

Auswirkungen von Speech Enhancement auf die automatische Spracherkennung

In dieser Bachelorarbeit werden Speech Enhancement Verfahren im Kontext der automatischen Spracherkennung untersucht. Ziel ist es, die Genauigkeit von automatischen Spracherkennungssystemen in Umgebungen mit Lärm und Nachhall zu verbessern.

Im besten Fall konnte bei einem Audiokorpus mit hoher Varianz die WER um 7.1% reduziert werden. Dies setzt voraus, dass der beste Präprozessor für eine bestimmte Audioaufnahme bekannt ist. Die Experimente mit SVM und neuronalen Netzen zeigen, dass mit den Metriken von Dolby.io und librosa der beste Präprozessor für einen bestimmten Audioclip nicht vorhergesagt werden kann. Allerdings erzielten die WER-Approximation Experimente vielversprechende Ergebnisse. So kann beispielsweise ein NN verwendet werden, um die WER vorherzusagen. Mit diesen Vorhersagen ist es dann möglich zu bestimmen, ob der Dolby.io-Enhancer auf einer Aufnahme angewendet werden sollte oder nicht, was zu einer WER-Reduzierung von 3.1% führte.

Das beste Ergebnis wurde mit Sentence Scoring erzielt. Durch Wahrscheinlichkeitsberechnungen der Transkripte konnten die fehlerhaften Transkripte erkannt werden und diese führten zu einer WER-Reduktion von 7%, bei einem Testdatensatz, bei dem der beste Fall 9.8% betragen würde. Die Vorhersagemodelle basieren auf DeepSpeech. In einer Folgearbeit soll untersucht werden, ob die Ergebnisse auch mit anderen STTs erzielt werden können.



Diplomierende
Manuel Berweger
Marvin Tseng

Dozent
Mark Cieliebak

Eingesetzte Komponenten um die WERs der Audios aus dem Korpus zu berechnen.

