

B-07

音源分離を用いた選択的聞き取り支援システムの開発

Development of A Selective Listening System Using Sound Source Separation

百代 楓† 酒井 ちひろ‡ 仲田 深紅‡ 清川清‡
Kaede Momoji Chihiro Sakai Miku Nakata Kiyoshi Kiyokawa

1. はじめに

聴覚過敏、聴覚処理障害、加齢による聴力の低下などのさまざまな要因により、会話での選択的聞き取りが困難になる場合がある。少子高齢化により聞き取りに困難を抱える人はさらなる増加が見込まれる。大人数がいる場所で会話をすることを選択的聞き取りが困難なため億劫に感じている人も多いと考えられる。

2. 目的

体内で音源分離をすることができず、選択的聞き取りに困難を抱えている人も人工的に話者を分離させるシステムを用いることで、円滑な会話ができるようになることを目指す。

複数話者環境においても発話の弁別が可能であるとユーザーが実感できるようなシステムの開発を目指す。

3. 音源分離の仕組み

音源分離とは混合音を個々の音源信号に分離する技術のことである。

音源分離のアプローチには独立成分分析に基づくもの、クラスタリングに基づくもの、マルチチャネルウィナーフィルタを用いるもの、深層学習に基づくものなどがある[1]。本研究では深層学習に基づくアプローチを主に行い、深層学習で得た話者特徴に対してクラスタリングを適用してクラスタリングに基づくアプローチで後処理を行っている。また、音源分離をする際には録音全体を音声セグメントと非音声セグメントに分ける音声分割と「誰が、いつ話したのか」を特定する話者ダイアライゼーション[2]を行う。

今回使用した pyannote[3]は音声か無音かを検出する音声区間検出、音声の中の話者が変わるタイミングを検出する話者交代検出、2人以上が同時に話している部分を検出する重複話者検出、発話セグメントを「誰が話したのか」でグループ化する話者のグルーピングを行って音源分離をしていた[4][5]。

4. 提案手法

ChatGPT[6]を利用して GitHub にある pyannote をもとに、リアルタイムで音声を分離できるシステムを作成した。

変更点

- (1) マイクでの入力を可能にする。

†奈良高等学校, Nara High School

‡奈良先端科学技術大学院大学, NAIST

- (2) リアルタイムでの音声の分離をできるようにする。

1 秒のチャンクを連続録音し、録音データをバッファリングし、フレーム単位で次の処理に渡す。

録音フレームをリアルタイムで処理する。

録音チャンクを一時ファイルに保存して推論し、話者区間（発話開始・終了時間+話者 ID）を取得する。

- (3) 分離度合いの感度の調整をできるようにする。

パイプライン設定で感度調整をする。

Flask 側にフォーム UI を作って、閾値の設定をユーザーが変更できるようにする。

- (4) スピーカーでの出力を可能にする。

分離された話者の音声データを wave ファイルに一時保存、または直接スピーカーへ出力する。

非同期的に再生処理をする。

再生キューを用意し、話者ごとの音声をキューに追加しながら連続再生をする。

- (5) web アプリでの利用を可能にする。

Flask を使ったルーティング処理のプログラムを作成する。

web 画面で話者名や時間を表示できるようにする。

HTML で web アプリの外観を整える。

5. 結果

pyannote をリアルタイムで音源分離をする話者分離プログラムに書き換えて、雑音のある場所で 3 人の会話を分離したところ抽出したい話者の発言を鮮明に聞くことができた。

普段、大人数の会話を聞き取るときはいろいろな音が混ざって聞こえてくるが、システムを介して聞くと耳に入る聴覚情報が取捨選択されていたので鮮明に聞き取ることができた。

外部の音を遮断しないとシステムを介してない聴覚情報も耳に入ってしまうため、話者分離の効果はあまり感じなかった。

6. 考察

選択的聞き取り支援システムの開発によって任意の話者の発言のみ抽出することができるといったことがわかった。

また、チャンクの秒数をなるべく少なくすることで一度に音源分離される秒数が減るため、よりリアルタイムで分離しているかのように聞こえると考えられる。

ヘッドホンやイヤホンでシステムの出力音を聞くようにすれば外部の音がある程度遮断され、聞き取りやすくなると考えられる。

web アプリでのシステムの利用をできるようにしたため、システムの外観を自由に整えることができ、操作性が上がったと考えられる。

7. 今後の展望

スマートフォンなどの電子機器を用いて使用するため音声を聞き取る際に、外部の音も一緒に聞き取ってしまう。そのことを防ぐためにヘッドホンやイヤホンを着用して外部の音を遮断する必要がある。完全に遮断すると会話が成立しない可能性があるため検討する必要がある。また、第三者からの評価を受けることで改善点を見つけることができ、よりよいシステムにすることができるだろう。さらに、製品としての使用が可能になれば、手軽にどこでも利用ができるようになり、選択的聞き取りが困難な人の豊かな会話を支援することができるだろう。そして、授業中の使用により、選択的聞き取りが困難な生徒への支援を可能にすることが見込まれる。

8. 謝辞

本研究は NAIST stella プログラムの援助により遂行されたものです。この場を借りて深く御礼申し上げます。

9. 参考資料

- [1]伊藤信貴,板東宜昭.”音源分離”.知識ベース知識の森.2024, pp. 1089-1091.
- [2] Hervé Bredin.” Speech segmentation and speaker diarization”. Universite du Mans.uMotion,2023, <https://umotion.univ-lemans.fr/video/9513-speech-segmentation-and-speaker-diarization/> ,(参照 2025 年 7 月 24 日).
- [3] Hervé Bredin 他. “pyannote-audio: Neural building blocks for speaker diarization”.GitHub.2025,<https://github.com/pyannote> ,(参照: 2025 年 7 月 24 日) .
- [4] Hervé Bredin. “pyannote-audio: neural building blocks for speaker diarization” . arXiv:1911.01255.2019, <https://arxiv.org/pdf/1911.01255> (参照: 2025 年 7 月 24 日) .
- [5] AssemblyAI, “Speaker Diarization using pyannote-audio | How to get started + Demo,” YouTube, 2023 年 2 月 14 日公開, https://www.youtube.com/watch?v=37R_R82lfwA (参照: 2025 年 7 月 24 日) .
- [6] OpenAI, “ChatGPT,” <https://chatgpt.com/>, (参照 : 2025 年 7 月 24 日) .