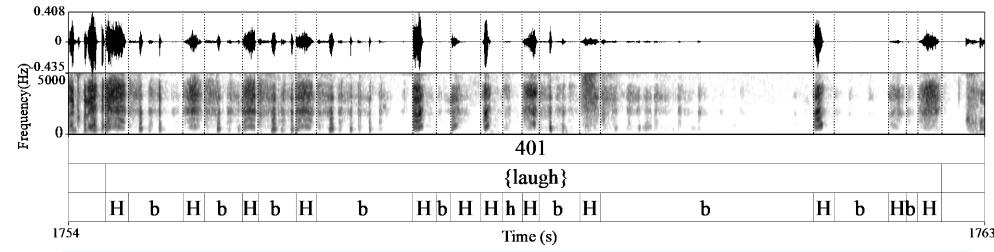
22-記-26

感情情報を利用した表現豊かな笑い声・叫び声合成

千葉工業大学 有本泰子

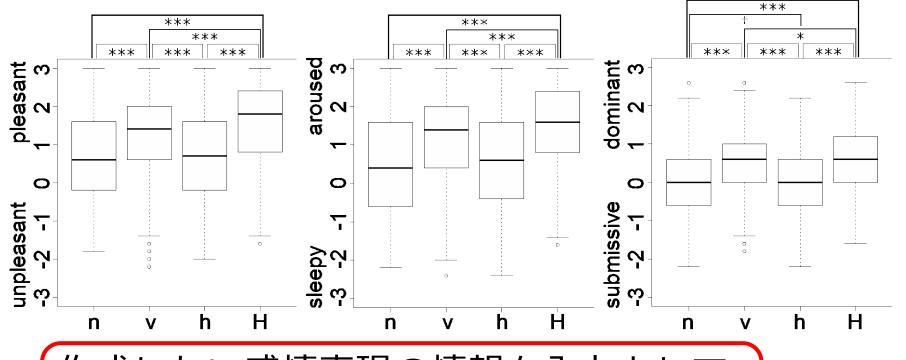
笑い声合成

- WaveNetによる笑い声合成[Mori+ 2019]
- DCTTS+Griffin-Limによる笑い声合成[Tits+2020]
 - 笑い声に音素を付与し、その音素を入力とする



そもそも笑い声の構成要素を音素として弁別可能か?

・ 笑い声の構成要素によって感情次元知覚特性が 異なる[今西2020]

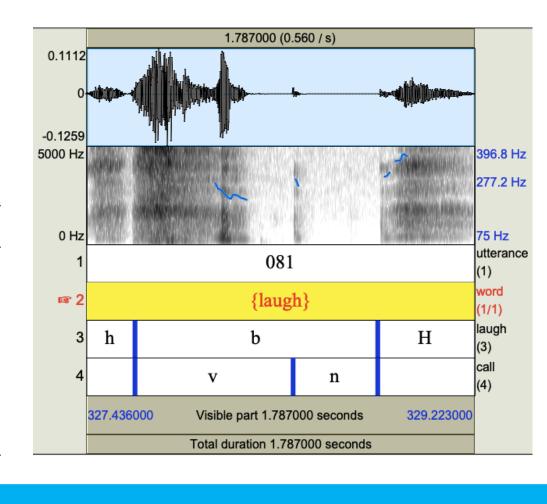


作成したい感情表現の情報を入力として, 笑い声合成を実現する

笑い声の構成要素

- 7名分の笑い声: 1,381個
- 笑い声に感情次元評価値を付与 [今西+ 2020]
 - Pleasantness
 - Arousal
 - Dominance
- ・4種類の構成要素

構成要素	個数
bout(b)	1,650
無声呼気(n)	1,309
有声呼気(v)	2,430
無声吸気(h)	1,308
有声吸気(H)	797



構成要素推定

- ディリクレ回帰による各構成要素の割合推定
 - ・目的変数1つ1つの値の合計が1なる

 $\{y_n, y_v, y_h, y_H\} = \beta_1 e_p + \beta_2 e_a + \beta_3 e_d + \beta_4 Dur + \varepsilon$ 笑い声の構成要素 感情次元値 笑い声の長さ

- 笑い声の長さおよび並びの決定
 - 同一の構成要素を複数含める場合

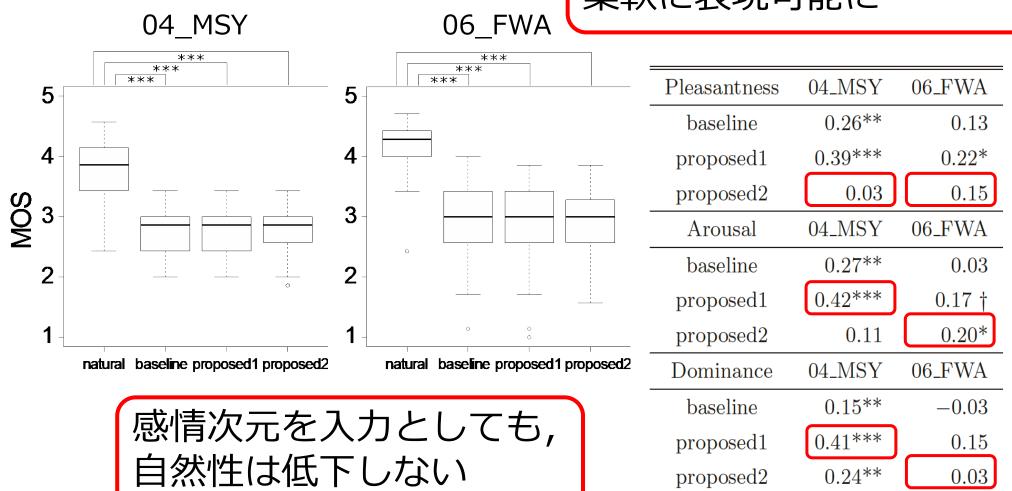
有声呼気の長さ:3.0s 2.0s 1.0s

各構成要素の継続時間長が平均±1σの範囲となるように分割

• 構成要素の並びはランダム

合成実験の結果と考察

4種類の構成要素で感情を 柔軟に表現可能に



- 合成器: WaveNet
- 学習データ:男性256個,女性237個
- 合成手法
 - baseline: 3種の構成要素[Mori+ 2019]
 - proposed1:4種の構成要素
 - proposed2:ディリクレ回帰 (4種の構成要素)

叫び声合成

- データ数:女性135個, 男性133個
- VITSで叫び声を合成
- single:特定話者モデルmulti:複数話者モデル
- proc: データ加工による拡張conv: 話者変換による拡張

合成実験の結果と考察

