

クラスタリングに基づく発話中における笑い声の変動要因の検討*

☆永田智洋, 森大毅 (宇都宮大)

1 はじめに

笑い是非ノンバーバル情報を伝達する代表的な方法の1つであり、音声においてはその行動は笑い声として現れる。笑い声の合成を実現することができれば人間対機械のコミュニケーションや、コミュニケーション支援を目的とした機器の実現に大きく貢献できる。

笑い声には様々な形態があるため、笑い声を合成するためにはそれらを合成し合わせる必要がある。統計モデルに基づく音声合成においては、音響特徴量の変動要因を記したコンテキストを用いた決定木クラスタリングを行うことで、任意のテキストの音声の合成を実現しており、この手法は笑い声に対しても有効であると考えられる。笑い声の音響的特徴の変動要因を明らかにし、笑い声における適切なコンテキストを定義することで、様々な形態の笑い声の合成が可能になると考えられる。

そこで本研究では、大規模な音響特徴量を用いた笑い声のクラスタリングを行い、クラスタリングに大きく寄与する音響特徴量を採り、その音響特徴量を変動させる要因を考察することで発話中の笑い声の音響特徴量の変動要因について検討する。

2 笑い声の特徴量

2.1 対象コーパス

本研究では、宇都宮大学パラ言語情報研究向け音声対話データベース (UADB)[1] と日本語話し言葉コーパス (CSJ)[2] の笑い声を分析対象とする。ここで、CSJ に収録されている笑い声は対話形式で収録されているものに限定し、言語音と独立な笑い声のみを対象とした。UADB および CSJ から対象とした笑い声の数はそれぞれ 280 個、1325 個であり、総数は 1605 個である。

2.2 音響特徴量の抽出

分析対象の笑い声に対して、openSMILE[3] を用いて基本的な音響特徴量と、そこから計算される汎関数を求めた。基本的な音響特徴量には INTER-SPEECH2010 のパラ言語情報認識チャレンジ [4] で使用されている静的・動的特徴量を用いた。それらの特徴量に対して Table 1 に示す汎関数を求めたものを特徴量として用いた。最終的に使用する特徴量は 1824 種類である。

Table 1 使用した汎関数

汎関数
算術平均, 標準偏差
歪度, 尖度
四分位数
第 N 四分位数と第 M 四分位数の幅
最大値, 最小値, 最大値と最小値の幅
最小値位置, 最大値位置
1%, 99%パーセンタイル
99%パーセンタイルと 1%パーセンタイルの幅
線形回帰直線の傾き, 切片
線形回帰直線からの二乗誤差
レンジの 75%を超えている時間の割合
レンジの 99%を超えている時間の割合

特徴量の抽出の際には、サンプリング周波数 16 kHz の音声に対してフレーム長 25 ms、フレーム周期 5 ms のハミング窓を用いて抽出した。

3 笑い声のクラスタリング

3.1 クラスタリング条件

前節で示した 1824 次元の特徴量ベクトルに基づき、構造的クラスタリングを行った。クラスタ間の類似度の尺度には標準化ユークリッド距離を使用した。クラスタリング手法には Ward 法を用い、クラスタの数は 15 とした。

3.2 クラスタリング結果

クラスタリングの結果として構築された木構造を Fig. 1 に示す。木構造の末端の数字はクラスタ番号を表し、縦軸は距離を表している。

この木構造は、クラスタ 1 からクラスタ 5 に属する笑い声の音響特徴量とクラスタ 6 からクラスタ 15 に属する笑い声の音響特徴量に大きな差があることを示している。また、クラスタ 6 からクラスタ 15 に属する笑い声のうち、クラスタ 6 からクラスタ 7 に属する笑い声の音響特徴量とクラスタ 8 からクラスタ 15 に属する笑い声の音響特徴量の差が大きく、クラスタ 9 からクラスタ 11 に属する笑い声の音響特徴量とクラスタ 12 からクラスタ 15 に属する笑い声の音響特徴量の差が大きいことも示している。

* Clustering-based identification of variability of laughter in speech . by NAGATA, Tomohiro, MORI, Hiroki (Utsunomiya University)

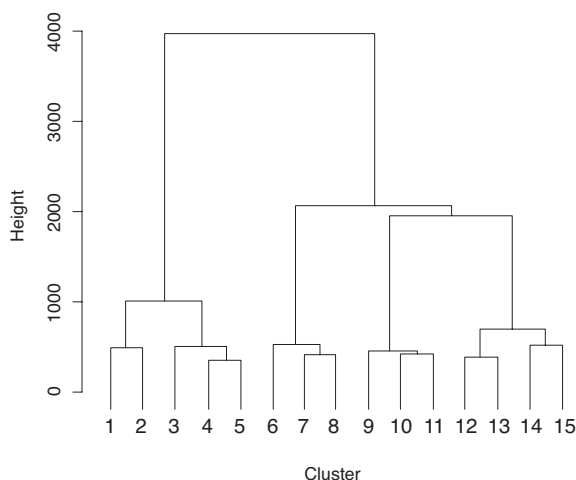


Fig. 1 クラスタリングによって構築された木構造

Table 2 クラスタ 1-5 とクラスタ 6-15 の差

特徴量	平均差の絶対値
F0 max	1.80
Δ F0 range	1.78
F0 stddev	1.78
Δ F0 1.0% percentile	1.77
Δ F0 min	1.76

クラスタ 1 からクラスタ 5 に属する笑い声とクラスタ 6 からクラスタ 15 に属する笑い声の平均差の大きい特徴量を Table 2 に示す。ここで、表は差の大きい順に 5 種類まで示している。表より、差の大きい特徴量は基本周波数に関する特徴量に集中していることがわかる。最も差があるのは基本周波数の最大値である。また、差の大きい音響特徴量には基本周波数の動的特徴量が多い傾向があった。

同様にクラスタ 6 からクラスタ 8 に属する笑い声とクラスタ 9 からクラスタ 15 に属する笑い声の平均差の大きい特徴量 5 種類を Table. 3 に示す。表中の \log MelFreqBand[N] はメル周波数帯域の対数パワーである。表より、差の大きい特徴量には低域のメル周波数帯域の対数パワーが集中していることがわかる。

更にクラスタ 9 からクラスタ 11 に属する笑い声とクラスタ 12 からクラスタ 15 に属する笑い声の平均差の大きい特徴量 5 種類を Table. 4 に示す。表より、差のある特徴量には高次のメル周波数ケプストラム係数の動的特徴が集中していることがわかる。

以上の結果から、笑い声のクラスタリングにおいては基本周波数、メル周波数帯域の対数パワー、高次のメル周波数ケプストラム係数が重要であることがわかった。基本周波数については、レンジや動的特徴量に関する特徴量の差が大きいという傾向から

Table 3 クラスタ 6-8 とクラスタ 9-15 の差

特徴量	平均差の絶対値
\log MelFreqBand[1] quartile1	1.51
\log MelFreqBand[1] mean	1.46
\log MelFreqBand[2] quartile1	1.45
MFCC[0th] quartile1	1.44
\log MelFreqBand[1] quartile2	1.40

Table 4 クラスタ 9-11 とクラスタ 12-15 の差

特徴量	平均差の絶対値
Δ MFCC[10th] range	1.25
Δ MFCC[12th] range	1.25
Δ MFCC[10th] min	1.23
Δ MFCC[13th] range	1.21
Δ MFCC[11th] range	1.19

笑い声のトーンに関する情報が変動要因として重要であると考えられる。また、メル周波数帯域の対数パワーやメル周波数ケプストラム係数に関する特徴量の差が大きいことから、母音/a/に似た笑い声、母音/e/に似た笑い声のような韻質に関する情報が変動要因として重要であると考えられる。

笑い声に対して、これらの情報をアノテーションすることによって笑い声の決定木クラスタリングを行うことができると考えられる。

4 おわりに

本研究では、UUDB および CSJ の笑い声から抽出された大規模な特徴量を利用して笑い声のクラスタリングを行った。クラスタリングの結果をもとに差のある音響特徴量を求め、クラスタリングに大きく寄与する音響特徴量を求めることで、その音響特徴量の変動させる要因について考察した。

今後の課題としては、見出した変動要因を笑い声に対して付与すること、付与した変動要因を用いた決定木クラスタリングによる笑い声の合成が挙げられる。

参考文献

- [1] Mori *et al.*, Speech Communication, 53, 36-50, 2011.
- [2] 前川, 日本語科学, 15, 111-133, 2004.
- [3] Eyben *et al.*, TU Munchen, MMK, 2013.
- [4] Schuller *et al.*, Proc. INTERSPEECH 2010, 2794-2797, 2010.