

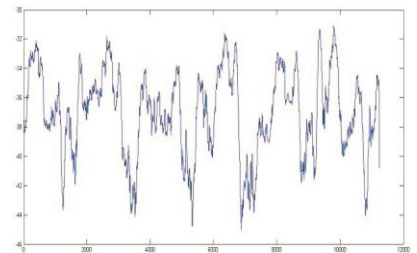
Talkalyzer: Mobile-App zur automatischen Sprecher-Erkennung

Um den eigenen Sprecheranteil bei einem Meeting oder einer Unterhaltung mit Freunden herauszufinden, lässt sich derzeit keine mobile Applikation in den Applikationsmärkten finden, welche eine gute Stimmenerkennung für Personen erreicht. Mit dieser Arbeit soll untersucht werden, ob ein Algorithmus implementiert werden kann, welcher eine Person in einer Unterhaltung verlässlich identifizieren und die Sprecheranteile fortlaufend in einem Diagramm darstellen kann. Es wird der Ansatz verfolgt, die Benutzerin oder den Benutzer beim ersten Applikationsstart laut Bilder beschreiben zu lassen, um ihre oder seine Stimme zu erfassen und davon ein Modell zu erstellen. Mit diesem Modell soll anschliessend in einer Unterhaltung die Stimme dieser Person wiedererkannt werden. Ebenfalls soll dargestellt werden, ob jemand anders gesprochen hat oder niemand. Um die Sprechererkennung durchführen zu können, wurden bestehende Bibliotheken oder Code-Ausschnitte sowie eigene Implementierungen von Komponenten verwendet. Abschliessend wurde ausgewertet, ob die implementierten Komponenten für ein gutes Resultat genügen. Andernfalls sollen die eingesetzten Komponenten effizient austauschbar sein. In der Applikation wird die CoMIRVA-Bibliothek verwendet, um die Audiodaten zu MFCCs zu verarbeiten. Anschliessend wird mit diesen Daten ein Sprechermodell mit einem k-Means++-Algorithmus von der Apache Commons Math-Bibliothek und in Java umgeschriebenen C++-Code für den EM-Algorithmus erstellt. Während einer Unterhaltung werden fortlaufend die zu MFCCs konvertierten Mikrofondaten mit dem Sprechermodell verglichen. Mit einem Schwellenwert wird entschieden, ob die MFCCs eine genug hohe Wahrscheinlichkeit zum Sprechermodell haben, womit die Sprecherbestimmung resultiert. Die Berechnung eines Sprechermodells variiert je nach Qualität und Menge der Daten. Für die vorausgesetzte Kalibrierungszeit von einer Minute wird das Sprechermodell in etwa fünf Sekunden erstellt. Die Sprecheranteile werden in einem sich aktualisierenden Kreisdiagramm dargestellt. Mit den implementierten Komponenten war es bei den durchgeführten Tests unter realen Bedingungen jedoch nicht möglich, die Stimme der Benutzerin oder des Benutzers eindeutig zu erkennen. Hierfür bedarf es daher weiterer Forschung, um die entscheidenden Komponenten durch verbesserte Verfahren zu ergänzen und dadurch die Erkennungsleistung zu steigern.

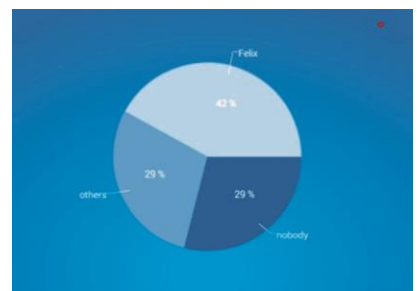


Diplomand
Lukas Kündig

Dozierende
Mark Cieliebak
Thilo Stadelmann



Log-Likelihood-Werte für die
Sprecherbestimmung



Kreisdiagramm mit den
Sprecheranteilen