

声と顔の“違い”を見える化

～ 全国話者・環境・顔表情データを活用した 音響分析とAI評価 ～

概要

全国3,700名超の話者による音素バランス文(ATR503文)の読み上げ音声を用いて、東京と大阪のf0傾向(ピッチ)差異を可視化し、言語特徴の地域差を定量的に提示します。顔表情画像と環境音を組み合わせたマルチモーダルデータを用いて、感情推定AIによる表情認識の精度変化を検証。背景音の有無が推論結果に与える影響について、定量的な比較結果を提示します。

特徴

- **全国規模の話者音声データを用いた方言イントネーション分析**
地域別f0差異(ピッチ)を定量的に可視化し、言語研究・教育支援に応用
TTSモデルの地域適応や自然性向上に貢献
- **環境音×顔表情のマルチモーダル評価**
背景音有無による感情推定AIの変化を比較し、現場ノイズ下での実用性を検証
- **高品質(48kHz音声、喜怒哀楽表情)で、実験・AI学習にすぐ使える整備済み製品**として提供中

今後の展開

- 現在提供中の音声・顔表情データを活用し、実運用環境を模した評価実験の拡充や、現場課題に基づく製品設計の可能性を検討中です。また、教育・福祉・遠隔対話といった分野での利活用を想定し、視線・動作などの情報との組み合わせによるマルチモーダル化も視野に入れています。

テーマ「社会課題と向き合う科学技術の最前線」との関連

- 高齢者や非母語話者への支援、感情や状況をくみ取る対話インタフェースなど、社会実装が求められる領域に向けて、コーパス製品を提供。多様な話者・表情・環境を網羅することで、現場で使えるAI設計の「共通基盤」として活用できることをコーパス事業部として提案しています。

