音声認識を用いた大正時代の歴史的音源の文字おこしに関する研究

Study on Transcription of Historical Sound Sources
in the Taisho Era using Speech Recognition

長谷川　祐士
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

キーワード：SP レコード, 音声認識, 文字起こし, ノイズ除去

1.　緒言

現代では音声認識技術が発展し、会議の議事録作成や音声入力システムなど多岐に渡って音声認識が用いられている。さらに、今後音声認識の技術は向上していくことが予想される。しかし、SP レコードに収録された音源などの歴史的音源に対して音声認識を適用した研究は他ではなされておらず、特にマイク録音が導入される前の大正時代以前の音源に対する音声認識の可能性は未知である。そこで本研究では大正時代の音源に対して音声認識を実施し、その評価を行う。

2.　手法

2.1.　使用エンジン、評価方法

本研究では使用する音声認識エンジンとして、オープンソースの Julius を使用した[1] 。音響/言語モデル、単語辞書はともに Julius のディクテーションキットに同梱されているものを使用することとした。さらに、比較対象として YouTube 自動字幕機能を用いた検証も伏せて実施した。

認識結果を評価する指標として、平仮名を使用のモーラ(拍)に着目して文字認識率を計算する指標を採用した。以下にその算出式を示す。

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

2.2.　入力音源

対象とした音源は「国立国会図書館ディジタルコレクション「歴史的音源」」[2]内に所蔵されている、「政治の倫理化」(後藤 新平、日本コロンビア、大正 13 年頃)を国立国会図書館の許可のもとに使用した。

2.3.　ノイズ除去手法

SP レコードは特性上、ノイズが発生するためノイズリダクションを実施する手法として Spectral Subtruction 法(以下 SS 法)を用いた。SS 法とは予め信号の無音声部分を分析窓で FFT 分析し、それを同様の処理を施した対象信号にスペクトル減算処理を施した後、IFFT を施してノイズが除去された音声を得る手法である[3]。

3.　認識結果

音声認識を実施し認識された結果において、ある一文をそれぞれ比較した結果を表 1 に示す。また表 1 の結果を含む認識結果をもとに算出した認識率等を表 2 に示す。

Julius を用いた認識結果であるが、ノイズ除去処理の有無にかかわらずある同じ単語が連続して出力されるなど、著しく不安定で不正確な認識結果となった。それに対して、YouTube の自動字幕機能を用いた認識結果は、同様に不正確な認識結果ではあるものの同じ単語が連続して出力されるような現象は出現しておらず、また最後の文字に限ってはかなり正確に認識されている。

ノイズ除去前と後を比較すると、除去前はサ行が比較的連続して出現したりしており、除去後はその点に限っては解消しているものの、ノイズ除去による認識精度の改善は確認できなかった。

また、全体的に認識精度が悪い原因として、音声信号の質が良くないこと、雑音が大きいこと、当研究で使用した言語モデル、単語辞書が現代の書き言葉仕様に合わせたものであることなどが考えられる。

表 1 認識結果

種類	結果
Julius(ノイズ除去前)	雑草の発想で裾呪詛 洪水、すそすそをおずおずスサノオずつ涼しさをわざわざだ 。
Julius(ノイズ除去後)	なぜ、清掃魚津は、魚住 ねはあ、それじゃ、魚津ファインS、魚座だは。
YouTube 自動字幕	あざっすあれリカオー数字は薄いこのもういいある一間アーマー失敗皿であります。
正解文	私が多年政治の倫理化を高唱する理由は、実に諸君と共にこの光栄ある歴史的大事業を完成し明治大帝にうべ奉りたいからであります

表 2 各結果の認識率（総文字数 927 文字）

種類	誤挿入 [個]	誤削除 [個]	誤置換 [個]	認識率 [%]
Julius(ノイズ除去前)	観測不能	観測不能	観測不能	算出不能
YouTube 自動字幕機能	33	422	230	26.1
Julius(ノイズ除去後)	観測不能	観測不能	観測不能	算出不能

Julius の認識結果については、上に示したように結果が正解文と非常にかけ離れており、認識結果の観測が不可能であったため、認識率の算出は行えなかった。対して、YouTube 自動字幕機能を用いた認識結果は 26.1[%]であった。しかし、これも音声認識率としては著しく低水準であり、現段階では大正時代の歴史的音源にこれらの音声認識手法を適用することは難しいことが分かった。

4. 結言

本研究では、大正期の歴史的音源の音声認識可能性を検証するため、Julius、YouTube の自動字幕機能を使用して実験を行った。結果としては、音声認識するにあたって十分な認識率を得ることができなかった。

今後はノイズ除去の手法を SS 法の代わりに一般化調和解析を用いた手法を検討する。

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』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] 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/）

[2] 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/）

[3] 松尾卓摩,長秀雄,”スペクトルサブトラクションを用いたノイズ環境下で AE 計測システムの開発,”非破壊検査第 58 巻 12 号,pp.549--550,2009

[4] 高見澤 龍児,片山 健司,神田 祥宏 他,”一般化調和解析(GHA)を用いた SP レコード再生音の雑音抑制の検討”,情報処理学会研究報告_システム LSI 設計技術,102 号,pp.1--3,2004