

音声認識技術を用いた歴史的音源のテキスト起こしに関する研究

Research on Text Transcription of Historical Sound Sources using Speech Recognition Technology

小泉 朝陽¹⁾

指導教員 三輪 賢一郎¹⁾

1) 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

キーワード：SP レコード，文字起こし，音声認識

1. 緒 言

SP レコード {Standard Playing Record} は、1887 年から 1950 年代後半まで使用されたレコード及び録音技術である。全盛期には、流行歌、演説、講談など様々な音源の録音に使用されており、その歴史的価値は高い。

SP レコードのジャンルに演説、講談などの「語りもの」がある。それらは、SP レコード自体は現存しているものの、当時用いられた原稿の存在は不明である場合が少なくない。一部の落語などは当時の速記本が現存しているケースもあるが、必ずしも実際の録音と文言が完全一致するわけではないことが分かっている[1]。したがって、それら膨大なる文化遺産を文字の形で後世に残すためには、何らかの方法で文字起こしを実施する必要がある。

近年では AI を用いた音声認識技術の進展により、高速かつ正確な文字起こしが可能になってきた。しかしながら、AI 技術を用いたシステムは高価であり、そのシステム規模は大掛かりなものとなる。また、このような文化遺産のアーカイブ化事業は、その性質上、国の予算で対応せざるを得ないことから、極力コストがかからない方法が望ましい。そこで、本研究では、安価でコンパクトなシステム構成にて音声認識を実行し、SP レコードに遺された「語りもの」の文字起こしの可能性を検証する。

2. 方 法

本研究では、音声認識エンジンには、オープンソースである Julius[2]を用いた。ハードウェアには安価でかつ小型の計算機である Raspberry Pi 4 を用いることとし、Julius を実装して使用した。また、音声認識に必要な音響モデル、発音辞書、及び言語モデルについては、Julius のディクテーションキットに同梱されているものを用いた。

認識精度の評価指標としては、今回は仮名ベースでの認識パフォーマンスを考えることとし、モーラ（拍）に着目した文字認識率を採用した。下記に文字認識率の算出式を示す。

文字認識率 =

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

3. 結 果

本実験で用いた音源は、「国立国会図書館デジタルコレクション「歴史的音源」」[2]に所蔵されている「ソロモン海戦に就いて 1」（日本コロムビア、昭和 19 年頃、収録時間 3 分 4 秒）であり、国立国会図書館の許可のもとに使用した。なお、音源には SP レコード特有のノイズが含まれるが、ノイズ除去などの処理はせずにそのまま使用した。

加えて、認識性能の比較対象として、YouTube の自動字幕付与機能を用いた認識実験を、同じ音

源に対して実施した。

図1に各音声認識システムの文字認識率を、図2に各認識結果の誤認識文字数をそれぞれ示す。結果から、提案方式（Juliusを使用）の認識精度はYouTubeと比べて現状では著しく低精度であり、誤挿入、誤削除、誤置換文字数も多いことが確認された。

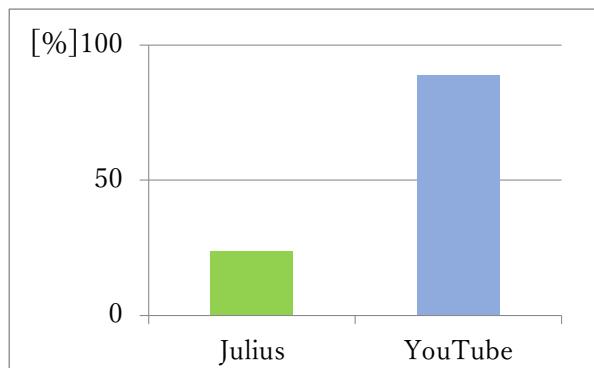


図1 各音声認識システムの認識率

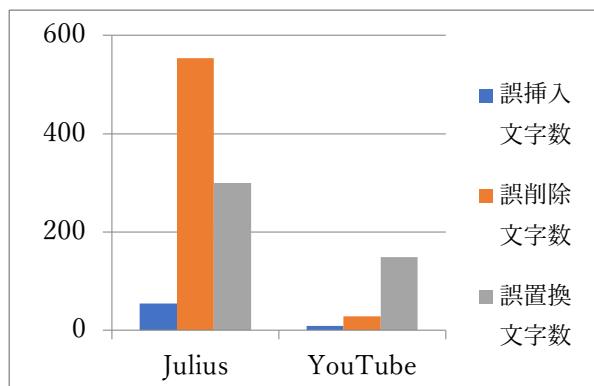


図2 各認識結果の誤認識文字数

4. 結 言

本研究では、安価でコンパクトな音声認識システムを構築し、SPレコードに遺された「語りもの」の文字起こしの可能性を検討した。結果から、提案システムの認識精度は現状では低水準であることが確認できた。この要因としては、SPレコード特有のノイズが、認識精度に少なからず影響していると考えられる。

今後は、認識精度の向上を図るべく、同じ音源に予めノイズ除去を施すことを検討する。また、さらに古い年代の音源（機械式録音）における文

字起こしの可能性についても併せて検討する。

謝 辞

本研究は、国立国会図書館のご厚意により、「国立国会図書館ディジタルコレクション 歴史的音源」[3]に所蔵の音源（同図書館ウェブサイトで公開中のもの）を用いております。

併せて、汎用大語彙連続音声認識エンジン Julius[2]、ならびに国立国語研究所の『現代日本語書き言葉均衡コーパス』(BCCWJ)を利用した言語モデルを同梱した Julius ディクテーションキットを利用しています。

参考文献

- [1] 金澤裕之, “現代に繋がる近代初期の口語的資料における言語実態：速記本と SP レコードによる東西の落語を対象として,” 国立国語研究所論集, no. 10, pp. 58–84, 2016 年 1 月
- [2] 汎用大語彙連続音声認識エンジン Julius プロジェクト Web サイト (<https://julius.osdn.jp/>)
- [3] 国立国会図書館ディジタルコレクション「歴史的音源」Web サイト (<https://rekion.dl.ndl.go.jp/>)