

## はじめに

### 国立国語研究所の共同利用型共同研究(登録型)

- 研究資料・言語資源・分析装置を活用して研究を行うことができる制度



### 本発表

- データベースを活用した言語研究手法の共有
- アーカイブIDの使用
- コロナ禍における研究の促進

## 分析対象のデータベースについて

国立国語研究所研究資料室収蔵資料にある250以上の研究プロジェクトのうち、「fo0245日本語音声における韻律的特徴の実態とその教育に関する総合的研究」の実験録音音声データ

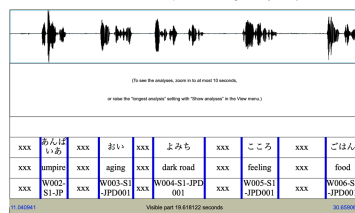
- 被験者**
  - 日本各地の方言話者
  - 小学生、中学生、若年層、壮年層、中年層、高年層
- 刺激**
  - 五十音、名詞、動詞、形容詞、文章、童話(桃太郎)の朗読、数字(1から9までと四桁の数字)
- 実験手順**
  - 練習
  - 紙に印刷された刺激の読み上げ

## データ処理について

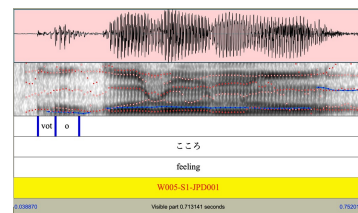
### 共同研究でのトピック: 日本語方言間の有声性の対立

- 被験者: 東北(M:F=10:0)・東京(M:F=5:5)方言話者10名ずつ(計20名)、高齢層(1910~1930年生まれ)
- 刺激リスト: 録音音声をもとに作成、名詞・文章のセクションに注目
- アーカイブIDの作成: “単語ID”-“繰り返し番号”-“実験内のセクションID”-“被験者ID”
  - 例: 「リスト上の一つ目の単語(W001)」で「セクション番号S1」、「被験者IDが001(JPD001)」→“W001-S1-JPD001”

### 2. アーカイブIDのラベリング

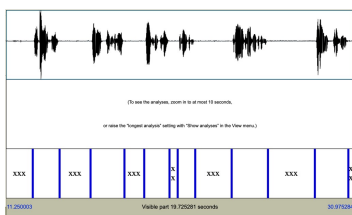


### 4. アノテーション

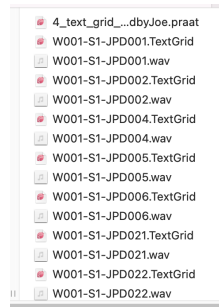


Praat (Boersma and Weenink, 2022)及びPraatスクリプトを使用することで半自動的な処理が可能

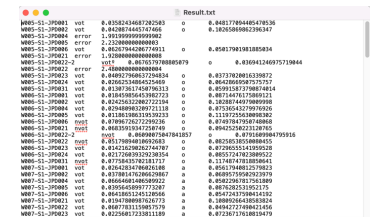
### 1. 境界の配置



### 3. 個別ファイルに分割



### 5. データの抽出



抽出されたデータ(.txt)はエクセルから開いて加工し、R(R core team, 2022)で読み取り可能な形式で保存する(.csv)

## おわりに

- アーカイブIDの使用が重要**
  - 将来の研究への貢献
  - 先行研究の再現性の確保
- 現在の社会情勢を踏まえると、まだ新たな実験を対面で行うことは簡単ではない
  - データベースを活用した研究の重要性が高まる
- 将来より多くの研究でデータベースが活用される際、
  - オリジナルのデータベース
  - 識別可能なIDが付与された処理データ
- がアーカイブされていたとしたら、研究トピックが異なっていたとしても、将来行われる研究の大きな促進につながる

## 謝辞

本発表は国立国語研究所の共同利用型共同研究プロジェクト「日本語の有声性の対立への複数の音響指標の影響」(研究代表者:李勝勲)の研究成果である。また、第一著者は国際基督教大学より「国際基督教大学博士研究員(A種)」の助成を受けている。

## 参考文献

- Boersma, P. and Weenink, D. (2022). "Praat: doing phonetics by computer [computer program] (version 6.2.14)".
- R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.