

車内アナウンスをすべての人に ～音声認識による車内アナウンスの文字起こし～ In-train Announcements for Everyone ～Transcription of in-train announcements using voice recognition～

永峯 蒼士†

Soshi Nagamine

前川知輝‡

Tomoki Maekawa

松本直弥

Naoya Matsumoto

峯山賢伸

Kenshin Mineyama

1. 研究背景

私たちの周りには車の走行音や信号機の音、救急車のサイレン、鳥の鳴き声、電車のアナウンスなど、日々さまざまな種類の音が飛び交っており、聴覚障害を持つ方々にとってそれらの音を正しく聞き取ることが容易ではない。私たちはこの「音」に焦点を当て、人々の支えになるような研究をしたいと考えた。その中でも、聴覚障害を持つ方々が公共交通機関を利用する際、走行音や人々の会話などにより、車内アナウンスを聞き取ることが難しく、緊急時などにアナウンスから情報を得ることが難しい。そこで、私たちは日々の生活の中で音に関して悩みを抱えている、聴覚に障害を持つ方々を対象に研究をすることにした。

2. 研究の目的

電車内では振幅の大きい電車の走行音が聞こえるため、それらの音声をノイズ除去し、正しいアナウンス音声を文字起こしする。

3. 研究方法

本研究では、聴覚に障害を持つ方々が電車を利用する場合を想定し、電車の走行音を「雑音」、アナウンス音声と雑音を含む音声を「全体音声」とする。また、本研究の音声データの編集、フーリエ変換、スペクトラルノイズ除去は全て Google Colaboratory で Python を用いて行った。

3.1 音声データの収集

ボイスレコーダーを用いて実際に電車内で雑音と全体音声を録音した。(ボイスレコーダーにはアナウンスやラジオなどの言語音声の振幅を大きく、雑音を小さく録音する性質がある)

3.2 スペクトラルノイズ除去

雑音と全体音声の音圧の差(Pa)を測定する。Python を用いて音圧差を音圧レベル(dB)に変換する。全体音声の音圧を固定し、雑音の音圧をさまざまに変化させることで 2 つの音声の音圧レベルを 0 から 50dB まで 5dB ごとに変化させて区切り、最適な音声認識が可能な音圧レベルを求めた。

* 音圧、音圧レベルについて

音圧とは、音による大気圧からの微弱な圧力の変化であ

り、振幅に比例する。音圧レベルとは、ある 2 種類の音声の音圧の比を常用対数を用いて表したものである。値が 0 から大きくなるほどアナウンス音声に対するノイズ音声の音圧が大きいことを示す。

(例) 0dB:大きさが等しい 10db:ノイズ音声のほうが大きい

3.3 フーリエ変換

全体音声と雑音を Google Colaboratory で Python を用いて「フーリエ変換」を行い、音声波形をスペクトルに変換した後、スペクトログラムに変換する。(音声データの情報を明確に可視化)

* スペクトルについて

ある 2 つの正弦波の合成波を表記するとき、合成波形ではその合成関数についての情報を得ることは容易ではないため、『スペクトル』(縦軸:振幅、横軸:角周波数)で表記する。スペクトルとは、数種類の周期関数で表記される合成関数を、複数の周期関数に分解し、(縦軸:振幅、横軸:角周波数)で表したものである。周期関数のスペクトルを「離散スペクトル」、非周期関数のスペクトルを「連続スペクトル」という。

* フーリエ変換について

非周期関数で表される(縦軸:振幅、横軸:時間)をスペクトル(縦軸:振幅、横軸:角周波数)に変換する操作のこと。

* スペクトログラムについて

スペクトルに時間要素を付け加えたもの。(縦軸:周波数、横軸:時間、色の濃さ:振幅の大きさ)

3.4 スペクトログラムの比較

全体音声と雑音のスペクトログラムを比較し、雑音が含まれている低周波数帯の音を全体音声から取り除く。

3.5 フーリエ逆変換

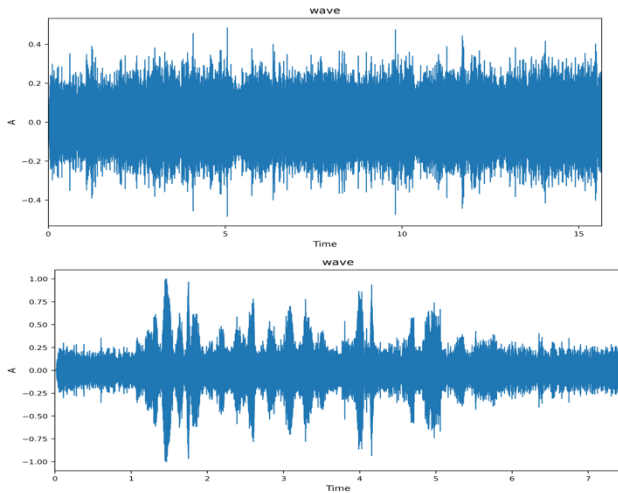
編集したスペクトログラムを波形に変換し、編集前と比

較し正しく認識できるかどうか音声認識を行う。

4. 実験結果

4.1 スペクトラルノイズ除去

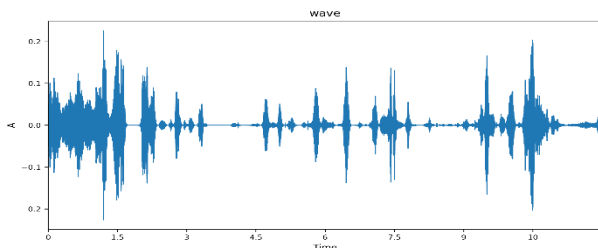
音圧レベルが増加するにつれ、認識の精度が向上し、20dBから25dBの範囲で最も正しく認識することができたが、25dBから増加するにつれ、認識の精度が低下し、文章として成立しなくなった。



雑音音

全体音

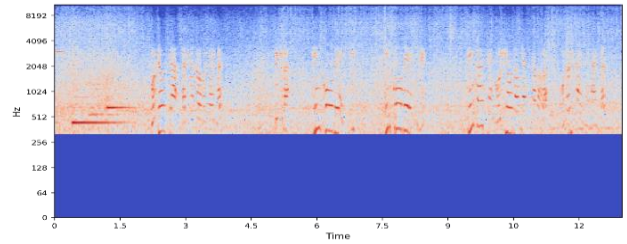
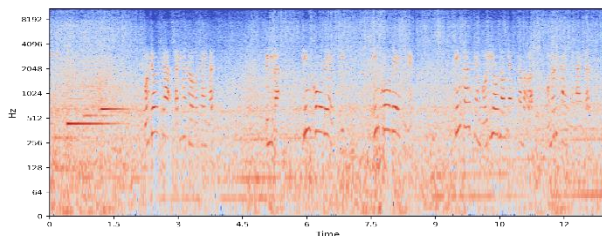
声
波形 編集前



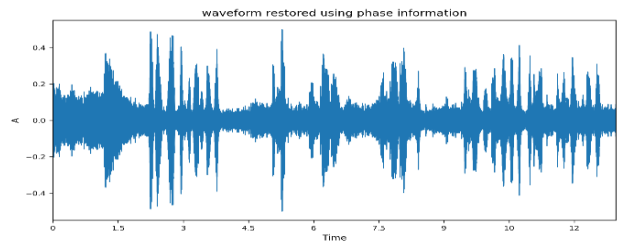
全体音声 波形 編集後

4.2 フーリエ変換

編集後は編集前と比較すると認識精度は向上したが、完全な認識は望めず、文章として成立しなかった。



全体音声 スペクトログラム 編集前
全体音声 スペクトログラム 編集後



全体音声波形 編集後

文字起こしの結果は以下ようになった。

- ・正しいアナウンス内容

「乗車ありがとうございます。次は新神戸駅 新神戸駅です。新幹線ご利用のお客様はお乗り換え願います。」

- ・編集前

「車ありがとうございます。変更でき エンコード 駅で新幹線をご利用のお客様はお乗り換えます。」

- ・編集後

「車ありがとうございます。次は新神戸 新神戸 新幹線を利用のお客様がご理解願います。」

5. 考察

5.1 スペクトラルノイズ除去について

アナウンス音声と雑音の音圧レベルが一定の環境下であれば、音圧レベルを変化させることにより正しい認識が可能であるが、電車内に雑音は走行音や乗客の会話など雑音の音圧は変動的であるため、一部雑音が残っていたと考えられる。

5.2 フーリエ変換について

周波数を基にしてアナウンス音声と雑音と区別

したため、雑音と同じ周波数帯のアナウンス音声も除去されてしまい、正しい認識が望めなかったと考えられる。

6. 結論と展望

今回の研究では音圧レベルと周波数の2種類の方法でアナウンスと雑音を区別したが、その結果完全な雑音除去は望めなかった。今後は全体音声のスペクトログラムにみられた電車の走行音と思われる周期的かつ振幅の大きな雑音を画像認識を用いて編集する「マッチング」という手法を検討する。

7. 謝辞

九州工業大学教授張力峰先生、中藤良久先生、株式会社ブレイン初田真幸氏には、本研究の遂行にあたり多大なご助言、ご協力頂きました。ここに感謝の意を表します。

8. 参考文献

- (1) <https://www.nli-research.co.jp/report/detail/id=69307?pageno=2&site=nli#:~:text=%E3%80%8C%E3%83%95%E3%83%BC%E3%83%AA%E3%82%A8%E5%A4%89%E6%8F%9B%EF%BC%88Fourier%20transform%EF%BC%9A,%E3%81%A8%E8%A8%80%E3%81%88%E3%82%8B%E>
- (2) <https://www.ai-shift.co.jp/techblog/1305>