

令和 4 年 6 月 8 日現在

機関番号：12201

研究種目：基盤研究(B)（一般）

研究期間：2019～2021

課題番号：19H01252

研究課題名（和文）日常会話の韻律モデル構築に向けた話者混在音声の分析基盤

研究課題名（英文）Infrastructure for analyzing the prosody of speaker-mixed speech for modeling daily conversation

研究代表者

森 大毅（Mori, Hiroki）

宇都宮大学・工学部・准教授

研究者番号：10302184

交付決定額（研究期間全体）：（直接経費） 13,700,000円

研究成果の概要（和文）：日常的な生活環境で収録した話者混在音声から、声の高さの情報であるF0を話者ごとに推定する技術の確立を目指した研究を実施し、以下に挙げる成果が得られた。

- (1) 各話者のICレコーダ収録音から当該話者以外の成分を抑圧し、次段のF0推定の精度を向上させるため、話者やマイクの移動に対処した音源分離を試み、日常会話コーパス中に対する分離音声の聞き取り実験を通じて有効性を確認した。
- (2) 指定した話者のF0情報を抽出する深層学習の手法を開発した。話者混在音声を対象とした評価実験の結果、分離前と比べて、F0推定誤りを60%以上削減できることがわかった。

研究成果の学術的意義や社会的意義

日常場面の中で当事者たち自身の動機や目的によって自然に生じた会話を収録したコーパスでは、各話者の音声音が音響的に分離されておらず、本人以外の声も入り込んでしまう。日常会話では複数の話者の発話が頻繁に重なっており、そのような部分では音声の韻律的特徴を正確に分析することができない。本研究の成果は、このような話者混在音声から各話者の韻律情報を分離する技術に道筋を付けるものであり、実環境で収録されたデータに応用することで、音声学・社会科学・心理学・音声情報処理をはじめとする広範な研究分野に貢献することが期待される。

研究成果の概要（英文）：This project aimed to establish a fundamental technology to estimate pitch independently for each speaker given overlapping speech recorded in everyday circumstances, and achieved the following:

- (1) Developed a speech separation that takes the movement of speakers or microphones into account. This suppresses components of speakers other than the target, and is expected to improve the accuracy of subsequent pitch estimation. A listening test for the Corpus of Everyday Japanese Conversation revealed its effectiveness.
- (2) Developed a novel deep learning method for extracting pitch information of specified speaker. Results of evaluation experiments on overlapping speech demonstrated that the proposed method could reduce the gross pitch error by more than 60% compared to the case for which the proposed method was not applied.

研究分野：音声言語情報処理

キーワード：ニューラルF0モデル 音源分離 話者埋め込み

1. 研究開始当初の背景

日常場面の中で当事者たち自身の動機や目的によって自然に生じた会話を収録したコーパスでは、各話者の音声が音響的に分離されておらず、本人以外の声も入り込んでしまう。日常会話では複数の話者の発話が頻繁に重なっており、そのような部分では音声の韻律的特徴を正確に分析することができない。このことは、日常会話コーパスを利用した韻律研究を推進する上で大きな障害となっていた。

2. 研究の目的

- (1) 本研究は、日常的な生活環境で収録した話者混在音声から、会話音声の韻律のモデル化に必要な F0(基本周波数)情報を話者ごとに分離することで、これまで不可能であった韻律の定量的分析を可能とする基盤技術を確立することを目的としていた。
- (2) さらに、この基盤技術を日本語日常会話コーパスの話者混在音声に対して評価するとともに、分析した話者ごとの F0 をコーパスの一部として収録・配布することで、音声学・社会科学・心理学・音声情報処理をはじめとする広範な研究分野において利用可能な大規模な会話の韻律データベースとして公開することも本研究の目的であった。

3. 研究の方法

- (1) [多チャンネル音源分離] 各話者の IC レコーダ収録音から当該話者以外の成分を抑圧し、次段の F0 推定の精度を向上させる。日本語日常会話コーパスにおいて収録された会話では、各話者の音声は別個の IC レコーダで収録されているため、音源分離技術の適用のためには精密な時間合わせが必要であるが、話者やマイクの移動に対処する必要がある。そこで、対話音声をブロックに分割し、独立ベクトル分析(IVA)に基づく方法で音源分離を行う。
- (2) [ニューラルネットワークを利用した目的話者の F0 推定] F0 推定の問題を、話者混在音声のスペクトル情報から目的話者の F0 への写像を推定する問題として定式化し、ニューラルネットワークによりモデル化する。ニューラルネットワークは、話者 ID から埋め込みベクトルに変換する部分、スペクトルから F0 情報の推定に有用な情報を抽出する部分、および話者情報と前後関係を考慮して F0 軌跡を推定する部分から構成される(図 1)。ネットワークの学習には、クリーン音声を混合した擬似話者混在音声を用いる。その際、指定した話者の単独発話音声から抽出した F0 を正解とし、話者ごとに損失を求めて累積した後、損失を最小化するようにネットワーク全体を学習する(図 2)。

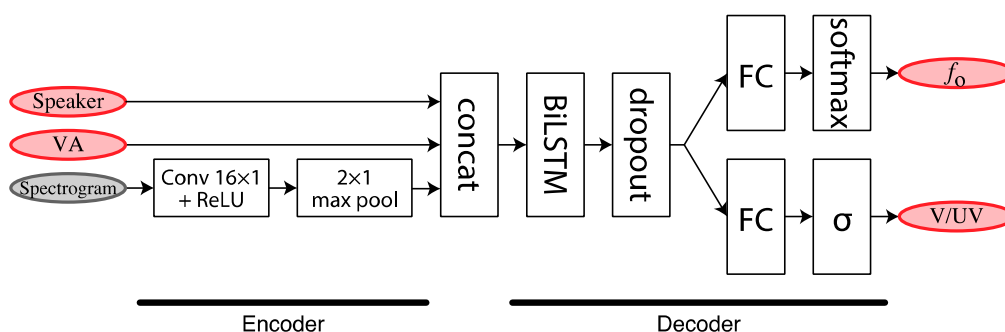


図 1

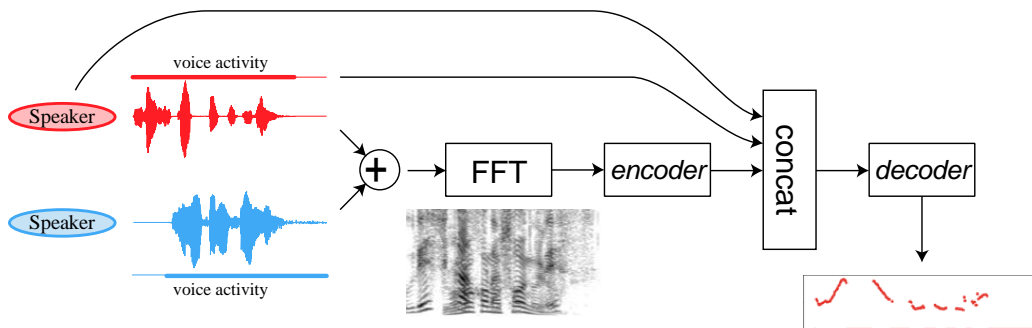


図 2

4. 研究成果

(1) [多チャンネル音源分離]日常会話コーパスは、話者やマイクが動く時変伝達系であり、また複数の IC レコーダによる非同期録音である点が、従来の音源分離手法と異なり困難な点である。令和 2 年度は、ブロック分割・同期化による補償に基づく日常会話コーパスの音源分離を検討した。音源分離手法としては、時間周波数マスクング、およびその結果を利用した独立ベクトル分析(IVA)を検討した。日常会話コーパス中の 2 シーンに対する分離音声の聞き取り実験の結果、時間周波数マスクングが有効であることがわかった。

(2) [ニューラルネットワークを利用した目的話者の F0 推定]クリーン音声のコーパスとして、日本語話し言葉コーパスの模擬講演音声を用いた。次に、これを訓練用セット(344 名)、適応用セット(訓練用セットに含まれない 20 名)、テストセット(適応用セットと同じ 20 名)に分割した。テストセットの話者埋め込みは、訓練用セットと適応用セットの和集合から話者埋め込みまで含めてモデルを学習し、そこから取り出すことで得た。

図 3, 図 4 に分離した F0 軌跡の例を示す。上段が正解、中・下段が各話者に対して推定された F0 軌跡である。どちらの例に対しても F0 軌跡の推定が精度良く行われていることがわかる。図 3 の例では女性話者と男性話者の F0 軌跡が交差している箇所があるが、分離は総じてうまく行っている。

F0 軌跡分離能力を定量的に評価するため、有声無声誤り率(VDE)、グロスエラー率(GPE)、ファインエラー率(FPE)を求めた。これらの誤り率は各話者の音声区間のみを対象として計算した。表 1 に結果を示す。表に示されるように、提案法によって F0 軌跡を分離することで、有声無声誤りとグロスエラーを大きく減少させることができた。

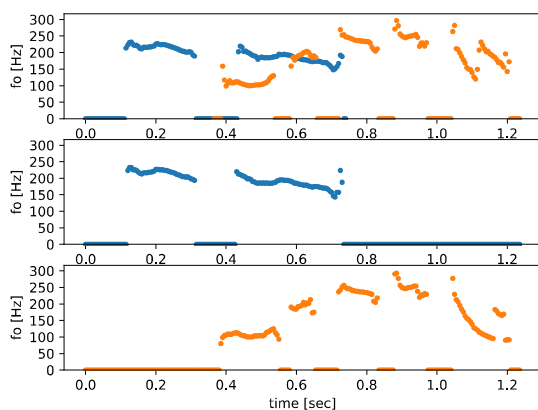


図 3

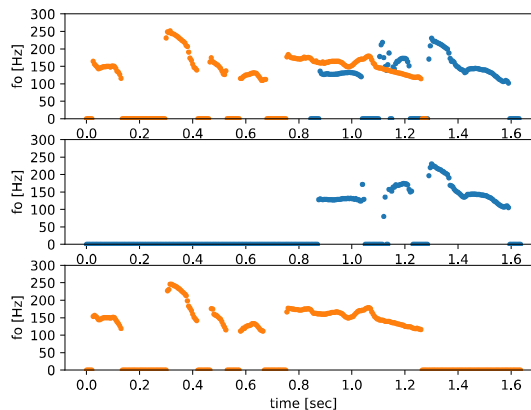


図 4

表 1

	有声無声誤り率 (%)	グロスエラー (%)	ファインエラー (st)
分離前	10.72	17.50	0.33
提案法	3.99	6.56	0.34

- (3) 日本語日常会話コーパスに含まれる未知の話者に対し、手作業および非重複区間を利用した再学習による話者埋め込み推定の結果を用いて F0 分離を試みたところ、有声無声誤り率 33.9%、グロスエラー 34.45%と、高精度な推定が行えなかった。この原因として、日本語日常会話コーパスでは各話者の音声の混合比がまちまちで、SN 比が悪い話者が多いこと、および日本語話し言葉コーパスとの収録環境の違いが考えられる。
- (4) 以上の成果を総括すると、当初目標 日常的な生活環境で収録した話者混在音声から F0 情報を話者ごとに分離する技術を開発する、に関しては、模擬話者混在音声に対しては達成することができた。ただし、実音声については未達成であった。また、当初目標 この技術を日本語日常会話コーパスの話者混在音声に対して評価し、分析した話者ごとの F0 をコーパスの一部として収録する、に関しては達成できなかった。

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 5件/うち国際共著 0件/うちオープンアクセス 5件）

1. 著者名 Hiroki Mori	4. 巻 -
2. 論文標題 Pitch contour separation from overlapping speech	5. 発行年 2021年
3. 雑誌名 Proc. Interspeech 2021	6. 最初と最後の頁 1209-1213
掲載論文のDOI（デジタルオブジェクト識別子） 10.21437/Interspeech.2021-2164	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Takahisa Iizuka, Hiroki Mori	4. 巻 -
2. 論文標題 Comparison of machine learning algorithms and acoustic features in emotion recognition from spontaneous speech	5. 発行年 2022年
3. 雑誌名 Acoustical Science and Technology	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 小磯 花絵	4. 巻 29
2. 論文標題 『日本語日常会話コーパス』本公開版の構築	5. 発行年 2022年
3. 雑誌名 自然言語処理	6. 最初と最後の頁 224-229
掲載論文のDOI（デジタルオブジェクト識別子） 10.5715/jnlp.29.224	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Hiroki Mori, Yuki Kikuchi	4. 巻 -
2. 論文標題 Gaming corpus for studying social screams	5. 発行年 2020年
3. 雑誌名 Proc. Interspeech 2020	6. 最初と最後の頁 3132-3135
掲載論文のDOI（デジタルオブジェクト識別子） 10.21437/Interspeech.2020-2553	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Kazuki Kaya, Hiroki Mori	4. 巻 E102-D
2. 論文標題 Effectiveness of speech mode adaptation for improving dialogue speech synthesis	5. 発行年 2019年
3. 雑誌名 IEICE Transactions on Information and Systems	6. 最初と最後の頁 2064-2066
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/transinf.2019EDL8024	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Hiroki Mori, Tomohiro Nagata, Yoshiko Arimoto	4. 巻 -
2. 論文標題 Conversational and social laughter synthesis with WaveNet	5. 発行年 2019年
3. 雑誌名 Proc. Interspeech 2019	6. 最初と最後の頁 520-523
掲載論文のDOI (デジタルオブジェクト識別子) 10.21437/Interspeech.2019-2131	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計22件 (うち招待講演 2件 / うち国際学会 0件)

1. 発表者名 森 大毅
2. 発表標題 『日本語日常会話コーパス』を対象とした話者混在音声の韻律の分析
3. 学会等名 シンポジウム 日常会話コーパスVII (招待講演)
4. 発表年 2022年

1. 発表者名 森 大毅
2. 発表標題 話者混在音声からのfo軌跡の分離
3. 学会等名 日本音響学会2021年秋季研究発表会
4. 発表年 2021年

1. 発表者名 飯塚 喬久, 森 大毅
2. 発表標題 自発音声コーパスを用いて合成した音声で話すエージェントが会話相手の行動に与える影響
3. 学会等名 日本音響学会2021年秋季研究発表会
4. 発表年 2021年

1. 発表者名 西野 広直, 森 大毅
2. 発表標題 複数の自然対話音声コーパスの併用によるend- to-end対話音声合成の高品質化
3. 学会等名 日本音響学会2022年春季研究発表会
4. 発表年 2022年

1. 発表者名 小磯 花絵
2. 発表標題 日本語日常会話コーパスの構築 - 日常のことばの特徴を探る -
3. 学会等名 第六回学習者コーパス・シンポジウム
4. 発表年 2021年

1. 発表者名 小磯 花絵
2. 発表標題 コーパスを通して日常のことばの特徴を探る
3. 学会等名 第16回NINJALフォーラム「ここまで進んだ！ここまで分かった！多様な言語資源に基づく日本語研究」
4. 発表年 2022年

1. 発表者名 小磯 花絵
2. 発表標題 話し言葉の分析：『日本語日常会話コース』CEJC
3. 学会等名 講演会「日本語コースの設計・構築・応用」
4. 発表年 2022年

1. 発表者名 小磯 花絵
2. 発表標題 『日本語日常会話コース』を活用した話し言葉研究の可能性
3. 学会等名 シンポジウム 日常会話コースVII
4. 発表年 2022年

1. 発表者名 小磯 花絵
2. 発表標題 『日本語日常会話コース』バランスの検証と研究の可能性
3. 学会等名 シンポジウム「ことば・認知・インタラクション10」
4. 発表年 2022年

1. 発表者名 小磯 花絵, 天谷 晴香, 石本 祐一, 居間 友里子, 白田 泰如, 柏野 和佳子, 川端 良子, 田中 弥生, 伝 康晴, 西川 賢哉, 渡邊 友香
2. 発表標題 『日本語日常会話コース』の設計と特徴
3. 学会等名 言語処理学会第28回年次大会
4. 発表年 2022年

1. 発表者名 前川 喜久雄
2. 発表標題 条件異音再考
3. 学会等名 日本音響学会音声研究会（招待講演）
4. 発表年 2022年

1. 発表者名 南木 春希, 山岡 洗瑛, 若林 佑幸, 小野 順貴
2. 発表標題 非同期会話録音に対する時間周波数マスキングと独立ベクトル分析を用いた音源分離の検討
3. 学会等名 音学シンポジウム
4. 発表年 2021年

1. 発表者名 前川喜久雄
2. 発表標題 データサイエンスが使えるようになるまで
3. 学会等名 日本語文法学会第21回大会
4. 発表年 2020年

1. 発表者名 前川喜久雄, 森大毅
2. 発表標題 アクセント句頭のFo上昇量はtone間距離で決まる
3. 学会等名 日本音声学会第341回研究例会
4. 発表年 2020年

1. 発表者名 高松屋 友翼, 森 大毅
2. 発表標題 録音音声とライブ音声による説明の違いが聞き手反応に及ぼす影響
3. 学会等名 HCGシンポジウム2020
4. 発表年 2020年

1. 発表者名 西野 広直, 森 大毅
2. 発表標題 韻律を考慮したend- to-end方式に基づく自発音声合成
3. 学会等名 日本音響学会2021年春季研究発表会
4. 発表年 2021年

1. 発表者名 飯塚 喬久, 森 大毅
2. 発表標題 自発音声に基づく合成音声で対話するシステムがユーザに与える影響の調査
3. 学会等名 日本音響学会2021年春季研究発表会
4. 発表年 2021年

1. 発表者名 高岸 勇斗, 森 大毅
2. 発表標題 感情表出系感動詞の鼻音化に関する種々の分析
3. 学会等名 日本音響学会2019年秋季研究発表会
4. 発表年 2019年

1. 発表者名 永田 智洋, 森 大毅
2. 発表標題 自発音声に対するニューラルF0モデリングの可能性
3. 学会等名 日本音響学会2019年秋季研究発表会
4. 発表年 2019年

1. 発表者名 高松屋 友翼, 森 大毅
2. 発表標題 原稿の有無が説明者の発話と被説明者の反応に与える影響の分析
3. 学会等名 日本音響学会2020年春季研究発表会
4. 発表年 2020年

1. 発表者名 菊池 英明, 山田 高明, 小磯 花絵
2. 発表標題 大規模日常会話コーパスにおけるイントネーションラベリング
3. 学会等名 日本音響学会2020年春季研究発表会
4. 発表年 2020年

1. 発表者名 小磯 花絵, 菊池 英明, 山田 高明
2. 発表標題 『日本語日常会話コーパス』への韻律ラベリング - ラベリングの設計と日常会話の韻律の特徴 -
3. 学会等名 人工知能学会第88回SLUD研究会
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	前川 喜久雄 (Maekawa Kikuo) (20173693)	大学共同利用機関法人人間文化研究機構国立国語研究所・音声言語研究領域・教授 (62618)	
研究分担者	小磯 花絵 (Koiso Hanae) (30312200)	大学共同利用機関法人人間文化研究機構国立国語研究所・音声言語研究領域・教授 (62618)	
研究分担者	小野 順貴 (Ono Kazutaka) (80334259)	東京都立大学・システムデザイン研究科・教授 (22604)	
研究分担者	永田 智洋 (Nagata Tomohiro) (80823450)	帝京大学・理工学部・助教 (32643)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------