

## 漫才番組における観客と漫才師の音声判別

中村 誠宏 阿部 孝司

近畿大学理工学部情報学科

## 1. まえがき

近年のお笑いブームの影響で若手漫才師は増加傾向にあり、養成学校も存在する。また、M-1グランプリのようなプロへの登竜門となっているTV番組もある。これらでは、熟練されたプロによる経験や印象など、人間により漫才の評価が行われているが、パフォーマンスの出来や漫才技術の成長過程に対する客観的評価の指標が無いため、漫才コンビは自身でパフォーマンスを評価することが容易でない。本研究では、2人1組で行う漫才映像に対して計算機により客観的評価を行う事を目的とする。漫才に対する客観的評価では、話の内容以外に、テンポ、タイミングや間でも感情を表現することが出来ると報告されている<sup>(1)</sup>。しかし、計算機でこれらを計測するには、漫才師が一人で発声している場所を認識する必要があり、現時点では歓声や拍手なども含む複数人の音声が入り混じっている状態の音声データから1人(漫才師)が話している箇所だけの音声を計算機で分ける事は容易ではない。本報告では、前述の計測を行うための第一段階として、漫才映像を用いて観客と漫才師の音声を判別する方法を提案する。

## 2. 観客と漫才師の音声に対する周波数特徴

図1のように漫才師発声時の周波数特性では、減衰振動が見られる。一方、観客の場合では、図2のように減衰振動の程度は図1に比べ有意に小さい。この違いは、観客は複数人による複合発声であるため、複数人の音声周波数が重なり合って1人の周波数特性よりも波の起伏が小さい特性になると推測する。従って、両者の判別を行うには、この違いを判別基準として両者を判別することが有効な手段の一つといえる。

## 3. 提案手法

2.で示した周波数特性の違いより、観客の音声と漫才師の音声の周波数特性から特徴抽出を行い、これを変数としてマハラノビス距離による判別分析を行うこと提案する。

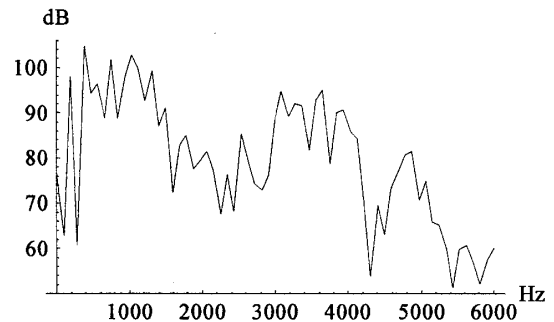


図1 漫才師の音声に対する周波数特性

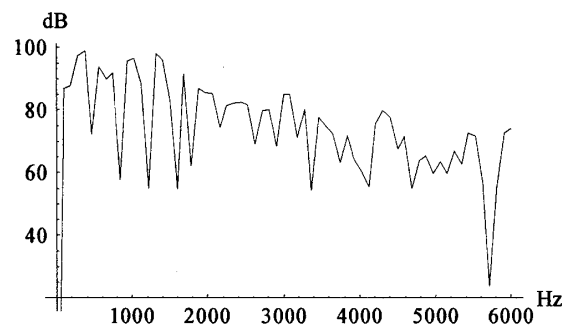


図2 歓声に対する周波数特性

## 3.1 サンプルデータの取得と変数定義

まず、漫才映像から音声データだけを抽出する。この音声データから漫才師 A, 漫才師 B, 観客だけの音声の場所を手動で約 1 秒間、各 1 箇所ずつ取り出す。取り出した約 1 秒の範囲内において、時間をランダムに 10 箇所を定め、その場所の周波数特性をそれぞれ求める。取り出したそれぞれの周波数特性から最小二乗法を用いて式1のように5次の近似式を作る。

$$y = a_0x^0 + a_1x^1 + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 \quad (1)$$

ここで、 $x$ は周波数、 $y$ はそのときの dB 値を表す。式1で求められた  $a_0 \sim a_5$  を判別を行う変数とする。一例として、図3と図4に図1と図2のそれぞれに対し式1で求めた近似曲線を示す。

## 3.2 判別関数の作成

$a_0 \sim a_5$  を変数として、式2に示すマハラノビス距離により、漫才師の音声か観客の音声かの判別を行う。取り出した漫才師の周波数特性を全て使うと漫才師のデータが観客のデータの倍の量になってしまうので、漫才師 A から 5 個、漫才師 B から 5 個の周波数特性をランダムに取り

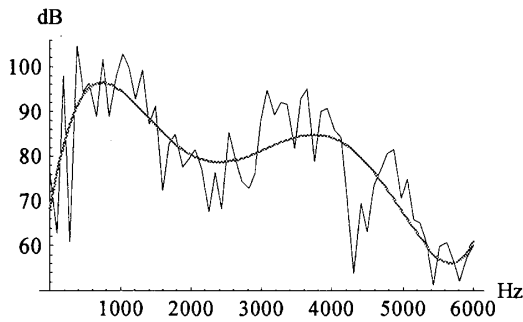


図3 漫才師の近似関数のグラフ

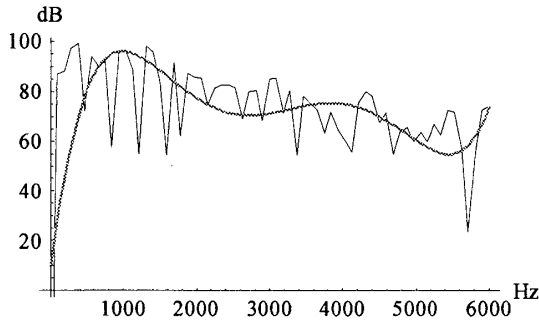


図4 観客の近似関数のグラフ

出し 10 個のデータセットを用意する。これを漫才師つまり一人の人間が発声している音声データに対するサンプルデータとする。観客に対するサンプルデータは、3.1 で抽出した 10 個の周波数特性である。

$$D_{(k)}^2 = \sum_{l=0}^5 \sum_{m=0}^5 (a_l - \bar{a}_{l(k)}) S_{(k)}^{lm} (a_m - \bar{a}_{m(k)}) \quad (2)$$

ここで  $k$  は群を表し、 $S_{(k)}^{lm}$  は第  $k$  群での分散・共分散行列の逆行列における  $l$  行  $m$  列の要素、 $\bar{a}_{l(k)}$  は変数  $a_l$  の第  $k$  群における平均を表す ( $l, m = 0, 1, 2, \dots, 5$ )。また、 $k$  は、1(漫才師)、2(観客)の 2 値を取り、対象となる周波数特性データはどちらかに分類される。式 2 において、 $D_{(1)}^2 \leq D_{(2)}^2$  のとき、そのデータは漫才師の音声と判別し、 $D_{(1)}^2 > D_{(2)}^2$  のときは観客と判別する。

#### 4. 実験

##### 4.1 実験方法

テストデータはサンプルデータと同様、漫才師 AB、観客だけの音声の場所を手動で約 1 秒間、各 1 箇所ずつ取り出し、取り出した約 1 秒から時間をランダム変更してそれぞれ 10 箇所の周波数特性を求め式 1 を作成し、変数  $a_0 \sim a_5$  を得る。ここでは、3 組の漫才において実験を行った。

##### 4.2 実験結果

表 1 には、漫才 3 組それぞれにおける漫才師の発話判別に対する再現率 (*recall*) と適合率

(*precision*) を示した。なお、再現率と適合率は、それぞれ、式 3 と式 4 により計算した。

$$recall = \frac{R}{C} \quad (3)$$

$$precision = \frac{R}{N} \quad (4)$$

ここで、 $R$  は本手法で実験を行いその結果が実際のカテゴリと同じであった数、 $C$  は実際のカテゴリの全数、 $N$  は本手法で判別された数を表す。

表 1 漫才コンビの判別結果

	再現率	適合率
漫才コンビ A	100%(10/10)	100%(10/10)
漫才コンビ B	100%(10/10)	100%(10/10)
漫才コンビ C	80%(8/10)	100%(8/8)

表 1 の判別結果で、漫才コンビ C に対する再現率が他の再現率の結果よりも低く、C に対する再現率の項目で漫才師の音声を使ったものが唯一判別を失敗していた。表 2 に距離  $D_{(k)}^2$  に対する分散値を示す。一つのデータから  $D_{(1)}^2$  と  $D_{(2)}^2$  が求まるが、ここでは小さい値を持つ方を選択し、そのデータに対する距離としている。この結果から、漫才コンビ C における漫才師に対する距離の分散が漫才コンビ AB に比べ有意に大きくなっているのがわかり、漫才師の発話中に他の音が紛れていたことが示唆される。漫才師 2 人が同時に発話している、あるいは、漫才師の発話と歓声が混じっている部分の判別を改善する事が今後の課題である。

表 2 テストデータにおける  $D^2$  の分散値

	漫才師	観客
漫才コンビ A	96	27
漫才コンビ B	2224	2032
漫才コンビ C	106509	92

#### 5. むすび

本報告では、漫才映像から漫才を計算機で客観的に評価するための前処理として、音声が入混在している状態のデータからの音声判別方法を提案した。実験の結果、再現率、適合率の両方において 80%以上の判別結果を得た。今後、漫才師 2 人が同時に発話している、あるいは、漫才師の発話と歓声が混じっている部分の判別を検討する必要がある。

#### 文献

- (1) 川嶋宏彰, スコギンズ・リーバイ, 松山隆司, “漫才の動的構造の分析—間の合った発話タイミング制御を目指して—,” ヒューマンインタフェース学会, vol.9, No.3, pp.379-390, 2007.