

# 音声認識API「Whisper」を用いた歴史的音源の文字起こしに関する研究

Research on Transcription for Historical Sound Sources  
using Whisper as Speech Recognition API

古川 亜澄馬  
指導教員 三輪 賢一郎

サレジオ工業高等専門学校 機械電子工学科 情報コミュニケーション研究室

SPレコードに録音されている「語りもの」の音源を文字の形で保存することには大きな意義がある。先行研究において音声認識API「Whisper」が大正期の歴史的音源に対して文字起こしシステムとしての可能性を持つことが確認できたことから、本稿では同手法の普遍性の評価について述べる。

キーワード：音声認識, 歴史的音源, 文字起こし, Whisper

## 1. 緒言

明治・大正期から昭和初期にかけての演説や講談などの「語りもの」の音源がSPレコード(Standard Playing Record)に数多く遺されている。これらの膨大な歴史的遺産を文字の形として保存することには大きな意義がある。当時用いられたSPレコードは現存しているものの、原稿の存在は不明である場合が少なくない。そのため、文字起こしを人間が行うとなると、莫大なコストや時間、労力がかかる。そのため、省力化に向けた文字起こしの方法を検証する必要がある。本研究室では令和3年度から音声認識API(Application Programming Interface)を用いた歴史的音源に対する文字起こしの研究を行っている。中でも大正期の音源は機械式録音による録音方式が用いられており、現在普通に用いられるマイクロフォンによる電気録音と比較して著しく周波数特性やSN比が低品質である。昨年度の研究で最新の音声認識APIである「Whisper」を用いることで、特定の音源に対して文字起こしシステムとしての可能性を持つことが確認できた[1]。しかし、多種多様な音源に対する普遍性については不明である。そこで本研究では「Whisper」による音声認識

処理を大正期の別の音源に対して実行し、同手法の普遍性を評価することを目的とする。

## 2. 方法

本研究では、音声認識にはOpenAI社が提供している音声認識API「Whisper」[2]を用いた。ハードウェアにはノートパソコンを用い、開発環境にはGoogle社が提供している「Google Colaboratory」[3]とMicrosoft社が提供している「Visual Studio Code」[4]を用いてPythonによりAPIを実行した。本実験で使用した音源は「国立国会図書館デジタルコレクション「歴史的音源」」[5]に所蔵されている「憲政に於ける世論の勢力(一)」(大隈重信、コロムビア、大正4年、収録時間3分7秒)を国立国会図書館の許可のもとに使用した。

各音源の略称の対応表を表1に示す。うち音源Bは後藤新平による大正期の演説で、昨年度実施された先行研究で用いられたものである。

表1 各音源の略称の対応表

音源A	憲政に於ける世論の勢力(一)
音源B	政治の倫理化(一)

実験結果の評価指標として、漢字ベースの各文字ごとに着目し文字認識率の算出を行う。

算出式は (1) 式に示す。

文字認識率 =

$$\frac{\text{正解文字数} - \text{誤挿文字数} - \text{誤削除文字数} - \text{誤置換文字数}}{\text{正解文字数}}$$

・・・(1)

### 3. 結果

図 1 に各音源の漢字ベースの文字認識率の比較結果を示す。音源によって、認識率に 5 % 程度の差異がみられることが確認できる。

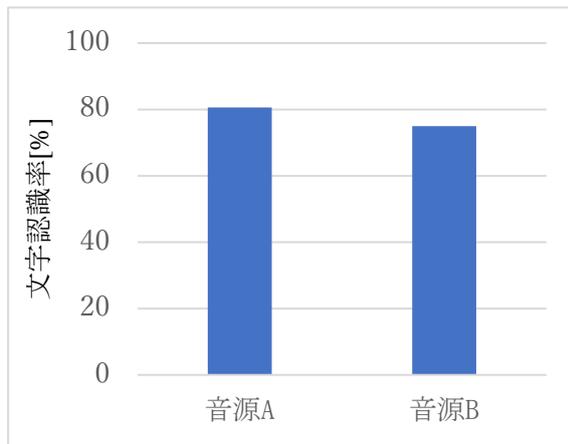


図 1 文字認識率の比較

### 4. 結言

本研究では、音声認識 API 「Whisper」を用いた歴史的音源に対する文字起こしの省力化の普遍性を検証した。結果から、音源により文字認識率に若干の差異は見られたものの、Whisper は大正期の音源に対してほぼ同水準にて文字起こしが実現できる可能性が確認できた。

### 5. 今後の予定

今後は同手法の普遍性を引き続き探るべく、同時代の歴史的音源に対してさらなる検証を実施する予定である。

### 謝辞

本研究は、国立国会図書館のご厚意により、「国立国会図書館デジタルコレクション歴史的音源」に所蔵の音源を用いております。

### 文献

- [1]山崎右京, 三輪賢一郎, “音声認識 API 「Whisper」を用いた大正期の歴史的音源の文字起こし”電子情報通信学会, 2024 年電子情報学会総合大会, D-14-01
- [2]OpenAI Whisper  
(<https://platform.openai.com/docs/guides/speech-to-text>)
- [3]Google LLC Google colab  
(<https://colab.google/>)
- [4]Microsoft Visual Studio Code  
(<https://azure.microsoft.com/ja-jp/products/visual-studio-code>)
- [5]国立国会図書館デジタルコレクション「歴史的音源」 Web サイト  
(<https://rekion.dl.ndl.go.jp/>)