新技術説明会
New Technology Presentation Meetings!

音声認識性能の予測技術

立命館大学　情報理工学部　情報理工学科
講師　福森 隆寛

2024年10月3日

はじめに

- 自己紹介：福森 隆寛
  - 2015.3 立命館大学 大学院博士後期課程 修了
  - 2015.4-2020.3 立命館大学 情報理工学部 助教
  - 2020.4-現在 立命館大学 情報理工学部 講師

- 専門分野：音声情報処理（主に**発話状態の推定**）
  - 叫び声検出、感情認識、**音声認識性能の予測**

【本日のテーマ】
音声認識システムの性能評価に関する作業負担を軽減する方法

音声認識

- コンピュータが人間の音声を理解し、テキストに変換したり、それに応じて特定の動作を実行する技術

- 応用例
  - スマートホンの音声アシスタント
  - 医療従事者による電子カルテの情報入力
  - リアルタイム自動翻訳

音声認識の難しさ

- 話し方や周囲の環境によっては
  人の声を正確に聴き取れない場合がある
- 音声を聴き取りにくくする要因
  - **話し方**
    - 方言、滑舌、イントネーションなど
  - **発話環境**
    - 雑音の有無、種類、量など

音声認識の難しさを表す指標

- 音声認識性能
  - 音声認識システムが人の音声を
    どれぐらい正確に認識できるかを定量化
  - 認識しやすい（あるいは認識しにくい）場所や話者を
    明らかにできる
    - 音声認識性能から適切な改善策を講じられる

- 音声認識性能の指標例
  - 正解率、誤り率など

音声認識性能の測定方法

【評価音声の収集】

- おはよう
- こんにちは
- さようなら

…

または

- おはよう
- こんにちは
- さようなら

…

評価対象の場所における音声を
<u>大量に収録</u> または **<u>計算機上で大量生成</u>**

【音声認識実験】

【音声認識性能】
正解率 80%

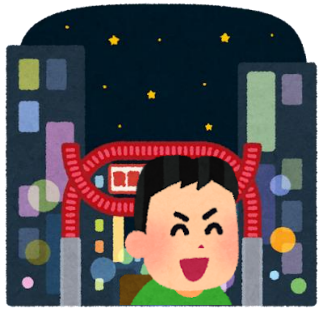
- 評価音声を音声認識システムに入力
- 実験結果から音声認識性能を算出

【問題点】音声認識性能の測定にかかる膨大なコスト

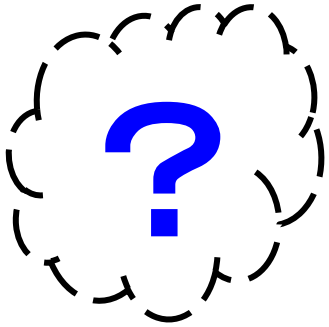
評価対象の話者や場所が変わる度に大量の音声を収録/模擬する必要がある

音声認識性能の予測

- 何かしらの方法で
  対象の場所や話者の音声認識性能を予測
- 音声認識性能を予測できると
  - 評価音声の収集の負担を軽減
  - 音声認識実験が不要



話者の音声や
発話環境に関するデータ



何かしらの方法

【音声認識性能】
正解率 60%（予測値）

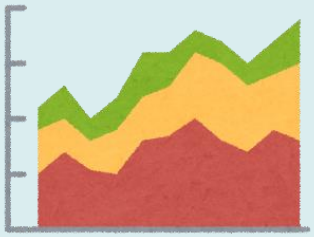
従来技術とその問題点

- 発話環境の雑音から音声認識性能を予測

【発話環境の雑音を計測】

- 対象の場所の雑音（騒音や残響など）を計測
- **音声収集が不要**

【音声認識性能の予測】



【音声認識性能】
正解率 65%
（予測値）

- 計測した雑音を分析
- 分析結果から音声認識性能を予測

【問題点１】 雑音を計測する**専用機材**や分析のための**専門知識**が必要
【問題点２】 認識しやすい/認識しにくい**話者**までは分析できない

新技術の特徴･従来技術との比較

- 深層学習技術を用いて
  少量の評価音声から音声認識性能を予測

【発話環境で音声収録】

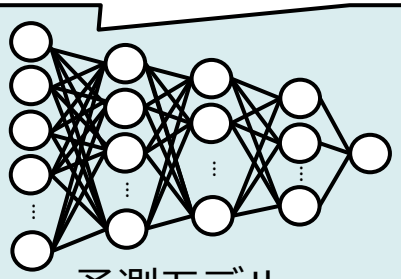
対象の場所と話者

携帯端末などで収録

対象の場所で**雑音を含む短い音声**を収集

【音声認識性能の予測】

様々な場所の音声とその認識性能を予め用意して
それらを用いて深層学習モデルを構築



予測モデル

【音声認識性能】
正解率 65%
（予測値）

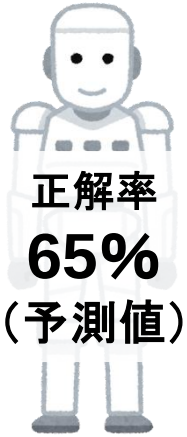
収録音声の認識性能を予測

従来技術と比較した新技術の強み

① **少量の音声**で評価できる（専用機器や専門知識が不要）
② 認識しやすい/しにくい**場所と話者**を分析できる

想定される用途

- 音声認識技術を搭載したサービス全般
  - 音声認識システムの性能を**簡便に評価**できる
  - 数ある音声認識システムの中から**話者や発話環境に最適**なものを発見できる

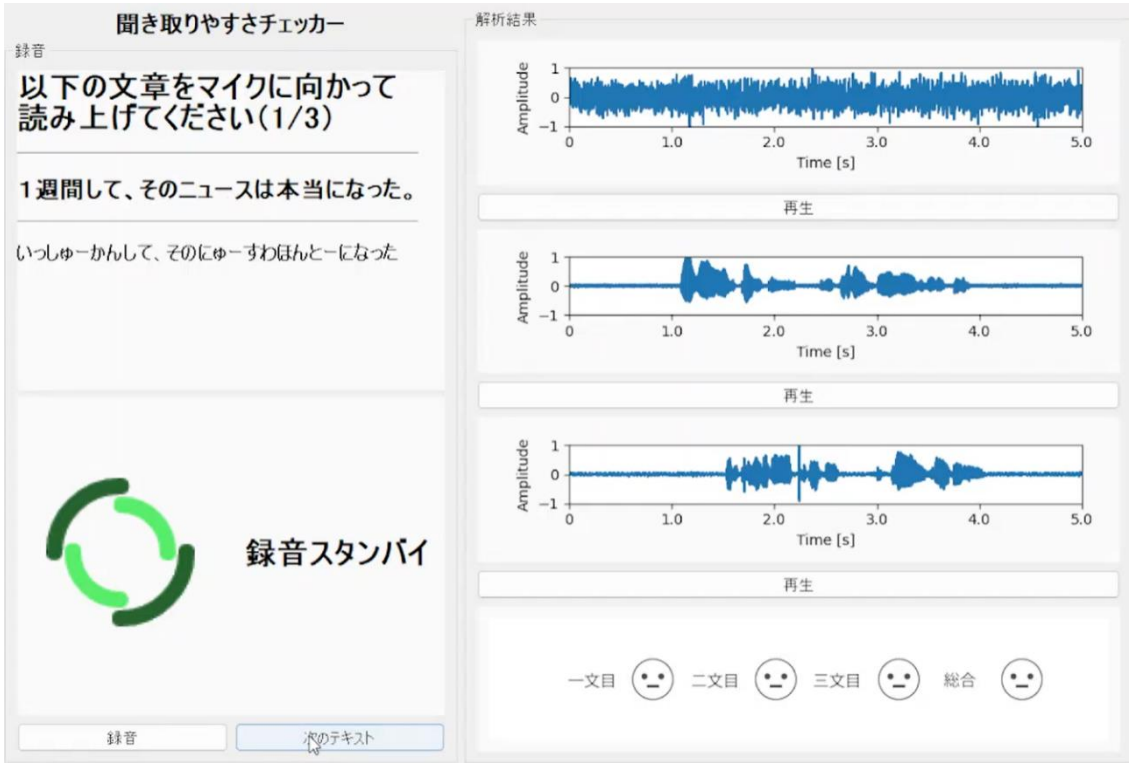


利用者や発話環境と相性の良い音声認識システムが簡単にわかる

想定される用途

- 発声訓練システム
  - 会話音声を多角的に評価
    - カラオケ採点の日常会話のようなイメージ
  - 明瞭な音声獲得に向けた発声の矯正指南
- 医学分野への展開
  - 音声の聴き取りやすさに起因する特徴の医学的な解明
  - 発音障碍者のためのリハビリテーションツール

音声認識性能の予測に関するシステム開発



3つの文を読み上げた音声を録音して
各音声と総合的な聴き取りやすさを分析

自分の声を最も正確に聴き取れる
音声認識AIを見つける

※ 下記の研究費や研究助成金などを活用してシステムを開発

- 2021年度 カワイサウンド技術・音楽振興財団 サウンド技術振興部門研究助成金 （代表）
- 2024-2026年度 日本学術振興会 学術研究助成基金助成金 基盤研究C（代表）

実用化に向けた課題

- 音声認識性能を予測する一連の処理は実装済みだが、実用化の可否を判断できるほど十分な実験データを確保できていない
  - 予測可能な音声認識システムは現状3種類のみ
    - **オープンソース**の音声認識システムのみ
    - 商用システムを利用した実験は**資金**が必要
  - 音声や発話環境の多様性が限定的
    - 予め決められたタスクの音声が中心（例：講演）
    - **日常生活の音声**は評価できていない

企業への期待

- 希望する共同研究先
  - 多様＆大量な音声を保有または収集可能な企業
  - 音声認識システムを開発している企業
  - 発声訓練など、音声の明瞭性に注目している企業

- 本技術の導入が有効と思われる企業
  - 音声認識システムを開発中/導入予定の企業
  - これまで評価したことのない場所や話者でも
    その場で音声認識性能を瞬時に算出できる
    - 例）営業先等で自社システムの性能を提示できる

企業への貢献、PRポイント

（本技術はあくまでも音声認識システムの開発プロセスの一部です）

- 音声認識に関する一連の技術や知見を保有
  - 音声データの収集・分析・前処理
  - 音声認識モデルの構築
    - 特徴抽出、モデル構造や学習方法の決定など
  - 音声認識モデルの性能評価
    - 音声認識性能の予測も含む
  - 音声認識に関するデモシステムの開発
- 音声認識の導入に向けた技術指導も可能

本技術に関する知的財産権

- 発明の名称　：
  音声認識性能の予測システム、学習モデルの構築方法、及び、音声認識性能の予測方法
- 出願番号　　：2019-114876
- 出願人　　：学校法人立命館
- 発明者　　：福森 隆寛、西浦 敬信

お問い合わせ先

立命館大学
研究部 OICリサーチオフィス

ＴＥＬ　072-665-2570
e-mail oicro@st.ritsumei.ac.jp