
ANWENDUNG DES MFCC-PLOTTERS ZUR ERFASSUNG CEPSTRALER UNTERSCHIEDE IN EMOTIONALER SPRACHE

Frederick Kukla, Vanessa Reichel

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

Kurzfassung: Diese Arbeit stellt einen Ansatz zur explorativen Analyse von Mel Frequency Cepstral Coefficients (MFCCs) vor. Es sollen Unterschiede in der gesprochenen Sprache zwischen den Emotionen Ekel, Wut, Angst, Freude und Trauer jeweils im Vergleich zu einer neutralen Emotion untersucht werden. Dieser Vergleich wird mit Mel Frequency Cepstral Coefficients, sogenannten MFCCs, durchgeführt. Da MFCCs als numerische Werte abstrakt und nicht intuitiv verständlich sind, werden die Abweichungen visuell mit dem MFCC-Plotter dargestellt. Der MFCC-Plotter ist ein von uns entwickeltes open source Werkzeug zur graphischen Aufbereitung von MFCC-Daten. Es erlaubt nicht nur eine Visualisierung einzelner MFCC-Datensätze, sondern ermöglicht es durch den sogenannten Merge-Modus schnell Unterschiede zwischen zwei Datensätzen erkennen zu können. Interaktiv können diverse Plots (Boxplot, ECDF-Plot, Histogramm) erzeugt werden, die eine detailreichere Analyse einzelner Dimensionen ermöglichen. Ergänzt werden diese Plots durch gängige statistische Kenngrößen, wie Durchschnittswerte, den Median und die Standardabweichung. Die Einbindung des cepstralen Raums soll bekannte phonetische Analysen unterstützen und neue Sichtweisen auf die Daten ermöglichen. Der MFCC-Plotter steht auf GitHub zum Download zur Verfügung und kann für die Untersuchung verschiedener phonetischer Fragestellungen verwendet werden. Ziel ist es, mit dieser beispielhaften Anwendung des MFCC-Plotters zur Analyse von emotionaler Sprache, die Vorteile des MFCC-Plotters aufzuzeigen und zu einer engeren Zusammenarbeit zwischen der Sprachtechnologie und der traditionellen Phonetik anzuregen.

1 Einleitung

Im Rahmen dieser Arbeit sollen Unterschiede von emotionaler Sprache anhand von MFCCs herausgestellt werden. Um diese Unterschiede visuell beschreiben zu können, wird die von uns entwickelte Software *MFCC-Plotter* [1] verwendet. Unser Ziel ist es, mit der Anwendung des MFCC-Plotters auf emotionale Sprache beispielhaft den Nutzen des Tools aufzuzeigen und dazu anzuregen, den MFCC-Plotter unterstützend zur traditionellen phonetischen Forschung zu verwenden. Die Arbeit versteht sich als Tutorial, das beispielhaft eine Anwendungsmöglichkeit des MFCC-Plotters zeigt.

Die Emotionen eines Menschen lassen sich nicht nur anhand von Mimik [2], Gestik [3] oder der Körperhaltung [4] erkennen. Auch die gesprochene Sprache verändert sich, je nachdem welche Emotion empfunden wird [5, 6, 7, 8]. Dabei sind sowohl Veränderungen auf der linguistischen Ebene [9] als auch in der Stimme selbst bemerkbar. Bei emotionaler Sprache lassen sich Veränderungen bezüglich verschiedener Parameter nachweisen. Dazu zählen Veränderungen in der durchschnittlichen Grundfrequenz (f_0) [6] sowie deren Range [7], Veränderungen in der Lage

des ersten Formanten (F1) [7], unterschiedliche Sprechgeschwindigkeiten [5, 8], Unterschiede in der Phonation [5], veränderte Lautstärken [6, 10, 11, 12, 13], aber auch in der Dauer von stimmhaften Frikativen [8] und der Dauer betonter Vokale [8]. [11] fasst zusammen, dass Merkmale mit Bezug zur Grundfrequenz (Minimum, Maximum, Durchschnitt), zur Lautheit, zu zeitlichen Aspekten (z.B. Dauern) sowie qualitative Parameter wie die Harmonics-to-Noise-Ratio zu den wichtigsten prosodischen Korrelaten bei emotionaler Sprache zählen.

Die oben genannten Merkmale und Parameter können aus dem Spektrum eines Audiosignals bzw. dem Signal selbst extrahiert werden [14]. Aus der automatischen Emotionsklassifikation ist bekannt, dass eine Kombination spektraler und cepstraler Merkmale zu besseren Klassifikationsergebnissen führt [15]. Die Verwendung cepstraler Merkmale bietet darüber hinaus einen weiteren Vorteil: während spektrale Eigenschaften wie die Grundfrequenz oder Formantwerte nur aus stimmhaften Segmenten extrahiert werden können, können MFCCs auch aus stimmlosen Segmenten berechnet und damit für eine Klassifikation verwendet werden [14]. Die Abkürzung MFCC steht für **Mel Frequency Cepstral Coefficient**. Diese sind als hochdimensionale, numerische Werte zu verstehen. Die Berechnung der MFCCs eines Sprachsignals erfolgt durch eine Fensterung des Signals und einer anschließenden Fourier-Transformation. Auf das daraus resultierende Spektrum wird die Mel-Skala zur Gruppierung der Frequenzbereiche angewandt. Diese orientiert sich am menschlichen Gehör. Das Spektrum wird dadurch zum Mel-Spektrum. Durch eine Anwendung der diskreten Kosinus-Transformation wird das Mel-Spektrum zum Mel-Cepstrum umgewandelt. Die Koeffizienten dieses Mel-Cepstrums entsprechen den MFCCs. Typischerweise werden die ersten 13 Koeffizienten (und deren erste und zweite Ableitung) für Analysen verwendet. MFCCs werden als hochdimensionale (meist 13-dimensionale Daten, mit deltas 39-dimensionale) Daten beschrieben. Für diese Arbeit ist der cepstrale Raum als 13-dimensionaler Raum definiert, in den MFCCs abgebildet werden [14]. Dieser cepstrale Raum soll mithilfe des MFCC-Plotters graphisch erschlossen werden. Dafür werden Unterschiede in den einzelnen Dimensionen analysiert.

2 Methode

2.1 Daten

Um emotionale Sprache zu untersuchen, wird der Korpus WaSeP [16] verwendet. Dieser enthält Aufnahmen von einer Sprecherin und einem Sprecher, die sowohl deutsche Substantive als auch Pseudowörter geäußert haben. Dabei haben beide Sprecher Äußerungen in unterschiedlichen Emotionen ausgesprochen: Ekel, Wut, Angst, Freude, Trauer sowie eine neutrale Emotion. Über die beiden Sprecher ist bekannt, dass es sich um eine weibliche Schauspielerin und einen männlichen Schauspieler handelt. Die verwendeten Aufnahmen haben eine hohe Qualität. Sie wurden mit einem Sony TCD-D7 Dat-Rekorder in einem Tonstudio aufgenommen. Bei den Äußerungen handelt es sich immer um Ein-Wort-Äußerungen. Die Aufnahmen wurden orthographisch transkribiert und mit MAUS [17, 18] automatisch segmentiert. Für die vorliegende Analyse wurden ausschließlich die Substantive betrachtet. Da nicht alle Wörter in allen Emotionen von beiden Sprechern gesprochen wurden, wurden nur Äußerungen ausgewählt, die von beiden Sprechern in jeweils allen Emotionen geäußert wurden. Dies sind 159 einzigartige Wörter (siehe auch https://github.com/frederick-m-k/mfcc_plotter/blob/essv_2023/resources/wasep_all_substantives.txt). Daraus ergeben sich pro Emotion 318 Aufnahmen (159 Aufnahmen der weiblichen Sprecherin und 159 Aufnahmen des männlichen Sprechers zu jeweils denselben Wörtern), die zur Kontrastierung einer Emotion zum neutralen Kontext geeignet sind.

Um den Untersuchungsgegenstand einzuschränken, wurden nur stimmlose Frikative und ein Affrikat betrachtet. Die stimmlosen Frikative und der Affrikat wurden zur Analyse ausgewählt, da

typische Merkmale zur Emotionsklassifikation (wie beispielsweise die f_0) nicht aus stimmlosen Segmenten berechnet werden können. Mittels MFCCs können neue Einblicke in Emotionsunterschiede bei stimmlosen Frikativen gewonnen werden. Die Frikative /ç/ und /x/ wurden von der Analyse exkludiert, da die Sampleanzahl gering und die Ergebnisse in der Folge wenig aussagekräftig sind. Von den 318 Aufnahmen werden die stimmlosen Frikative /f/, /s/ und /ʃ/ sowie der stimmlose Affrikat /ts/ in jedem Emotionskontext zur neutralen Emotion verglichen. Da eine graphische Analyse aller oben genannter Phoneme den Rahmen dieser Arbeit übersteigen würde, werden an dieser Stelle nur die Ergebnisse des Frikativs /f/ präsentiert. Die Ergebnisse der weiteren Phoneme sind im repository des MFCC-Plotters zu finden: https://github.com/frederick-m-k/mfcc_plotter/tree/essv_2023/resources/wasep_public_results.

2.2 Installation des MFCC-Plotters

Um emotionale Sprache auf cepstraler Ebene untersuchen zu können, kann der MFCC-Plotter angewandt werden. Der MFCC-Plotter ist ein Visualisierungstool, mit dessen Hilfe MFCCs graphisch analysiert werden können. Das Tool bietet sich zur explorativen Datenauswertung an, da eine Vielzahl an Settings auf eine unkomplizierte Weise untersucht werden können, ohne einen nennenswerten Arbeitsaufwand auf Seiten des Nutzer. Es muss jedoch betont werden, dass der MFCC-Plotter keine statistische Auswertung durchführt. Der Einsatzbereich des Tools besteht in der visuellen Auswertung und Erstellung von Hypothesen, um diese in einem nächsten Schritt mit geeigneten Programmen durch statistische Tests auszuwerten. Die Software kann auf der Plattform GitHub heruntergeladen werden, wo auch weitere Hinweise zur Installation zu finden sind: https://github.com/frederick-m-k/mfcc_plotter. Die Software wurde in Python (Version 3.8.0) [19] implementiert und nutzt die Library TKinter [20] zur Erstellung der graphischen Oberfläche. Ein Nutzer des MFCC-Plotters muss im Allgemeinen über keinerlei Python- oder Programmierkenntnisse verfügen, um die Software anwenden zu können.



Abbildung 1 – Hauptmenü des MFCC-Plotters

2.3 Hauptmenü: erste visuelle Analyse

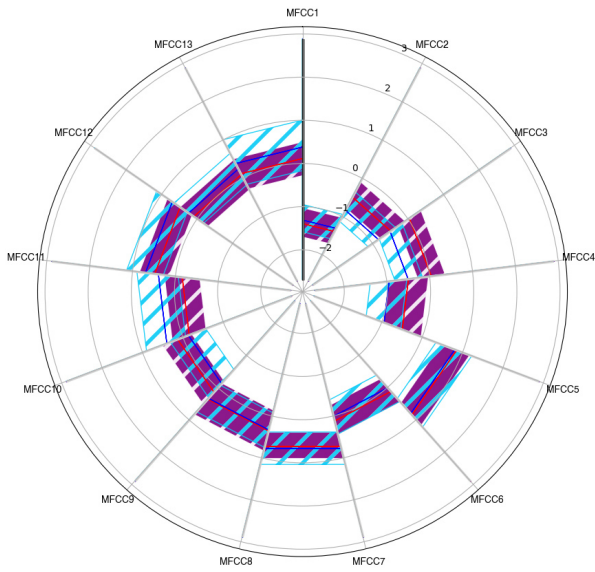
Im Folgenden soll gezeigt werden, wie der MFCC-Plotter bei der Bearbeitung einer phonetischen Fragestellung unterstützen kann. Hierfür werden die zuvor erläuterten Daten des WaSeP-Korpus in den MFCC-Plotter importiert, um emotionale Sprache untersuchen zu können. Vor der ersten Nutzung des MFCC-Plotters muss der Pfad zur Praat-Installation angegeben werden. Dieser wird benötigt, da die MFCCs in der Standardkonfiguration mit einem Praat-Skript berechnet werden. Der Pfad zur Praat-Installation wird über das Informationsmenü gesetzt und bleibt für spätere Anwendungen des MFCC-Plotters festgelegt. Das Menü öffnet sich durch Klick auf den **i**-Button, der sich am oberen rechten Rand des Hauptmenüs befindet (siehe auch 1). Nachdem der Pfad eingegeben wurde, können die zu analysierenden Daten geladen werden. Dafür werden durch Klick auf den Button *Provide wav and annotation* Dateienpaare bestehend aus Audio- (.wav) und Annotationsdateien (.TextGrid oder _annot.json) ausgewählt. Über die Annotationsdateien kann der MFCC-Plotter die aus der Audiodatei berechneten MFCCs den jeweiligen Phonemen zuordnen. Damit diese Zuordnung erfolgen kann, müssen zusammengehörige Audio- und Annotationsdateien denselben Namen haben. Nachdem die Dateien importiert wurden, werden die MFCCs automatisch berechnet. Für den Nutzer unsichtbar wird ein Praat-Skript ausgeführt, das die MFCCs für die gewählten Dateien berechnet. Sobald die Berechnung abgeschlossen ist, zeigen sich im Hauptmenü zwei Plots.

Abbildung 1 zeigt das Hauptmenü des MFCC-Plotters. Die Darstellung ist zweigeteilt: auf der linken Seite sind die Daten der beiden Sprecher in der neutralen Emotion abgebildet (blau); auf der rechten Seite sind die Daten der beiden Sprecher beim Empfinden von Ekel visualisiert (violett). Beide Radarplots zeigen alle Frames des Phonems /f/ aus allen Input-Dateien. Oberhalb der Plots befindet sich ein Auswahlm Menü, über das die Input-Daten, das betrachtete Phonem, verschiedene Normalisierungsoptionen und unterschiedliche Berechnungsmöglichkeiten zur Erstellung der MFCCs angepasst werden können. Die einzelnen Radarplots sind in 13 Sektoren unterteilt, wobei jeder Sektor die Daten einer MFCC-Dimension umfasst. Die farbigen Flächen sind als Boxplots zu verstehen, die 50% der MFCC-Daten der jeweiligen Dimension enthalten. Bereits in dieser Ansicht ist erkennbar, dass sich einige MFCCs des Phonems /f/ in Abhängigkeit zur empfundenen Emotion unterscheiden. Diese Unterschiede treten umso stärker hervor, wenn der MFCC-Plotter im sogenannten Merge-Modus verwendet wird (siehe beispielhaft Abbildung 2).

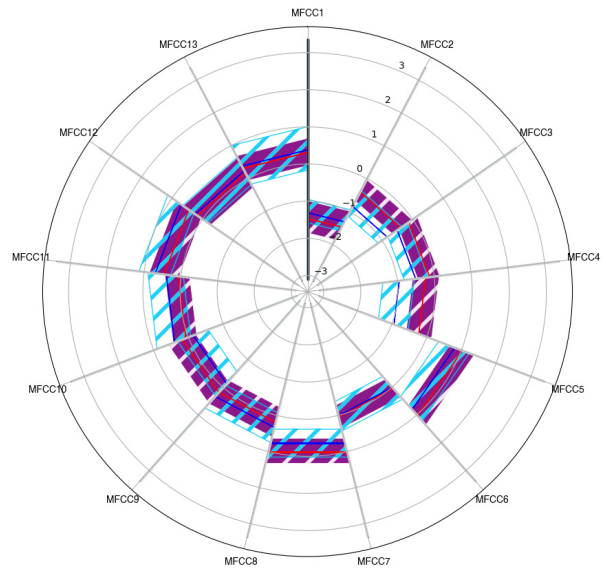
2.4 Merge-Modus: Aufzeigen von Gemeinsamkeiten und Diskrepanzen

Der Merge-Modus eignet sich, um die einzelnen MFCC-Dimensionen zu kontrastieren. Er wird im Hauptmenü über den *merge*-Button, der zwischen den beiden Plots liegt, aktiviert (siehe Abbildung 1). Hierbei werden die Plots übereinandergelegt und die flächige Farbgebung durch eine Schraffur ersetzt. In diesem Setting zeigen sich Gemeinsamkeiten und Diskrepanzen im cepstralen Raum. In Abbildung 2 sind die gemergten Ansichten aller untersuchten Emotionskontraste dargestellt. In allen Abbildungen wird das Phonem /f/ visualisiert. Die neutrale Emotion ist in den Abbildungen in blau dargestellt, während die fünf Emotionskontexte Ekel, Wut, Angst, Freude und Trauer durch violette Flächen dargestellt werden. Konsistente graphische Unterschiede über die verschiedenen Emotionskontexte hinweg fallen für den zweiten und dritten MFCC auf. Der dritte MFCC ist für jeden Emotionskontext höher als die neutrale Emotion (violett hat höhere Werte als blau). Der zweite MFCC ist mit Ausnahme von Trauer für jeden Emotionskontext niedriger als die neutrale Emotion (violett hat für alle Emotionen außer Trauer niedrigere Werte als blau).

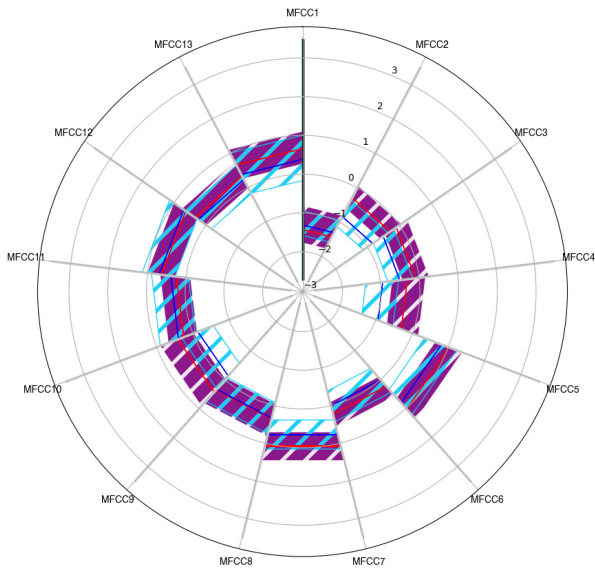
Dies ist ein mögliches Indiz dafür, dass sich /f/ allgemein dafür eignet, eine der Emotionen Ekel, Freude, Angst, Trauer oder Wut zur neutralen Emotion zu kontrastieren.



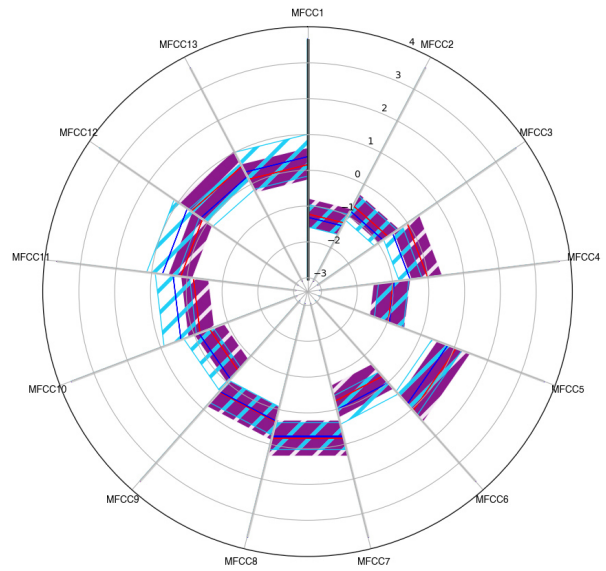
a) Kontrast von Ekel (violett) und neutral (blau)



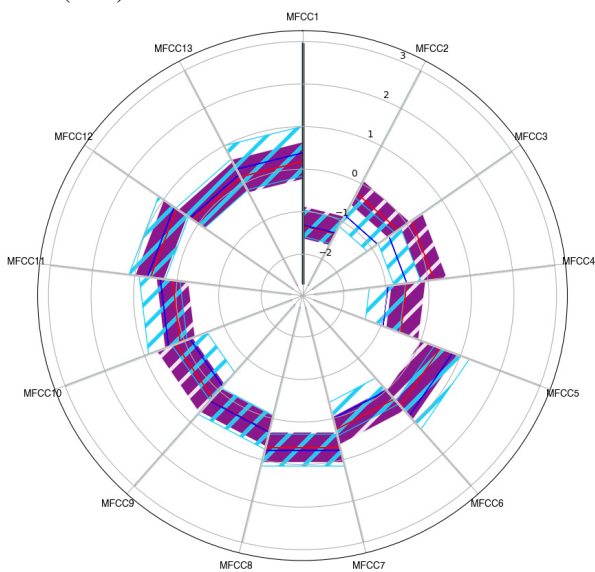
b) Kontrast von Freude (violett) und neutral (blau)



c) Kontrast von Angst (violett) und neutral (blau)



d) Kontrast von Trauer (violett) und neutral (blau)



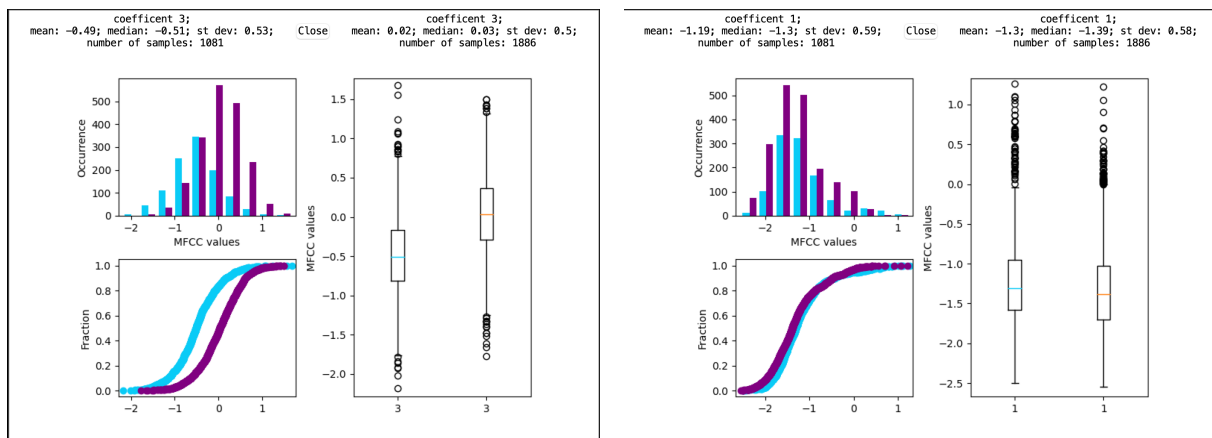
e) Kontrast von Wut (violett) und neutral (blau)

Abbildung 2 – 13-dimensionale MFCCs von allen /f/-Phonemen im a) Ekel-, b) Freude-, c) Angst-, d) Trauer- und e) Wut-Emotionskontext (jeweils violett) im Kontrast zum neutralen Emotionskontext (blau). Alle MFCCs sind CMVN-Sprechernormalisiert. Die Daten stammen aus dem WaSeP-Korpus (<https://clarin.phonetik.uni-muenchen.de/BASRepository/>).

2.5 Statistikmenü: Analyse eines einzelnen Koeffizienten

Um die Daten eines einzelnen Koeffizienten detaillierter zu analysieren, kann das Statistikmenü verwendet werden. Dieses öffnet sich, wenn mit der Maus in einen der Sektoren des Radarplots geklickt wird. Das Statistikmenü zeigt die Daten des ausgewählten Koeffizienten. Es umfasst insgesamt drei verschiedene Plots: einen Boxplot, einen ECDF-Plot und ein Histogramm. Oberhalb der Plots werden der Durchschnitt, Median und die Standardabweichung für die jeweilige Dimension aufgelistet. Neben diesen Angaben ist dem Statistikmenü zudem zu entnehmen, wie viele Datenpunkte in die Auswertung aufgenommen wurden. Das Histogramm, der ECDF-Plot und der Boxplot ermöglichen eine tiefere Analyse der einzelnen Dimensionen. Alle Plots basieren auf denselben Daten, wie die farbige Fläche eines Sektors im Hauptmenü des MFCC-Plotters. Während im Hauptmenü im Radarplot nur 50% der Daten pro Dimension gezeigt werden, zeigt der Boxplot im Statistikmenü 100% der Daten.

Abbildung 3 zeigt zwei Abbildungen des Statistikmenüs. In diesen Beispielen wurden die Statistikmenüs geöffnet, während der Merge-Modus aktiviert war. Daher sind die Daten beider Emotionen gezeigt. Die Farbgebung ist unverändert: blaue Abbildungen beziehen sich auf die neutrale Emotion, in violett ist die Emotion Ekel dargestellt.



(a) Statistikmenü von MFCC3: graphischer Kontrast zwischen den beiden Emotionen

(b) Statistikmenü von MFCC1: große Ähnlichkeit zwischen den Emotionen

Abbildung 3 – Statistikmenüs des MFCC-Plotters des Phonems /f/ im Ekel-Emotionskontext (violett in den Abbildungen) kontrastiert mit der neutralen Emotion (blau in der Abbildung).

Abbildung 3a zeigt die Daten von MFCC3. Sie ist ein Beispiel für einen Koeffizienten, dessen Werte sich in den beiden Emotionen graphisch unterscheiden. Der durchschnittliche MFCC-Wert dieses Koeffizienten liegt in der neutralen Emotion bei -0.49; der Median bei -0.51. Die Berechnung basiert auf 1081 Sample-Werten. Im Vergleich dazu liegen die Werte desselben Koeffizienten im Ekel-Emotionskontext höher. Der Durchschnitt beträgt hier 0.02 während der Median bei 0.3 liegt. Insgesamt sind in die Berechnung in der Ekel-Emotion 1886 Sample-Werte eingegangen. Die Unterschiede im Durchschnitt und im Median lassen sich auch visuell erkennen. Aus dem Histogramm kann abgelesen werden, dass Ekel in dieser Dimension einen höheren Durchschnittswert erreicht als die neutrale Emotion. Diese Beobachtung spiegelt sich auch im ECDF-Plot wider. Aus dem ECDF-Plot lässt sich ablesen, wie viel Prozent der Daten unterhalb einem gewissen Messwert liegen. So sind beispielsweise in Abbildung 3a im ECDF-Plot 50% (y-Achse Fraction) der violetten Daten unter einem Wert von etwa 0.1 (x-Achse). In den Boxplots kann gesehen werden, dass sich die Mediane, Quantile, Whiskers und Outlier der beiden Emotionen graphisch unterscheiden.

Im Gegensatz zu den klaren Unterschieden, die sich in MFCC3 gezeigt haben, unterscheiden sich die Daten in MFCC1 kaum. Die durchschnittlichen MFCC1-Werte der neutralen Emotion und Ekel sind in dieser Dimension kaum zu differenzieren. Auch die drei Plots zeigen für beide Emotionen je eine ähnliche Verteilung der MFCC1-Werte. Bezüglich des Histogramms muss an dieser Stelle auf die unterschiedliche Anzahl an Samples hingewiesen werden (neutral = 1081 Sample-Werte; Ekel = 1886 Sample-Werte), die erklären, weshalb die violetten Balken höher reichen.

Mithilfe des Statistikmenüs und den darin befindlichen Plots können die einzelnen MFCC-Dimensionen verglichen werden und Hypothesen generiert werden, welche Dimensionen für eine Emotionsklassifikation geeignet scheinen. Auch hierbei handelt es sich um ein Beispiel, das verdeutlichen soll, wie das Statistikmenü angewendet werden kann, um visuell Unterschiede im cepstralen Raum herauszustellen.

3 Diskussion und Ausblick

Im Rahmen dieser Arbeit wurde die Anwendung des MFCC-Plotters zur Untersuchung emotionaler Sprache dargestellt. Anhand von Beispieldaten aus dem WaSeP-Korpus wurden graphische Interpretationsmöglichkeiten aufgezeigt. Die Phoneme /f/, /s/, /ʃ/ und /ts/ wurden graphisch analysiert. Dabei konnten visuelle Unterschiede festgestellt werden. Es gilt noch zu überprüfen, ob die visuellen Unterschiede statistische Signifikanz erreichen.

Der vorgestellte MFCC-Plotter ist ein laufendes Software- und Forschungsprojekt. Dementsprechend werden fortlaufend neue Funktionen, Darstellungsformen und Ideen implementiert. Eine Funktion, die als nächstes implementiert werden soll, ist die Downloadmöglichkeit der generierten Plots. Eine zweite, komplexere Funktionalität ist die Darstellung der deltas. Deltas sind die zeitlichen Ableitungen der MFCCs, die neben den eigentlichen Cepstren eine elementare Größe in der Sprachtechnologie sind [14]. Dementsprechend kann die graphische Analysemöglichkeit dieser Werte weitere Perspektiven auf Forschungsdaten erlauben.

Literatur

- [1] KUKLA, F. und V. REICHEL: *Mfcc-plotter - ein graphisches analysetool für cepstrale daten*. In *Phonetik und Phonologie im deutschsprachigen Raum*. Bielefeld, Deutschland, 2022. doi:<https://doi.org/10.11576/pundp2022-1033>.
- [2] RANI, J. und K. GARG: *Emotion detection using facial expressions-a review*. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(4), 2014.
- [3] TRIPATHI, S., S. TRIPATHI, und H. BEIGI: *Multi-modal emotion recognition on iemocap dataset using deep learning*. *arXiv preprint arXiv:1804.05788*, 2018.
- [4] SANEIRO, M., O. C. SANTOS, S. SALMERON-MAJADAS, und J. G. BOTICARIO: *Towards emotion detection in educational scenarios from facial expressions and body movements through multimodal approaches*. *The Scientific World Journal*, 2014.
- [5] BURKHARDT, F. und W. F. SENDLMEIER: *Verification of acoustical correlates of emotional speech using formant-synthesis*. In *ISCA Tutorial and Research Workshop (ITRW) on speech and emotion*. 2000.
- [6] DASGUPTA, P. B.: *Detection and analysis of human emotions through voice and speech pattern processing*. *arXiv preprint arXiv:1710.10198*, 2017.

-
- [7] HARTMANN, K., I. SIEGERT, D. PHILIPPOU-HÜBNER, und A. WENDEMUTH: *Emotion detection in hci: from speech features to emotion space*. *IFAC Proceedings Volumes*, 46(15), S. 288–295, 2013.
- [8] KIENAST, M., A. PAESCHKE, und W. SENDLMEIER: *Articulatory reduction in emotional speech*. In *Sixth European Conference on Speech Communication and Technology*. 1999.
- [9] SHIVHARE, S. N. und S. KHETHAWAT: *Emotion detection from text*. *arXiv preprint arXiv:1205.4944*, 2012.
- [10] BANDELA, S. R. und T. K. KUMAR: *Stressed speech emotion recognition using feature fusion of teager energy operator and mfcc*. In *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, S. 1–5. IEEE, 2017.
- [11] LAUSEN, A. und K. HAMMERSCHMIDT: *Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters*. *Humanities and Social Sciences Communications*, 7(1), S. 1–17, 2020.
- [12] JUSLIN, P. N. und P. LAUKKA: *Communication of emotions in vocal expression and music performance: Different channels, same code?* *Psychological bulletin*, 129(5), S. 770, 2003.
- [13] VERVERIDIS, D. und C. KOTROPOULOS: *Emotional speech recognition: Resources, features, and methods*. *Speech communication*, 48(9), S. 1162–1181, 2006.
- [14] PFISTER, B. und T. KAUFMANN: *Sprachverarbeitung - Grundlagen und Methoden der Sprachsynthese und Spracherkennung*. Springer, Berlin, Heidelberg, 2017. doi:10.1007/978-3-662-52838-9. URL <http://link.springer.com/10.1007/978-3-662-52838-9>.
- [15] DUMOUCHEL, P., N. DEHAK, Y. ATTAABI, R. DEHAK, und N. BOUFADEN: *Cepstral and long-term features for emotion recognition*. In *Tenth annual conference of the international speech communication association*. 2009.
- [16] FOR SPEECH SIGNALS (BAS), B. A.: 2023. URL <https://clarin.phonetik.uni-muenchen.de/BASRepository/>.
- [17] SCHIEL, F.: *Automatic phonetic transcription of non-prompted speech*. *Proceedings of the ICPhS*, S. 607–610, 1999.
- [18] KISLER, T., U. REICHEL, und F. SCHIEL: *Multilingual processing of speech via web services*. *Computer Speech & Language*, 45, S. 326–347, 2017.
- [19] FOUNDATION, P. S.: *Python.org*. 2022. URL <https://www.python.org/>.
- [20] FOUNDATION, P. S.: *tkinter — python interface to tcl/tk — python 3.11.0 documentation*. 2022. URL <https://docs.python.org/3/library/tkinter.html>.