社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

# 音声の分節的特徴に着眼したパラ・非言語情報推定に関する実験的検討

浜野　紘一†　　峯松　信明††　　広瀬　啓吉†

† 東京大学大学院新領域創成科学研究科
†† 東京大学大学院情報理工学系研究科
〒 113–0033 東京都文京区本郷 7–3–1
E-mail: †{k-hamano,mine,hirose}@gavo.t.u-tokyo.ac.jp

**あらまし**　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・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，音声の分節的特徴に着眼して，幾つかのパラ・非言語情報の自動抽出を検討した．従来パラ・非言語情報の抽出は音声の韻律に着眼した研究例が多いが，本研究では近年提案された，音声の分節的特徴に対する新しい分析手法である，音声の音響的普遍構造に基づく自動抽出を試みた．本分析手法は，音声に不可避的に混入する静的な歪み（年齢，性別などの話者性や，マイク・伝送特性などの音響条件など）を表現する次元を消失させた形で定義される音声の物理表象である．本報では，この音声の音響的普遍構造のサイズが凡そ調音努力に相当する情報を担っていることに着眼し，音声中のパラ・非言語情報と構造サイズとの関係について検討したので報告する．更には構造的な歪みと各種情報との対応についても実験的に検討した．
**キーワード**　パラ言語情報，非言語情報，分節的特徴，音響的普遍構造，構造サイズ，構造歪み

# Experimental Study on Estimation of Para- and Non-linguistic Information in Speech based upon Segmental Features

Koichi HAMANO†, Nobuaki MINEMATSU††, and Keikichi HIROSE†

† Graduate School of Frontier Sciences,
†† Graduate School of Information Science and Technology
7–3–1, Hongo, Bunkyo-ku, Tokyo 113–0033 Japan
E-mail: †{k-hamano,mine,hirose}@gavo.t.u-tokyo.ac.jp

**Abstract**　Speech communication between humans transmits/receives many kinds of para- and/or non-linguistic information as well as linguistic information. In this paper, automatic extraction of several kinds of the para- and/or non-linguistic information from speech was investigated. Conventionally, the extraction was often examined with prosodic features, however, only the segmental features were focused in this study. This is because a new method of speech analysis was proposed with respect to the segmental features, where a speech event is characterized as a structure composed of a set of distributions and the structure has completely no dimensions to represent static distortions such as age, gender, individuality, microphone, room acoustics, and so on. It was already reported that size of the structure corresponds to articulatory efforts in speech production and, in this paper, relations between size of the structure and para- and/or non-linguistic information were focused. Further, it was also examined whether the structural distortions also had some indications on the para- and/or non-linguistic information in speech.
**Key words**　para-linguistic & non-linguistic information, segmental features, acoustic universal structure, size of the structure, structural distortion

## 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・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，幾つかのパラ・非言語情報に関してその自動抽出を分節的特徴に着眼して検討した．従来音声の韻律に着眼した自動抽出が検討されている [1]〜[3] が，本研究では，分節的特徴のみに着眼する．

音声は不可避的に種々の歪みを保有した形でのみ存在できる．発声者の年齢，性別などに代表される話者性，マイク，伝達特性に代表される音響機器特性，更には聴取者の聴覚特性も厳密には聴取者毎に異なる．これらの歪みは音声というメディアには（静的ではあるが）不可避的に存在する．これらの静的な歪み（乗算性及び線形変換性の歪み）を表現する次元を消失させた形で定義される音声の物理表象が近年提案されている [4], [5]．音声事象を確率論的に有限個の状態として記述し，状態間距離を情報論的に算出し，最終的に音声事象を相対論的に状態群が成す構造として捉えると，その構造は乗算性歪みによって平行移動し，線形変換性歪みによって回転するだけであり，構造そのものはこれら歪みに一切影響を受けないことが数学的に導出される．筆者らの一部は，この構造のサイズが英語の強勢・弱勢の差異を明確に表現することを示している [6]．英語の母音は弱勢化すると，最も調音的に “楽な” 発声である schwa に変貌することを考えると，構造のサイズは個々の音をどれだけ明確に区別しようとしているのか，即ち，調音努力を反映していると解釈できる．本研究ではこの知見を更に発展させ，構造サイズの差異がパラ・非言語情報によっても変化すると仮定し，この仮定の実験的検証を目的とする．更には構造的な歪みとパラ・非言語情報との対応についても検討する．なお，本研究では分析を容易にするため，母音のみを扱うことにした．

## 2. 音声に内在する音響的普遍構造

音素(注1)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する．任意の二音素間の距離（距離行列）を以下の式で与えられるバタチャリヤ距離で算出する．

$$
\begin{aligned}
&BD(u,v)\\
&= -\ln \int_{-\infty}^{\infty} \sqrt{P_u(\boldsymbol{x})P_v(\boldsymbol{x})}d\boldsymbol{x}\\
&= \frac{1}{8}\mu_{uv}\left(\frac{\sum_u+\sum_v}{2}\right)^{-1}\mu_{uv}^T + \frac{1}{2}\ln\frac{|(\sum_u+\sum_v)/2|}{|\sum_u|^{\frac{1}{2}}|\sum_v|^{\frac{1}{2}}}
\end{aligned}
$$

$\mu_u$ は $u$ の平均ベクトル，$\mu_{uv}$ は $u_u - u_v$ を，$\sum_u$ は $u$ の分散共分散行列を意味する．上式より分かる様に，バタチャリヤ距離は二確率密度分布に対して，その独立性を仮定した上で両事象の同時確率に対する自己情報量として定義される．空間内の $n$ 点に対して ${}_nC_2$ 個だけ存在する対角線の長さのみを抽出することは，$n$ 点で張られる構造を考えることに等しい（図 1）．さて，バタチャリヤ距離は，2 つの分布に対して共通の如何なる一次変換 $Ax+b$ を施しても距離は変わらない性質を持つ（図 2）．つまり，構造は不変となる．$b$ を足す演算（乗算性歪み）は，収録環境の音響的差異，更には話者性の一部を表現するがこれは，構造の平行移動となる．$A$ を掛ける演算（線形変換性歪み）は，例えば周波数ウォーピングを意味し，これは声道長の差異による音響的差異，聴取者間による聴覚特性の違いを表現する(注2)が，これは構造の回転となる．以上の議論より構造化された音声事象は，音声の生成 · 収録 · 伝送 · 再生 · 聴取の過程において不可避に混入する静的歪みに一切影響を受けないことが数学的に導出される（音声の音響的普遍構造）．音響的普遍構造は，回転及び移動に関して明確な物理的意味を有するが，構造のサイズに関しても実験的にその意味付けが行われている．英語の強勢 · 弱勢母音群を各々構造化すると，構造のサイズの大小が，強勢 · 弱勢に相当することが示されている．「弱勢母音が schwa（最も発話エネルギーを使わない母音，弱母音）に近づく」という音声学の知見を考慮すると，構造のサイズは調音努力を表すものとして解釈できる．以下，この構造のサイズの大小とパラ・非言語情報との関係を実験的に検討する．

(注1)：言語的に意味のある音響事象である必要は無い．純粋に音響的に定義される音響事象を対象としても以下の議論は成立する．

(注2)：バーク尺度に代表される聴覚特性を変数変換によって，線形周波数軸空間における音声変形として捉えると，これも周波数ウォーピングとなる．

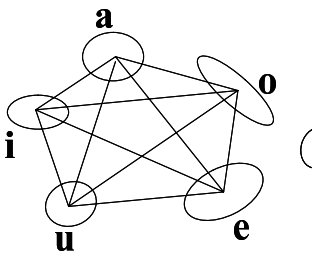


図 1　5 母音で構成される構造

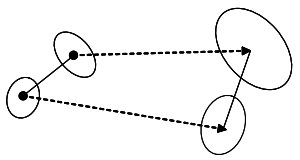

図 2　分布間距離と一次変換

表 1　その表出を依頼した感情の種類と度合い

| | | 感情の種類 | | | |
|---|---|---|---|---|---|
| 度合 | 大 | 明瞭 | 激怒 | 大喜び | 深い悲しみ |
| | 中 | 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 |

## 3. 音響的普遍構造に基づくパラ・非言語情報分析

### 3.1 音声試料

プロの声優（女性）による数種類のパラ・非言語情報を込めた音声を以下の手順に従って収録した．

**3.1.1** 感情を込めた孤立五母音発声

表 1 に示す各感情及びその度合いに対して，日本語五母音の孤立発声を依頼した．なお，各感情とも発声は人格を統一してもらい，具体的な感情の表出方法は声優に一任した．

**3.1.2** その他のパラ・非言語情報を込めた孤立五母音発声

上記に加え，以下の状況を想定し，同様の音声収録を行なった．

**ごまかし**： 言った内容をごまかしたいという感じで
**ため息**　： 落胆したときのため息のような声で
**恐怖**　： 幽閉されたとして，恐怖におののいた感じで
**囁き**　： 耳もとで囁くように
**驚き**　： 驚きのあまり声が洩れてしまった感じで
**震え**　： 声が震えてしまうような悲しみで
**明確**　： 耳の遠いおばあちゃんに明確に伝えるつもりで
**仕方なく**： 仕方なくどうでもいい感じで
**思いきり**： 心にとどめていた思いを，思いきり吐き出すように
**目一杯**　： 目一杯声が届くようにはっきりと
**恥じらい**： 言うのが恥ずかしいという感じで
**自慢**　： 自慢気に

なお，以下では見出しに書いてある言葉を使用する．

表 2　分 析 条 件

| | |
|---|---|
| データ | 各種パラ・非言語情報母音/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 (1〜24 次元) |
| 母音モデリング | 全角分散共分散行列による GMM |

### 3.1.3　種々の感情を込めた文発声

種々の感情を表現しやすいと考えられる 7 文を選び,「平静」「怒り」「喜び」「悲しみ」の 4 感情を込めて発声させた．この時，特に感情を強く表現する箇所を 2，3 箇所特定して発声を依頼した．その箇所を変えたものを 3 種類用意した．

**・文の例 (太字の箇所により強い感情を込める)**

紀ちゃんの言う/**青春を謳歌するってことと**/ちょっと違うかもしれないが/燃えているような/**充実感は**/今まで何度も/味わってきたよ．

### 3.2　構造サイズの算出

得られた音声試料を表 2 のような条件のもと，分析した．バタチャリヤ距離によって算出された音素間距離行列を Ward 法によって樹型図化する．Ward 法は累積歪みが最小となるようにマージ対象の 2 要素を選択するボトムアップクラスタリング手法であり，最終的に得られるその樹型図の高さは，全要素を一点で代表させたときの歪み，即ち VQ 歪みとなる．これは，音素群を構造として見た場合の，構造の半径に相当する物理量である．以下，この構造の半径を構造サイズと定義し，第 3.1 節によって収録された各音声に対し，その抽出を検討した．図 3〜6 に第 3.1.1 節で収録された各感情（中度合）の五母音から作成された樹型図を，図 7〜10 に第 3.1.2 節で収録された各パラ・非言語情報の五母音から作成された樹型図の例を示す．図 11，12 には得られた各種パラ・非言語情報の構造サイズの平均値をグラフ形式で示す．

### 3.3　各パラ・非言語情報と構造サイズの関係

構造のサイズを見てみると，小さい度合の「喜び」のサイズが「悲しみ」よりも小さくなっている．また「怒り」のサイズも「平静」より小さくなっているものがある．実際に音声を聞いてみると，前者は，ささやくような感情の表現方法であったことが，その原因だと考えられる．一方後者は憤った感じではなく，軽蔑したような怒り方であった．その他のパラ・非言語情報の構造のサイズを見てみると，言語から予想される構造サイズと実験結果とのよい対応が見られるものの，調音努力が大きくなると予想される「思いきり」が，反対に小さくなると予想される「ごまかし」や「恥じらい」,「仕方なく」よりも小さくなっていることが分かる．実際に「思いきり」の音声を聞いてみると，他の発声に比べて発話速度が速かった．話速が上がることにより，スペクトルが安定せず，分散が大きくなった結果として構造サイズが小さくなったものと考えられる．

構造サイズの大小から得られる順位付けは，一部を除いて各スタイルに対して人間が推察する積極性の度合と非常に整合性のとれた順位付けとなっており，本構造サイズ推定が発声の積極性推定に有効であることを示唆している．またその一方で，例に挙げたような予想と異なる結果も見られた．音響的普遍構造で観測されるものは音声の一側面であり，例えば，音源の制御手法の違いによって非言語情報を表現することも可能であり，構造のサイズに表れない表現方法もあることが分かる．

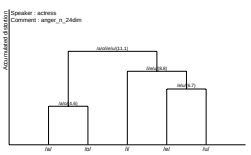
図 3　怒　　り

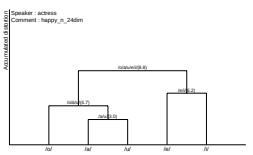
図 4　喜　　び

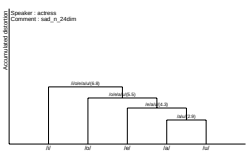
図 5　悲　し　み

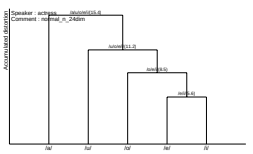
図 6　平　　静

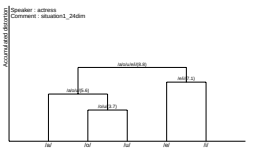
図 7　ごまかし

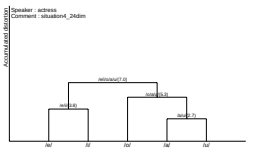
図 8　嘯　　き

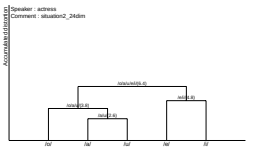
図 9　た　め　息

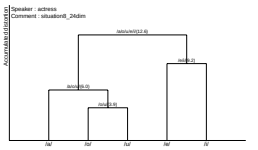
図 10　明　　確

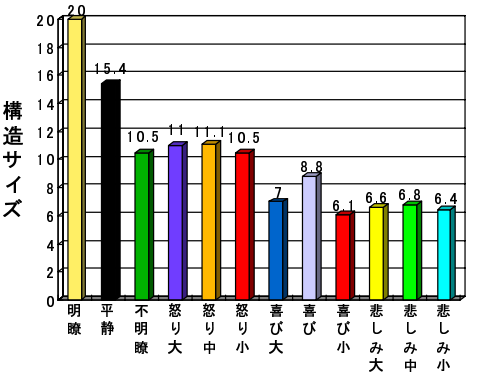

図 11　構造サイズ (感情)

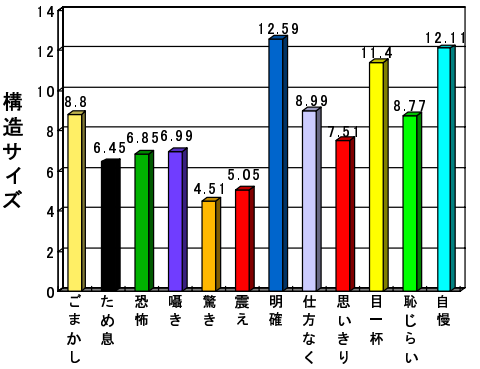

図 12　構造サイズ (その他)

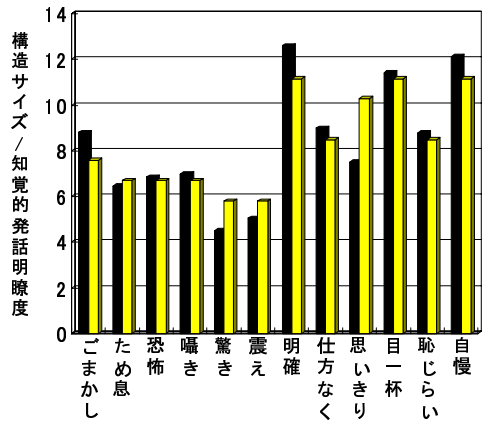


図 13　構造サイズと知覚発話明瞭度との関係

### 3.4　人間の知覚との比較

五母音構造のサイズと，その音声を聞いた時に感じる発話明瞭度の大小との関係を調べるために，聴取実験を行なった．各種パラ・非言語情報の音声データ 5 回発声のうちの一つ (/aiueo/) を聞かせ，以下に示す評価基準をもとにして 5 人の被験者に選んでもらった．評価得点を構造サイズの合計値により正規化し，それぞれを比較した結果を図 13 に示す．図中の棒グラフの左側が構造サイズ，右側が知覚的発話明瞭度である．

**評価基準**

5　各母音が正確に発声され，その違いが明確に伝わってくる．
4　各母音の違いがどちらかと言えば明確に伝わってくる．
3　どちらとも言えない，普通の発声．
2　各母音の違いがどちらかと言えば，不明瞭で伝わりにくい．
1　各母音の違いが不明瞭で伝わりにくい．

知覚的発話明瞭度と構造サイズを比較してみると，非常に良好な対応がとれていることが分かる．しかし，やはり「思いきり」における構造サイズと知覚的発話明瞭度との間に他との違いが見られる．全体の相関は 0.903 となり,「思いきり」を除くと 0.978 となった．分散項の効果は分析条件に依存する．人間の知覚表象との整合性を考慮した分析条件の設定が必要である．

### 3.5　構造サイズに基づいた有意差検定

得られた各スタイルにおける 5 回の孤立母音発声により抽出された 5 つの構造のサイズに対して分散分析を行ない，有意差検定を行った．表 3，4 に得られた有意水準 [%] を示す．1.0%以下は太字で示している．構造サイズと各感情 · パラ・非言語情報間に常に明確な有意差が観測される結果とはならなかったが，発話様式を「調音努力」という一次元で分析していることを考えると，十分な情報抽出が出来ていると考えている．

### 3.6　他話者による発声との比較

第 2. 節で述べたように，音響的普遍構造は，性別・年齢といった静的な話者性による歪みに一切依存しないことが示されている [7]．そこで女性声優によって発声された各種パラ・非言語情報が込められた音声データを別の男性話者に聞かせ，その発声方法を模倣させた．2 話者の構造サイズを比較したものを図 14，15 に示す．図中の棒グラフの左側が女性声優，右側が男性話者である．2 話者の構造サイズの相関は，感情で 0.882，その他のパラ・非言語情報で 0.804 となり，比較的良好な対応がとれていることが分かる．しかし，今回の収録ではアニメやゲームでの音声収録を本業とするプロの女性声優の発声（非常に特徴的な発声をしている）を，一般の素人男性に模擬させたものであり，その能力には限界があるのも事実である．演劇関係者など，発声スタイルを模擬することに慣れた発声者を用いた分析を行なう必要があると考えられる．

表 3　有意水準（感情）

| | 明瞭 | 平静 | 不明瞭 | 喜大 | 喜中 | 喜小 | 怒大 | 怒中 | 怒小 | 悲大 | 悲中 | 悲小 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 明瞭 | - | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** |
| 平静 | - | - | **0.0** | **0.0** | **0.0** | **0.0** | **0.3** | **0.8** | **0.5** | **0.0** | **0.0** | **0.0** |
| 不明瞭 | - | - | - | **0.0** | 1.5 | **0.0** | 55.8 | 55.2 | 95.0 | **0.0** | **0.0** | **0.0** |
| 喜大 | - | - | - | - | **0.3** | **0.0** | **0.1** | **0.3** | **0.8** | 30.8 | 43.9 | 24.3 |
| 喜中 | - | - | - | - | - | **0.0** | 3.6 | 6.3 | 15.4 | **0.4** | **0.2** | **0.4** |
| 喜小 | - | - | - | - | - | - | **0.0** | **0.1** | **0.2** | 23.6 | **0.5** | 45.3 |
| 怒大 | - | - | - | - | - | - | - | 90.9 | 74.3 | **0.1** | **0.1** | **0.1** |
| 怒り | - | - | - | - | - | - | - | - | 69.5 | **0.2** | **0.2** | **0.2** |
| 怒小 | - | - | - | - | - | - | - | - | - | **0.6** | **0.6** | **0.5** |
| 悲大 | - | - | - | - | - | - | - | - | - | - | 56.3 | 81.5 |
| 悲中 | - | - | - | - | - | - | - | - | - | - | - | 43.1 |
| 悲小 | - | - | - | - | - | - | - | - | - | - | - | - |

表 4　有意水準 (その他)

| | ごまかし | ため息 | 恐怖 | 囁き | 驚き | 震え | 明確 | 仕方なく | 思いきり | 目一杯 | 恥じらい | 自慢 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ごまかし | - | **0.1** | **0.6** | **0.5** | **0.0** | **0.0** | **0.2** | 76.0 | 3.8 | **0.1** | 90.8 | **0.3** |
| ため息 | - | - | 13.1 | 5.3 | **0.0** | **0.0** | **0.0** | **0.1** | 1.1 | **0.0** | **0.2** | **0.0** |
| 恐怖 | - | - | - | 90.6 | **0.0** | **0.0** | **0.0** | **0.5** | 19.4 | **0.0** | 1.3 | **0.0** |
| 囁き | - | - | - | - | **0.0** | **0.0** | **0.0** | **0.4** | 18.4 | **0.0** | 1.2 | **0.0** |
| 驚き | - | - | - | - | - | 7.3 | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** |
| 震え | - | - | - | - | - | - | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** | **0.0** |
| 明確 | - | - | - | - | - | - | - | **0.3** | **0.0** | 16.6 | **0.3** | 65.6 |
| 仕方なく | - | - | - | - | - | - | - | - | **2.8** | **0.3** | 87.5 | **0.5** |
| 思いきり | - | - | - | - | - | - | - | - | - | **0.0** | 5.8 | **0.0** |
| 目一杯 | - | - | - | - | - | - | - | - | - | - | **0.4** | 34.9 |
| 恥じらい | - | - | - | - | - | - | - | - | - | - | - | **0.5** |
| 自慢 | - | - | - | - | - | - | - | - | - | - | - | - |

### 3.7　音響的普遍構造の局所的歪みに関する分析

前節までは構造のサイズを見てきたが，本節では構造サイズの正規化を施した後の，構造の部分的歪みについて分析を行なった．即ち，各母音間の正規化音素間距離を，各パラ・非言語情報に対して求めた．結果を表 5，6 に示す．表 5 を見ると,「平静」と各感情の間では全体的に大きな歪みが表れていることが確認できる．特に/a/とその他の母音の距離が狭くなっている．表 6 からもサイズの大きくなるパラ・非言語情報とそれ以外では特徴が二分されている．特に目立つ特徴は/u-e/間距離はは全体を通してほとんど変化がなく,「平静」の/a-o/，/e-o/,「喜び」の/i-u/,「明確」の/a-i/，/u-o/,「恥じらい」の/i-u/には他と大きな違いが表れている．このように音声にパラ・非言語情報が加わることで，母音間距離行列内に様々な部分的歪みが生じることが確認できる．この結果は，これらの情報をパラ・非言語情報推定に有効であることを示唆している．

### 3.8　音響的普遍構造の大局的歪みに関する分析

前節では構造内の局所的な歪みに着眼したが，本節では異なる二構造間の大局的な差異を定量化することで普遍構造とパラ・非言語情報との関係について分析する．前節と同様に構造

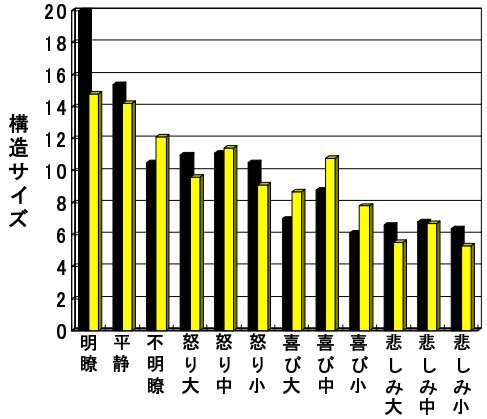


図 14　2 話者の構造サイズの比較（感情）

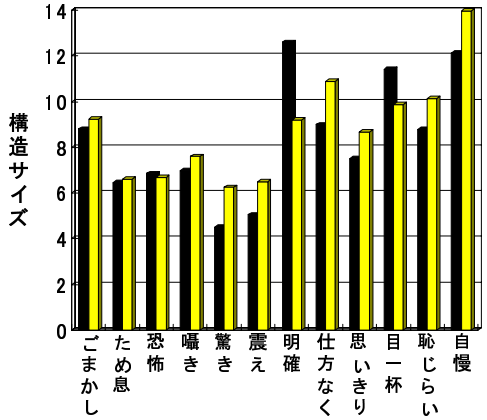


図 15　2 話者の構造サイズの比較（その他）

表 5　母音間距離（感情）

| | 明瞭 | 平静 | 不明瞭 | 怒大 | 怒中 | 怒小 | 喜大 | 喜中 | 喜小 | 悲大 | 悲中 | 悲小 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a-i | 0.76 | 0.91 | 0.86 | 0.80 | 0.80 | 0.82 | 0.80 | 0.87 | 0.80 | 0.73 | 0.85 | 0.80 |
| a-u | 0.86 | 0.79 | 0.71 | 0.64 | 0.68 | 0.69 | 0.53 | 0.48 | 0.53 | 0.62 | 0.59 | 0.49 |
| a-e | 0.66 | 0.81 | 0.78 | 0.78 | 0.75 | 0.80 | 0.60 | 0.60 | 0.66 | 0.66 | 0.67 | 0.66 |
| a-o | 0.75 | **0.79** | 0.71 | 0.72 | 0.59 | 0.73 | 0.52 | 0.53 | 0.61 | 0.65 | 0.62 | 0.62 |
| i-u | 0.67 | 0.64 | 0.65 | 0.62 | 0.64 | 0.59 | 0.80 | **0.95** | 0.78 | 0.75 | 0.73 | 0.82 |
| i-e | 0.59 | 0.51 | 0.57 | 0.54 | 0.71 | 0.63 | 0.70 | 0.64 | 0.62 | 0.55 | 0.67 | 0.60 |
| i-o | 0.63 | 0.59 | 0.82 | 0.83 | 0.76 | 0.78 | 0.84 | 0.89 | 0.90 | 0.83 | 0.82 | 0.85 |
| u-e | 0.59 | 0.64 | 0.57 | 0.51 | 0.61 | 0.57 | 0.74 | 0.58 | 0.59 | 0.57 | 0.62 | 0.58 |
| u-o | 0.76 | 0.62 | 0.63 | 0.70 | 0.67 | 0.59 | 0.61 | 0.57 | 0.68 | 0.79 | 0.68 | 0.72 |
| e-o | 0.66 | **0.54** | 0.64 | 0.77 | 0.77 | 0.74 | 0.82 | 0.72 | 0.79 | 0.82 | 0.75 | 0.79 |

表 6　母音間距離（その他）

| | ごまかし | ため息 | 恐怖 | 囁き | 驚き | 震え | 明確 | 仕方なく | 思いきり | 目一杯 | 恥じらい | 自慢 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a-i | 0.80 | 0.82 | 0.74 | 0.78 | 0.76 | 0.78 | **1.07** | 0.83 | 0.89 | 0.90 | 0.98 | 0.80 |
| a-u | 0.67 | 0.57 | 0.64 | 0.54 | 0.66 | 0.57 | 0.49 | 0.67 | 0.48 | 0.57 | 0.68 | 0.55 |
| a-e | 0.73 | 0.74 | 0.60 | 0.68 | 0.67 | 0.71 | 0.81 | 0.68 | 0.57 | 0.75 | 0.59 | 0.96 |
| a-o | 0.63 | 0.60 | 0.54 | 0.64 | 0.65 | 0.55 | 0.49 | 0.70 | 0.55 | 0.47 | 0.52 | 0.58 |
| i-u | 0.76 | 0.74 | 0.76 | 0.71 | 0.70 | 0.68 | 0.66 | 0.65 | 0.83 | 0.78 | **0.90** | 0.64 |
| i-e | 0.68 | 0.62 | 0.64 | 0.56 | 0.60 | 0.56 | 0.77 | 0.60 | 0.65 | 0.70 | 0.50 | 0.61 |
| i-o | 0.77 | 0.86 | 0.88 | 0.95 | 0.80 | 0.91 | 0.72 | 0.85 | 0.99 | 0.96 | 0.75 | 0.74 |
| u-e | 0.66 | 0.70 | 0.68 | 0.66 | 0.72 | 0.58 | 0.62 | 0.58 | 0.59 | 0.57 | 0.67 | 0.62 |
| u-o | 0.60 | 0.60 | 0.72 | 0.72 | 0.65 | 0.72 | **0.44** | 0.75 | 0.60 | 0.52 | 0.60 | 0.58 |
| e-o | 0.71 | 0.74 | 0.70 | 0.73 | 0.77 | 0.85 | 0.70 | 0.67 | 0.73 | 0.61 | 0.60 | 0.81 |

のサイズを正規化した上で，各種パラ・非言語情報により構成される構造間の差異を次式により導出した．ここで $M$ は $N$ 次元音響空間に存在する音素数であり，$\{P_i\}, \{Q_i\}$ は構造を表す

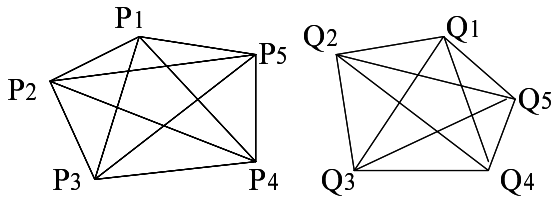


図 16　構造の大局的比較に基づく音響的照合

(図 16)．即ち次式は，$P$，$Q$ を表現する距離行列をベクトルと見なした場合のユークリッド距離に相当するが，この値が近似的に，構造 $Q$ に対して，乗算性及び線形変換性歪みに着眼して構造 $P$ へと適応化をかけ，適応化後の両構造間の対応する音素間距離の平均値となることが示されている [7]．得られた構造間距離の結果をまとめたものを表 7，8 に示す．

$$\sqrt{\frac{1}{M^2}\sum_{i<j}(\overline{P_iP_j}-\overline{Q_iQ_j})^2}$$

度合による違いを除いた同一感情内では，構造距離は比較的小さな値になっている．これは，同じ感情においては構造間の差異が小さいということを表している．しかし，表 7 で最も距離が近いものは「喜び大」と「悲しみ小」,「喜び小」と「悲しみ小」であった．今回の分析では構造サイズを正規化した上で（即ち調音努力を無視した上で）二構造間の比較を行なっている．その意味において「喜び」と「悲しみ」が似てきても不思議ではない．一方で，最も距離が離れているのは,「平静」と

表 7　構造間距離 (感情)[単位 $10^{-3}$]

| | 明瞭 | 平静 | 不明瞭 | 怒大 | 怒中 | 怒小 | 喜大 | 喜中 | 喜小 | 悲大 | 悲中 | 悲小 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 明瞭 | - | 64 | 64 | 73 | 70 | 69 | 111 | 125 | 97 | 75 | 80 | 98 |
| 平静 | - | - | 59 | 84 | 89 | 69 | 128 | **134** | 115 | 107 | 96 | 116 |
| 不明瞭 | - | - | - | 37 | 51 | 31 | 91 | 96 | 67 | 66 | 52 | 73 |
| 怒大 | - | - | - | - | 51 | 36 | 91 | 99 | 59 | 48 | 53 | 62 |
| 怒中 | - | - | - | - | - | 42 | 64 | 88 | 55 | 56 | **35** | 62 |
| 怒小 | - | - | - | - | - | - | 89 | 104 | 71 | 70 | 65 | 77 |
| 喜大 | - | - | - | - | - | - | - | 54 | 43 | 40 | 28 | **17** |
| 喜中 | - | - | - | - | - | - | - | - | 49 | 81 | 60 | 51 |
| 喜小 | - | - | - | - | - | - | - | - | - | 40 | 28 | **17** |
| 悲大 | - | - | - | - | - | - | - | - | - | - | 45 | 38 |
| 悲中 | - | - | - | - | - | - | - | - | - | - | - | 35 |
| 悲小 | - | - | - | - | - | - | - | - | - | - | - | - |

表 8　構造間距離 (その他)[単位 $10^{-3}$]

| | ごまかし | ため息 | 恐怖 | 囁き | 驚き | 震え | 明確 | 仕方なく | 思いきり | 目一杯 | 恥じらい | 自慢 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ごまかし | - | 23 | 22 | 35 | **5** | 40 | **77** | 19 | 63 | 61 | 60 | 45 |
| ため息 | - | - | 19 | 26 | 24 | 20 | 59 | 20 | 48 | 43 | 44 | 34 |
| 恐怖 | - | - | - | 17 | 23 | 29 | 57 | 19 | 48 | 47 | 44 | 31 |
| 囁き | - | - | - | - | 37 | 28 | 47 | 27 | 43 | 42 | 37 | 29 |
| 驚き | - | - | - | - | - | 39 | **77** | 19 | 62 | 61 | 60 | 45 |
| 震え | - | - | - | - | - | - | 47 | 30 | 32 | 28 | 30 | 25 |
| 明確 | - | - | - | - | - | - | - | 66 | 36 | 37 | 38 | 42 |
| 仕方なく | - | - | - | - | - | - | - | - | 49 | 49 | 44 | 35 |
| 思いきり | - | - | - | - | - | - | - | - | - | 19 | 22 | 24 |
| 目一杯 | - | - | - | - | - | - | - | - | - | - | 21 | 31 |
| 恥じらい | - | - | - | - | - | - | - | - | - | - | - | 29 |
| 自慢 | - | - | - | - | - | - | - | - | - | - | - | - |

表 9　感情文における母音構造サイズの平均値

| 平静 | 怒り | 喜び | 悲しみ |
|---|---|---|---|
| 22.14 | 23.17 | 22.96 | 20.07 |

「喜び中」であった．表 8 で一番構造間差異が大きいのは「ごまかし」と「明確」,「驚き」と「明確」で，一番小さいのは「ごまかし」と「驚き」であった.「ごまかし」と「驚き」は構造のサイズのみを見た場合には，それほど近いとは言えなかった．また第 3.2 節では,「思いきり」の構造サイズが予想される値より小さなものとなったが，本分析では，構造サイズの近かった他のパラ・非言語情報との有意差が観測された．このように構造サイズを正規化した上で構造間差異を見ることで，サイズとは異なる観点から発話スタイルを分析できることが確認された．

### 3.9　文音声試料を用いた構造解析

第 3.1.3 節により収録した文音声を用いて，これまでと同様に，音声中に内在する音響的普遍構造を利用した分析を行なった．

#### 3.9.1　分 析 手 順

（1） monophone を用いて音素アライメントを行なう．

（2） パラメータとして 24 次元の改良ケプストラムを利用し，母音分析と同様に，各音素の音素間距離行列を求める．

（3） 収録の際に指定した区間毎に，その中に存在する母音を用いて構造抽出を行なう．

表 9 に得られた母音構造サイズを示す．構造サイズが大きい順に並べると,「怒り」 ＞ 「喜び」 ＞ 「平静」 ＞「悲しみ」となり，調音努力が大きいと予想される順位付けとなった．しかし各感情文ごとに，感情を強く表現した箇所とそうでない箇所の構造サイズを比較した場合，必ずしも感情を強く表現した箇所がそうでない箇所より大きくなるという結果は得られなかった．これは孤立母音による実験結果と同様であり，構造サイズ以外の要因に基づく感情表出が行なわれていることが原因の一つであると考えている．

## 4. ま と め

音声に含まれるパラ・非言語情報の一部が，音響的普遍構造のサイズやその歪みとして表現されるとの仮説の下，種々の分析を行ない，この仮説の妥当性について検討した．本分析手法が分節的特徴のみに着眼した手法であることを鑑みれば，種々の実験結果は，提案手法がパラ・非言語情報の表象として十分機能していることを示している．第 3.2 節で述べた「喜び小」,「怒り」等の構造サイズが予想されたサイズと異なるものであった原因は，このデータ話者のパラ・非言語情報の表現方法が音源制御など，調音努力以外の要因に基づくものであったことなどが考えられる．第 3.2 節，第 3.9 節での感情の度合いを変えた時の結果に対しても，同様の考察が可能である．人間はある感情を表現するのに幾つかの手段を持っているが，段階を表現する際，徐々に各手段を増強するのではなく，利用する手段，あるいは割合を変更する，と考えられる．これは感情表出の手段に大きな個人差があることに対応している．しかし普遍構造解析で観測されるものは音声のごく一側面であり，その意味において他要因との融合が今後必要であろう．しかし，聴取実験による知覚的発話明瞭度と構造サイズとの比較では，非常に良好な対応が見られたことから，本構造サイズ推定が発声の明瞭度・積極性の推定には十分に有効に寄与できると考えている．

今後の課題としては，上記した基本周波数や，パワー，話速といった音源情報を導入することの他に，音響的普遍構造は話者性によっても不変であるという特性を用いて，話者性を取り除いたパラ・非言語情報推定の手法をより詳細に検討することなどが挙げられる．その際，母音だけでなくその他の音素を使用することで，違った結果が得られると考えられる．更には音素という枠を越えて，パラ・非言語情報を抽出するために適した音響単位を模索することも検討事項の一つである．更には，構造間の局所的歪み，大局的歪みを入力としたパラ・非言語情報推定を機械学習的に実装することも興味深い．

## 文　　献


[1] Oh-Wook, Kwon, Kwokleung.Chan, Jiucang, Hao, Te-Wonl.Lee, “Emotion Recognition by Speech Signals,” Proc. EUROSPEECH’03, pp.125–128, GENEVA, Switzerland, Sep. 2003.

[2] Vladimir.Hozjan,Zdravko.Kacic, “Improved Emotion Recognition with Large Set of Statistical Features,” Proc. EUROSPEECH’03, pp.133–136, GENEVA, Switzerland, Sep. 2003.

[3] Sherif Yacoub, Steve Simske, Xiaofan Lin, John Burns, “Recognition of Emotions in Interactive Voice Response Systems,” Proc.EUROSPEECH’03, GENEVA, Switzerland, pp.729–732, Sep. 2003.

[4] 峯松信明, “音声に内在する音響的普遍構造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, Jan. 2004.

[5] 峯松信明, 松井健, 広瀬啓吉, “音声に内在する音響的普遍構造とそれに基づく音声コミュニケーション,” 第 3 回話し言葉の科学と工学ワークショップ講演論文集, Feb. 2004. （発表予定）

[6] 朝川智, 峯松信明, 広瀬啓吉, “音声の音響的普遍構造に着眼した英語強勢・弱勢母音の分析,” 日本音響学会春季講演論文集, Mar. 2004.（発表予定）

[7] 峯松信明, “音声の音響的普遍構造の歪みに着眼した外国語発音の自動評定,” 電子情報通信学会音声研究会資料, SP2003-181, Jan. 2004.

[8] 浜野紘一, 峯松信明, 広瀬啓吉, “音声中に内在する音響的普遍構造に着眼した非言語情報推定に関する実験的検討,” 日本音響学会春季講演論文集, Mar. 2004. （発表予定）