

音声認識 API「Google Chirp」を用いた 大正時代の歴史的音源に対する文字起こしに関する研究

Research on Transcription for Historical Sound Sources in Taisho Era
using Speech Recognition API “Google Chirp”

塩崎 惣一郎
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た数多くの歴史的音源を文字情報として後世に遺すために、音声認識技術を用いた歴史的音源の効率的な文字起こしを検討している。本研究では音声認識 API「Speech-to-Text」内で利用できる「Chirp」モデルの歴史的音源に対する基本的な性能と実用性を検証する。

キーワード：音声認識, 文字起こし, SPレコード, Chirp

1. 緒言

SP(Standard Playing)レコードは、1887年から1963年まで生産されていた録音・再生メディアである。その中には当時の演説や落語などの歴史的価値のある音源が録音され保存されている。これらの膨大な歴史的音源を文字情報として後世に伝えるために、近年 AI によって技術が向上している音声認識技術を活用し、歴史的音源の文字起こしを効率的に行うことを目指している。

本研究室では令和4年度より音声認識 API (Application Programming Interface)を用いた文字起こしの研究を行っている。昨年度は音声認識 API「Whisper」を用いた方法により、従来の音声認識システムより大正期の音源に対する文字認識精度が向上できたことを確認した[1]。しかしながら、「Whisper」は漢字ベースでの文字認識率が閾値とされる90%に達しておらず、実用システムには程遠い。そこで本研究では、Google社より提供されている音声認識 API「Speech-to-text」で利用できる「Chirp」モデルを歴史的音源に対して適用し、Chirpの基本的な性能とその実用性を検証する。

2. 方法

本研究では、音声認識を行うにあたり、Google社より提供されている「Speech-to-Text」APIの

「Chirp」モデルを採用した[2]。ハードウェアはノートパソコンを用い、Google Cloud のサイトにて wav 形式のファイルを送信し、文字起こしされた文章を得ることにより実施した。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」」に所蔵されている「政治講演：政治の倫理化(一)」(後藤新平、ニッポンノホン、大正13年、収録時間3分39秒)を国立国会図書館の許可のもと使用した[3]。

実験結果の評価指標として、平仮名のモーラ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 =

$$\frac{\text{正解文文字数} - (\text{誤置換文字数} + \text{誤挿入文字数} + \text{誤削除文字数})}{\text{正解文文字数}}$$

…(1)

3. 結果

平仮名ベースの文字認識率の結果を図1に示す。結果から、文字認識率は72%となり、Whisperの89%を下回った。また、図2の誤差の内訳と表1の認識結果の比較から、誤削除が多く発生していることが確認された。

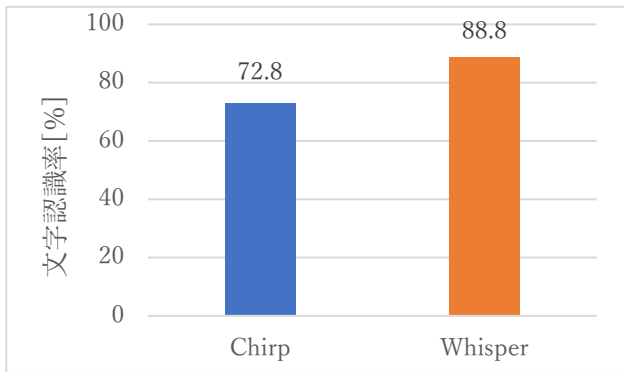


図1 文字認識率の比較

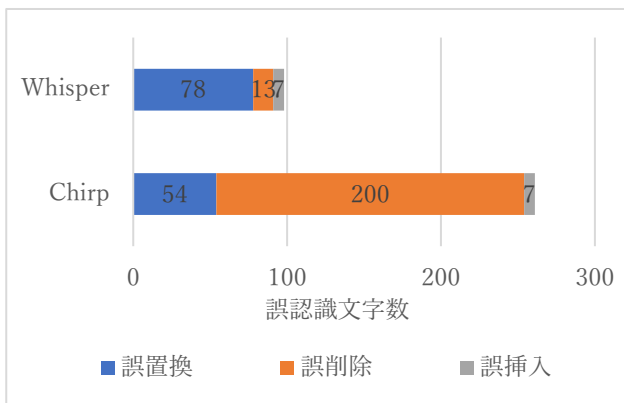


図2 誤認識の内訳

表1 認識結果の比較

Chirp	よがしゆくせきのしゅちようたるせいじのりんりかをていしようすることはもっともきんかいとするところ であります。
Whisper	よがしゆくせいのしゅそうたるせいじのりんりかをけいしようすることがもっともきんかいとするところ あります。
正解文	よがしゆくてきのしゅちようたるせいじのりんりかをていしようすることはもっともきんかいとするところ であります。

4. 結 言

本研究では、音声認識 API「Speech-to-Text」の「Chirp」モデルを用いて歴史的音源の文字起こしを行い、性能を検証した。その結果、「Chirp」モデルの文字認識率は 72%程で、歴史的音源の文字起こ

しの省力化にはあまり期待できない結果となった。今後は、別の同時代の音源による検証も進める予定である。

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」を用いた歴史的音源に関する研究”,令和 5 年度サレジオ工業高等専門学校卒業論文
- [2] Google Cloud Speech-to-Text Chirp (<https://cloud.google.com/speech-to-text/v2/docs/chirp-model?hl=ja>)
- [3] 国立国会図書館 歴史的音源 Web サイト (<https://rekion.dl.ndl.go.jp/ja/>)