5U-6

# 種々の発話印象を表現する音声合成のための音響的特徴量の検討

上野 吉弘 † 政倉 祐子 †† 大野 澄雄 ††
† 東京工科大学大学院バイオ・情報メディア研究科 †† 東京工科大学コンピュータサイエンス学部

## 1 はじめに

近年人と機械が接することが多くなり音声合成技術の研究が盛んになってきている. 既存研究により人間の話す音声には韻律的特徴に加え, 分節的特徴も影響を与えており, 自然な音声合成を行うためには韻律的特徴に加え, 分節的特徴も考慮する必要があることが考えられる. また, 発話内容を正確に伝えるためには, 言語情報以外に場面や話者の状態, 感情などの感性情報 (以下発話印象) も重要であることがわかっている.

そこで本研究では, 自然な音声合成の実現を目指し, 発話印象を考慮した韻律的特徴, 及び分節的特徴を分析を行い音声合成の発話印象付加による品質向上を図ることを目的とし, そのための特徴量の検討を行う.

そのためにまず録音した自然発話に SD 法を用い,9 軸の評価軸に対して評価実験を行った. その結果を元に因子分析を行い, 因子軸を 2 つにしぼった. その後, その 2 軸に対して音声の特徴量の比較を行い発話印象ごとの変化を調べた.

## 2 研究内容

音声を録音し, その音声から発話印象を考慮した合成規則を見つけ, 音声合成の質を高める.

### 2.1 音声コーパス

本研究における分析対象として,277 発話を対象にしている.

音声は Mister_O という絵本を用い,1 対 1 の対話方式で録音を行った. Miste_O は台詞のない絵が 60 コマ並んでいる外国の作者の絵本である. 絵本の中の 1 つの話から間の 24 コマを抜き出しランダムに並べる. 正しい順番を二人で話し合い, その様子を録音した.

録音は Skype による会話を callbuber というソフトを用い 16bit,42kHz で行った. また話者は男性 3 名, 女性 1 名の計 4 名の音声を録音した.

### 2.2 対象とする音声の特徴量

本実験で対象とする音声の特徴量を表 1 に挙げる. 特徴量はそれぞれの発話単位の音声から抽出している. $F_0$ は音声の高さを表しており, 一つの音声発話内の 10%,50%,90%と平均, 標準偏差を求めている.

Power は音声の強さを表している.$F_0$ と同じく一つの音声発話内の 10%,50%,90%と平均, 標準偏差を求めている.MFCC(Mel-frequency cepstral coefficients) は第 1 次メルケプストラム係数であり, 声質を表している. 一つの音声発話内の最大値, 最小値, 平均, 標準偏差を求めいている. また, それぞれの値は praat を用いて抽出している.

表 1: 特徴量抜粋

| | | |
|---|---|---|
| $F_0$ | 10% | $F_0$ の 10%の値 |
| | 50% | $F_0$ の 50%の値 |
| | 90% | $F_0$ の 90%の値% |
| | 平均 | $F_0$ の平均値 |
| | 標準偏差 | $F_0$ の標準偏差 |
| Power | 10% | Power の 10%の値 |
| | 50% | Power の 50%の値 |
| | 90% | Power の 90%の値 |
| | 平均 | Power の平均値 |
| | 標準偏差 | Power の標準偏差 |
| MFCC | 最小値 | MFCC の最小値 |
| | 最大値 | MFCC の最大値 |
| | 平均 | MFCC の平均 |
| | 標準偏差 | MFCC の標準偏差 |

### 2.3 評価実験

上記のコーパスに対し, 評価実験を行った. 評価方法は SD 法を用い, 一つの音声に対して 9 軸,5 段階評価で行った. SD 法を用いた理由として, 対立する印象によって一つの発話ごとの特徴量の変化傾向をみることができると考えたからである. SD 法に用いた軸は, 声質によって判別できると考えられるものを取り扱った. また評価者は男性 8 名である.

(1) 冷静(-2) – (-1) – (0) – (1) – (2)興奮
(2) 穏やか(-2) – (-1) – (0) – (1) – (2)激しい
(3) 暗い(-2) – (-1) – (0) – (1) – (2)明るい
(4) 快(-2) – (-1) – (0) – (1) – (2)不快
(5) 軽快(-2) – (-1) – (0) – (1) – (2)重厚
(6) 暖かい(-2) – (-1) – (0) – (1) – (2)冷たい
(7) 威圧(-2) – (-1) – (0) – (1) – (2)謙虚
(8) 馴れ馴れしい(-2) – (-1) – (0) – (1) – (2)余所余所しい
(9) 親しみにくい(-2) – (-1) – (0) – (1) – (2)親しみやすい

図 1: 評価軸と 5 段階評価

Examination of the acoustic features of speech synthesis for impressive speech to express a variety of
† Yoshihiro UENO(Tokyo University of Technolosy)
†† Yuko MASAKURA(Tokyo University of Technolosy)
†† Sumio OHNO(Tokyo University of Technolosy)

## 3　結果

### 3.1　因子分析

評価実験に用いた 9 軸に対して因子分析を行った. その結果を図 2 及び表 2 に示す.

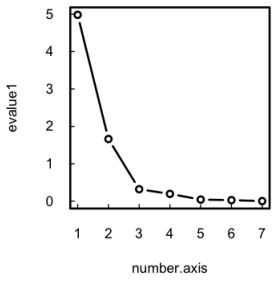


図 2: 因子分析結果

表 2: 因子得点表

| axis | fact1 | fact2 |
|---|---|---|
| 1 | 0.83 | 0.327 |
| 2 | 0.676 | 0.581 |
| 3 | 0.903 | -0.034 |
| 4 | -0.727 | 0.433 |
| 5 | -0.854 | 0.201 |
| 6 | -0.773 | 0.419 |
| 7 | -0.16 | -0.76 |
| 8 | -0.746 | -0.382 |
| 9 | -0.763 | 0.299 |

図 2 から第 1, 及び第 2 因子が影響を与えていることがわかる. そこで第 1, 第 2 因子の因子得点を求めた. 表 2 はそれぞれの軸ごとの因子得点の値を示している. 図 1 と表 2 から高い因子得点係数を示している発話印象がわかる.fact1 は”興奮”,”明るい”,”軽快”,fact2 は”威圧”,”激しい”などが挙げられる.

### 3.2　音響特徴量の比較

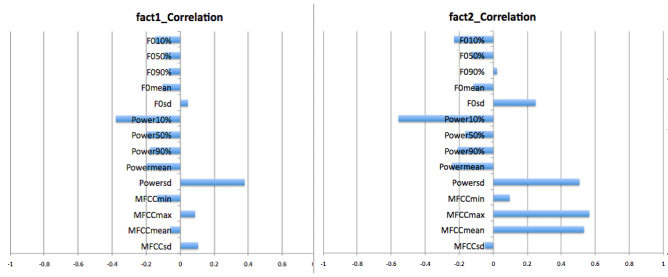


図 3: 第 1 因子, 第 2 因子と特徴量の相関係数

第 1, 第 2 因子と各特徴量の相関関係を調べた. 図 3 に第 1 因子, 第 2 因子と各特徴量との相関のグラフを示す.

第 1 因子と特徴量の相関を見たところ,Power の最小値, 標準偏差がやや高い. これは第 2 因子との相関にも言えることから,Power における最小値, 標準偏差は因子との関係が深いことがわかった. また第 2 因子と特徴量の相関をみると, 最小値, 最大値の他に MFCC の最大値, 平均が高い相関を示している. 第 2 因子は”威圧”,”激しい”などの発話印象を表していると思われるため, これらの発話印象は MFCC の値が上昇する傾向があることがわかった.

$F_0$ に関して, 大きな相関が得られなかった. これは今回個人差, 男女差の正規化を行わずに特徴量を扱ったためと考える.

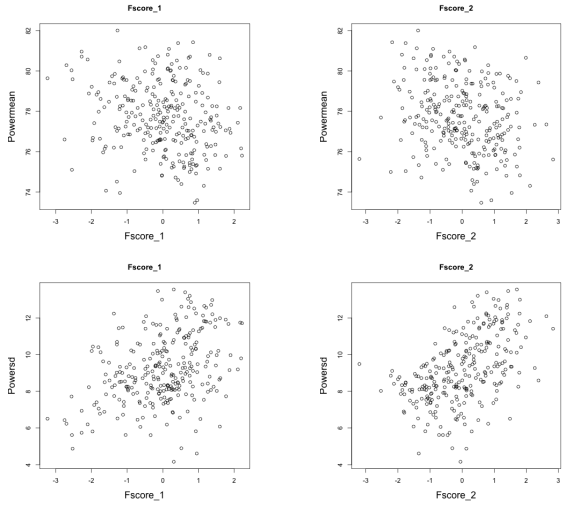


図 4: Power と因子得点の関係図

図 4 は求めた特徴量の中の Power の値の中から平均, 標準偏差を抜粋した. また左が第 1 因子得点, 右が第 2 因子得点をそれぞれ対象にした際のグラフである. 横軸が因子得点, 縦軸が Power を示している.

平均において, 第 1, 第 2 因子ともにばらつきは見られるが, 全体的に右下がりの傾向があることがわかる. これより Power の平均値は因子得点の値が大きくなると平均値が小さく, 因子得点が小さくなると平均値が大きくなることがわかる. また平均とは逆に, 標準偏差は因子得点の係数が大きいとき, 標準偏差の値も大きく, 因子得点が小さくなると標準偏差も小さくなることがわかる. また全体的にみて第 1, 第 2 因子において Power の値の分布に大きな変化は見られない.

## 4　まとめ

本研究では音声合成の発話印象付加を目指し, 因子分析と特徴量の比較, 検討を行った. その結果求めた因子の差による特徴量の変化はあまり見られなかった. しかし, 因子得点の係数が大きくなることで値が大きく, または小さくなる特徴量がいくつか見られた.

今回正規化を行わなかったために低い相関であった $F_0$ などの特徴量の取り扱う方法や, 今回検討しなかった別の特徴量などの導入を検討する必要があると考えられる. また, 今後は分節的特徴をより詳しく分析を行い, 音声合成に反映させる予定である.

## 参考文献

[1] 小川 順平, 西田 昌史, 堀内 靖雄, 黒岩 眞吾,”書き起こしへの付与を目指した発話印象の表現法に関する分析”, 情報処理学会, 2007-SLP-69(16)