

複数の会話コーパスを対象とした笑い声イベントのアノテーション*

○森 大毅 (宇都宮大), 有本 泰子 (帝京大/理研), 永田 智洋 (宇都宮大)

1 はじめに

笑い是对話参加者の心理的・認知的状態を伝達する社会的シグナルである。機械による笑い声の形態的特徴の認識や、様々な形態の笑い声の合成は、音声による人間と機械のコミュニケーションに新たなチャンネルを導入する点で重要な意義を持つが、その研究のためには笑い声のデータが必要である。

大規模なコーパスに対する信頼性の高いアノテーションを実現するためには作業マニュアルの整備が欠かせない。しかし、音声言語と異なり、笑い声のアノテーションにはその基盤となる理論があまりなく、参考となる作業マニュアルも存在しない。そこで我々は、過去のコーパス構築の事例 [1] にならい、これまでの笑い声アノテーション [2] の経験を基に定めた当初仕様に従って作業を進め、発見された問題に対処する過程で仕様を修正して行く試行錯誤的な方法で作業マニュアルを作成してきた。本論文では、この過程で顕在化した笑い声のアノテーションにおける問題点を示すとともに、策定した作業ガイドラインを述べることで笑い声アノテーションの標準的方法の確立を目指す。

なお、笑い声の分節音 (call) レベルのアノテーションについては稿を改めて取り上げることにする。

2 笑い声の構造とアノテーション

2.1 Episode, bout, call, 吸気/呼気

本研究では, Trouvain [3] の図式に従い笑い声を階層的に記述する (図 1)。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 [4]。

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 [5]。一部の吸気音は, 前後の bout と一体となって笑い声の特徴づけている。

2.2 笑い episode 区間の同定

UADB など既存の対話コーパスの多くは, 笑いに付随する吸気音を発話区間に含めていない。このため, 笑い声の後に言語音が続く場合と, 笑い声で発話

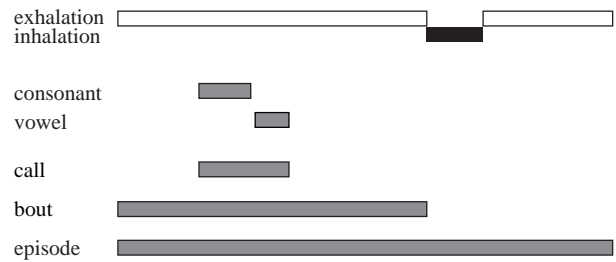
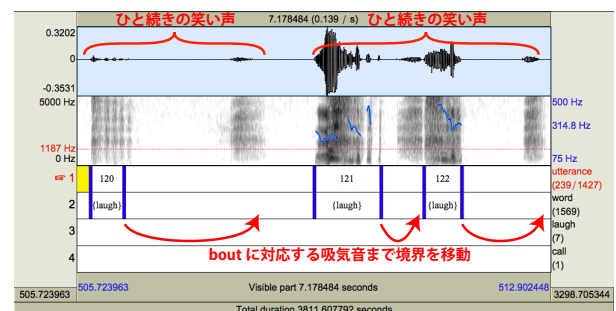


Fig. 1 笑い声の時間構造 ([3] を一部改変)

が終わる場合とで, 吸気音が笑い声に含まれるか否かが異なる問題があった。そこで本ガイドラインでは以下のように記述した。

笑いに付随する吸気音は, それが一と続きの笑い声を形成していると判断できる限り, episode に含める。

吸気音を episode に含めることにより, 既存のコーパスでは発話区間の変更が多数生じることとなる。



本研究では言語音と独立した笑い声のみを対象とし, 笑いながら言語音を生成している部分 (speech-laugh) は episode に含めない。言語音と bout が切れ目なく続いている場合は, その bout の最初の call の開始時刻を episode の開始時刻とする。また, 当初は episode は 1 つ以上の bout を含むと定義したが, これでは speech-laugh の直後の吸気音を笑い声の一部として記述することができない。そこで, このような吸気音は単独で episode を形成し得ることとした。

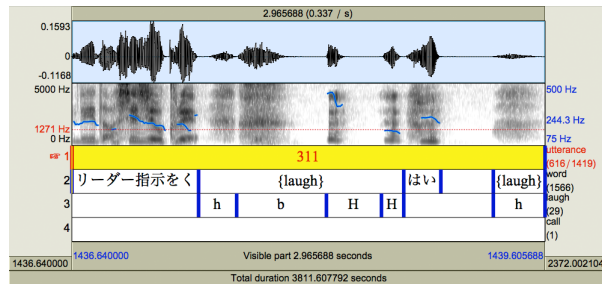
2.3 無音区間

本研究が対象としている会話コーパスでは, 400 ms 以上の無音区間が発話の境界となっている (ただし語中の促音は 400 ms 以上の無音でも発話境界ではな

* Laughter event annotation for different conversational speech corpora. by MORI, Hiroki (Utsunomiya University), ARIMOTO, Yoshiko (Teikyo University / RIKEN), and NAGATA, Tomihiro (Utsunomiya University)

い)。一方、発話内部の分節化された無音をポーズと呼んでいる。

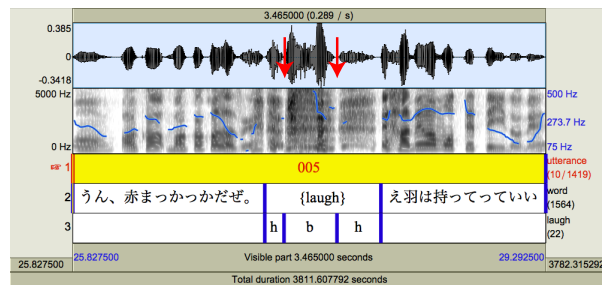
笑い episode は bout と吸気音からなるが、ポーズは含まないものとした。これは、ポーズの認定は作業効率を著しく下げるだけでなく、経験上、笑い声の形態を規定する上でポーズの有無は本質的でないと予想したためである。Bout の直後の呼気を使い切って無音になる区間は、bout の一部とすることとした。この無音区間は 400 ms を超える長さを持つことがある。同様に、吸気音の直後の息を吸い切って無音になる区間は吸気音の一部とすることとした。



2.4 呼気音と吸気音の判別

図 1 に示したように、bout は call すなわち呼気による音から構成される。逆に、吸気による音は bout ではない。これは、笑い声のアノテーションにおいて呼気音と吸気音を区別する必要があることを意味する。これら 2 種類の発声は生成機構において顕著な違いを有する [6] が、音響的な違いは顕著でない場合が多く、正確に判別するのは困難である。

吸気音 (下図 h) と呼気音 (下図 b) の間には、しばしば特徴的な非常に短時間の低パワー区間が見られ、これを吸気音と呼気音の判別に利用することができる。



2.5 顕著な吸気音

笑い声の中では、吸気音はしばしば声帯振動を伴う。応用上重要と考えられるため、無声 (h) と有声 (H) の吸気音は区別して記述することとした。

吸気音の中には、呼気による主要な音の後に付随するものだけでなく、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)。典型的には、複数回の吸気音が連続してひと続きの笑いを構成する。このような顕著な吸気音と、bout に付随する吸気音の判別は困難であり、区別なく記述することとした。

Table 1 作業員 1 との笑い声アノテーション一致度

	episode		bout		吸気音	
	R	P	R	P	R	P
作業員 2	0.98	0.97	0.83	0.69	0.76	0.88
作業員 3	0.78	0.98	0.76	0.75	0.51	0.88

3 作業マニュアルの有効性検証

作成した作業マニュアルにより複数作業員が一致した笑い声アノテーションを行えることを検証するため、笑い声を豊富に含む以下のコーパスについて笑い声のアノテーションを試行している。

- UADB
- OGVC
- アクションゲーム音声コミュニケーションコーパス (AGSC) [7]

本稿では、このうち OGVC の女性話者 1 名の音声 (約 24 分) に対して行った笑い声アノテーションの一致度を示す。作業員 1 が付与した episode, bout, 吸気音のラベルを ground truth とし、別の作業員 2 名が付与した各ラベルの再現率 (R) および適合率 (P) を表 1 に示す。作業員 1 が付与したラベルの個数は、episode 106 個, bout 89 個, 吸気音 (有声+無声) 178 個である。Episode, bout のアノテーションについては作業員間で比較的一致していることがわかる。Episode の不一致の主な原因の 1 つは、speech-laugh の直後に微弱な息の音がある場合に、そこを笑い声の一部とするかどうかの判断の揺れであった。また、吸気音の再現率が比較的低い。この主な原因は、微弱な吸気音を笑いの一部とみなすかどうか、また「クックッ」などの微弱な音を吸気音と判断するかどうかの判断の違いであった。

4 おわりに

笑い声アノテーションの作業ガイドラインを作成し、作業員間一貫性を検証した。今後は不一致の原因となる箇所の記述を改良し、合成などの応用において信頼できる箇所の選択的使用を可能にしたい。

謝辞 本研究は JSPS 科研費 26280100 の助成を受けている。

参考文献

- [1] 国立国語研究所編, “日本語話し言葉コーパスの構築法,” 国立国語研究所, 2006.
- [2] 森, 音講論 (春), 2-Q-43, 2015.
- [3] J. Trouvain, Proc. ICPhS 03, 2793–2796, 2003.
- [4] Bachorowski et al., J. Acoust. Soc. Am. **110**, 1581–1597, 2001.
- [5] Urbain et al., Proc. LREC 2010, 2996–3001, 2010.
- [6] K.-I. Sakakibara, ICVPB 2012, 2012.
- [7] 菊地, 森, 音講論 (秋), 3-P-31, 2017.