

機械学習を用いたマスク着用における明瞭な声質の再現

1230323 久保田 成美 【ネットワーク信号処理研究室】

1 はじめに

昨今、新型コロナウイルス感染症により感染症対策としてマスク着用が推奨されている [1]. マスク着用による影響としてマスク未着用時に比べて音声聞き取りにくくなっている. 理由として、口を大きく開けにくくなることで声の大きさが小さくなることや、マスクによって声が籠ることなどが挙げられる. この中でも声がかもってしまう理由としてはマスクによって高い周波数帯が減衰してしまうからである. 従って、マスクを着用していても聞き取りやすい音声にするためには減衰している高い周波数帯を増幅する必要がある.

本研究では、マスク着用の認識をし、マスク着用時の音声をマスク未着用時の音声に変える方法を提案する.

2 マスクの着用による声質の変化

1人がマスク有とマスク無の2パターンでそれぞれ20分程度ずつ小説を朗読する. その音声をボイスレコーダーで録音し、今回の音声データとして扱う.

録音した音声のデータを周波数領域に変換し、メルスペクトログラムとして可視化する. 図1と図2はマスク有の音声データとマスク無の音声データをそれぞれ表示した結果である. マスク有の音声はマスク無の音声と比較すると高い周波数帯が減衰していることが確認できる.

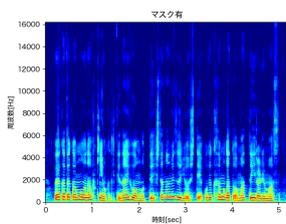


図1 マスク有

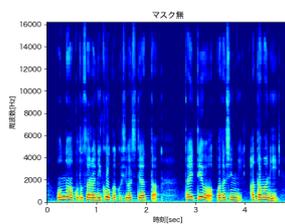


図2 マスク無

3 マスク着用における明瞭な声質の再現

マスク着用における明瞭な声質の再現としてフィルタや増幅回路を用いた再現と、機械学習を用いた再現を検討する. また、音声データを機械学習を用いて画像でマスク着用について認識させる.

3.1 フィルタや増幅回路による再現

声質を再現する方法としてフィルタや増幅回路を用いて再現することを提案する. フィルタではFIRフィルタを用いて高域を強調させる. 増幅回路では反復増幅回路を用いて信号自体を強調させる. 強調した結果、マスク無のメルスペクトログラムに近い結果が得られること

を確認した. しかし、FIRフィルタや反復増幅回路を用いて作成した音声データは雑音が多かった. これは、減衰してしまった音を増幅することで、雑音も一緒に増幅してしまうからであると考えられる.

3.2 機械学習による声質の再現

次に機械学習を用いて声質を再現することを提案する. 方法は、CycleGAN-VCを用いてマスク有の音声をマスク無の音声に声質変換することで再現させる. CycleGAN-VCではマスク無の音声の特徴量を抽出し、マスク有の音声に合成を行う [2]. 特徴量を抽出し与えることで、マスクによって減衰してしまった高域を補うことができる. 高域を補うことで、マスク無の声質を再現することが可能となる.

3.3 機械学習によるマスク着用の認識

録音した音声はマスク着用しているかどうかを認識させる. 方法は、メルスペクトログラムの画像を用いて畳み込みニューラルネットワークによって認識させる. 学習に用いたデータはマスク有、マスク無のそれぞれ130枚のメルスペクトログラム画像である. 機械学習の結果、認識率99%以上の学習結果を得られた. マスク有の音声データをフィルタや増幅回路を用いてマスク無に再現した音声データはマスク無としての認識することを確認した.

4 まとめ

本研究ではマスクによってこもった音をマスク未着用時のような声質に再現する方法を提案した. マスク有の音声を増幅しても雑音が強くなるという結果が得られた. そこで、マスク無の音声データから高域を抽出し、マスク有のデータに足すことにより、籠らない音を再現する方法を提案した. 今後の課題として、学習に用いたデータは1人の音声データのみである. そのため第三者による音声の場合ではどのように変化するか検討する必要があると考える.

参考文献

- [1] 厚生労働省, "マスクの着用について" https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kansentaisaku_00001.html, 2023年1月30日閲覧
- [2] 房福明, 山岸順一, 越前功, "CycleGANを用いた高品質なノンパラレル声質変換", 情報処理学会, 2017年, IPSJ-SLP17119009.pdf, 2023年2月2日閲覧