

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 11 日現在

機関番号：12201

研究種目：基盤研究(B) (一般)

研究期間：2014～2017

課題番号：26280100

研究課題名(和文) Affect burst 音声対話における無意識な感情表出の分析および合成

研究課題名(英文) Affect burst: Analysis and synthesis of unconscious exposition of emotion

研究代表者

森 大毅 (Mori, Hiroki)

宇都宮大学・工学(系)研究科(研究院)・准教授

研究者番号：10302184

交付決定額(研究期間全体)：(直接経費) 9,700,000円

研究成果の概要(和文)：(1) 叫び声を誘発しやすいオンラインゲームをプレイする状況のコーパスを開発した。このコーパスには既存コーパスの10倍以上の頻度で叫び声が含まれている。叫び声の音響分析により、通常語彙や感動詞との音響的特性の違いを明らかにした。
(2) 感情表出系感動詞の形態を分類し、多様な形態を持つ「あ」を合成した。合成音声を用いた知覚実験により、形態とパラ言語情報との関係を明らかにした。
(3) 自然対話コーパスから笑い声の構成要素の変動要因を明らかにするとともに、コーパスベース音声合成を応用した多様な笑い声合成を実現した。知覚実験により、定義した変動要因を考慮することにより自然性が向上することがわかった。

研究成果の概要(英文)：(1) A multimodal corpus of gaming interaction that easily induces interlocutors' shouts was developed. This corpus contains more than ten times of shouts compared to existing corpora. Analysis of the shouts revealed the acoustical differences to regular words or interjections.
(2) A taxonomy of expressive interjections was developed, which enabled the synthesis of the interjections "a" with various forms. A perceptual experiment using the synthesized interjections revealed the relationship between the forms and paralinguistic information.
(3) Factors that affect the acoustical properties of laugh calls were identified. Incorporating these factors into the definition of context for the framework of the HMM-based speech synthesis enabled a flexible laughter synthesis. A perceptual experiment revealed the advantage of incorporating these contextual factors with respect to the naturalness of synthesized laughter.

研究分野：音声言語情報処理

キーワード：笑い声 叫び声 感情表出系感動詞 パラ言語情報 合成 知覚 対話音声

1. 研究開始当初の背景

Affect burst (感情の噴出)とは、表情や声によるノンバーバルで急激な感情表出のことであり、不随意的な笑いはその典型例である。Affect burst には対話参加者の心理的・認知的状态を伝達する社会的シグナルとしての側面があり、人間と音声でインタラクションを行う機械にも、やがて affect burst の生成および認識能力が求められることになると考えられる。しかし、本研究の開始以前、その基礎となる言語的・パラ言語的および語用論的性質はほとんど何もわかっていなかった。

研究代表者は、音声で伝達する感情・意図・態度などのパラ言語情報の重要性に着目し、その基礎資料としての音声対話コーパス(UUDB)を構築・整備・公開するとともに、自然(spontaneous)な対話音声で伝達するパラ言語情報の音響的特性と知覚との関係を解明してきた。また、このコーパスに基づく表情豊かな対話音声合成の研究に取り組んできた。これらの一連の研究の中でも感情表出系感動詞の多様性が改めて明らかになったが、言語音ではない笑い声や叫び声は分析の対象とはなっていなかった。また、笑い声や叫び声を合成することはできず、感情表出系感動詞のパラ言語的多様性を表現することはできていなかった。

2. 研究の目的

本研究は、UUDB、OGVC など複数の自発音声対話コーパスの横断的分析により、affect burst の構造および機能を解明し、対話音声合成へ応用することを目的とした。

この目的を達成するため、以下の研究目標を設定した。

- ・ Affect burst を類型化し、日本語における affect burst のレキシコンを構築する
- ・ Affect burst の音響的特徴および聴覚的声質との関係を分析する
- ・ Affect burst の存在やその特徴が話者の状態の知覚に与える影響を分析する
- ・ 多様な affect burst を含んだ対話音声合成を実現する

3. 研究の方法

(1) Affect burst の類型化

言語／非言語現象の類型化のための観点には、形態的側面と機能的側面がある。本研究では、まず形態的側面からの類型化を試みた。

① 感情表出系感動詞の形態的記述

UUDB に含まれる感情表出系感動詞「あ」694 例を対象に、パラ言語的な違いに着目し聴覚的に形態的分類を行った。形態的特徴の記述は、強-弱、長-短、口音-鼻音の各軸により行った。

② 叫び声コーパスの収録と叫び声の同定

本研究において、叫び声は、感情表出系感動詞の中でも話者の制御下にある度合いが低く、言語音としての様式化の程度が低いものと定義した。この定義に従って既存コーパス中の叫び声の同定を行った結果、UUDB および OGVC に含まれる叫び声はそれぞれ 4 例、53 例となり、予想よりも少ないことがわかった。このため、叫び声を誘発しやすいタスク設定、および対話の収録を行った。具体的には、分離された防音室内で学生同士がオンラインゲームをプレイする状況のコーパスを収録した(図 1)。また、収録したコーパスに対し、上記の定義に基づき叫び声の同定を行った。



図 1 叫び声コーパスの収録

③ 笑い声の形態的記述

対話中に生じる笑い声の統計モデルを構築するため、UUDB および OGVC に含まれる笑い声の詳細なセグメンテーションを行った。笑い声は、音節に対応する call, 1 つ以上の call から成り 1 回の呼吸に対応する bout, bout および吸気の連続からなる 1 連の笑いエピソードのように、階層的構造を有する。各笑いエピソードに対して bout/吸気および call レベルのセグメンテーションを行った。各 bout の特徴は、「ハ」「フ」などの日本語の音節による表記を基本にし、有声/無声、鼻音、引き伸ばしなどのバリエーションを補助記号で記述することにより行った。

(2) Affect burst の音響的／聴覚的分析

① 感情表出系感動詞の形態とパラ言語知覚との関係

形態の違いを反映して合成した感情表出系感動詞「あ」を被験者 8 名に呈示し、知覚されるパラ言語情報を「理解・同意」「驚き」「落胆」「気づき」「戸惑い」「意外」「聞き返し」「否定的態度(不満・非難・嫌悪)」「フィラー」から選択させる実験を行った。

② 叫び声の音響的特徴の分析

収録したコーパスに含まれる叫び声 (1412 例) の音響的特徴 (継続時間、F0 平均値、強度最大値、スペクトル傾斜、基本周期ゆらぎ、振幅ゆらぎ) を、叫び声以外の感情表出系感動詞 (3510 例) および通常語彙と比較した。

③ 笑い声の音響的特徴の分析

Bout 内の call の位置における音響特徴量に関する分析を行った。Call 位置 (先頭/最後/それ以外) と 0 次メルケプストラム係数、対数基本周波数、call 継続長の分布との関係を調べた。

(3) Affect burst を含む対話音声の合成

① 対話音声合成の品質向上

UUDB のような自発音声対話コーパスは言語的・パラ言語的不均衡を有する上に発話数も十分でないため、対話音声合成に利用する際に合成音声の品質が十分でない問題があった。そこで、対話音声ではないがそれに近い性質を有する大量のデータが収録されたコーパスによりベースモデルの合成音声の品質を確保し、自発音声対話コーパスの対話音声を目指したモデル適応を行うことにより、対話音声らしい音声を合成することを試みた。

② 感情表出系感動詞の合成

多様な形態を持つ感情表出系感動詞を合成するため、強-弱、長-短、口音-鼻音の各軸に基づき感情表出系感動詞の形態を 7 種類に分類し、それぞれに別個のモデルを学習し、そこから合成を行った

③ 笑い声の合成

HMM (隠れマルコフモデル) 音声合成と類似の枠組で笑い声の合成を行った。音素に対応する合成単位として call のモデル化を行った。Call の多様性を記述するコンテキストとして、当該/先行/後続 call の音素表記のほか、笑い声の音響分析の結果から得られた知見に基づき、bout の位置、call の位置などの要因を考慮した場合について、笑い声を合成し、その自然性を評価した。

4. 研究成果

(1) アクションゲームオンラインゲームコーパス (AGSC) の開発

収録したゲーム中の音声チャットのデータに基づき、発話単位への切り出し、書き起こし、affect burst の種別と時間情報を人手により実施し、映像データと同期を取って、ELAN で閲覧可能なコーパス (AGSC) として整備した。対話参加者は女性 12 名男性 12 名で、のべ収録時間は 888 分である。新たに構築したコーパスに対するアノテーションの結果、1427 例

の叫び声を同定することができ、構築したコーパスにより叫び声を収集することが有効であることがわかった。本コーパスは、コミュニケーションにおける叫び声研究に利用可能な世界初の音声言語資源である。実験参加者との契約に基づき、映像を除いたデータを公開コーパスとして国立情報学研究所音声資源コンソーシアムより公開予定である。

AGSC に収録されている叫び声の音響的特徴を、叫び声以外の感情表出系感動詞および通常語彙の音響的特徴と共に図 2 に示す。韻律の特徴としては、叫び声の強度は叫び以外の感情表出系感動詞および通常語彙に比べ平均で 15 dB 程度大きいことが示された。また声質の特徴としては、声の張りに関する音響パラメータであるスペクトル傾斜は叫び以外の感情表出系感動詞に比べ平均で 20 dB 程度小さく、より高域のエネルギーを強く持つ声質を有することが示された。

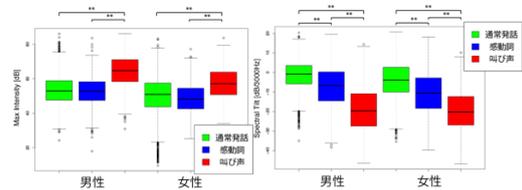


図 2 叫び声の音響的特徴

(2) 笑い声の形態的特性

UUDB 中の笑い声の特性を検討した。その結果は以下のようにまとめられる。

- ・ほとんどの笑いは単一 bout からなる。最も長い笑いは 4 個の bout からなる。
- ・Bout を構成する call 数は、1 から 4 までが多い。
- ・30%強を占める単一 call から成る bout のうち、半数以上は全体が無声であった。これに対し、複数 call からなる bout のほとんどは有声 call を含んでいた。すなわち、bout の構造に関する非対称性が存在していた。
- ・call の素性としては、90%以上が口音であり、80%が有声であり、u 系、a 系、e 系の順に多かった。

(3) 合成された感情表出系感動詞の形態とパラ言語知覚および自然性との関係

5 種類の形態 (強短・弱短・超長・強長・弱長) で合成した感情表出系感動詞「あ」から知覚されたパラ言語情報の割合を図 3 に示す。

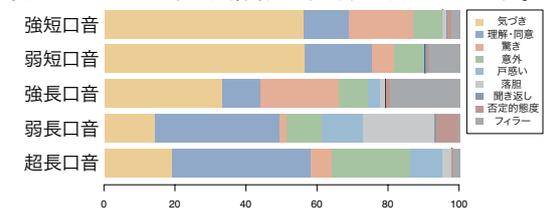


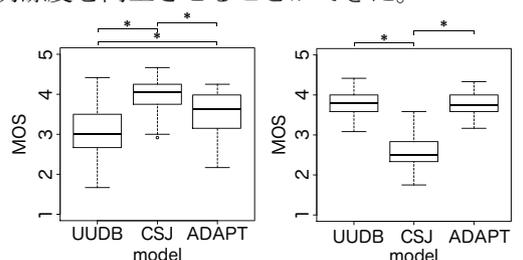
図 3 合成した「あ」の形態とパラ言語知覚

「あ」からは一般に「気づき」「理解・同意」が知覚されやすいことがわかる。また、長さの違いにより、「気づき」と「理解・同意」の相対的知覚されやすさが影響を受けている。さらに、強い形態の「あ」は「驚き」に知覚されやすい傾向がある。

UUDBに含まれる発話のラベルを用いて合成した文音声の感情表出系感動詞「あ」を提案手法により合成した音に置き換え自然性を5段階で評価させる実験を行った。この際、原音声の感情表出系感動詞に対して付与された形態ラベルと同じ形態の合成音と置き換える条件(matched)と、異なる形態の合成音と置き換える条件(unmatched)を変数の1つとした。分散分析の結果、unmatched条件の自然性平均評価値(3.41)に比べmatched条件の自然性平均評価値(4.03)が有意に高く、文脈に合っていない形態の合成音は文全体の自然性を下げることがわかった。

(4) 発話モード適応による対話音声合成の高品質化の検証

提案法の有効性検証のための学習データとして、UUDBの対話音声(女性話者1名、11分)および日本語話し言葉コーパス(CSJ)の独話音声(女性話者54名、5時間)を用いた。UUDBのみから学習したモデル(従来法)、CSJのみから学習したモデル、およびCSJから学習した後にUUDBを適応先とした発話モード適応を行ったモデル(提案法)を作成し、テスト用の50文に対し合成音声を作成した。12名の被験者は、それぞれの刺激の明瞭度および対話音声らしさを5段階で評価した。結果を図4に示す。提案法により、対話音声らしさはUUDB単体から学習した場合と同程度ながら、明瞭度を向上させることができた。



(a): 明瞭度評価実験 (b): 対話音声らしさ評価実験
図4 発話モード適応の効果

(5) 合成笑い声の自然性評価

UUDBおよびOGVCの笑い声109 boutsから学習したモデルを基に、3条件で笑い声を合成した。

・ ベースライン(BL)

当該callのみを考慮したコンテキストを用いて合成された笑い声

・ ベースライン+グループA (BL+A)

BLにおけるコンテキストに加えて、先行・後続セグメントの音韻を考慮したコンテキストを用いて合成された笑い声

・ ベースライン+グループA+グループB (BL+AB)

BL+ABにおけるコンテキストに加えて、発話におけるbout位置、boutにおけるcall位置、boutを構成するcall数を考慮したコンテキストを用いて合成された笑い声

10名の被験者に、各刺激の自然性を5段階で評価させた。BL, BL+A, およびBL+ABの自然性評価値の平均はそれぞれ2.54, 2.74, 3.01であった。分散分析の結果、合成条件の主効果が有意であり、Tukey HSD法による多重比較の結果、合成条件の全ての対に対して差が有意であった。これらの結果から、本研究で提案した笑い声の変動要因をコンテキストとして加えることにより、合成笑い声の自然性が向上することがわかった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計14件)

- ① Tomohiro Nagata, Hiroki Mori, "Defining laughter context for laughter synthesis with spontaneous speech corpus", IEEE Transactions on Affective Computing, 査読有(掲載決定)
DOI: 10.1109/TAFFC.2018.2813381
- ② Kikuo Maekawa, Hiroki Mori, "Comparison of voice quality between the vowels in filled pauses and ordinary lexical items", 音声研究, 査読有, Vol.21, No.3, 2018, pp. 53-62
- ③ 永岡 篤, 森 大毅, 有本 泰子, "感情音声コーパス共通化のための新たな感情ラベル推定における既存感情ラベル併用の効果", 日本音響学会誌, 査読有, Vol.73, No.11, 2017, pp. 682-693
DOI: 10.20697/jasj.73.11_682
- ④ Yoshiko Arimoto, Hiroki Mori, "Emotion category mapping to emotional space by cross-corpus emotion labeling", Proc. Interspeech 2017, 査読有, 2017, pp. 3276-3280
DOI: 10.21437/Interspeech.2017-994
- ⑤ Tomohiro Nagata, Hiroki Mori, Takashi Nose, "Dimensional paralinguistic information control based on multiple-regression HSMM for spontaneous dialogue speech synthesis with robust parameter estimation", Speech Communication, 査読有, Vol.88, 2017, pp. 137-148
DOI: 10.1016/j.specom.2017.01.002

- ⑥ Emika Takeishi, Takashi Nose, Yuya Chiba, Akinori Ito, "Construction and analysis of phonetically and prosodically balanced emotional speech database", Oriental COCOSDA 2016, 査読有, 2016, pp. 16--21
DOI: 10.1109/ICSDA.2016.7918977
- ⑦ Kikuo Maekawa, Hiroki Mori, "Voice-quality difference between the vowels in filled pauses and ordinary lexical items", Proc. Interspeech 2016, 査読有, 2016, pp. 3171-3175
DOI: 10.21437/Interspeech.2016-1309
- ⑧ Hiroki Mori, Atsushi Nagaoka, Yoshiko Arimoto, "Accuracy of automatic cross-corpus emotion labeling for conversational speech corpus commonization", Proc. LREC 2016, 査読有, 2016, pp. 4019-4023
<http://www.lrec-conf.org/proceedings/lrec2016/summaries/198.html>
- ⑨ Shuhei Yamada, Takashi Nose, Akinori Ito, "A study on tailor-made speech synthesis based on deep neural networks", Proc. 12th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 査読有, 2016, pp. 159-166
DOI: 10.1007/978-3-319-50209-0_20
- ⑩ Hiroki Mori, "Morphology of vocal affect bursts: Exploring expressive interjections in Japanese conversation", Proc. Interspeech 2015, 査読有, 2015, pp. 1309-1313
https://www.isca-speech.org/archive/interspeech_2015/i15_1309.html
- ⑪ Takashi Nose, Yusuke Arao, Takao Kobayashi, Komei Sugiura, Yoshinori Shiga, Akinori Ito, "Entropy-based sentence selection for speech synthesis using phonetic and prosodic contexts", Proc. Interspeech 2015, 査読有, 2015, pp. 3491-3495
https://www.isca-speech.org/archive/interspeech_2015/i15_3491.html
- ⑫ Yoshiko Arimoto and Kazuo Okanoya, "Multimodal features for automatic emotion estimation during face-to-face conversation", 音声研究, 査読有, Vol. 19, No. 1, 2015, pp. 53-67
- ⑬ Yoshiko Arimoto, Kazuo Okanoya, "Mutual emotional understanding in a face-to-face communication environment: How speakers understand and react to listeners' emotion in a game task dialog", Acoustical Science and Technology, 査読有, Vol. 36, No. 4, 2015, pp. 370-373
DOI: 10.1250/ast.36.370
- ⑭ Yoshiko Arimoto and Kazuo Okanoya, "Emotional synchrony and covariation of behavioral/physiological reactions between interlocutors", Oriental COCOSDA 2014, 査読有, 2014, pp. 100-105
DOI: 10.1109/ICSDA.2014.7051428
- [学会発表] (計 29 件)
- ① 戸井田 裕貴, 森 大毅, "パラ言語的に多様な感情表出系感動詞の音声合成", 2018 年電子情報通信学会総合大会, 2018
- ② 鈴木 圭, 森 大毅, "アクセント句 F0 の統計モデルに基づくアクセント句境界推定", 2018 年電子情報通信学会総合大会, 2018
- ③ 横山 雅季, 永田 智洋, 森 大毅, "自然対話音声コーパスを用いた DNN 音声合成におけるパラ言語情報制御", 日本音響学会 2018 年春季研究発表会, 2018
- ④ 永田 智洋, 森 大毅, "自然対話における発話の文脈を考慮した笑い声合成の検討", 電子情報通信学会音声研究会, 2017
- ⑤ 藍原 瞭, 森 大毅, "話し手の修復行動を伴う聞き手の復唱の分析", 人工知能学会言語・音声理解と対話処理研究会資料, 2017
- ⑥ 森 大毅, 有本 泰子, 永田 智洋, "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", 日本音響学会 2017 年秋季研究発表会, 2017
- ⑦ 嘉屋 和樹, 森 大毅, "発話モード適応による対話音声合成の高品質化", 日本音響学会 2017 年秋季研究発表会, 2017
- ⑧ 菊地 佑樹, 森 大毅, "音声コミュニケーションにおける叫び声", 日本音響学会 2017 年秋季研究発表会, 2017
- ⑨ 横山 雅季, 森 大毅, 有本 泰子, "対話音声の感情認識における音響的特徴と言語的特徴の効果", 日本音響学会 2017 年春季研究発表会, 2017
- ⑩ 花房 伸哉, 能勢 隆, 伊藤 彰則, "話し言葉音声合成のための Web 上のテキストデータを用いたコーパス設計の検討", 日本音響学会 2017 年春季研究発表会, 2017
- ⑪ 武石 笑歌, 能勢 隆, 山田 修平, 千葉 祐弥, 伊藤 彰則, "感情音声データベース"

- スJTESの音声合成による評価”, 日本音響学会 2017 年春季研究発表会, 2017
- ⑫ 藍原 瞭, 森 大毅, “話し手の説明戦略に影響する聞き手行動の分析”, HCG シンポジウム 2016, 2016
- ⑬ 有本 泰子, 森 大毅, “クロスコーパスラベリングによる感情空間への感情カテゴリのマッピング—音声知覚および音響分析による検討—”, HCG シンポジウム 2016, 2016
- ⑭ 嘉屋 和樹, 森 大毅, “HMM 音声合成における複数の自然対話コーパス (UADB, OGVC) 併用の有効性”, 日本音響学会 2016 年秋季研究発表会, 2016
- ⑮ 田中 勇翔, 森 大毅, “ニューラルネットワークを用いた基本周波数パターンのモデル化”, 日本音響学会 2016 年秋季研究発表会, 2016
- ⑯ 有本 泰子, 森 大毅, “音声からの知覚による感情カテゴリと感情次元の関係”, 日本音響学会 2016 年秋季研究発表会, 2016
- ⑰ 有本 泰子, 森 大毅, “クロスコーパス感情ラベリングによる対話音声の比較”, 日本音響学会 2016 年春季研究発表会, 2016
- ⑱ Kikuo Maekawa and Hiroki Mori, “Voice quality analysis of Japanese filled pauses: A preliminary report”, Disfluency in Spontaneous Speech Workshop, 2015
- ⑲ 森 大毅, “Affect burst の音声学的分類—感情表出系感動詞の言語的・パラ言語的特徴—”, 日本音響学会 2015 年秋季研究発表会, 2015
- ⑳ 永田 智洋, 森 大毅, “クラスタリングに基づく発話中における笑い声の変動要因の検討”, 日本音響学会 2015 年春季研究発表会, 2015
- 21 森 大毅, “Affect burst の形態論的分類—UADB を対象とした検討—”, 日本音響学会 2015 年春季研究発表会, 2015
- 22 前川 喜久雄, 森 大毅, “フィラーの声質上の特徴に関する予備的分析”, 日本音響学会 2015 年春季研究発表会, 2015
- 23 森 大毅, 森本 郁代, 大場 美和子, 吉田 悦子, 伝 康晴, “多様な会話コーパスを対象とした発話連鎖ラベリングの試み”, 第 6 回コーパス日本語学ワークショップ, 2014
- 24 森 大毅, 高橋 俊介, 永田 智洋, “HMM に基づく対話音声合成におけるパラ言語情報制御手法の比較”, 電子情報通信学会音声研究会, 2014
- 25 永田 智洋, 森 大毅, “対話音声合成を目的とした発話中の笑い声の変動要因の検討”, 電子情報通信学会音声研究会, 2014
- 26 有本 泰子, “対話場面におけるマルチモーダル感情コミュニケーション”, ヒューマンストレス産業技術研究会第 27 回講演会, 2014
- 27 森 大毅, 渡邊 諒馬, “藤崎モデルに基づく対話音声の句末音調の操作がパラ言語情報の知覚に与える影響”, 日本音響学会 2014 年秋季研究発表会, 2014
- 28 永岡 篤, 森 大毅, “サポートベクター回帰による自発音声の感情次元推定”, 日本音響学会 2014 年秋季研究発表会, 2014
- 29 有本 泰子, 江田 康太郎, 菊池 英明, 岡ノ谷 一夫, “オンラインコミュニケーションにおけるモダリティ統制下の情動理解”, 日本音響学会 2014 年秋季研究発表会, 2014.
- [図書] (計 1 件)
- ① 森 大毅, 前川 喜久雄, 粕谷 英樹, コロナ社, 音声は何を伝えているか—感情・パラ言語情報・個人性の音声科学—”, 2014, 207
6. 研究組織
- (1) 研究代表者
森 大毅 (MORI, Hiroki)
宇都宮大学・工学研究科・准教授
研究者番号: 10302184
- (2) 研究分担者
有本 泰子 (ARIMOTO, Yoshiko)
独立行政法人理化学研究所・情動情報連携チーム・客員研究員
研究者番号: 60586957
- 能勢 隆 (NOSE, Takashi)
東北大学・工学研究科・准教授
研究者番号: 90550591
- (3) 研究協力者
永田 智洋 (NAGATA, Tomohiro)