

先端科学技術研究科 修士論文要旨

所属研究室 (主指導教員)	知能コミュニケーション (中村 哲 (教授))		
学籍番号	1911005	提出日	令和 3年 1月 18日
学生氏名	東 佑樹		
論文題目	Automatic Speech Recognition Recovering Missing Words due to Sudden Noises 突発的な雑音に対する影響を考慮した音声認識システム		
要旨			
<p>音声認識は入力された音声から文字情報を出力する技術であり、スマートスピーカーなど、普段の生活で触れる多くのシステムの基盤技術となっている。</p> <p>本論文では、外部からの突発的な雑音による音声認識の性能の低下という問題に着目し、その問題を解決する手法について述べる。</p> <p>実生活において、音声認識システムは空港や街中など多様な環境での使用が想定される。そのため、そのような外部環境の種類に左右されずに、認識精度を高い品質に保つ必要がある。一般に、音声認識は入力対象の音声以外の音響信号の影響を受けやすく、雑音の多い環境下では認識率が低下するという問題がある。</p> <p>そこで、本論文では、突発的な雑音による認識性能の低下を抑えるために、3つの手法を提案する。</p> <p>1つ目の提案手法では、音声認識の出力結果をBERTを用いて修正する。BERTは周辺の単語情報からマスクされた単語を復元する事前学習を行うモデルであるため、雑音による誤認識トークンを復元する能力があることが期待できる。実験により、任意の文章に対し高頻度に出現する単語に関してはBERTを用いた単語の修復が可能であることが明らかになった。</p> <p>2つ目の提案手法では、生成モデルの1種であるVQ-VAEを用いて入力音声に含まれる雑音の影響を抑えた音声を再構成する。VQ-VAEは生成モデルの1種であり、潜在変数が強力なデコーダーにより無視されるPosterior Collapseという問題をベクトル量子化により解決することで、高品質な画像やビデオ、音声のサンプリングを可能にしたモデルである。</p> <p>本研究では、VQ-VAEが突発的な雑音により破損した音声情報を復元する能力があると仮定し、入力音声の前処理に使用した。実験により、破損した箇所の音声情報が周辺の音声情報により復元され、音声認識の認識精度が改善することが明らかになった。</p> <p>最後に、VQ-VAEから得られたコードブックから直接文章を変換する手法を提案する。VQ-VAEは入力音声をコードブックという離散シンボルに変換した後、そのシンボルから入力音声を再構成するため、コードブックには音声の特徴付ける情報が縮約されていることが期待される。このコードブックから音声の書き起こしテキストへ直接変換するモデルを構築した。実験により、Character単位で翻訳を行い、ストライド幅を2にした場合に最も良い性能となることが明らかになった。</p>			