

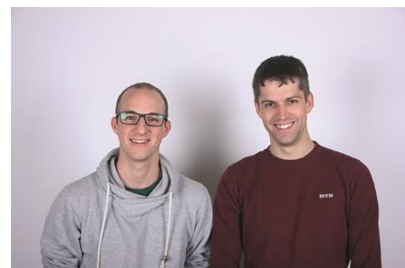
## Speaker Clustering mit Metric Embeddings

Das Ziel dieser Arbeit liegt darin, auf bestehende Ansätze aufzubauen und diese zu verbessern. In vorhergehenden Arbeiten wurden verschiedene Netzwerkarchitekturen evaluiert und bilden eine solide Ausgangslage. Zusammen mit einer neuen Loss-Funktion soll ein Modell trainiert werden, welches direkt Embeddings aus den Audiosegmenten erzeugt. Diese Embeddings stellen die Featurevektoren der Audiosegmente dar und werden in einem zweiten Schritt mit klassischen Verfahren geclustert.

Die bestehende Architektur wurde in TensorFlow implementiert. Zusammen mit der neuen Loss-Funktion wurde das Netzwerk in Betrieb genommen. Für die Experimente wurde der bewährte TIMIT Datensatz verwendet.

Die Experimente bestätigten, dass dieser Ansatz funktioniert und Resultate erzeugt, welche vergleichbar mit denen aus vorhergehenden Arbeiten sind. Damit eignet sich dieser Ansatz für zukünftige Arbeiten, in welchen die Hinzunahme von nicht annotierten Daten geprüft oder die Netzwerkarchitektur weiter verbessert werden könnte.

Im Rahmen dieser Arbeit wurde das bestehende experimentelle Setup kritisch hinterfragt. Um zukünftig die Reproduzierbarkeit und Vergleichbarkeit solcher Experimente zu gewährleisten, werden neue Definitionen und Konzepte vorgeschlagen. Für die Berechnung der Misclassification Rate (MR) wurde eine klare Definition verfasst, welche keinen Interpretationsspielraum lässt. Zudem wurde mit der Durchschnitts-MR ein neues Mass eingeführt, welches unabhängiger von einzelnen Messpunkten ist. Für die Generierung von Sprecherlisten auf dem TIMIT Datensatz wurde ein Konzept entwickelt, welches erlaubt, Trainings- und Testlisten von besserer Diversität und beliebiger Grösse zu erstellen.

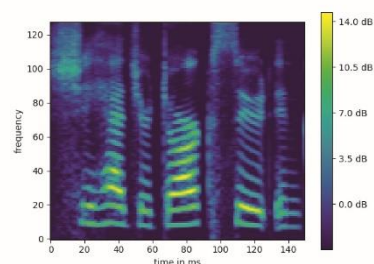


### Diplomierende

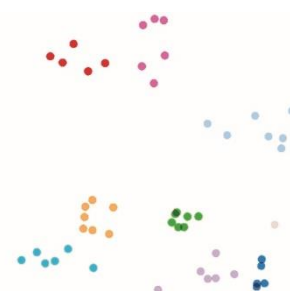
Jörg Egli  
Timon Gygax

### Dozierende

Oliver Dürr  
Thilo Stadelmann



Spektrogramm einer Sprachäusserung für die Verarbeitung mit dem Convolutional Neural Network.



PCA Plot von Embeddings, welche bereits Cluster erkennen lassen.