.4

木のサイズを見てみると, 感情パターンで一番サイズの小さくなっている, 度合の小さい喜びの音声を実際聞いてみたとき, ささやくような感情の表現方法であったことが, サイズが一番小さくなった原因ではないかと思われる. 怒りの場合も, 憤った感じではなく, 軽蔑したような怒り方であった. 不変構造で見れるものは音声の一側面であり, この話者のように感情・意図の表現には音源のコントロールの方法の違いなどにより, 構造に表れない感情表現もあることが分かる.

3.3.2 有意差の有無

1 パターンにつき 5 つの木ができるが, そこから得られた 12 パターンの木構造のサイズを用い, 各感情の有意差を ANOVA を用いて分析する. この値が 5%より小さければ有意な差がある, つまりデータは異なった母集団に属していると言うことができる. 表 5 に普通の度合の 4 感情の有意差をまとめたものを示す. 上の表から, 各パターンごとに明確な差があるわ

表 5. 有意水準 (感情)

	平静	怒り	喜び	悲しみ
平静	-	0.77	0.01	0.00
怒り	-	-	6.30	0.24
喜び	-	-	-	0.23
悲しみ	-	-	-	-

表 6. 有意水準 (意図)

	明確	ごまかし	囁やき	ため息
明確	-	0.16	0.01	0.00
ごまかし	-	-	0.47	0.67
囁やき	-	-	-	5.34
ため息	-	-	-	-

けではないが, 非言語情報を positive/negative で分類するなら, 両者には有意な差があると言える.

3.3.3 母音間距離の変化

各パターンの構造のサイズを正規化し, 各母音間の距離を求め [表 7], どのような歪みが表れるのかを分析した.

表 7. 母音距離の比較, 構造の歪み

	平静	怒り	喜び	悲しみ	ごまかし	囁やき	ため息	明確
a-i	0.91	0.80	0.87	0.85	0.80	0.78	0.82	1.07
a-u	0.79	0.68	0.48	0.59	0.67	0.54	0.57	0.49
a-e	0.81	0.75	0.60	0.67	0.73	0.68	0.74	0.81
a-o	0.79	0.59	0.53	0.62	0.63	0.64	0.60	0.49
i-u	0.64	0.64	0.95	0.73	0.76	0.71	0.74	0.66
i-e	0.51	0.71	0.64	0.67	0.68	0.56	0.62	0.77
i-o	0.59	0.76	0.89	0.82	0.77	0.95	0.86	0.72
u-e	0.63	0.61	0.58	0.62	0.66	0.66	0.70	0.62
u-o	0.61	0.67	0.57	0.68	0.60	0.72	0.60	0.44
e-o	0.54	0.76	0.72	0.75	0.72	0.73	0.74	0.70

コメント

4 おわりに

音声に内在する音響的普遍構造に着眼することで, 感情や発話意図を含む音声に対しその母音の構造に様々な変化が表れることを示した. 特に positive or negative といった分類では, 構造のサイズに大きな違いが表れ, 構造を見ることの有効性が伺える. 今後の課題としては …

参考文献

- [1] 峯松信明. "音声に内在する音響的特徴とそれに基づく語学学習者モデリング", 電子情報通信学会音声研究会資料, SP2003-179, 2004.
- [2] Hartmut R. Pfitzinger. "The /i/-/a/-/u/-ness of Spoken Vowels", Eurospeech2003.