

音声感情認識を用いた AR 感情可視化システム

2531154 山本梓莉 成見哲研究室

1 はじめに

話者と聞き手の感情の受け取り方には差が生じることが時々あるものの、自分の感情表現の癖や不足している点に気付く機会は多くない [1]。プレゼンテーションや接客などの場面では、話の内容だけでなく、声に込められた感情が聞き手の印象や理解に大きな影響を与える。また、日常的なコミュニケーションにおいても、感情を適切に表現することは円滑な意思疎通や信頼関係の構築につながる。そのため、自身の感情表現を客観的に確認し、改善できる環境が求められている。

一方で、現在の発声練習や話し方のトレーニングでは、指導者からの評価や自己評価に頼る場合が多く、リアルタイムに感情表現を客観的にフィードバックする仕組みは限られている。特に、音声から推定した感情を利用者自身が直感的に理解できる形で提示し、繰り返し練習に活用できるシステムは十分に整備されていない。また、他者から直接評価を受ける環境では、周囲の目を気にして自然な発話が難しくなる場合もある。

そこで本研究では、自己の発言に対する評価を利用者本人のみに提示することで、他人からの評価を気にすることなく、自身の感情表現を客観的に確認・改善できる環境を提供することを目指す。音声感情認識と AR 技術を用いて、話者の感情をリアルタイムに可視化し、感情表現のトレーニングを支援する。

2 関連研究

2.1 スピーチのトレーニング支援システム

Wang[2] らは、スピーチにおける音声変調（ピッチ、音量、速度、間など）のスキル向上を支援するトレーニングシステム「VoiceCoach」を提案している。本システムは、ユーザが発声練習を行う際には、手本となる事例と自身の音声のピッチや音量の差をリアルタイムかつ定量的に視覚化し、直感的なフィードバックを提供する。専門家へのインタビューおよびユーザスタディの結果、具体例の提示と視覚的なフィードバックが、初心者には効果的なスキル習得と反復練習を促進し、プレゼンテーションにおける音声表現の向上に寄与することが示されている。

2.2 音声の感情を可視化した字幕システム

Pataca[3] らは、聴覚障害者のオンラインコミュニケーション支援を目的として、音声の韻律や感情を文字の色・大きさ・太さ・配置などで視覚化する字幕システムを提案した。韻律や感情を反映した複数の字幕表現を比較した結果、感情表現を取り入れた字幕は従来の字幕よりも話者の感情を理解しやすくし、特に感情モデルは可読性を大きく損なうことなく高い効果を示した。

2.3 本研究との違い

これらの研究では、音声情報を視覚化することで発話スキルの向上やコミュニケーション支援を実現している。本研究では、音声感情認識の結果を AR 空間上にリアルタイムで可視化することで、利用者が自身の感情表現を客観的に把握し、繰り返し練習を通して感情表現の向上を支援するシステムを提案する。

3 提案システム

システムの概要を図 1 に示す。本システムでは、利用者がマイクに向かって発話すると、発話が途切れたタイミングまでを一つの音声ファイルとして取得する。発話の終了は一定時間無音状態が続いたことを検知することで判定し、発話ごとに処理を実行する。取得した音声に対して、感情解析と文字起こしを並列に行い、可視化アプリで表示する。

3.1 感情解析アプリ

感情解析には、日本語音声基盤モデル「くしなだ」[4]を用いており、文字起こしには Whisper[5]を用いている。感情解析では、「怒り」「喜び」「悲しみ」「中立」の4種類の感情について、それぞれの感情が含まれる割合を推定する。最も高い割合を示した感情を、その発話の代表的な感情として可視化に利用する。(図 2)

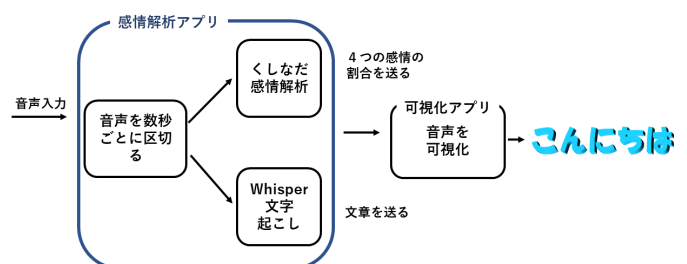


図 1 システム概要



図2 感情解析アプリ

くしなだは、大量の日本語音声データを用いた自己教師あり学習により事前学習された音声基盤モデルであり、音声認識や音声感情認識などのタスクに適用できる。本研究では、感情認識用に学習されたモデルを用いて発話音声から感情を推定し、その結果を Unity 上で視覚化する。

Whisper は、OpenAI が開発した Transformer ベースの自動音声認識モデルである。多言語・多様な環境の音声データを用いて学習されており、日本語を含む複数言語の音声を高精度にテキストへ変換できる。本研究では、発話音声を文字起こしし、くしなだによる感情認識結果と組み合わせて Unity 上で視覚化するために利用する。

3.2 可視化アプリ

感情解析の結果と文字起こしされた文章は、UDP 通信を用いて Unity で作成した可視化アプリへ送信される。Unity 側では受信したデータをもとに、文字列の生成や色の変更、アニメーションの制御を行う。解析された感情に応じて 3D 文字を生成し、「怒り」であれば赤、「喜び」であれば黄色、「悲しみ」であれば青、「中立」であれば白で表示する(図3～図6)。これらの色の割り当ては、感情と色彩の対応関係に関する先行研究 [6], [7] を参考に設定した。

さらに、可視化された 3D 文字は、利用者が発話した方向へ飛び出すようなアニメーションを付与することで、音声の感情を直感的に把握できるようにしている。字が飛び出す速度は発話速度に応じて変化し、速く話した場合は文字も速く移動し、ゆっくり話した場合は文字もゆっくり移動する。利用者は自身の発話と可視化結果をリアルタイムに確認することで、感情表現が意図したとおりに伝わっているかを客観的に振り返り、繰り返し練習に活用することができる。

4 おわりに

本研究では、自身の感情表現を客観的に理解し、感情表現の改善を支援することを目的として、発話音声から推定した



図3 怒りの文字



図4 喜びの文字



図5 悲しみの文字



図6 中立の文字

感情をリアルタイムに可視化するシステムを提案した。アンガーマネジメントなどの感情制御や、感情表現が苦手な人の円滑なコミュニケーションに寄与することが期待される。

今後の予定として、制作したシステムを実際に利用した際に、自分の音声に含まれる感情を理解できるかを検証するための被験者実験を行う。実験では、こちらで用意したシナリオに沿って被験者に感情を込めて発話してもらい、本システムを用いることで自分の声に感情を乗せやすくなるか、またコミュニケーションにおいて感情を理解しやすくなるかを調査する。また実際の対人コミュニケーションの場で使うために、AR での実装も行う。

参考文献

- [1] Carlos Busso and Shrikanth S. Narayanan. The expression and perception of emotions: comparing assessments of self versus others. In *Interspeech 2008*, pp. 257–260, 2008.
- [2] Xingbo Wang, Haipeng Zeng, Yong Wang, Aoyu Wu, Zhida Sun, Xiaojuan Ma, and Huamin Qu. Voicecoach: Interactive evidence-based training for voice modulation skills in public speaking. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, pp. 1–12, 2020.
- [3] Caluã de Lacerda Pataca, Matthew Watkins, Roshan Peiris, Sooyeon Lee, and Matt Huenerfauth. Visualization of speech prosody and emotion in captions: Accessibility for deaf and hard-of-hearing users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.
- [4] Intelligent Media Processing Research Team. kushinada-hubert-large-jtes-er. <https://huggingface.co/imprt/kushinada-hubert-large-jtes-er>, 2025. Accessed: 2026-07-17.
- [5] OpenAI. Whisper. <https://github.com/openai/whisper>, 2022. Accessed: 2026-07-17.
- [6] Mitsuhiro Hanada. Correspondence analysis of color – emotion associations. *Color Research & Application*, Vol. 43, No. 2, pp. 224–237, 2018.
- [7] Fumiyo Takahashi and Yasuhiro Kawabata. The association between colors and emotions for emotional words and facial expressions. *Color Research & Application*, Vol. 43, No. 2, pp. 247–257, 2018.