

対話中に表出した笑い声の声質分析*

○有本泰子 (帝京大), 森大毅 (宇都宮大)

1 はじめに

ヘッドマウントディスプレイなどで VR を利用したゲーム開発が人気を集める中、ゲーム内のアバターとの一体感を向上させ、よりリアルなバーチャル空間を実現することが求められている。本研究では、対戦型ゲームのプレイ中に無意識に表出する感情を入力とし、アバターの状態を、プレイヤーの意図に関わらず強制的に変更（ヒットポイントの増減や対戦相手への攻撃、場面転換など）する「感情入力インタフェース」の開発を目指す。本システムにおける感情表出は音声に含まれる Affect Burst を対象とする。Affect Burst とは音声の非言語的使用による感情の急激な表出のことで、無意識な叫び声や笑い声、泣き声や、「ゲッ」「おお！」など言語音として意識的な制御下にあるものを含む [1]。プレイヤーが思わず発する様々な Affect Burst を、その音響的・言語的特徴から機械学習によって検出する仕組みを構築する。

本報告では、様々な Affect Burst のなかでも、まずは笑い声に着目し、笑い声を構成する吸気および呼気による要素の声質の特徴を音響的に明らかにする。将来の研究目標を念頭に、実環境における笑い声が収録されている感情評定値付きオンラインゲーム音声チャットコーパス (OGVC) [2] に対して、笑い声をラベリングした結果についても述べる。

2 笑い声ラベリング

OGVC の自発対話音声に表出する笑い声に対し、セグメンテーションと笑い声のラベリングを行った。7名分（男性5名、女性2名）の音声（合計7時間）をラベリング対象とし、音声分析ツール Praat を使用して笑い声ラベリングを実施した。

笑い声に対するラベリングの規準は、森らにより整備されたガイドラインに準じている [4]。笑い声の定義は Trouvain の定義に基づき、Episode, bout/吸気, および call の3層によって構成されるものとした。Episode とは、1回以上の吸気・呼気から構成される音響イベントであり、これが知覚的なひとかたまりの笑い声を形成する。bout/吸気とは、ひとつの Episode を構成する複数の吸気・呼気 (=bout) である。bout も吸気も、声帯振動を伴う場合と伴わない場合がある。call とは、ひとつの bout を構成する音響的要素である。現時点では、対象とする音声へのラ

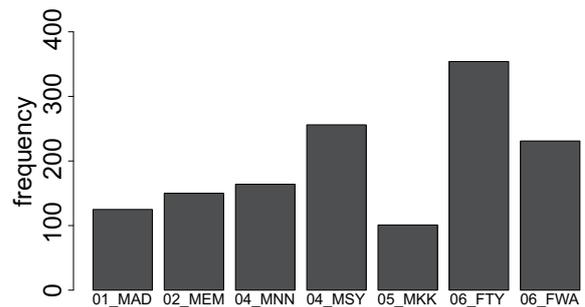


Fig. 1 Number of laughter episodes for each speakers

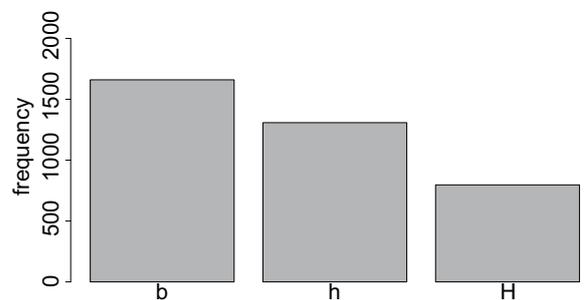


Fig. 2 Number of laughter components

ベリングは Episode レベルと bout/吸気レベルが完了しており、本報告では、笑い声の構成要素のレベル (bout (b), 無声吸気 (h), 有声吸気 (H)) における声質分析を行う。

話者ごとの Episode 数を Fig. 1 に、bout および吸気の総数を Fig. 2 に示す。7名分の対話音声中には 1,381 個の笑い声が含まれていた。また、bout と吸気などの笑い声の要素数は 3,765 であった。

3 声質分析

OGVC 音声ファイルは 48kHz で収録されているが、分析対象とする周波数領域を限定するため、10kHz にダウンサンプリングした。ダウンサンプリングした音声ファイルは、笑い声ラベルのセグメンテーションに従い、構成要素の単位に分割し、音声分析ツール Praat を用いて、声質に係る音響特徴量の抽出を行った。声質に係る特徴量として、第1フォルマント周波数 (F1)、第2フォルマント周波数 (F2)、MFCC の1次係数 (C1) を抽出した。分析対象のフレームは笑い声の要素の開始から 100ms とした。F1 および F2 は、25ms の分析フレームを 5ms ずつシフトさせ、

*Voice quality analysis of laughter components during spontaneous dialog. by ARIMOTO, Yoshiko (Teikyo University), MORI, Hiroki (Utsunomiya University)

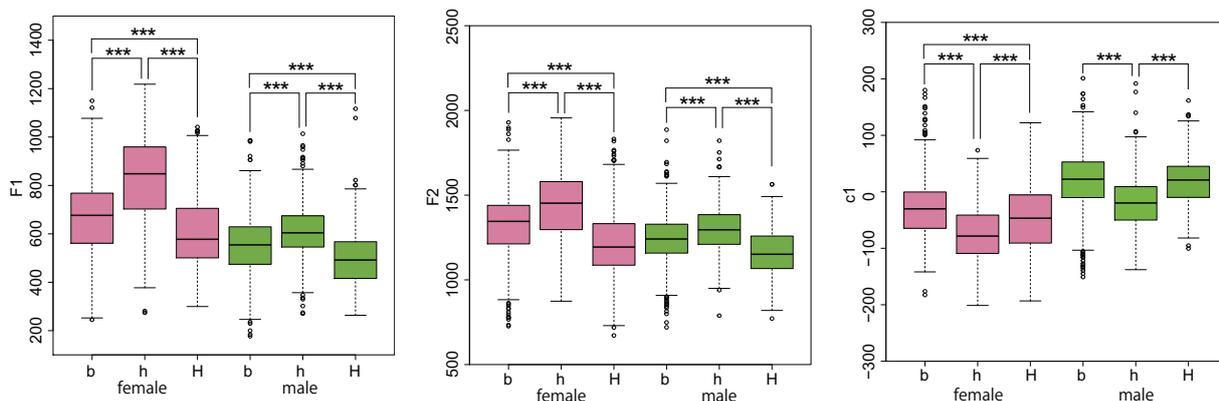


Fig. 3 Distribution of each voice quality parameter for each component (***) $p < 0.001$

LPC 分析により 16 次元の係数を求めて、抽出した。C1 の分析フレーム長は 15ms、シフト長は 5ms である。このとき、C1 が抽出できなかった音声があったため、その音声は分析から除外した。

4 呼気と吸気の声質比較

呼気による笑い声と、吸気による笑い声の声質を比較するため、求めた声質に係る音響的特徴量に対し、性別 × 笑い声の構成要素を要因とした二元配置分散分析を行った。また、その下位検定として、多重比較を Tukey の HSD 法で行った。結果を Fig. 3 に示す。

F1 における分散分析の結果、性別と笑い声の構成要素に主効果があった（それぞれ、 $F(1, 3759) = 1028.04, p < 0.001$ および $F(2, 3759) = 380.67, p < 0.001$ ）。多重比較の結果を見ると、男女ともに、b (b) および有声の吸気による笑い声 (H) の間で有意な差を認めた ($p < 0.001$)。F1 は口の開大に相当する特徴量である。この結果より、男女ともに、有声の吸気による笑い声の方が呼気による笑い声に比べ開大が狭い可能性がある。

F2 も F1 と同様、性別と笑い声の構成要素に主効果を認めた（それぞれ、 $F(1, 3759) = 336.29, p < 0.001$ および $F(2, 3759) = 298.97, p < 0.001$ ）。多重比較の結果も、男女ともに、b および H の間で有意な差を認めた ($p < 0.001$)。F2 は舌の前後の位置を表す特徴量である。この結果より、男女ともに、有声の吸気による笑い声の方が呼気による笑い声に比べ後舌よりである可能性がある。

C1 における分散分析の結果も、性別と笑い声の構成要素に主効果があった（それぞれ、 $F(1, 3759) = 1038.447, p < 0.001$ および $F(2, 3759) = 281.003, p < 0.001$ ）。多重比較の結果は、女性では b, h, および H の間で有意な差を認めた ($p < 0.001$) が、男性は b と h, h と H の間でのみ有意差を認めた ($p < 0.001$)。C1 はスペクトル

包絡の傾きに関する特徴量である。この結果から、無声の吸気による笑い声 (h) は、呼気・吸気に関わらず、有声の笑い声より高周波成分を多く含むこと、女性においては呼気による笑い声より有声の吸気による笑い声の方が高周波成分を多く含むことが分かる。また、女性の b および H は負の値を示していることから、有声であるにもかかわらず高周波成分を多く含むことが示唆された。

5 おわりに

対戦型ゲームをプレイ中に無意識に表出する感情を入力とし、アバターの状態を、プレイヤーの意図に関わらず強制的に変更する「感情入力インタフェース」の開発を念頭に、実環境で表出したリアルな笑い声の声質の特徴を音響的に分析した。

その結果、有声の吸気による笑い声は開大が狭いこと、有声の吸気による笑い声は呼気による笑い声より後舌よりであることが示唆された。さらに、無声の吸気による笑い声は、呼気・吸気に関わらず、有声の笑い声よりスペクトルの傾斜が急峻であること、女性においては呼気による笑い声より有声の吸気による笑い声の方が高周波成分を多く含むことが分かった。

謝辞 本研究の一部は、公益財団法人中山隼雄科学技術文化財団平成 29 年度助成研究 A-2、および JSPS 科研費 26280100 の助成を受けた。

参考文献

- [1] Schröder, *Speech Commun.*, 40 (1-2), 99–116, 2003.
- [2] Arimoto, *et al.*, *Acoust. Sci. Technol.*, 33(6), 359–369, 2012.
- [3] Trouvain, *Proc. ICPHS '03*, 2793–2796, 2003.
- [4] 森, 有本, 永田, *音講論 (秋)*, 217–218, 2017.