

日常生活における私的発話識別法の提案

友広 純々野^{1,a)} 塚田 浩二¹

概要：人々は他者との意思疎通を目的としない発話、いわゆる独り言を日常の中で発している。本研究ではこの発話を私的発話と呼ぶ。私的発話の中でも発話者が意識的に発するとされるセルフトークがスポーツ界において運動パフォーマンスを向上させる手段として注目されている。意識的な私的発話がある一方で、無意識的に発せられている私的発話も存在し、どちらも発話内容が記憶に残りにくいという特徴がある。本研究は、これらの私的発話を、2種類のマイク（単一指向性/全指向性）を用いて記録し、音声認識ライブラリを用いて文字起こしされたテキストデータと、音声データを発話者のライフログとして振り返ることができるシステムを提案する。本システムにより、発話時間や発話内容から、話者本人のその日の出来事を想起させたり、隠れたアイデアの発掘ができると考える。本稿では、日常生活の発話を記録し、私的発話の分析を行った。分析結果から、本研究の応用について議論する。

Detecting Method of Private Utterance for Daily Life

TOMOHIRO SUZUNO^{1,a)} TSUKADA KOJI¹

1. はじめに

他者との意思疎通を目的としない発話、いわゆる独り言を人々は日常の中で発している。本研究ではこの発話を私的発話と呼ぶ。

近年、意識的な私的発話であるセルフトークやライフログが注目されている。セルフトークとは自分自身と会話するという意味を持ち、感情をコントロールするメンタルトレーニングの1つ [1] とされている。例えば今川ら [2] は、セルフトークが脳活動を活性化させ、自信の向上や不安の解消といった精神に影響を与える働きがあると示唆している。また、外国語学習におけるシャドーイングが、学習者の聞き取り能力向上に有効な指導法であることが分かっている [3]。このような意識的な私的発話に加え、無意識的な私的発話も存在する。例えば、話者のひらめきや何か用事を思いだした時、話者自身の考えや感情を知らぬ間に発話しているものがある。

ライフログとは、蓄積された個人の生活の行動履歴 [4] であり、近年個人の活動に関する情報を、自動的にデジタル

データとして記録する事例が存在する。例えば、カメラやセンサを用いた個人の行動を記録する研究事例 [5], [6] や、スマートウォッチを用いた健康を管理する事例 [7] である。

このように意識的な私的発話は、セルフトークやシャドーイングといった研究事例が存在するが、無意識的な私的発話を利用した事例は少ない。さらに私的発話は、発話者の記憶に残りにくいといった問題がある。

そこで本研究は、こうした私的発話の特徴とライフログを組み合わせ、音声ライフログの記録システムを提案する。私的発話をライフログとして記録することで、個人が無意識的に発したアイデアの発掘や、セルフトークのような意識改革や行動変容を促すことができると考える。

本論文では、音声ライフログ構築のために、個人の発話から私的発話を識別するアプローチに焦点を当てる。

2. 関連研究

本章では、本研究に関連する研究事例として、「音声ライフログの事例」、「発話の検出に関する事例」、「音声認識を用いた事例」の3つの観点から説明する。

¹ 公立はこだて未来大学
Future University Hakodate

^{a)} g2121040@fun.ac.jp

2.1 音声ライフログの事例

川島ら [8] は、ユーザーの発話を記録し、発話データを形態素解析、係り受け解析、概念解析することで、ユーザーの経験に関する情報を抽出している。この経験に関する情報を利用し、ユーザーの過去の経験に沿った雑談内容を提示する雑談対話システムを構築している。このシステムを使って生成された雑談の提示は、ユーザーに親近感を与える事を明らかにした。

綾部ら [9] は、話者の発話時のピッチなどから感情を推測し、その感情タグを用いて過去に起こった出来事の種類を試みる手法を提案している。

また、横山ら [10] は、話者が自覚しづらい口癖や言葉遣いを改善するための音声返戻システムを提案している。このシステムは利用者自身が発言を控えたい単語を設定し、発言中にその単語を発するとリアルタイムで発話音声を利用者にフィードバックされるシステムである。その結果、設定した単語や口癖に対する意識づけに効果があることを明らかにした。

私的発話を利用した研究事例として、長利ら [11] は、コーディング中の発話音声と文章を紐づけて、コードを読み返す際にその発話音声をフィードバックするシステムを制作している。このシステムを利用することで文字としては記録されなかった話者のアイデアや考察などを確認することを可能にしている。

このように発話を対象とした研究 [8], [9], [10] は多く取り組まれているが、私的発話を対象とした研究事例は少ない。また、[11] の事例のように私的発話を利用した研究事例が存在するが、発話時の場面が限定されており、特定の作業に特化した取り組みである。本研究では、発話時の場面を限定せず、発話データすべてをライフログの対象とする。

2.2 発話の検出に関する事例

熊谷ら [12] は、カメラを用いて他者の顔を検出、マイクを用いて発話区間を検出し、この両方のデータを組み合わせることにより、他者との対話区間を検出する手法を提案している。

木田らは [13]、ステレオマイクから得られた音響特徴量を利用した声の性質から話者を特定することに加え、複数人の発話者の位置関係を割り出し、発話者のクラスタリング精度を上げる技術を開発している。

Yan ら [14] は、2つの同じマイクを用いて、それぞれのマイクで記録した音声差分を利用することで、口元を手で覆うジェスチャーを検出する手法を提案している。本研究ではこれを参考にし、特徴の違う2つのマイクを用いて、それぞれの音声差分を利用することで私的発話の識別を試みる。

2.3 音声認識を用いた事例

現在、音声認識の精度は向上かつ広く普及しており、スマートフォンのキーボードで音声入力为标准で使えるようになっている。さらに、さまざまな場面で音声認識を用いた文字起こしの事例が存在する。例えば、リアルタイムにテレビや講義の字幕を作成 [15], [16] したり、議事録作成ツールなどに利用されている。また、聴覚障害者のコミュニケーションを支援するために、他者の会話内容を音声認識で取得するといったアプリケーションの開発にも利用されている [17]。

このような音声認識精度の向上や、利用可能なライブラリの普及を受け、本研究では発話時の音声データを文字起こしする際に利用する。文字起こしをすることで、ライフログシステムの構築の際に、視覚的な発話の振り返りを可能にしたり、発話内容の分析を可能にする。

3. 提案

私的発話の識別に向けて、複数種類のマイクを用いた発話の記録検証を行った。その結果、単一指向性マイクと全指向性マイクを組み合わせたシステムを構築することとした。単一指向性マイクで記録した音声データは、ノイズ（生活音・他者の発話）が少なく、発話内容が聞き取りやすいといったメリットがあった。一方、全指向性マイクで記録した音声データは、発話前後の状況が分かり、発話内容を振り返りやすいというメリットがあった。

ここでは、単一指向性マイクから得られた発話を個人の発話とし、全指向性マイクから得られた発話を周辺音も含む発話とする。個人の発話とは個人が発している発話すべてであり、対話中における相手の発話などは含まれない。周辺音も含む発話とは、個人が発している発話に加えて、対話中における相手の発話や、ざわざわとした環境音も含まれる発話のことである。本研究では、私的発話の識別にこの2種類の発話から得られたデータ差分が利用できると考えた。さらに、ライフログとしての手軽さを考え、2つのマイクを Android 端末に接続して発話を記録する。

3.1 私的発話識別の流れ

予備検証の結果から、個人の発話から私的発話を識別するために以下の3手順を踏む。

- (1) 2種類のマイク（単一指向性マイク/全指向性マイク）を用いた個人の発話と周辺音も含む発話の記録
- (2) 音声データと音声認識を用いて文字起こしされたテキストデータの取得
- (3) 個人の発話と周辺音も含む発話データの比較

まず、単一指向性マイクを用いて個人の発話を記録し、全指向性マイクを用いて周辺音も含む発話を記録する。

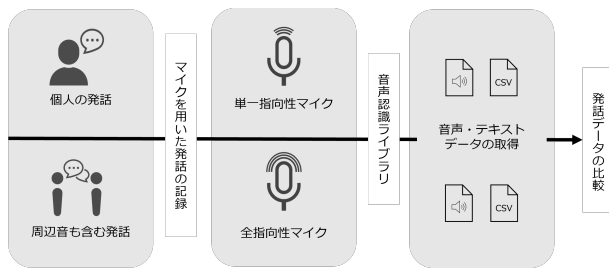


図1 私的発話識別の流れ

次に、音声データと、音声認識ライブラリを利用して発話内容の文字起こしをし、2種類の発話に関するテキストデータを取得する。テキストデータを用意する理由は、音声ライフログのシステムとして、効率的な発話の振り返りを可能にするためである。

最後に、個人の発話と周辺音も含む発話データを比較し私的発話の識別を行っていく。ここでは取得した発話データの中でもそれぞれのマイクで認識されたテキスト量に注目する。現在、2章で示した通り、音声認識の精度が向上している。そのため、2つのマイクで認識されたテキスト量を比較した際に、一定の差が出ると考えた。また、音響ベースの比較よりもテキスト量の比較といった実装の方が容易であると考えた。

本研究では、日常生活における私的発話を定義するために、3つの発話時の状況（A：周りに他者が存在しない、B：周りに他者は存在するが交流を取っていない、C：発話者が他者と交流を取っている）を設定した。本研究では、A、Bの状況下で発せられた発話を私的発話とする。

4. 実装

本研究のシステム構成を図2に示す。まず、Android端末に2つのマイクを接続し、2種類の音声データを同時に記録していく。次に、記録した音声データをPC側に転送し、Pythonの音声認識ライブラリSpeechRecognitionを利用して、音声データを文字起こししたテキストデータの作成する。なお、このライブラリは複数の音声認識エンジンを切り替えて実行できるライブラリであり、例として「Microsoft Bing VoiceRecognition」、「IBM Speech to Text」などがある。本研究では、無料で使用できる「Google Speech Recognition」を利用する。最後に、作成したテキストデータを用いてテキストデータ量の比較を行う。具体的な実装方法は図2も参考に次章より記す。

4.1 発話の記録端末構成

3.1章で示した通り、私的発話の識別のために2種類の発話に着目し、個人の発話は単一指向性マイクで、周辺音も含む発話は全指向性マイクを用いて記録する。また、これらの発話データをステレオ音声データとして保存するためにデュアルチャンネルオーディオインターフェース

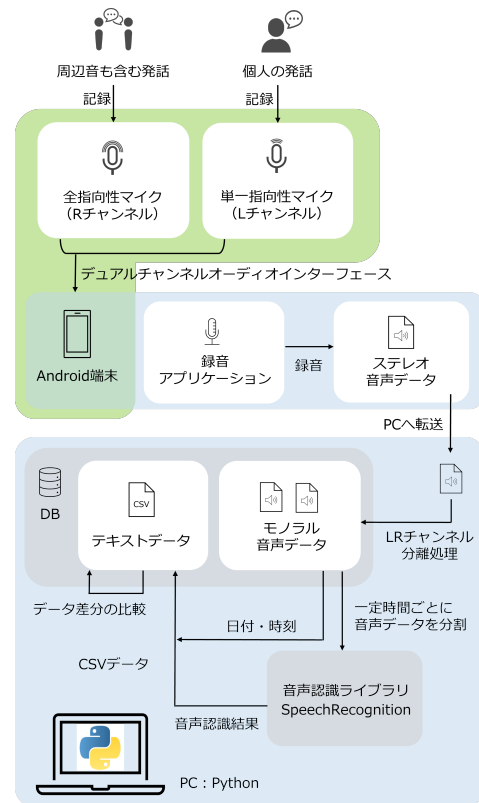


図2 システム構成図

を利用する。発話を記録するためのデバイスはそれぞれ、Android端末（RazerPhone：RZ35-0215）と、デュアルチャンネルオーディオインターフェース（BOYABY-DM20）、単一指向性マイク（単一指向性モノラルマイクロホンセット：ME52W）、全指向性マイク（BY-DM20 付属マイク）を利用した。図3はAndroid端末に2つのマイクを接続した様子である。



図3 単一指向性マイクと全指向性マイクをAndroid端末に接続した様子

マイク装着位置の予備検証として、2つのマイクを胸元もしくは口元に装着して発話の記録をした。この予備検証

の結果、2つのマイクの装着位置は口元の方が良いことが分かった。理由は、口元に装着した方が、発話者の音源と近くなり、個人の発話を不足なく取得できるためである。これにより、胸元では取得しづらかった、私的発話にあるぼそっと発話や小声などが取得しやすくなった。また、現在の新型コロナウイルス蔓延の影響でマスクの着用が一般化していることも踏まえて、マスクにマイクを取り付けることで口元近くへの装着を実現した。

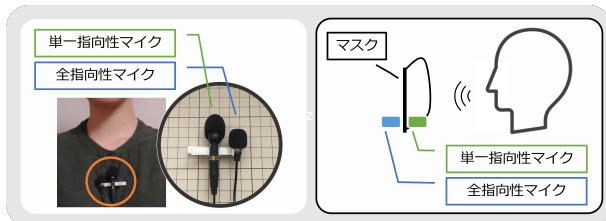


図4 左から、胸元に装着する様子と口元のマイクの位置関係

4.2 音声認識部分

ここからは、記録した音声データの処理、テキストデータの取得、テキストデータを用いた私的発話識別の提案について記す。なお、テキストデータ取得までの流れを図5に示す。

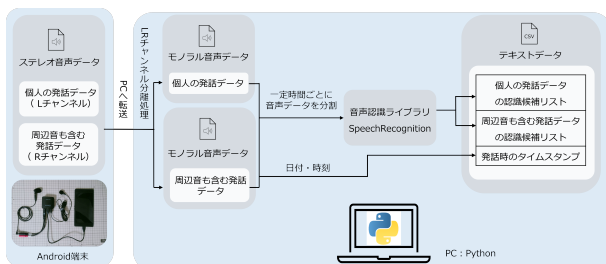


図5 テキストデータ取得までの流れ

まず、Android 端末側にある2種類の発話が含まれたステレオ音声データをPC側に転送する。そして、Pythonを用いて音声データのチャンネルを分離し、2種類の発話のモノラル音声データとして取得する。

次に、このモノラル音声データを、Pythonの音声認識ライブラリであるSpeechRecognizerを用いて発話内容のテキストデータ(CSVファイル)を取得する。なお音声認識に投入する音声データは、記録時間が数時間とデータ量が多く、この状態のまま音声認識ライブラリに入力することができない。そのため、元の音声データ一定時間(例:10秒や15秒)ごとに分割して、音声認識ライブラリへデータを入力する。これにより、テキストデータには音声認識ライブラリから得られた個人の発話データの認識候補リストと、周辺音も含む発話データの認識候補リストに加えて、音声データの分割時間を利用した発話時のタイムスタンプ

が記録される。

最後に、上記で記録したテキストデータの、個人の発話と周辺音も含む発話データ差分を比較し、私的発話の識別を行う。さまざまな比較手法が考えられるが、ここではまず一定時間あたりにおけるテキストデータ量の差分を比較する。具体的には、単一指向性マイクから個人の発話を検出した際に、前後の個人の発話と周辺音も含む発話のテキスト量を比較する。この時、両者のテキスト量がほぼ同一であれば私的発話と解釈する。一方、対話場面においては、周辺音も含む発話のテキストには対話者の発話内容が含まれると考えられる。よって、両データのテキスト量を比較して、周辺音も含む発話のテキスト量が一定以上多ければ私的発話ではないと解釈することができると思う。

4.3 振り返りのプロトタイプシステム

音声ライフログの構築に向けたプロトタイプとして、簡易的な発話の振り返りシステムを実装した。



図6 プロトタイプのシステム構成図

現在は、発話時の音声データの再生や、テキストデータの一覧を時間ごとに表示する機能がある。今後はプロトタイプシステムを基に、私的発話の振り返りに必要な要素の分析や、応用に向けた機能の拡張を行っていく。

5. データ収集と分析

5.1 私的発話内容の記録

音声ライフログ構築にむけ、日常生活の中でどのような私的発話を発しているか記録し、私的発話の用途について分析した。この分析では、著者の発話データを1日数時間程度、6日分収集し、データの記録後、発話内容の振り返りを行った。収集時は周りに他者が存在しない静かな状況、および論文執筆の作業や、動画などのコンテンツを視聴している状況であった。結果のを表1に示す。無意識的な私的発話が41発話記録され、「説明」「体調」「感嘆」「状況非依存」の4つに分類することができた。なお分類は、音声ファイルを元に著者が発話時の状況を想起しながら行った。また発話内容は、音声認識を用いて文字起こしし、認識結果が誤っていたものは、著者が音声ファイルを元に手動で修正した。また、発話内容の振り返り時に、発話内容の意味と、発話時の状況を想起することが出来なかった3

表 1 私的発話の分類とその発話数

発話内容	発話時の状況や行動に依存			発話時の状況や行動に非依存
	説明	体調	感嘆	状況非依存
発話例 (一部抜粋)	割れてしまった ○○じゃなくて やっちゃった 毎回ダメですね 何しよう 干し忘れた	疲れた 眠たい 何もできない人になる 疲れた帰りたいよ だるだるだる	あれ～ わかりました はい マジか ○○ばかりじゃん	洗濯しなきゃ いーわせてー(歌) 清く正しく(歌) 声をきかせて(歌)
発話数	21	7	8	5

発話は、分析の結果から抜いた。

以下は分類項目の説明である。「説明」は発話時の状況や発話者の行動に対して説明をしている発話である。例にある「割れてしまった」とは、物を床に落して割れてしまった状況に対して説明をしている様子であった。また、「○じゃなくて」とは、作業中に間違った行動を取ったことに対して、ツッコミを入れている様子であった。

「体調」とは、発話者の体調に関する発話をしたものである。例にある「疲れた」、「眠たい」といった発話は、作業中やコンテンツ視聴中に唐突に発話しているのが観察された。

「感嘆」とは、作業内容やコンテンツに対する疑問や応答として反応している発話である。「あれ～」などは、作業に失敗した時の反応であった。「わかりました」といった発話は、講義のような動画コンテンツ内容を理解したときの反応であった。

「状況非依存」とは、発話時の状況や行動に依存しない発話全般のことである。「洗濯しなきゃ」というのは、唐突に発話者が用事を思い出した時の発話であった。またこの他にも歌を口ずさんでいる発話が含まれた。

5.2 考察

私的発話を振り返る際、このような分類を行うことで、発話時の状況や状態を多く想起することができた。例えば、「説明」の分類項目から、1日の活動を振り返るというライフログの目的に貢献できる可能性がある。また「体調」の分類項目を振り返ることで、その日の気分の浮き沈みといったメンタルに関する予測することができた。また、「体調」に関してネガティブな発話が多かったことがある。ネガティブな発話とポジティブな発話を測ることが出来れば、メンタルヘルスケアなどに活用できるのではないかと考える。さらに、「状況非依存」の分類項目から、その日の用事といった、忘れがちな出来事の振り返りもできると考える。

このように私的発話そのものを分類することは、今後の応用として価値があると考えられる。しかし、今回の分析では発話内容から分類項目を考え、手動で発話内容を分類した。今後の展望として、自動的な私的発話の分類のために、健

康管理として用いられるスマートウォッチなどを組み合わせた実装も視野に入れていく。

6. 終わりに

日常生活における私的発話の識別を目標に、提案システムと識別手法の提案、および私的発話の調査について述べた。提案システムでは、個人の発話と周辺音も含む発話に注目し、それぞれを単一指向性マイク/全指向性マイクで記録するシステム構成について述べた。提案手法では、2種類のマイクから得られた音声データを音声機認識技術を用いてテキストデータ化し、その認識された文字数の差分を利用することを述べた。最後に私的発話の調査では、発話時の状況説明や、発話時の体調に関する情報が音声ライフログとして有用である可能性があることが分かった。

現在、私的発話の識別に向けて実装を進めてきたが、発話のデータ量が少なく、十分な検証ができていない。今後は、データ収集やシステムの改良を進め、本研究の識別手法を確立していく。また、日常生活の中で利用する応用例の構築を目指す。

参考文献

- [1] 高妻容一：基礎から学ぶ!メンタルトレーニング，ベースボール・マガジン社，2008.
- [2] 今川 新悟，松本 清，佐久間春夫：セルフトークの精神生理学的効果について，バイオフィードバック研究，第45巻，第1号，pp. 25-32，2018.
- [3] 玉井 健：シャドーイングの効果と聴解プロセスにおける位置づけ，時事英語学研究会/日本時事英語学会 [編]，第1997巻，第36号，pp. 105-116，1997
- [4] 総務省：利用者視点を踏まえた ICT サービスに係る諸問題に関する研究会 - 第二次提言 (案)，平成 22 年 5 月，2010.
- [5] 木村繁基，沖真帆，塚田浩二：生活を「彩る」ライフログデバイスの提案，情報処理学会，インタラクティブ 2017，pp. 207-209，2017.
- [6] 奥野茜，角康之：一人称ライフログ映像からの顔検出に基づいた社会活動計測，情報処理学会論文誌，第62巻，第2号，pp. 607-616，2021.
- [7] fitbit スマートウォッチ：fitbit，<https://www.fitbit.com/global/jp/home> (閲覧日：2021年7月27日)。
- [8] 川島 嵩弘，西村 隆志，安江 駿亮，六沼 元貴，和田 史織，杉本 徹：ライフログ雑談対話システムに関する研究，人工知能学会研究会資料，pp.120-121，2016.
- [9] 綾部 櫻子，田野 俊一，市野順子，岩田 満，橋山 智訓：イベントの内容，感情をロギングするリッチなサウンドライフログの提案，第28回ファジィシステムシンポジウム，pp. 625-630，2012.
- [10] 横山 紗菜，鈴木優：言葉遣いを改善する音声返戻システムの開発，情報処理学会，インタラクティブ 2020，pp.936-939，2020.
- [11] 長利 慎吾，寺田実：独り言と文字を紐付けて自動記録するテキストエディタ，情報処理学会，インタラクティブ 2016，pp.710-712，2016.
- [12] 熊谷 怜史，全 炳東：ライフログのための対話検出，第10回情報科学技術フォーラム，pp.861-862，2011.
- [13] 木田 祐介，丁 寧，広畑誠：会議の効率的な振返りを支援

- する話者クラスタリング技術, 東芝レビュー = Toshiba review, pp.26-pp.29, 2015.
- [14] Yukang Yan, Chun Yu, Yingtian Shi, Minxing Xie : PrivateTalk: Activating Voice Input with Hand-On-Mouth Gesture Detected by Bluetooth Earphones, UIST '19: Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, pp. 1013–1020, 2019.
- [15] 今井亨, 奥 貴裕, 小林彰夫 : 音声認識によるリアルタイム字幕放送の進展, 情報処理学会研究報告, pp.1-pp.6, 2011
- [16] 河原 達也, 秋田 祐 : 哉聴覚障害者のための講演・講義の音声認識による字幕付与, 日本音響学会誌 74 巻 3 号, pp.156–pp.162, 2018.
- [17] 服部 哲, 柴田 邦臣 : 聴覚障害児・者のコミュニケーションを支援する Android アプリの開発, ワークショップ 2014 (GN Workshop 2014) 論文集, pp.1–pp.6, 2014.