MFCC-Plotter
— Ein graphisches Analysetool für cepstrale Daten —

Frederick Kukla & Vanessa Reichel

{*Frederick.Kukla, Vanessa.Reichel*}*@campus.lmu.de*
Institut für Phonetik und Sprachverarbeitung, LMU München

Menschen fällt es leicht eine andere Person anhand ihrer Stimme zu erkennen. Für eine Maschine ist diese Aufgabe nicht trivial. Um eine solche Erkennung realisieren zu können, muss zunächst gelernt werden, was die Stimme eines Sprechers auszeichnet. Diese Information steckt in den Mel Frequeny Cepstral Coefficents (MFCCs), die mittels mehrerer Signalverarbeitungsprozesse aus einer Sprachaufnahme eines Sprechers gewonnen werden. MFCCs können als mehrdimensionale numerische Werte verstanden werden, die Informationen über den Sprecher sowie über das Gesprochene erhalten. Aufgrund dieser beiden Eigenschaften werden sie häufig in der Sprach- und Sprechererkennung verwendet [1] [2].
Allerdings sind MFCCs sehr abstrakt und für Menschen schwer verständlich. Aus den rein numerischen Daten Informationen darüber herauszufiltern, welches Geschlecht der Sprecher hat oder ob ein Vokal oder ein Konsonant gesprochen wurde, ist für Menschen kaum möglich. Leichter verständliche graphische Darstellungen gibt es bereits vereinzelt. [3] fokussiert sich dabei beispielsweise auf die Darstellungen von MFCCs auf zeitlicher Ebene. Da es in unseren Augen bisher keine Möglichkeit gibt, MFCCs ganzheitlich und interaktiv zu analysieren, haben wir ein Tool entwickelt, das die Darstellung sowie den Vergleich der abstrakten Daten in verschiedenen Formen ermöglicht.
Die Abbildung 1 stellt eine beispielhafte Anwendung des Programms dar. Die Darstellung zeigt zwei Plots. Oberhalb des Plots befindet sich je ein Auswahlmenü, über welches die Graphik unterhalb des jeweiligen Menüs modifiziert werden kann.
Dateienpaare, bestehend aus je einer .wav- und einer .TextGrid-Datei, können in die Anwendung geladen werden. Nach einem erfolgreichen Upload können einzelne Phoneme, Berechnungsoptionen und Normalisierungsverfahren zur Darstellung ausgewählt werden. Im Informationsmenü können weitere Informationen über die verschiedenen Auswahlparameter und Darstellungen erfragt werden.
Allgemein liegt der primäre Anwendungsbereich des entwickelten Programms in der Lehre. Wie bereits dargestellt wurde, sind MFCCs sehr abstrakt und als numerische Werte für den Menschen wenig aussagekräftig ([4], S. 339). Dies macht es gerade für Studierende schwer, MFCCs zu begreifen. Eine visuelle Darstellung kann das Verständnis fördern. Durch den interaktiven Charakter des Programms können in der Sprachtechnologie übliche Verfahren wie unterschiedliche Normalisierungsarten besser nachvollzogen werden. Zudem können die cepstralen Daten einzelner Phoneme verglichen werden.
Abgesehen von der Lehre kann das Programm dazu verwendet werden, einen Überblick über die Daten zu gewinnen. So kann die Anwendung Forschenden, die an Sprach- oder Sprechererkennungssystemen arbeiten, eine kompakte Ansicht auf ihre Daten bieten.
Weitere Funktionen sind derzeit in Planung. Beispielsweise ein Prognosetool, das anzeigt, wie gut sich die eingegebenen Daten eignen, um darauf basierende Sprecher- oder Spracherkenner zu entwickeln.

doi:10.11576/pundp2022-1033

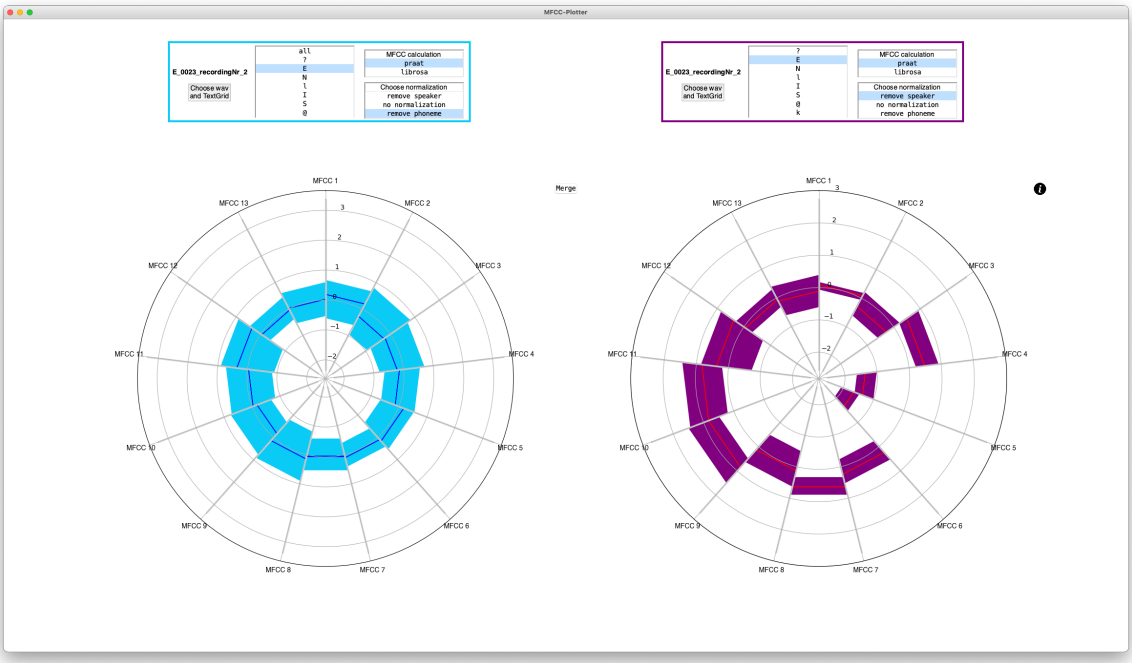

Abbildung 1: MFCC-Plotter: Linke Abbildung: 13 MFCCs von /E/ mit Praat berechnet, Phonem-Normalisierung; rechte Abbildung: 13 MFCCs von /E/ mit Praat berechnet, Sprecher-Normalisierung

Literatur

[1] Nidhi Desai, Kinnal Dhameliya, and Vijayendra Desai. Feature extraction and classification techniques for speech recognition: A review. *International Journal of Emerging Technology and Advanced Engineering*, 3(12):367–371, 2013.

[2] Tomi Kinnunen and Haizhou Li. An overview of text-independent speaker recognition: From features to supervectors. *Speech communication*, 52(1):12–40, 2010.

[3] OpenGenus IQ. Mfcc (mel frequency cepstral coefficients) for audio format, 2022.

[4] Beat Pfister and Tobias Kaufmann. *Sprachverarbeitung: Grundlagen und Methoden der Sprachsynthese und Spracherkennung*, volume 2. Springer, 2017.

doi:10.11576/pundp2022-1033