



# Deep feature-based speech emotion recognition for smart affective services

Seminararbeit an der Universität Ulm

**Vorgelegt von:**

Salih Bedelce  
salih.bedelce@uni-ulm.de

**Gutachter:**

Prof. Dr. Friedhelm Schwenker

**Betreuer:**

Prof. Dr. Friedhelm Schwenker

2021



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>3</b>
1.1	Was ist speech emotions recognition (SER) . . . . .	3
1.2	Spektrogramme . . . . .	3
1.2.1	STFT und FFT . . . . .	3
<b>2</b>	<b>Convolutional Neural Networks</b>	<b>5</b>
<b>3</b>	<b>Modellarchitektur und Ablauf</b>	<b>7</b>
3.1	2 Phasen der SER . . . . .	7
3.1.1	Verarbeitungseinheit (processing unit) . . . . .	7
3.1.2	Klassifikator (classifier) . . . . .	7
3.2	Aufbau der Modellarchitektur . . . . .	7
3.3	Der Ablauf bei SER . . . . .	8
<b>4</b>	<b>Schlussfolgerung</b>	<b>9</b>
	<b>Literaturverzeichnis</b>	<b>11</b>



# Zusammenfassung

Diese kleine Einleitung soll dem Nutzer helfen selbst die eigene Arbeit mit  $\text{\LaTeX}$  zu schreiben.  
Sie enthält zu den wichtigsten Themen Beispiele.

[2] [4] [5] [1]



# 1 Einleitung

Die Sprache des Menschen ist die natürlichste Art und Weise um miteinander zu kommunizieren. Durch die Entwicklung der Technologie kommen immer mehr Interaktionen zwischen Mensch und Maschine durch die Sprache zustande [2]. Die Bezeichnung dafür sind die sogenannten intelligent personal assistants (IPAs) wie Amazon Alexa, Apple Siri und Google Assistant. Google Home, Amazon Echo und Apple HomePod sind Home-Assistant Systeme, die primär Sprachsignale als Interaktionsmöglichkeit besitzen. Diese IPAs sind sehr stark verbreitet und auf vielen Geräten verfügbar [3].

## 1.1 Was ist speech emotions recognition (SER)

## 1.2 Spektrogramme

Spektrogramme spielen bei SER eine wichtige Rolle, denn sie dienen als Input.

### 1.2.1 STFT und FFT





## 2 Convolutional Neural Networks

Convolutional Neural Network (CNN) ist ein hierarchisches neuronales Netz, welches aus unterschiedlichen Schichten (layers) besteht. Diese Schichten kann man in drei Hauptkomponenten aufteilen.[2]

- convolutional layers  
diese Schicht ist für das Filtern des Inputs zuständig
- pooling layers
- fully connected layers



## 3 Modellarchitektur und Ablauf

Diese kleine Einleitung soll dem Nutzer helfen selbst die eigene Arbeit mit  $\text{\LaTeX}$  zu schreiben. Sie enthält zu den wichtigsten Themen Beispiele.

### 3.1 2 Phasen der SER

Für diese Arbeit lassen sich als Überschriften die Überschriften in verschiedenen Stufen verwenden.

#### 3.1.1 Verarbeitungseinheit (processing unit)

#### 3.1.2 Klassifikator (classifier)

### 3.2 Aufbau der Modellarchitektur

Mit Hilfe eines Labels kann man sich dann im Text auf diese Grafik (3.1) beziehen.

Durch `subfigure` lassen sich auch zwei kleine Bilder nebeneinander setzen. In Abbildung 3.2(a) ist ein fettes u auf der linken und in 3.2(b) ein dünneres auf der rechten Seite zu sehen.

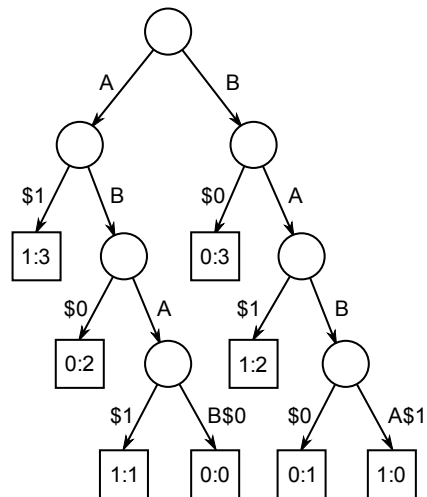


Abbildung 3.1: Beschreibung des Bilds



(a) Ein fettes u



(b) Ein dünneres u

Abbildung 3.2: Die *us* aus der Wortmarke

### 3.3 Der Ablauf bei SER

Hier nur ein kurzes Beispiel, in jedem  $\text{\LaTeX}$  Buch finden sich gute Anleitungen zum Erstellen von Tabellen.

## 4 Schlussfolgerung

Diese kleine Einleitung soll dem Nutzer helfen selbst die eigene Arbeit mit  $\text{\LaTeX}$  zu schreiben.  
Sie enthält zu den wichtigsten Themen Beispiele.



## Literaturverzeichnis

- [1] BADSHAH, Abdul M. ; AHMAD, Jamil ; RAHIM, Nasir ; BAIK, Sung W.: Speech Emotion Recognition from Spectrograms with Deep Convolutional Neural Network. In: *2017 International Conference on Platform Technology and Service (PlatCon)*, 2017, S. 1–5
- [2] BADSHAH, Abdul M. ; RAHIM, Nasir ; ULLAH, Noor ; AHMAD, Jamil ; MUHAMMAD, Khan ; LEE, Mi Y. ; KWON, Soonil ; BAIK, Sung W.: Deep features-based speech emotion recognition for smart affective services. In: *Multimedia Tools and Applications* 78 (2019), Nr. 5, S. 5571–5589
- [3] CLARK, Leigh ; DOYLE, Philip ; GARAIALDE, Diego ; GILMARTIN, Emer ; SCHLÖGL, Stephan ; EDLUND, Jens ; AYLETT, Matthew ; CABRAL, João ; MUNTEANU, Cosmin ; EDWARDS, Justin ; R COWAN, Benjamin: The State of Speech in HCI: Trends, Themes and Challenges. In: *Interacting with Computers* 31 (2019), 09, Nr. 4, 349-371. <http://dx.doi.org/10.1093/iwc/iwz016>. – DOI 10.1093/iwc/iwz016. – ISSN 0953–5438
- [4] HUANG, Zhengwei ; DONG, Ming ; MAO, Qirong ; ZHAN, Yongzhao: Speech Emotion Recognition Using CNN. In: *Proceedings of the 22nd ACM International Conference on Multimedia*. New York, NY, USA : Association for Computing Machinery, 2014 (MM '14). – ISBN 9781450330633, 801–804
- [5] LI, Wu ; ZHANG, Yanhui ; FU, Yingzi: Speech Emotion Recognition in E-learning System Based on Affective Computing. In: *Third International Conference on Natural Computation (ICNC 2007)* Bd. 5, 2007, S. 809–813

*Literaturverzeichnis*

Name: Salih Bedelce

Matrikelnummer: 1038226

**Erklärung**

Ich erkläre, dass ich die Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Ulm, den .....

Salih Bedelce