

合成音声を人間の声に近づけるために

物理班

背景・目的

先行研究¹より従来の音声合成方法(GMM-HMM合成)に表現力に限界があることが指摘されており、我々も実生活で合成音声に違和感を感じる事が多い。

本実験では、

- ①NNを用いた音声を作成
- ②NNと人間の声の比較
- ③なぜ違いが生まれるのかの把握を目標として行う。

仮説

従来の音声合成方法では、我々は違和感を覚えることが多い

→NNを用いることによってより精度の高い合成音声を作成できるのではないか。

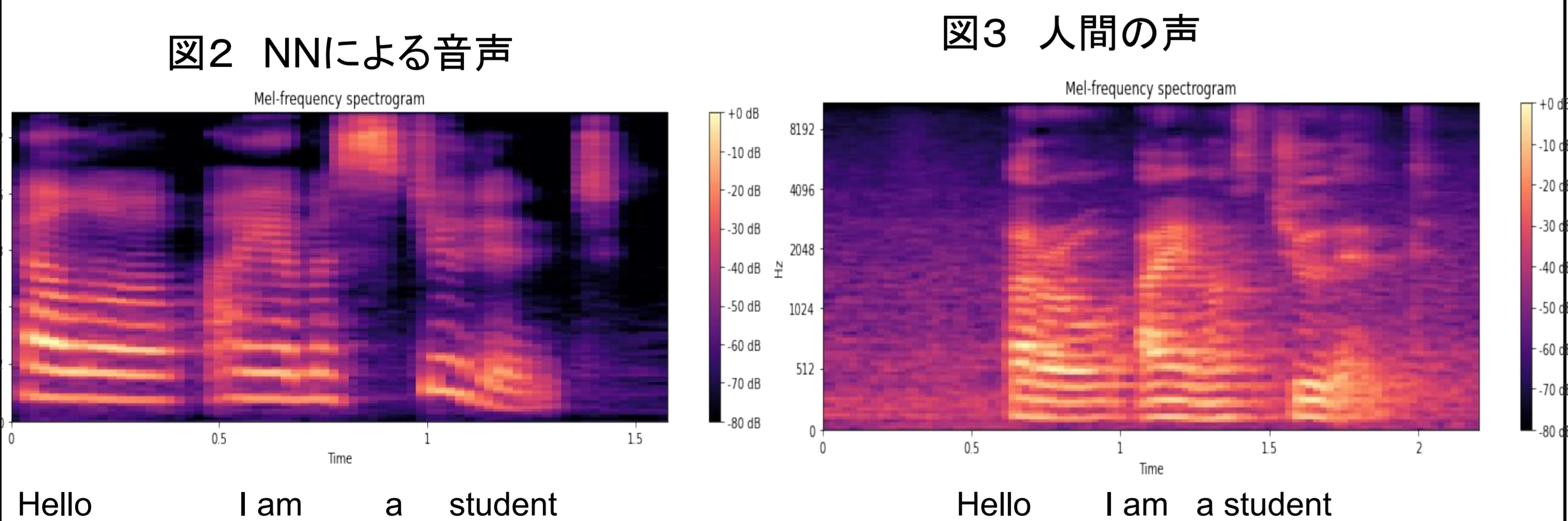
- ・GMM-HMM合成
→・確率予測により音声を作成
- ・NN合成
→・多量の計算、データ、機械学習を用いて作成

実験1 NNと人間の比較

Tactron2を用いてNN音声合成を行い人間の声との比較を行う

結果

・以下のグラフを得た



・二つのグラフにおいてHello, I am の二部分が似ているとの結論を得た

考察

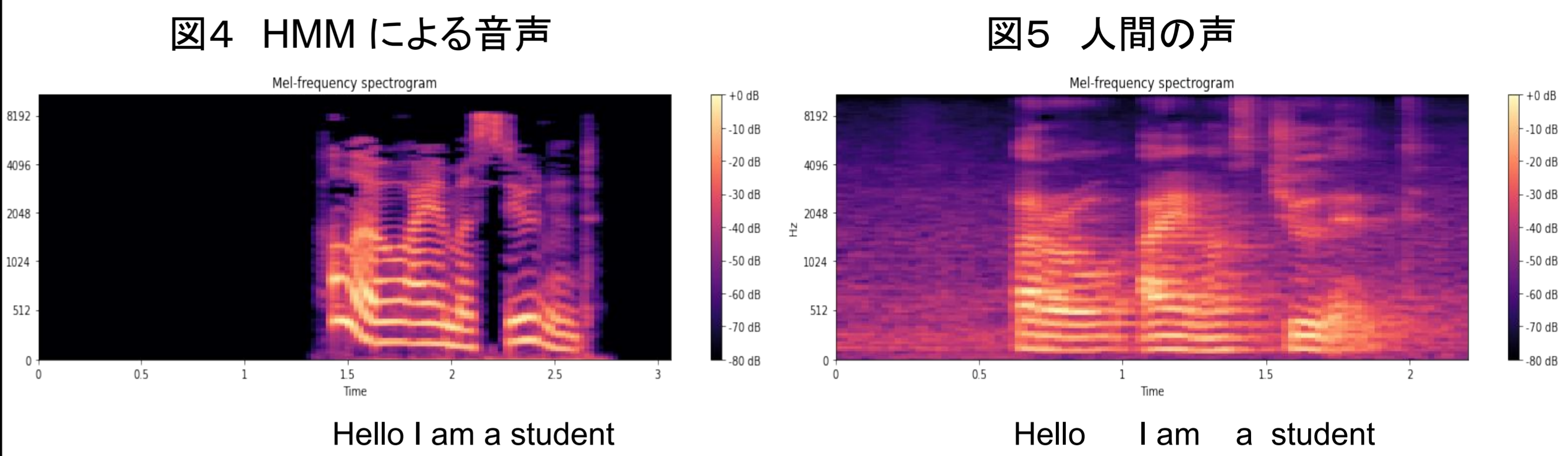
Hello, I am の二部分以外が似ていなかったのは学習に用いたデータがネイティブスピーカーのものであったのでその影響である話者依存性が原因ではないかと思われる。そのためネイティブスピーカーでない人のデータを使用すればグラフに類似点がさらに多く見られると考えられる。

実験2 HMMと人間の比較

GMM-HMM合成を行い人間の声との比較を行う

結果

・以下のグラフを得た



・HMMのグラフにおいてグラフの跳ね上がりが多くあるなど差異が多く見られる

考察

実験1とは異なりグラフの差異が多くみられた。原因としてはGMM-HMM合成において音声特徴量がGMM(混合ガウス分布)に従わなかったためと考えられる。また、このことよりGMM-HMM合成は音声特徴量とGMMの間でミスマッチが発生した場合表現力が著しく低下すると考えられる。

まとめ

本実験では実験1、2より

- ①NNを用いた合成は学習の性質上安定した表現力を持つ
- ②GMM-HMM合成は特定の条件下において著しく表現力を失う

という2つの結論を得た

展望

以降の実験では

- ①より多くの音声データの作成と比較
- ②表現力に差が出る詳しい要因の解明

この2つを主目標として実験を行っていきたいと思う。

参考文献

- 1: <https://www.gavo.t.u-tokyo.ac.jp/~mine/japanese/nlp+slp/IPSJ-MGN451003.pdf>
- 2: 音声言語処理と自然言語処理(13年、コロナ社)