(招待講演) 会話エージェントは，いつ，どのように笑うべきか: ヒトの笑い声研究からの示唆 *

○森　大毅 (宇都宮大)

1　はじめに

スマートスピーカーの登場以来，人と音声でインタラクションを行う機械は見慣れた存在になった。さらに，ここ1年ほどの間の大規模事前学習モデルの発展は劇的であり，音声研究者の長年の目標であった，音声による人と機械の自然なインタラクションの実現もそう遠くないように思える。一方，大規模言語モデルの事前学習に利用されるテキストデータの大半は書かれた (written) ものであって，話し言葉に見られる諸現象を網羅しているわけではない。

書き言葉には見られない話し言葉の要素の1つに，笑い声などのノンバーバルな感情表出 [1] がある。AI という言葉に象徴されるような最新の会話エージェントでさえも，人間と同じように笑いを表出し，また会話という相互行為の中で笑いを利用する能力はまだ持っていない (乳幼児さえ持っているのに！)。

一方で，機械が笑う必要なんてない，という意見は根強くある。建設的な議論のためには，人がなぜ笑うか (=笑う必要があるのか)，いつ笑うか，どのように笑うかについての認識を共有するとともに，「機械に笑わせる」という研究目標をもっと具体化する，すなわち「何がどこまで達成できれば，機械に笑わせることができたと言えるのか」を明確にすることが必要である。

本講演では，今後の会話エージェントに求められる笑い声合成研究の目標を明確化すべく，人の笑い声研究を概観するとともに，そこから学ぶべきものについて議論する。

2　笑い声の起源

2.1　笑い声の進化論

明確に笑う動物は人間だけである。しかし，チンパンジー，ゴリラ，オランウータンなどの霊長類が笑いのような音を発することは既にダーウィンの頃から知られていた [2]。**2.2** で述べるように，この音は笑い声とは似ていても異なる。119名の学生にチンパンジーの音を聴かせ，何の音か記述させた実験では，笑いと回答したのは2名であり，最も多かった回答 (36名) は犬などの動物のあえぐ声であったという [3]。

比較・系統発生学研究は，笑いが進化的に獲得された生得的行動であることを示している。多くの霊長類は，取っ組み合いのような体を使った遊びの中で，プレイ・フェイスと呼ばれる，口を開けて下の歯を出した表情を呈する。大型類人猿，特にチンパンジー属の「くすぐり」や「追いかけっこ」遊びでは，プレイ・フェイスがしばしばあえぎ声のような発声として聴かれる激しい呼吸を伴う [2, 3]。笑いは，体を使った遊びの荒い息遣いから進化し，楽しい遊びに特殊化されたシグナルとしてステレオタイプ化したと考えられている。

2.2　ヒトとサルの笑い声の違い

人間とチンパンジーの笑いは重要な点において異なる [3]。まず，チンパンジーの笑いは速い。人間の笑いの call (**5** を参照) の立ち上がりの間隔がおよそ 210 ms であるのに対し，チンパンジーではおよそ 120 ms である。これは，人間が呼気のみによって笑うのに対し，チンパンジーが吸気と呼気の交替によって音を発するためである。また，人間の笑い call は徐々に小さくなって行くが，チンパンジーの場合には変化しないとされており，これも呼吸の違いによる。Provine は，ヒトが1回の呼息で断続気流を生成できる神経筋機構を備えるに至った背景にはヒトの二足歩行への進化があり，さらに言語音の生成能力の獲得とも関連していると述べている [3]。

2.3　情動発達と笑い声

乳児は生後1ヶ月ほどたつと，快なる社会的インタラクションにおいて笑顔を示すようになるが [4]，笑い声を出すようになるのはもっと遅い [5]。Kret らは，乳児が最初は呼気だけでなく吸気によっても笑い，発達とともに呼気の割合が増加して行くことを示した [6]。この説明とし

---

* When and how should conversational agents laugh? Implications from human laughter studies.
by MORI, Hiroki (Utsunomiya Univ.)

て，1つには乳児の発声機構が解剖学的および機能的にヒト以外の霊長類に近いこと，もう1つには世話をする人の笑う声の真似によって環境からのポジティブな反応を最大限に引き出すことを学ぶことを挙げている。

3　笑いの情動コミュニケーション

笑いは社会的である。人は，独りでいる時に比べ，誰かといる時の方が30倍多く笑う [3]。また，笑いには高い伝染性があり，誰かの笑いは別の誰かの笑いを誘発する。

情動伝染はヒト以外の動物にも見られる。ミラーニューロン，すなわち特定の行動で活性化するだけでなく，他者によるそれらの行動を知覚する際にも活性化するという特性を持つ視覚運動ニューロンは，最初サルの運動前野で発見された。笑いの伝染を，人間にも存在すると目されるミラーシステムによって理解しようとする試みがある [7]。

笑いは単なる快感情の表出ではなく，インタラクションにおける情動状態のマネジメントとしても使われる。笑いには，不快感情に伴うストレス反応を抑制する生理学的働きがある [8]。自然災害で深刻な被害を受けた人が，インタビューアに対して笑っているかのようにふるまう場面をよく目にするが，これも社会的な情動状態のマネジメントの例のように思われる。

4　笑い声の収集

笑い声を研究するためには，まず多くの笑い声を収集する必要がある。しかし，これは思ったほど簡単ではない。よくあるやり方は笑いを喚起する映像刺激を見せるもの [9, 10] であるが，Provine は，隔絶された実験室環境での収録はうまく行かないと主張する [3]。Provine はまた，過去の笑い研究がこのような観客指向の受動的笑いのみを対象としていたことを批判し，笑いが社会的ふるまいであること，そして実は話し手の方がよく笑うことを示している。

こうしたことを踏まえれば，笑い研究においても自発対話音声コーパスを研究の材料にすることは自然であろう。Fig. 1 に，アクションゲーム音声コミュニケーションコーパス (AGSC [11])，オンラインゲーム音声チャットコーパス (OGVC [12])，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 (UUDB [13]) 中に出現する笑い

Fig. 1　会話コーパス中の笑い声の数

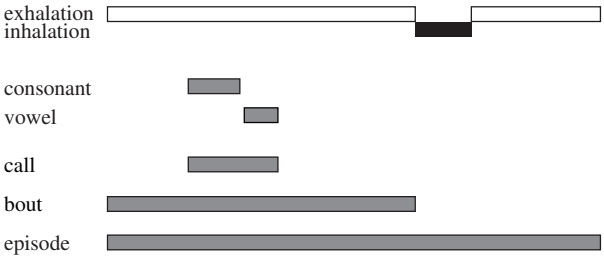

Fig. 2　笑い声の時間構造 ([14] を一部改変).

声の数を示す。UUDB は感情表現に富んだコーパスであるが，笑い声の数は多い話者でも 50 に満たない。一方，AGSC や OGVC のようなオンラインゲーム中の会話は，笑いを多く含む傾向があり，オンラインゲームは笑い声の収集に適していると言える。ただし，図からわかるように，個人差がたいへん大きい。

5　笑い声の形態学

笑い声の構造は，Fig. 2 のように階層的に理解されている [14]。笑い声は，1回以上の呼気／吸気に対応する音響イベントからなり，これらが知覚的なひとかたまりを形成する。音声コーパスの転記では<笑>などと記述されるこのまとまりを episode と呼ぶ。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 [9]。例えば，「あはは」という bout は3つの call からなっている。人間の笑いは呼気のみから構成されると言われている (**2.2** を参照) が，呼気による主要な音の後には，しばしば吸気音が聞かれる [10]。一部の吸気音は，前後の bout と一体となって笑い声を特徴づけていると考えられるが，多くの音声コーパスではこれらの吸気音を笑い区間に含めていない。著者らは，bout に先行または後続する吸気音を生成・知覚の面から重要だと考え，笑い声研究に用いるコーパスのラベリングマニュアル策定に反映させている [15]。

人間は呼気のみで笑うというのは本当だろう

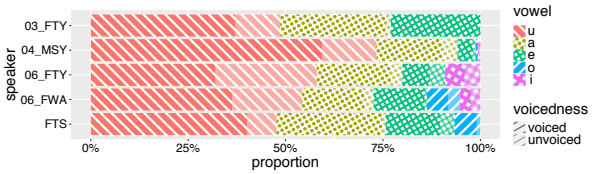

Fig. 3　各 call の母音の割合

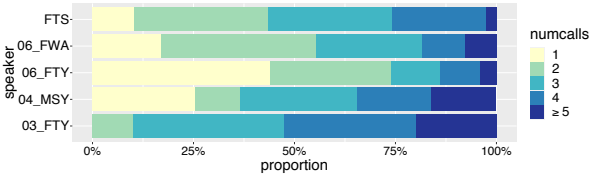

Fig. 4　各 bout を構成する call 数の割合

か。笑いの吸気音はしばしば声帯振動を伴い (有声)，中にはそれ自身が主要な笑い声を形成しているとみなせるものがある (いわゆる引き笑い)。OGVC 中の笑いが多い話者 7 名について，先頭あるいは末尾以外に有声吸気音を含む episode の割合を調べたところ，うち 2 名は 1%前後，3 名は 10%前後，残り 2 名はそれぞれ 21%, 27%であった。オンラインゲーム中のように感情的な覚醒度が高い場面では，ほとんど呼気だけで笑う人と，吸気笑いを多用する人の両方がいることがわかった。今西らは OGVC を対象とした研究で，有声吸気音を含む笑い声はより快寄り・覚醒寄りに知覚されることを示している [16]。吸気笑いは動物や乳児の笑いの特徴であった。このような笑い方をする人が，あらゆる社会的場面で同じように笑うことは少し考えにくく，日常会話場面での笑い声の分析が望まれる。

Fig. 1 中の 5 名 (04_MSY は男性，他は女性) について，各 call の母音の割合を Fig. 3 に示す。笑い声の構成音は，「あはは」のような a 系よりも u 系の方が多いことがわかる。ただし，これらの違いは明瞭ではなく，実際にはどれも中舌中央母音 [ə] 付近で実現する [9]。

三省堂国語辞典には，第七版より「あはは」「うふふ」「おほほ」に加え「いひひ」「えへへ」が採録された。私たちは，実際にこのように笑っているのだろうか。各 bout の call 数の割合を Fig. 4 に示す。話者にもよるが 2 call 以下が多数派となっている。a 系 call の出現頻度が次点であること (Fig. 3) も併せて考えると，「あはは」のように笑うことは実はそれほど多くないのである。ついでに，三省堂国語辞典の「うふふ」の語釈「口を閉じたまま軽く笑う声」には第七版から「(女性が)」が括弧書きで追加されたが，u 系 call が最多であるのは男性話者でも同じである。

Fig. 4 はまた，単一 call の bout が意外に多いことも示している。このような call は無声の割合が 35%と大きい (全体では 20%)。笑い声の有声/無声の別は喚起される快-不快に関連することが知られている [17]。(声帯を振動させないで)「フッ」と笑うのが適切な場面もあるだろう。

6　会話の構成要素としての笑い

会話分析の基本的な方法に，話し手の発話と聞き手の反応の間の関係，すなわち連鎖構造の分析がある。笑いが連鎖構造の中で果たす役割を知ることは，会話エージェントがいつ笑うべきか，また人間の笑いに対して会話エージェントがどのようにふるまうべきかについてのヒントをもたらしてくれる。

笑いは隣接ペア第 1 部分 (先行要素。「質問」「依頼」など) にも第 2 部分 (後続要素。「答え」「拒否」など) にも含まれ得る。UUDB の観察では，第 1 部分に笑いが含まれている場合に，第 2 部分で第 1 部分の一部が繰り返されていることが多いようだ。

また，笑いだけで隣接ペア第 2 部分を構成する場合がある。UUDB にある，一方の話者が「年賀状は (D なに)(D どう) どうつながるのこれ {laugh}」と問いかけている箇所を例に挙げる。もう一方の話者には，この第 1 部分への応答義務が生じているが，この話者にもまだ謎が解けていない。もちろん「わからない」と明示的に答えることもできたわけだが，この話者は言葉は発さず，第 1 部分に含まれる笑いに呼応する形で笑いだけを発している。この笑いを第 2 部分とみなすべきか，それとも回答を保留しているだけとみなすべきかはやや議論があるところであるが，少なくとも第 1 部分に対する応答の不在を避ける目的ないし機能があることは確かだろう。

隣接ペア第 2 部分の後方に，第 2 部分に対する了解や評価などによって連鎖を終わらせるような発話がなされる場合があり，連鎖終結の第 3 部分と呼ばれる。第 2 部分の後の笑いはしばしば連鎖終結の機能を持つ。会話分析では，考えた末の回答が短すぎ，ちょっと気まずい場面で笑うことによって，もう話すことがないことを示す例が知られている。

隣接ペアとは異なるが，話し手が聞き手の笑

いを誘う場面がある。話し手は，何かを話した後に自ら笑うことによって「ここは笑うところ」だと示し，聞き手の笑いを引き出す [18]。笑いの共有は人同士のつながりを示すシグナルであり，会話エージェントにとっても重要である [19]。

実際には，笑いを伴う連鎖は，条件的に適切な応答として笑いを期待した話し手の意図によるものよりも，むしろ聞き手が話し手の発話に笑うべきところを発見し，実際に笑った結果として，笑われるべき (laughable) 発話であったことが事後的に決まるものの方が多い。実際，笑いに先行する発話それ自身が何か滑稽な内容であることは少ない。Provine は，市中の会話を観察して得た 1200 個の笑いに先行する発話のうち，かろうじて滑稽だと思われたのは 10 から 20 パーセントであったと述べている [3]。

7　笑う会話エージェントに向けて

音声合成研究で利用される音声資源は，音韻バランス文に代表される脈絡のない文の集合である。笑い声合成研究でも，笑い声を脈絡なく集めたデータセット [10, 20, 21] が用いられてきた。会話中の笑い声を用いた合成研究 [22, 23] も存在するが，評価は個々の合成笑い声を対象としている。だが，ここまで見てきたように，笑いは社会的文脈の中に存在する。「いつ，どのように笑うべきか」の評価を社会的文脈から切り離された笑いに対して行うことには，やはり無理がある。

文脈が笑い声の巨視的形態に与える影響は，生成モデルにより再現できる [23]。また，笑い声を言語音と共に一連のテキストにして音との対応を学習する end-to-end モデル [22] の枠組は，合成笑い声 (と，言語音) の微視的形態に文脈を自然に反映することができる。しかし実際には，生成モデルとテキスト音声合成が解決できるのは「どのように笑うべきか」までである。これに比べ，「いつ笑うべきか」は難問である。大規模言語モデルがこの難問を解決するか？　今のところ，著者は否定的である。

ここまで見てきた笑いの生成と機能の生物学的・心理学的・社会学的側面を考慮し，どのようなテストを設計すれば，社会的文脈の中で合成笑い声の良さを評価できるのか。これからの重要な課題である。

参考文献

[1] 森, 音講論 (春), 405–406, 2015.
[2] Darwin, “The Expression of the Emotions in Man and Animals,” John Murray, 1872.
[3] Provine, “Laughter: A Scientific Investigation,” Viking, 2000.
[4] Messinger et al., Dev. Sci., **5**, 48–54, 2002.
[5] Washburn, Genetic Psychology Monographs, **6**, 403–537, 1929.
[6] Kret et al., Biol. Lett., **17**, 1–25, 2021.
[7] Gervais and Wilson, Q. Rev. Biol., **80**, 395–430, 2005.
[8] Scott et al., Trends Cogn. Sci., **18**, 618–620, 2014.
[9] Bachorowski et al., J. Acoust. Soc. Am., **110**, 1581–1597, 2001.
[10] Urbain et al., Proc. LREC 2010, 2996–3001, 2010.
[11] Mori and Kikuchi, Proc. Interspeech 2020, 3132–3135, 2020.
[12] Arimoto et al., Acoust. Sci. & Tech., **33**, 359–369, 2012.
[13] Mori et al., Speech. Commun,, **53**, 36–50, 2011.
[14] Trouvain, Proc. ICPhS ’03, 2793–2796, 2003.
[15] 森, 有本, 永田, 音講論 (秋), 217–218, 2017.
[16] 今西，小山，有本，音講論 (秋)，787–788, 2020.
[17] Bachorowski and Owren, Psychol. Sci., **12**, 252–257, 2001.
[18] Jefferson, in “Everyday Language: Studies in Ethnomethodology,” Irvington, 1979, 79–96.
[19] Inoue et al., Front. Robot. AI, **9**(933261), 1–11, 2022.
[20] El Haddad et al., in “Statistical Language and Speech Processing,” Springer International Publishing, 2017, 229–240.
[21] Xin et al., Proc. Interspeech 2023 (to appear).
[22] Mitsui et al., Proc. Interspeech 2022, 2328–2332, 2022.
[23] Mori and Kimura, Proc. Interspeech 2023 (to appear).