

部 門	競 技 部 門	No.1 登録番号	30008
-----	---------	-----------	-------

No.2

1) 予定開発期間：5ヶ月

2) 予定開発人数：3人

	4月	5月	6月	7月	8月	9月	10月
問題分析							
設計							
実装							
試用・トレーニング							

sNo.3	<p>実現方法</p> <p>1) 音声の解析アルゴリズム</p> <p>全体の方針 合成された音声データを学習した機械学習のプログラムにかけて、どの音声合成されたのかを出力する。 機械学習を行う前に、波形データを画像化する音声の前処理を行い、画像による機械学習を行う。</p> <p>・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(画像)に変換する。（*1）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し、時系列データでの処理が不要になる。</p> <div> </div> <p>・機械学習の方針 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。その際にどの音声合成されたかを表す教師データも作成する。（*2） スペクトログラム化された問題データを入力データとして渡し、CNN(畳み込みニューラルネットワーク)で画像の機械学習を行い、どの音声が使われたのかを表す0~1の88個の配列を出力する。その後、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。（*3）</p> <div> </div> <p>・本番(本戦)での出力 CNNの出力は、使われた音声は1,使われていない音声は0に近い値になっている88種類の配列である。これを1,0に近いほうに変換し、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。</p> <p>2) その他(独創的なところ)</p> <ul style="list-style-type: none"> ・ 音声波形データとして扱わず、画像化して処理を行うこと。 ・ 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。 	
	<p>No.4</p> <p>開発言語：Python 開発環境：Anaconda, Visual Studio Code, TensorFlow, FFmpeg, Neural Network Console, Git Hub</p>	