

笑い声合成における音声記号表現と音響特徴量の感情次元による制御*

☆木村駿野, 森大毅 (宇都宮大)

1 はじめに

コーパスに基づいて笑い声を合成するためには、音声記号により笑い声の転記を行い、テキスト音声合成と同様の枠組でモデル化するのが一般的である [1]。この場合、合成時には合成したい笑い声の音声記号列を与える必要がある。しかし、与えられた文脈にふさわしい笑い声の音声記号表現をどのように作り出すかが課題となっている。

笑い声から喚起される感情は有声性に関連することが知られており [2]、有声性は笑い声の call (音節) 数とも関連する [3]。今西・有本らは、笑い声の構成音パターンと、話者の感情状態との関係を分析するとともに [4]、感情次元からの笑い声の構成要素の推定を提案している [5]。

本稿では、笑い声の音声記号表現、および音響特徴量と、その笑い声から知覚される感情状態との関係を、機械学習によりモデル化することを試みる。得られたモデルを用いて感情次元から笑い声を合成し、合成笑い声から知覚される感情状態の制御が可能か検証する。

2 コーパス

笑い声 (laughter episode) は、1 回以上の呼気／吸気で構成された音響イベントである。また、1 回の呼気に対応する笑い“句”(bout) は 1 個以上の笑い“音節”(call) からなる [6]。

本研究では、多くの笑い声が含まれる自然対話コーパスである OGVC [7] を用いる。これまでの研究で、OGVC に含まれる笑い episode に対し、call のアノテーションを行い、それらを音声記号で表現した [1]。また、各笑い episode に対し、著者らが感情次元 (快-不快、覚醒-睡眠) の評価を 7 段階 (1: 非常に不快／睡眠、7: 非常に快／覚醒) で行い、それぞれの平均スコアを感情ラベルとして付与した。今回は、男性話者 04.MSY の 125 episodes、女性話者 06.FWA の 100 episodes が実験データとして選ばれた。

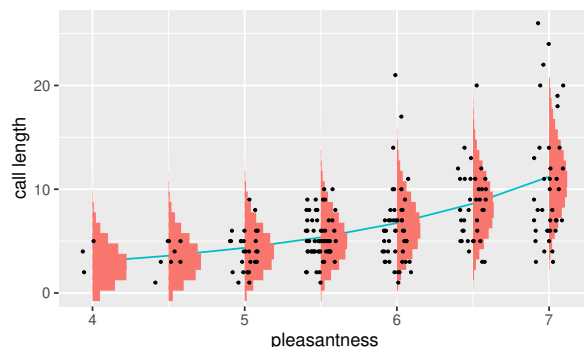


Fig. 1: 快-不快次元に対する call 系列長の分布

3 笑い声合成

3.1 音声記号表現生成モデル

提案する笑い声の音声記号表現のモデルは、call の系列長を予測する一般化線形モデルと、call 系列を予測するニューラルネットから成る。

データから求められた系列長モデルの回帰式を以下に示す。

$$\log(\lambda_i) = 1.105 + 0.987x_i^{\text{ple}} + 0.32x_i^{\text{ple}}x_i^{\text{aro}} \quad (1)$$

$$y_i \sim \text{Poisson}(\lambda_i) \quad (2)$$

ただし y_i は笑い声の系列長、 $\text{Poisson}(\lambda)$ は λ を平均パラメータとして持つポアソン分布、 x_i^{ple} および x_i^{aro} は [1, 7] を [-1, 1] に線形変換した笑い声の感情次元 (快-不快および覚醒-睡眠) である。

Fig. 1 に、快-不快と系列長の分布 (黒点)、式 (1) で与えられる平均パラメータ (緑線)、およびポアソン分布の確率質量関数 (赤) を重ねて示す。ただし、図の簡略化のため理論プロットにおける覚醒-睡眠の値は快-不快と同じとした。図から、上記モデルが快-不快の違いによる系列長分布の違いをよく表現していることがわかる。

Call 系列予測モデルには LSTM (ユニット数:128) を用いた。Call の音声記号は 16 次元の埋め込みで表現され、感情次元 2 次元、call 位置 1 次元を付与した合計 19 次元をモデルへの入力とした。その後、線形層+softmax により出力された確率分布に従い、22 種の call または終端記号をランダムに 1 つ生成する。これを終端記号が出力されるまで繰り返す。

*Generating laughter components and acoustic features with emotion manipulation. by KIMURA, Shunya, MORI, Hiroki (Utsunomiya University)

pleasantness: 4 arousal: 4	pleasantness: 5 arousal: 5	pleasantness: 6 arousal: 5
hu hu H hu hu he ha hu hu hu hu hN hN hN hN h hu hu	hu hu hu hu ha H hu hu hu hu hu ha h a ha hu ha H	hu hu H na hu hu hu hu hi e ha ha ha H hu H he h hu hu ha H hu H he hu u ha hu ha H hu H hu h hu hu hu ha H h
pleasantness: 6 arousal: 6	pleasantness: 6 arousal: 7	pleasantness: 7 arousal: 7
ha H h a ha ha H hu h hu he he hu H hi hu ha ha N h hu ha u hu hu hu ha ha H	hu wa ha H hi u hu hu hu ha H hi H hi h a ha H hi H hi H hu H H hu hu hu hu ha H hu u H h he hu H hu u ho ho ho ho ho ho ho ho ho ho ho ho H	hu ha a ha H hu h hu h hu ha H hu H hi hi H hu H hu u hu h H hi he he he he he h a ha H na ha H hu H he u hu ha H hi H hu H hu u H u ha H hu H hu H hi H hu h ha H u hu H

Fig. 2: 生成された call 系列の例

3.2 音響特徴量生成モデル

Call 系列から得られたコンテキスト [1] から、笑い声の音響特徴量を生成する。この際、感情次元による制御の有効性を評価するため、コンテキストとして感情次元 2 次元を付与する／しないの 2 種類のモデルを作成する。

モデル構造は、双方向 LSTM (ユニット数:512) 3 層+線形層 1 層である。笑い声合成モデルは話者 04.MSY の笑い声のみを用いて学習された。

4 感情次元評価実験

感情次元 (ple, aro) を、(4, 4), (4, 5), (5, 4), (5, 5), (5, 6), (6, 5), (6, 6), (6, 7), (7, 6), (7, 7) の 10 組に設定し、call 系列生成モデルによって計 200 個の call 系列を生成した。Fig. 2 に生成された call 系列の例を示す。系列の長さだけでなく、出現する call の種類にも違いがあることが見て取れる。

音響特徴量生成モデルへの入力として感情次元を付与しない条件 (phones) と、付与する条件 (phones+acoust) の 2 条件で比較する。被験者は 7 名で、ランダムな順番で呈示された合成笑い声を聞き、快-不快と覚醒-睡眠をそれぞれ 7 段階で評価した。

phones 条件での相関係数は $r = .252$ (快-不快)、 $r = .143$ (覚醒-睡眠) であり、いずれも相関は有意であった。このことから、感情次元による call 系列生成の制御は、合成笑い声から知覚される感情状態の制御に有効であった。また、phones+acoust 条件での相関係数は $r = .478$ (快-不快)、 $r = .466$ (覚醒-睡眠) であり (Fig. 3)、2 条件での相関係数の差はいずれも有意であった。したがって、call 系列だけでなく音響特徴量の生

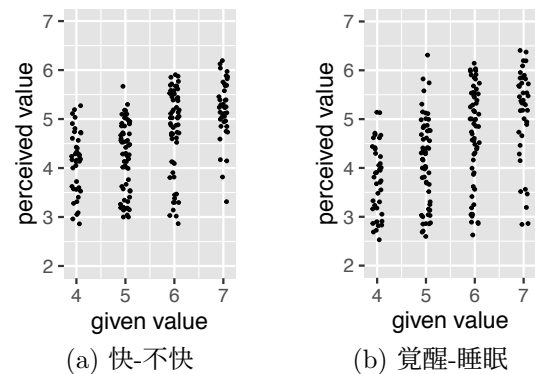


Fig. 3: phones+acoust 条件における感情次元の設定値と評価値の関係

成においても感情状態を入力とすることが有効であると言える。

5 おわりに

本研究では、感情次元による合成笑い声の制御について検討した。評価実験によって、call 系列および音響特徴量の感情次元による制御が、合成笑い声から知覚される感情状態の制御に有効であることが示された。

参考文献

- [1] Nagata and Mori, IEEE Trans. Affect. Comput., **11**, 553–559, 2018.
- [2] Bachorowski et al., Psychol. Sci., **12**, 252–257, 2001.
- [3] 森, 音講論 (春), 2-Q-43, 2015.
- [4] 今西, 小山, 有本, 音講論 (秋), 3-2-4, 2020.
- [5] 有本, 今西, 森, 情報処理学会論文誌 Vol.63 No.4, 1159–1169, 2022.
- [6] Bachorowski et al., J. Acoust. Soc. Am., **110**, 1581–1597, 2001.
- [7] Arimoto et al., Acoust. Sci. Tech., **33**, 359–369, 2012.