部 門

# 競技部門

No.1 登録番号

予定開発期間:5ヶ月
予定開発人数:3人

No.2

No.3

	4月		5月			6月			7月			8月			9月		10月		]	
問題分析				•								<b>\</b>								
設計						•						<b>\</b>								
実装										<b>†</b>					<b>\</b>					
試用・トレーニング										<b>†</b>										<b>\</b>

## 実現方法

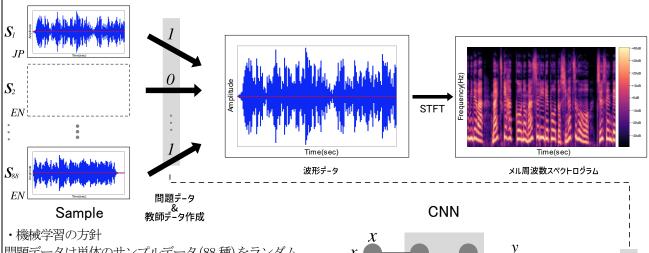
## 1) 音声の解析アルゴリズム

#### 全体の方針

機械学習を行い、合成された音声を分離する。

## ・音声データの前処理

合成された音声データ (波形データ) を STFT (短時間フーリエ変換) を用いてメル周波数スペクトログラムに変換する。本来必要であった時系列ごとの処理や、入力データである 48,000\*t(s) 個のデータを、画像化することによって、最大 縦\*横 (ピクセル) のデータ数まで削減することができ、時系列ごと (リカレントニューラルネットワーク) の処理が不要になり、CNN(畳み込みニューラルネットワーク) での処理が可能になる。



問題データは単体のサンプルデータ(88種)をランダム(3~20種)で合成して作成し、分割やずらしなども表現する。その際にどの音声が合成されたかを表す教師データも作成する。

スペクトログラム化された問題データを入力データと スペクトログラム して渡し、CNN で画像の機械学習を行い、どの音声が (画像) 使われたのかを表す 0or1 の88 個の配列を出力する。 問題データを作成した際に生成された教師データと出 力データとの差で誤差をとり、学習を行う。

 $X_1$   $X_2$   $X_3$   $X_4$   $X_5$   $X_$ 

#### ・本番(本戦)での出力

CNN の出力は、使われた音声が 1, 使われていない音声が 0 になっている 88 種類の配列である。 取り札が 44 種類のため、日本語・英語のどちらかが 1 だと、その音声に対応した札を 1 として出力する。

### 2) その他(独創的なところ)

・予め、どこまでダウンサンプリングを行うとスペクトログラムの品質が落ちるのかを調べておく。

これを行うことで、CNNの処理を高速化でき、かつあまり劣化していないスペクトログラムを使うことができる。

・スペクトログラムにメル周波数スペクトログラムを用いることで、人間の聞き取れる周波数をスペクトログラム化する。これにより、音声の特徴などを見つけやすくなるのではないかと考えた。

No.4 開発

開発言語:Python

開発環境:Anaconda, Visual Studio Code, Neural Network Console, Git Hub

#### 第33回 全国高等専門学校 プログラミングコンテスト: 群馬大会