

# 日本の古文用音声認識システムの構築と活用に関する研究

古文用音声認識システム研究会

研究代表者：文学研究科 戸塚史織

## 1. 研究背景

### □ くずし字

- 古典籍、古文書などの前近代資料に使われてきた文字
- 日本には数百万の古典籍・古文書があり、近年多くの機関がインターネットを通じてこれを公開し始めているが、くずし字で記されているため現代の多くの人には読めない
- **くずし字翻刻**
  - くずし字を現在使用されている文字に変換する

### □ 現状

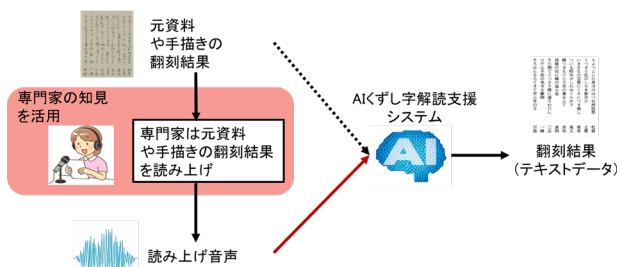
- 翻刻の負担が大きいため、解読支援システムが求められている
- OCR(光学文字認識)くずし字解読技術がよく使われている
  - 性能が十分でない場合がある
    - ・ データの不足や不均衡、原本の損傷などが、OCRくずし字解読技術の性能に大きく影響している
  - 対応が難しい領域がある
    - ・ くずし字解読を得意とする人々の多くは、使い慣れたアナログ的な方法の方がより効率的に翻刻できる
    - ・ 翻刻に慣れた人が過去に自分の手元に手書きで残した翻刻を、テキストデータ化することが求められている

### 目的

- アナログ的手法による翻刻作業の支援向上
- AIくずし字解読支援システムの高精度化および効率化

### □ 音声を活用したくずし字解読支援システムを提案

- 画像情報だけでなく、音声情報を活用した支援手法を検討
  - ✓ アナログ的な翻刻方法もより良く支援
  - ✓ 専門家の知見(読み上げ音声)の活用によるシステム性能の向上



## 2. マルチモーダルくずし字翻刻支援システムの提案(今年度)

### □ 画像および音声情報を併用してくずし字翻刻を支援

- OCRくずし字認識
  - 視覚的な曖昧さ(字形の類似など)により、画像情報のみでは誤認識が発生する場合がある
- 音声認識(ASR)
  - 同音異義語などにより、音声情報のみでは正しい文字を特定できない場合がある
- OCRから得た文字候補を音声情報で絞り込む、かなレベルでの融合手法の提案

#### ◆ Step 1: 画像・音声ペアの準備

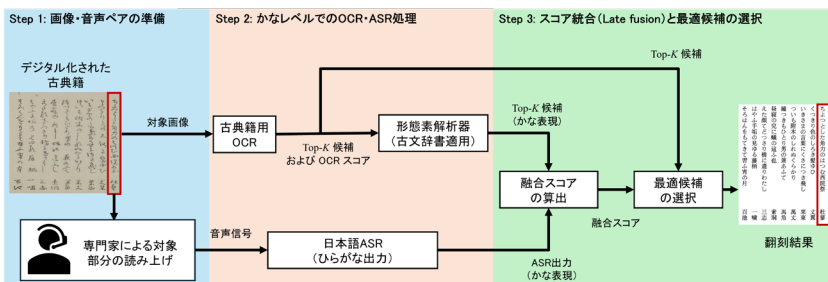
- 専門家がデジタル化された古典籍から対象箇所を選定し、読み上げ音声データを収録

#### ◆ Step 2: かなレベルでのOCR・ASR処理

- 対象画像を認識し、OCR推論時にビームサーチによりTop-K候補文字列とOCRスコアを出力
- 形態素解析ツールにより、Top-K候補文字列を自動的にひらがな列へ変換
- 収録音声をもASRで認識し、ひらがな列結果を出力

#### ◆ Step 3: スコア統合と最適候補の選択

- OCR各候補とASR出力の発音類似度を算出
- 融合スコアを算出
- 最適候補を選出



## 3. 評価実験

- 提案手法の有効性を検証する
- テストデータ: 洒落本作品「聖遊廊」の一部(専門家が対象箇所を読み上げ、音声を収録)
- 評価指標: 文字誤り率(CER) (単位:%、値が小さいほど高精度)

### ● Methods

- OCR(Baseline): OCRシステムのみ利用
- Fusion(Proposed): 提案するマルチモーダル手法
- OCR Top-5(Oracle): Top-5におけるOCRの理論上限

Method	Mean	Std	Median
OCR (Baseline)	27.1	12.6	26.0
Fusion (Proposed)	25.3	12.8	25.0
OCR Top-5 (Oracle)	24.3	12.9	23.5

## 4. 今後の計画

- 古文音声学習データの追加収録
- 構築した古文用ASRモデルを活用し、OCR Top-K候補に依存しない文字レベルの融合を検討
- 翻刻現場での実用化に向け、ソフトウェア(アプリケーション)を開発