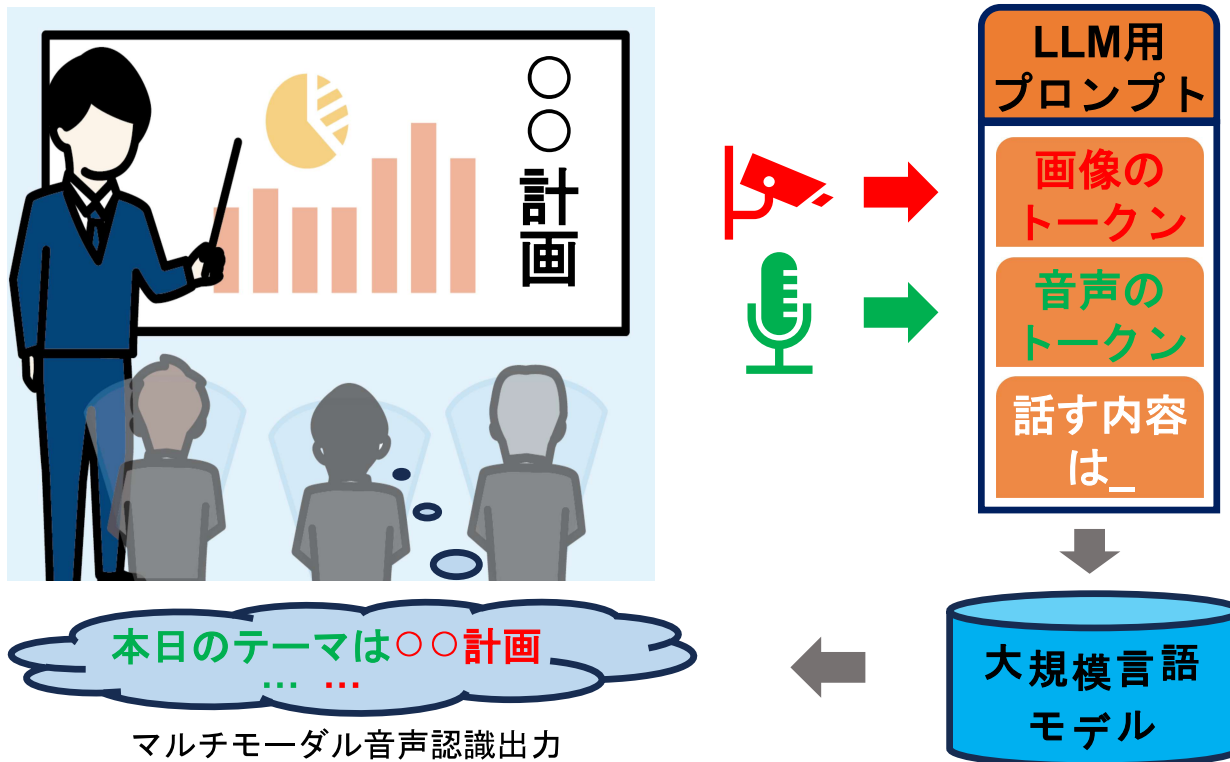


マルチモーダル音声認識システム

～ 音声だけでなく他の情報も活用して音声認識の精度を向上 ～

概要

マルチモーダル音声認識システムは、大規模言語モデルを活用することにより、音声だけでなく会場の映像に含まれる文字情報等他の情報も活用することで、発話内容を高精度に認識するシステムです。



特徴

- 発音が似ている言葉の自動聞き分けが可能
- 難解な語や使用頻度の低い語を他の情報から認識
- 方言、外国アクセント、発音ミスなどに対応

ユースケース

- 学会等のプレゼンテーションの音声認識
- 公式会見など別の資料がある発表での音声認識
- 筆談と音声会話を組み合わせたコミュニケーション

今後の展開

- 話者の話し方や表情による発話内容の予測
- 認識精度の改善
- 実環境でのテスト

【お問合せ先】

ユニバーサルコミュニケーション研究所 マルチモーダルAIコミュニケーション研究センター 多言語・多文化コミュニケーション研究室
Mail : {xiaolin.wang, mutiyama}@nict.go.jp 王 曉林 内山 将夫

NICTオープンハウス2026

Copyright © 2026 NICT All Rights Reserved.