感情音声の研究を始める人のための音声コーパス入門*

○森　大毅 (宇都宮大)

1　はじめに

「感情音声コーパス？　喜び、怒り、悲しみの感情をこめて、声優さんに文章を読んでもらえばいいんだよね？」

音声から感情を認識したい。合成音声で感情を表現したい。感情音声の研究はますます多くの関心を集めるようになっている。

ことばと異なり感情それ自体は測定できない (仮説的構成概念)。このため、感情音声研究のための音声コーパスの構築および利用においては、ことばそのものの処理にはなかった方法論上の問題に対応する必要がある。

本講演は、音声が伝達する感情に関心を持ってきた講演者が、自らの経験および関連研究の調査に基づき、この分野の研究に関心を持つ人を対象に、コーパスの構築および利用において注意すべきと考える点をまとめたものである。

2　目的を明らかにする

「感情音声」という語が指す概念は、個々の研究によって異なる。逆に、研究の目的がはっきりすれば、その枠組の中で感情とは何かを定義でき、コーパスに求められるものも明確になる。

感情音声の研究には、3 つの典型がある。

① 感情音声の音声学／心理学
音声の音響的特徴が感情の影響をどのように受けるか (生成)、および音声から伝わる感情が音響的特徴の影響をどのように受けるか (知覚)。網羅的でなく少数の特定の感情に注目したものが多い。

② 音声からの感情の認識
音声の音響的特徴を説明変数、話者の感情を目的変数としたパターン認識。(1) の知覚を模倣する機械の実現ということもできる。

③ 感情を伝達する音声の合成
(1) の生成を模倣する機械の実現。話者の感情を説明変数、音声の音響的特徴を目的変数としたパターン生成 (回帰) ということもできる。

認識および合成の問題設定において重要なのは、あなたの目的を達するために、伝統的な心理学研究と同じように特定の感情だけ扱えばよいのか、それとも網羅的でなければならないか、を見極めることである。例えば、音声から相手が怒ったことを検出したいという場合には、コーパスには「怒り」に関する情報が記述されていればよい。そうでなく、相手がどんな感情状態なのか認識したいという場合には、コーパスにはあらゆる感情状態を記述できるような枠組が必要である。

3　感情を定める

感情心理学の主流派は、普遍的で文化に依存しない感情 (＝基本感情) が存在すると考えている。基本感情の中でも、「怒り」「喜び」「悲しみ」「驚き」「恐れ」「嫌悪」はこれまでの感情研究の中で繰り返し取り上げられている。表情を中心とする心理学研究では、感情喚起または刺激により特定の感情を誘発する方法を取ることが多い。

特定の感情を対象とした感情音声研究では、これに類似の方法を採るか、対象となる感情を演技した音声を収録する。この場合、感情の記述は自明となるが、事後の主観評価 (5 節) は必要である。

基本感情は感情の全てではない。コミュニケーションに必要な感情は、上記 6 感情よりもはるかに多様だと考えられる。コミュニケーションツールとしての感情生成では、表現すべき感情の種類を決定することがまず課題となる。(SNS 等で使われる感情絵文字 emoticons は 80 種類。)

一方、感情の種類を決めない方法もある。感情心理学には、感情のカテゴリーは感情空間における布置を示すだけであり、重要なのはその空間を定義する次元であるとする考え方がある。感情の記述には、valence (感情価, 快さ) と activation (活性, 覚醒度) の 2 次元が標準的に用いられる。第 3 の次元として dominance (支配性) を加える

---

*Beginner's guide to speech corpora for the studies of speech and emotion.
by MORI, Hiroki (Utsunomiya Univ.)

ことも多い。自発音声の感情を網羅的に評価するときは、用意した感情カテゴリーの中にぴったり当てはまるものが存在しないことも多いため、次元による感情記述が有効である。

4　音声を入手する

感情音声コーパスには、感情を演技したもの (例: JTES, 声優統計コーパス, OGVC Vol.2) と、コミュニケーション場面で自然に生じた感情を収録したもの (例: UUDB, OGVC Vol.1) がある。

声優が感情豊かに演じるアニメのキャラクターの声を合成音声で再現することが目的ならば、声優が感情を演技した音声コーパスが必要となる。反面、そのようなコーパスは自発音声の感情認識には向かない。感情を演技した音声の性質は, 自発的な感情表出とは異なるからである [1]。認識の対象とすべき感情を含む自発音声の入手が難しい場合には、かわりに演技音声のコーパスを利用する選択肢もあるが、実運用では満足な認識性能が出ないおそれがある。

目的に合致した感情音声コーパスがない場合には、自分で作ることもあり得る。設計において重要な点は、コーパスの利用目的を明らかにすること (2 節)、目的に合致した感情記述の枠組を決めること (3 節)、目的に合致した感情表出 (演技／喚起／自発) および収録の方法 (話者への指示) を決めることである。

架空の単語や感情的中立文を感情を込めて朗読するという行為は、冷静に見れば異常である。感情を演技させる場合には、複数人によるロールプレイ、感情と整合した文などの工夫により、収録した感情音声が現実離れしたものとならないよう注意すべきである。

5　感情を評価する

話者の感情は直接測定できないため、主観評価が用いられる。音声からの感情の評定では、刺激呈示方法、言語情報、作業者の特性など多くの剰余変数が影響するため、条件の統制が重要である。特に、作業者へのインストラクション (感情の理論、記述すべき感情の説明、音声のどの側面を評価するか、など) はぜひ文書の形で作成しておきたい。

感情評定の安定性は個人差が大きいため、作業者の事前スクリーニングが望ましい。同一の刺激に対する一貫性、評定の多数決に対する一致率などが基準となり得る [2]。また、最近はクラウドソーシングによる感情評定も現実的な選択肢である。信頼できる作業者により評定済の発話を作業対象に忍ばせておくことで、作業者のスクリーニングを動的に行うことも可能である [3]。

6　感情ラベルを利用する

感情音声コーパスを統計処理または機械学習にかける際は、複数作業者の評定結果を代表値に集約することが多い。

感情カテゴリーラベルが与えられている場合には、多数決により 1 つの感情に決めてしまうのが標準的である。過半数の感情カテゴリーが存在しなかった発話はデータから除外することが多い。OGVC Vol.1 では、10 カテゴリの感情ラベルに対し、3 名の作業者の完全一致率は 11.0%, 2 名以上一致率は 58.5%である。

感情次元ラベルが与えられている場合には、評定平均値を使うのが単純明快である。

RECOLA データベースは感情次元ラベルを持つコーパスであるが、UUDB のように発話単位のラベルではなく、作業者が映像を見ながらスライダーを操作することにより各次元の値を連続的に評価したものとなっている。操作の遅延時間は作業者に依存するため、単純に平均を取ると情報が失われてしまう。音響的特徴と評定値との相関を最大にする遅延時間を作業者ごとに求め、感情ラベルの時間整合を取ってから平均を求めることで、ラベルの持つ情報を最大限に生かすことができる [4]。

7　おわりに

研究の枠組は、ある程度コーパスにより決まってしまう現実がある。しかし、どのような目的にも使える感情音声コーパスは存在し得ない。

感情研究は、もっと多様な現象に目を向けて行かねばならない。「感情音声とは何か」の常識にとらわれない、新しい研究の土壌となるようなユニークなコーパスの開発に期待したい。

参考文献

[1] Douglas-Cowie et al., Speech Commun. **40**, 33–60, 2003.
[2] 森, 相澤, 粕谷, 音響学会誌 **61**, 690–697, 2005.
[3] Burmania et al., Proc. Interspeech 2017, 152–156, 2017.
[4] Mariooryad and Busso., IEEE Trans. Affect. Comput., **6**, 97–108, 2014.