Hochschule Rhein-Waal
Rhine-Waal University of Applied Sciences
Faculty of Communication and Environment

Degree Program Information Engineering and Computer Science, M. Sc.
Prof. Dr. Timo Kahl

Voice emotion detection for teaching review - A case study at the Rhine-Waal University of Applied Sciences

Term Paper

WS 2018/2019

Module “Applied Research Project B”

by

Ronit Saha (24939),

Viet-Hung Vu (26078),

Sachin Kumar (24914),

Aishwarya Narkar (24854), &

Vincent Finn Alexander Meyer zu Wickern (25142)

Abstract

Voice Emotion Detection is a technology, which allows to determine the emotion that a person feels at a given time by analyzing the voice of this person recorded at that time. Application areas of this technology can be found for example in lie detection or Human-Computer interaction (HCI), but also the analysis of participant emotions during interview questions may be a viable application scenario, since hidden emotions may be revealed and the information gain through interviews may be increased. The University of Applied Sciences Rhein-Waal regularly assesses the learning experience of its students with written forms. As this is a potentially optimizable process with voice emotion detection, this paper aims to assess the learning experience of a sample of current students with the help of voice emotion detection. Interviews are conducted with these students on five areas of their learning experience.

It is found that emotions among the students of the Rhein-Waal University of Applied Sciences are wide-spread and that not one emotional reaction prevails for all students or across all answers. The emotions to each question range from positively connotated emotions, such as “happy” or “calm”, to negatively connotated emotions such as “sad” or “angry”. Therefore, further research into the different factors influencing each of these emotions is proposed.

Emotion classification models are developed, of which the support vector classification even reaches a median of 63% accuracy in the train-test-split among the training data. The emotions in the interviews can also be analyzed, however, in this case the interview use case poses challenges that are revealed in this paper. The results of voice emotion detection depend on the strength of the emotion and the intensity that the interviewed person expressed the emotion with, but the neutral and undirected trigger of emotions is a challenge, for which further research is necessary. Additionally, the annotation of interviews is difficult due to a lacking clarity of emotions in many cases, which impedes the valid numbering of an accuracy. In a review of the classifier generated emotion predictions, 60% of the results have been found reasonable.

The use case of interview analysis for the technology of voice emotion detection is found to pose challenges, which are required to be solved in the future for it to be a viable solution. However, in this analysis of student emotions it brought about additional insights into the student state of mind and encouraged a deeper analysis of vocal interview participant behavior.

Table of Contents

List of Figures

Abbreviations

Abbreviation	Long Form
API	Application Programming Interface
ACD	Automatic Call Distribution
BP	Backpropagation
CTI	Computer Telephony Integration
DCT	Discrete Cosine Transform
DNN	Deep Neural Network
DTM	Dynamic Time Warping
GUI	Graphical User Interface
HNR	Harmonics-to-Noise Ratio
HMM	Hidden Markov Model
HCI	Human-Computer Interaction
HSRW	Hochschule Rhein-Waal
IDE	Integrated Development Environment
IEMOCAP	Interactive emotional dyadic motion capture
IVR	Interactive Voice Response
KNN	K-Nearest Neighbors Algorithm
MFCC	Mel-frequency cepstral coefficient
MLP	Multilayer perceptron
PCA	Principal components analysisy
RAVDESS	Ryerson Audio-Visual Database of Emotional Speech and Song
SAIL	Speech Analysis and Interpretation Laboratory
SAL	Sensitive Artificial Listener
SFS	Sequential Forward Selection

SSI	Social Signal Interpretation
SVM	Support Vector Machine

1 Introduction

Founded in 2009, Rhine-Waal University of Applied Sciences aims to provide an innovative, interdisciplinary and international academic environment for students coming from different regions of the world. With more than 7000 students studying in 25 undergraduate and 11 graduate programs as of 2018 (Rhein-Waal, 2017), it is essential for the university to gather and analyze students' feedback on its academic services. The university collects feedback in the form of paper-based service every semester. However, as Mathai found out, text is less useful for the detection of emotions compared to voice (Mathai). Therefore it would be difficult to perform text-based analysis on the collected data from the university for the purpose of having a better understanding of the students' reception expectation for the university's academic services. Voice-based analyses, on the other hand, have been proved to be more suitable and effective for such applications (Chavan and Gohokar, 2012).

The focus of this project is, therefore, to apply Voice Emotion Detection technology, combined with the use of machine learning techniques, on spoken interviews about the university's academic services to determine students' emotions with respect to the students' answers. This analysis is expected to help the university administration to explore these areas to improve the student life and experience by finding out the acceptance of students towards the conduct of study units like lectures and seminars. Since learning experiences are subjective and emotion influenced, emotion analysis may be beneficial to this area. The results must be evaluated critically and may provide a basis for suggestions on university improvement measures in future work. However, it is not presumable that every student knows, what will be beneficial to her/his academic learning and not every suggestion may be applicable for the whole body of students. Nevertheless, acceptance is an important consideration in academic planning of the university

To prepare the voice emotion analysis, a questionnaire was designed with a consideration of both contextual and psychological aspects to gather students' feedback on the university's academic services in addition to triggering the students' emotions in their answers. Data preprocessing, feature extraction and classification techniques were applied to provide accurate predictions of the emotions and provide possible insights into the students' experiences at Rhine-Waal University of Applied Sciences.

This paper is divided into the following chapters:

- Chapter 1 states the aim and motivation of the project.
- Chapter 2 provides a review of the literature related to Voice Emotion Detection.

- Chapter 3 outlines the work carried out in the project, including the questionnaire design, interview process, data preprocessing of interview audio files, feature extraction and classification of the emotions.
- Chapter 4 discusses the evaluation including different annotation processes and cross-validation on the classification output.
- Chapter 5 covers the results including both predicted and academic analyses.
- Chapter 6 discusses the challenges and limitations of the project.
- Chapter 7 provides conclusions on the research and discusses possible future works.

2 Literature Review

2.1 Technical Review

2.1.1 Emotion Models

Voice emotion detection strives to predict human emotions, therefore it is necessary to define both nature and description techniques of emotion specifications.

Emotions are defined in a variety of ways, however, two criteria of emotions, prevail and are agreed upon in the scientific community (Brave and Nass 2003). First, an emotion is a reaction to occurrences, which are considered significant to the respective person and appeal to necessities, aspirations and preoccupations. Second, an emotion incorporates parts from physiology, affect, behavior and cognition.

The term “emotion” is often used as a synonym to other terms, such as “mood” or “sentiment”, which does not reflect the details of each concept. While an emotion is directed at certain objects, such as a person, topic or event, moods are broader and not directed (Brave and Nass 2003). In temporal comparison, emotions are shorter than moods, which, in contrast, extend over a longer time-period. Moods and emotions are highly interrelated and influence each other. While emotions can initiate a mood, moods in return can affect and cause a type of emotion. A verbal classification in languages such as English often does not distinguish between moods and emotions and the same words could detail both concepts, e.g. the word “happy” is not limited to either a mood or emotion. Emotions and moods both are temporal in nature, opposed to sentiments, which may hold permanently. Consequently, sentiments are rather attributes to an individual than a condition, which the individual is in.

Emotions may be separated by various characteristics, such as the individual’s appraisal, precedent actions and circumstances, reaction, physiology and expression, e.g. in form of voice or facial expression (Ekman 1992). To segment combinations of these characteristics, emotions may be

grouped into emotion families, generally referred to as “basic emotions”, which comprise a multitude of similar emotional conditions. As an example, Ekman could distinguish 60 different anger expressions, which resemble each other for example in muscular patterns that led to the upper eyelid being raised, brows being lowered and drawn together, while the lips were pressed together. Oatley and Johnson-Laird could narrow down the count of emotions to five basic emotions, which are expressed independently of culture: Happiness, sadness, anxiety, anger and disgust (Oatley and Johnson-Laird 1987). While there are multiple models and names for the distinct emotions, these five basic emotions in precisely these or synonym words are part of most models (Vogt 2010). It is, however, seldom that these emotions occur in pure form, but usually are mixtures of different basic emotions.

Emotions are not always described in distinct classes, but may also be described as points in a dimensional space that mostly is defined by two or three characteristics (Vogt 2010). The most common variables in use are arousal and valence, while sometimes dominance is added as a third attribute. On one hand these dimensional may add more detail to the emotion specification, on the other hand they may be too simple to encompass all different features, which represent an emotion. Another challenge, specifically with regards to voice emotion detection is the annotation of data into statistical classifiers, that are better understood and can be edited faster in distinct emotions than in points in the dimensional space. Until now, Vogt found discrete emotional classes to return the most useful outcomes.

2.1.2 Features

An abundancy in emotional indicators through speech signals is essential in the detection of speaker emotions (Prakash, Gaikwad et al. 2015). To extract appropriate features is therefore crucial in any voice emotion detection project. The selection of features is a vibrant research subfield since the beginning of speech emotion recognition. Selected features should remain stable under influences of sounds in the environment as well as differences through culture and language (Schuller 2018). The increase in number of features stands in contrast to the small base of training data, which is available in voice emotion, however, a current trend enlarges the set of features to multiple thousand features extracted by brute-force.

In general, two groups of attributes consisting of long-term features and short-term features can be separated (Prakash, Gaikwad et al. 2015). Short-term attributes, such as pitch, energy or formants are measured in short time slots of 20-30 ms. (Vogt 2010). Long-term dimensions form means, maximums, minimums and other statistical functions in longer time frames, since the produced suprasegmental observations often lead to more interesting results in this field. Generating new features through the application of statistical functions may on the one hand add information to the

analysis, on the other hand it increases calculation effort and it may lead to an overfitting of the data, described as the “curse of dimensionality”. With respect to short-term and long-term features, the break point in the voice recording must be considered and decided upon. It is stated prevalent to cut recordings into longer units, as for example turns, utterances, phrases or words. Differently, equal size time intervals of the recordings could be used and are considered more applicable by autonomous machines.

As stated, many different features may be used in voice emotion detection, of which very important ones are presented in the following.

The dimension pitch refers to the listener’s ear perceiving tone height (Vogt 2010). In most applications, this would be represented by the fundamental frequency, although these terms differ in that pitch is transformed and lowered in the human perception of the sound. Especially changes in pitch may indicate emotions, for example an uplift in the pitch may indicate greater arousal. This feature is relevant, however, usually found to be less relevant than suspected.

Resonance in speech produces the attribute of formants, which is measurable as local maxima in the frequency spectrum (Vogt 2010). The global maximum is generally given by the fundamental frequency and consecutive local maxima ordered by frequency compose the formants. In contrast to pitch, formants are frequently suspected to show low significance to emotions, however, higher first or third formants are found to indicate positive emotions (Biersack and Kempe 2005, Waaramaa, Alku et al. 2006).

The sound strength specifically as experienced by the listener is referred to as loudness (Vogt 2010). This is mostly used as a related feature, as it is difficult to be determined directly. A Fourier transformation may be used to find out the energy, which is a combination of all available noises that resembles the actual perception of the loudness in the ear. A high arousal can be related to high energy and the nature and pattern of changes in the energy level. The form of energy curve depends not only on emotion, but moreover on influences such as phonemes, speaking style and type of content.

The feature of mel-frequency cepstral coefficients (MFCCs) is mainly analyzed in automatic speech recognition, since MFCCs are applicable for filtering of linguistically unrelated elements (Vogt 2010). Although linguistics is irrelevant to voice emotion detection and this should render the information less useful, MFCCs are found to contribute significantly to correct predictions. MFCCs are extracted by transforming a windowed signal with the Fourier transformation followed by the application of a Mel-scale filter bank. The logarithmic spectrum is converted into a cepstrum using a discrete cosine transform (DCT). The amplitudes of the generated cepstrum present the MFCCs, of which most often the first 12 coefficients are taken into the analysis.

Wavelet transformations as features can represent a functional transformation of a signal, such as in the Fourier transformation (Daubechies 1992). Typically used wavelets termed “mother-wavelets” are Haar-, Daubechies-, or “Mexican Hat”-wavelets (Vogt 2010). Wavelets integrate time additional to the most times solely used frequency. Wavelets are currently seldomly used but hold potential for further research. Next to the already described features related to frequency, more calculations can be drawn from it, e.g. spectral slope, mean and center of gravity.

Another important dimension is the extent of speech units in form of utterance length, word length or syllable length as well as speaking rate, which refers to the quota of words or syllables to time (Vogt 2010). When speakers were identified as happy, this for example related to a faster speaking rate than of sentences spoken with different emotions.

Measurements of jitter, shimmer and harmonics-to-noise ratio (HNR) allow the inclusion of the attribute voice quality, which refers to a speaker speaking in different manners like whispering, breathing little, much, short or long as well as speaking harshly or softly and creaky or firm speech (Vogt 2010). Voice quality as an attribute is often expected to support the prediction significantly but is mostly discarded from the feature set.

Feature values may differ between different speaker types, e.g. when comparing across age, gender, language and culture (Vogt 2010). Despite these differences, features and related techniques remain equal. Voices of children show a higher variability in values that can be treated with a vocal tract length normalization. Scherer found speech emotions to be majorly pancultural and influenced by universal psychobiological mechanisms (Scherer 2000). Predictions across cultures could be made with a more significant accuracy than random correlation. If, however, segmental and suprasegmental analyses with semantic values opposed to solely nonlinguistic interpretations are performed, language may influence predictions.

Until now, there is no decision about the most adequate feature set found in the scientific community (Schuller 2018). To determine the feature set for a classifier, one may either select features with a subject-of-matter approach considering the influence of each indicator to the voice and emotions in the given scenario or decide on the features with a feature reduction algorithm, as for example the sequential forward selection (SFS). Additionally, the number of attributes may be reduced by limiting the feature space through transformations, such as the principal component analysis (PCA).

2.1.3 Voice Emotion Databases

There are several databases available for voice emotion detection in different languages (English, German, French, Chinese, Spanish, etc.). Also, some database voices are recorded in multiple languages like voice material in databases recorded at a luggage belt of Geneva Airport (Scherer and

Ceschi, 1997). Most of these databases comprise voices with different emotions from professional actors in their respective languages and a few databases contain natural voices which are recorded in different scenarios like stress during amusement park roller-coaster rides, flights in helicopter cockpits etc. The most common emotions available in these databases are anger, sadness, happiness, fear, disgust, joy, surprise, boredom, stress, etc.

Some famous databases are:

1. **Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) Database** (Livingstone and Russo, 2018): This is a Canadian database in English language where two statements, namely 'Kids are talking by the door' and 'Dogs are sitting by the door', are used to show different emotions.The voices were recorded are in a professional recording studio at Ryerson University. It contains voices of 24 professional actors (12 female actors and , 12 male actors) in speech and song format with two different intensities, 'neutral' & 'strong'. The actors produced 104 distinct vocalizations consisting of 60 spoken utterances and 44 sung utterances. The recordings are conducted in 3 different modes - "Audio-video", "audio only" and "video only". The database contains 7,356 recordings , built from 4,320 speech recordings and 3,036 song recordings. The recorded emotions are neutral, calm, happy, sad, angry, fearful, disgust and surprised.
2. Interactive emotional dyadic motion capture **(IEMOCAP)** Database (Busso et al. 2008): This database is available in English language and recorded at the Speech Analysis and Interpretation Laboratory (SAIL) at the University of Southern California. It contains audio-visual recordings of 10 professional actors (five female actors, five male actors) in dyadic sessions. Additionally, facial expression and hand movement were recorded. The database contains approximately 12 hours of data with 10,039 utterances (scripted: 5,255 & spontaneous: 4,784) and the average duration of utterances is 4.5 seconds with an average length of 11 words in each utterance. The recorded emotions are "anger", "frustration", "happiness", "sadness" and "neutral".
3. **Berlin Database** (Burkhardt et al. 2005): This database is available in German language. It was recorded by 10 German professional actors (five males, five females). It has approximately 800 utterances and contains short and long sentences. It contains 7 different emotions and these emotions are "anger", "fear", "boredom", "disgust", "joy", "sadness" and "neutral".
4. **SUSAS Database** (Hansen et al. 1998): This database is in English. The recordings are done under simulated or actual stress in different conditions like amusement park roller-coasters,

helicopter cockpits, etc. It contains approximately 1,600 utterances from 36 speakers (13 females, 23 males). The different emotions recorded are “anxiety”, “fear”, “depression” and “angry”.

5. **Geneva Airport Luggage belt Database** (Scherer and Ceschi, 1997): This database is available in multiple language (English, French, German and Italian). is the voices were recorded at the lost luggage desk of Geneva airport and it contains both audio and video recordings. It consists of recordings of passengers' interaction at desk from 112 passengers and 12 airline employees and the length of each recording is between 10 and 20 minutes. The emotions available in this database are “worry”, “anger”, “resignation”, “indifference” and “good humor”.
6. **Belfast Naturalistic Database** (Douglas-Cowie et al. 2000): The language of this database is English. It contains audio and visual recordings of people debating on television programs. It has 239 clips between 10 and 60 seconds from TV recordings (209) and interview recordings (30). These clips have 125 speakers in which 94 are females and 31 are males and covers a wide range of emotions.
7. **DARPA Database** (Walker et al. 2001): This is available in English language and the recordings are done when users are trying to make air travel arrangements through call centers over a phone. It has a total of 13,187 utterances, of which 1,750 could be linked with emotions. The emotions recorded are frustration and annoyance.
8. **Sensitive Artificial Listener (SAL) Database** (Cowie et al. 2004): The language of this database is English. It has audio-video recordings of 10 speakers conversing with an artificial intelligence machine. The recording is available for approximately 10 hours and it covers wide range of emotional states.
9. **LDC Database** (Liscombe et al. 2003 and Yacoub et al. 2003): This database is available in English language. This was recorded by 8 professional actors (5 female actors, 3 male actors) reading short dates and numbers of 4-syllables each. The emotions recorded in this database are “anxiety”, “cold” “anger”, “contempt”, “boredom”, “despair”, “elation”, “disgust”, “hot anger”, “interest”, “happy”, “pride”, “panic”, “shame”, “sadness” and “neutral”.

2.1.4 Modelling techniques

There are different basic methods applied for the prediction emotions from voice. The following fundamental approaches are taken into consideration.

2.1.4.1 Acoustic Phonetic Approach

In the acoustic phonetic approach, the labelling of audio signals is crucial. In this process, different types of sounds are identified and labelled. In any spoken language, there exist a finite number of phonemes which are exclusive and described by the acoustic properties which form the basis for this approach.

Steps involved in this approach are:

a) Spectral analysis of voice: A range of acoustic properties of various phonetic units is described here.
b) Segmenting and labelling of voice: In this step, the voice is segmented and labelled.
c) Determining the string of words: Finally, the string of words is identified from segment-labelled voice.

2.1.4.2 Pattern Recognition Approach

Pattern Recognition approach is the most popular approach in determining the emotion from voice. It involves two steps:

a) Pattern training: In any audio signal, the acoustic properties helps to identify patterns in different sounds. Training assists in identifying repeated patterns for labelling to increase recognition quality and is the first step in supervised learning methods.

b) Pattern comparison: This step compares the voice to be recognized (unknown voice) with the different patterns observed in training stage to find the identity of the unknown voice on the basis of the best matched pattern.

This approach is further classified into two methods

1. Template based: In this approach templates having prototypes of various voice patterns are made for reference which acts as a dictionary of words. Then voice to be recognized (unknown voice) is matched with these reference templates and the best matched pattern is selected.

2. Stochastic based: This is the probabilistic model approach which deals with uncertain and incomplete information. This is the most suitable approach for voice recognition.

Various popular methods based on stochastic approach are:

Support Vector Machine (SVM) is a supervised machine learning method used for classification problems (Vogt, 2010). This algorithm uses data items as points in an n-dimensional space and finds the best hyperplane for segregating different classes of data points. The hyperplane is searched to maximize the shortest distance between the data points of two classes and this is achieved by the help of support vectors. The support vectors are the inner-boundary lines of different classes (Tan, 2007).

The benefit of SVM is that it can solve non-separable class problems through the transformation into a higher-dimensional space. SVM mainly deals with two class problems but can handle multi-class problems by splitting into a set of two class problems. However, there are some demerits to SVMs, for example will the increase to a higher dimension raise the number of coefficients related to it, which may result to an overfitting of the model, if not taken care of.

b) The Hidden Markov Model (HMM) is mainly applied for reinforcement learning, but is also used for voice recognition. They consist of latent variables (hidden variables) for each of the N possible states that a hidden variable at time t can be in. There is a transition probability from this state to each of the N possible states of the hidden variable at a certain time (Vogt, 2010).

c) Dynamic Time Warping (DTW) is used for finding similarities between sequences of observations that vary in time or speed. To apply it adequately for a voice recognition system, it should be able to handle or deal with different speeds of multiple speakers.

Furthermore, for an optimal sequence match in DTW, certain rules and restriction must follow:

i. Sequence matching of all the indices of first with at least one index of other and vice versa
ii. The first index of both the sequences must be matched (but it should be the single match)
iii. The last sequence of both the sequence must be matched (but it should be the single match)
iv. Indices mapping from first to other sequences must be in increasing order and vice versa

2.1.5 Artificial Intelligence Approach (Knowledge based approach)

The Artificial Intelligence Approach (Knowledge based approach) is a mixture of pattern recognition and the acoustic phonetic approach. In this approach, the recognition of voice is conducted similar to the way how human beings apply intelligence to read emotions of others. This approach uses linguistics, phonetics and spectrogram information of the voice. In the context of voice emotion detection, the knowledge-based approach requires a detailed study of spectrograms and is incorporated using rules and procedures which helps in its pure form of knowledge engineering design.

However, due to the problem faced in the quantification of expert knowledge, the success of this approach is limited. Further, the integration of different levels of human knowledge like phonetics, syntax, phonotactics, semantics, lexical access and pragmatics has many difficulties. For this approach, artificial neural networks method are more reliable which are based on a set of operations to a set of variables with each operation in the network having a weight attached.

2.1.5.1 Classification Algorithms

1. SVM: The SVM algorithm considers a set of features as data points in the hyperspace. A hyperplane is divided in such a way that it maximizes the distance between this hyperplane and each support vector, while the objective function is minimized. Sometimes, dividing the hyperspace is a difficult. Hence, SVM allows to misclassify a few data points using a margin but increases the overall performance. (Òscar, 2017)
2. Naïve Bayes: It is a simple learning method based on the Bayes conditional probability rule. The probability of a set of observations to belong to a certain class bases on the method that each single feature contributes independently to the final probability result to be a class and each feature has its own distribution.
3. K-Nearest Neighbors Algorithm (KNN): The K-Nearest Neighbors algorithm is a non-parametric, instance-based learning method used both in classification and regression. In classification, an object in space is identified by a majority vote of its nearest neighbors. The value of k is always positive and requires a relative weighting so that the nearest neighbors can contribute more than the distant objects. In any training dataset, it looks for similar feature to predict values whenever new data points are added to test. It implies that the new point is assigned a value based on its closeness to data points in the training set. Hence this algorithm is very effective and robust for a large training data.
4. Multilayer perceptron (MLP): The multilayer perceptron (MLP) is a neural network using feed-forward having up to two hidden layers. It is a function where one or more independent variables are used to minimize the prediction error of one or more target (dependent) variables. It belongs to a class of Deep Neural Network (DNN) and is based on the Backpropagation (BP) learning algorithm. However, major drawbacks include heavy computational burdens of densely connected multilayers, structures and deep iterations. Hence, it requires a considerable amount of time for recognition accuracy.
5. HMM: The HMM is one of the most important machine learning models in speech and language processing. A Markov Model is a stochastic state space model involving transitions (randomly) between states where the probability of the transitions is only dependent upon the current state, rather than any of the previous states. The model is known to possess memoryless markov properties.

   Hidden Markov Models are Markov Models where the states are hidden from view, rather than being observable. Instead there are a set of output observations, related to the states, which are directly visible.

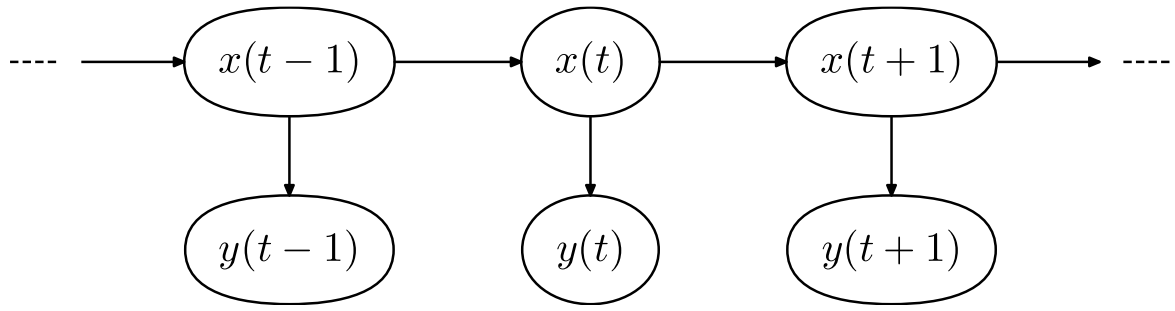


Figure 1 Hidden states in HMM

Source: Wikipedia contributors. (2018). Hidden Markov model. In *Wikipedia, The Free Encyclopedia*. Retrieved 15:37, February 7, 2019, from https://en.wikipedia.org/w/index.php?title=Hidden_Markov_model&oldid=872390786

The above diagram shows the architecture of hidden states in HMM. A random variable x(t) is the hidden state for time t and y(t) which is also a random variable is the observation at time t. Conditional dependencies are denoted by the arrows in the diagram. The conditional probability distribution of x(t) at any given time t depends on value of the hidden variable x(t-1). Any other preceding value from time (t-2) has no influence in this model. This property is extended in the hidden model where the state space of the hidden variables is discrete with the observations can be both discrete and continuous. Two types of variable exist in HMM:

a) Transition probabilities: They control the way the hidden state at time t is considered, given the hidden state at time (t-1)
b) Output probabilities: It controls the distribution of observed variable at a particular time provided the state of the hidden variable is given at that time

There are N possible values for a hidden state space. Given for each of these N possible states, there is a transition probability from the current state to each of the N possible states of the hidden variable at time (t+1), thereby providing N2 transition probabilities. This NxN matrix of transition probabilities is called a Markov matrix.

In speech recognition, HMM provides a robust framework for modelling the relationship between a small set of phonemes (hidden states) and acoustic features (observations). On the contrary, there is a limitation of HMM on the assumption of conditional independence. It means that the observation is independently and similarly distributed for a given hidden state. However recent developments in Deep Neural Network (DNN) for acoustic modelling has somewhat removed this problem.

2.1.6 Voice Emotion Detection Tools

A variety of tools for emotion recognition from speech have been developed and are constantly being updated as well as maintained. As for the academic nature of this project, this section only discusses open-source frameworks that are freely available rather than the commercialized counterparts.

2.1.6.1 EmoVoice

EmoVoice is a framework aimed at voiced emotion detection developed by the Institut für Informatik, Universität Augsburg. It is part of a bigger umbrella framework, Social Signal Interpretation (SSI) framework, which is also open-source (Vogt et al., 2008). EmoVoice includes the following modules:

- Database creation
- Feature extraction
- Classifier building and testing
- Online recognition

The Application Programming Interface (API) exposed to the end-user of EmoVoice is in Python 3 which enables rapid development and testing. The framework is well-documented, well-organized and is actively maintained. Under the SSI umbrella, EmoVoice can also be used in combination with other libraries, namely:

- LIBSVM – A Library for Support Vector Machines
- LIBLINEAR – A Library for Large Linear Classification
- openSMILE – The Munich Versatile and Fast Open-Source Audio Feature Extractor
- Emo-DB – Berlin Database of Emotional Speech

2.1.6.2 OpenSMILE

OpenSMILE is an open-source, real-time-capable feature extraction tool developed for research-oriented applications (Eyben et al., 2010). It was initially developed at Technische Universität München and is currently supported by audEERING. It provides the following modules:

- Audio signal processing
- Feature extraction
- Visualization

OpenSMILE does not expose an API for developers, instead the source code which was written in C++ has to be manually compiled. While programs in C++ are inherently faster and more powerful than those written in Python/R, C++ has a much steeper learning curve.

2.1.6.3 MIRtoolbox

MIRtoolbox is a Matlab library developed by the University of Jyväskylä, originally developed to investigate the relation between musical features and emotions (Lartillot, 2011). The library includes different functions for:

- Audio I/O and processing
- Segmentation
- Feature extraction
- Post-processing and statistics

MIRtoolbox would be best suited for those familiar with Matlab. The tool is well-documented with tutorials and research papers published related to it.

2.1.6.4 Praat

Praat is an open-source program for analyzing speech in linguistic research which was developed by the University of Amsterdam (Boersma, 2002). The program is cross-platform and provides a friendly graphical user-interface for speech analysis. Features of the program include:

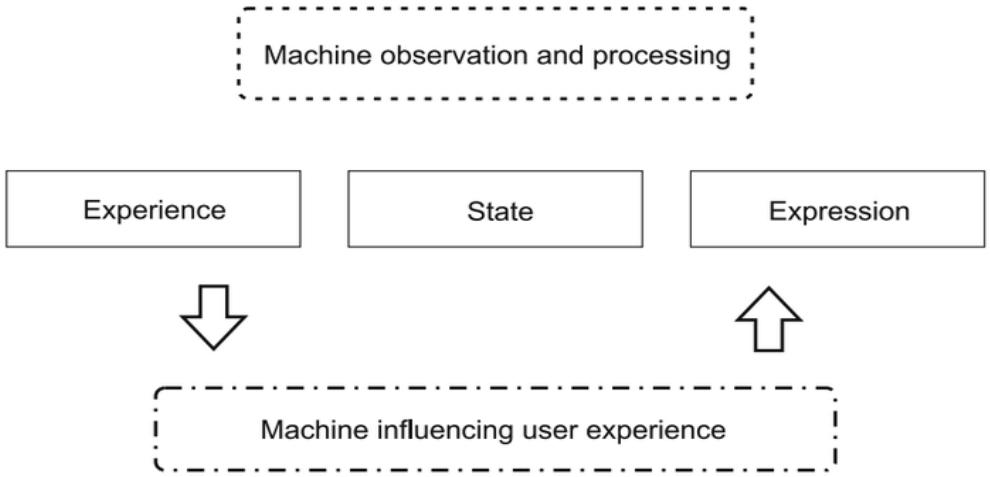
- Recording sounds, opening and saving sound files
- Sound file manipulation, i.e. converting among file types, cropping, slicing, filtering.
- Spectrogram
- Measuring duration
- Measuring pitch, formants, intensity/amplitude

2.2 Application Areas Review

2.2.1 Applications of Voice Emotion Detection Systems

2.2.1.1 Robotics

Robots act as assistants or companions and have been staple diet of science since many years. Robots can be used for industrial and domestic tasks and are complete package of electrical, mechanical, automation and computing field of technology. With the rapid advancement of technology, robots can now be controlled by human voices. This allows the users to work on other tasks and free up their hands. “Affective Computing” is the research field in Computer



Source:Lugović, S., Dunđer, I. and Horvat, M., 2016, May. Techniques and applications of emotion recognition in speech. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2016 39th International Convention on* (pp. 1278-1283). IEEE.

Science which focuses on improvement of human communication with the machines (Brave, 2003). The basic framework of Affective computing is given above:

Many researches in this field are carried out in a scientific manner such as robot arm control, door lock control, mobile vehicle systems control and control of wheel chair using voice control systems. The main goal is to establish natural communication between humans and robots. This would eliminate the needs of devices such as keyboards and mouse and leverage suitable and intuitive ways of communication for non-technical and disabled users.

The field of social robotics is a growing branch in the field of robotics which is integration of technical disciplines, psychology and sociology. A robot could modify its behavior according to a user's humor, or even more, express a related emotion, resulting in an increase of synergy with the owner in the cooperation.

2.2.1.2 Automated Call Centers

A call center is a popular term for a service organization that handles large number of telephone calls. In today's world, all large companies have at least one call center. Calls are classified into inbound and outbound. Inbound calls are calls that are initiated by customers to query regarding product or to find out about new schemes. Inbound calls a sender initiates a call to the customer to inform the customers regarding the new services or sell the products. Call centers have been aided by a range of telecommunications and computer technologies including automatic call distribution (ACD), interactive voice response (IVR) and computer telephony integration (CTI) allowing the actions of the computer to be synchronized with actions happening on the phone (Maglio, 2006).

IVRis a popular technology designed to provide callers with verbal, fax and online inquiries without the assistance of people. Following are the advantages of using IVR in call centers:

- **Improved customer satisfaction**

The usage of IVR can reduce queue times. Customers spend less time on hold and need only few attempts to get in touch with the company. Another benefit are extended service hours. Self service applications also help to maintain privacy. There are some transactions which the customer would not prefer to discuss with an agent, which may be used in this case with IVR.

- **Reduced costs**

Staffing expenses can be highly reduced. Service phone call with no human interaction costs less than 50 %. Salaries are not only expense associated with staffing. Hence, if the agents do not have the skills that they need they cost money.

In summary automated call centers can be used in banking, brokerage, insurance, sales and catalog houses and they prevent costs from rising, reduce costs, improve service levels and hence overall success of the company.

There are some limitations to this section for example the customer might not always be satisfied by the service of automated call centers and might get annoyed listening to audios.

2.2.1.3 Automated Cars

Cars become increasingly rich with various interactive systems that are installed in the car. New cars provide speech enabled communications such as voice dial as well as control over the car cockpit including entertainment systems, climate and satellite navigation. In addition there is the potential for a richer interaction between driver and car by automatically recognizing the emotional state of the driver and responding intelligently and appropriately (Litman, 2017). Driver emotion and driving performance are often intrinsically linked and knowledge of the driver emotion can enable to the car to support the driving experience and encourage better driving. The limitations to automated cars is that the car insurance companies can have entire information about the car holder and use that information in deceptive manner.

2.2.1.4 Lie detection Systems

While body language has been in focus for detecting lies, voice may be more accurate to detect lies than merely observing the behavior. Voice analysis software systems perform layered voice analyses in order to determine the different stress levels associated with the voice, emotional reactions and cognitive processes associated with the subject’s voices (Ekman, 1991). Different researches carried out show that frequencies within the human voice affect honesty.

2.2.2 Applications of Text Emotion Detection

Text emotion detection is used in different life situations depending upon the goal and concept. Given below are major application areas where text emotion detection is used:

2.2.2.1 Text emotion detection in Business

The applications of text emotion detection in business cannot be overlooked. The company can get actionable insights by exploiting the unstructured data and thus improve their business. By knowing the emotions from competitor's text, company can gain valuable insights and perk up their performance. Text emotion detection plays a crucial role in detecting ongoing trends in the market. Customers are truly responsible in making any business successful.in internet savvy area, emotions of the customer can be analysed from text through their social posting and online feedback. Emotions can be analysed in text and put in to different categories which will help companies to detect which part of business needs improvement.

2.2.2.2 Text emotion detection in Politics

With the help of text emotion detection, the campaign managers can track how the voters feel about different issues and problems. People's opinions can be detected and analysed by conducting different surveys and based upon these opinions the related speeches and actions can be taken care of. Text emotion detection also helps to know the effect of new policies and regulations on people.

2.2.2.3 Text emotion detection in Psychology

In this era of social media, people express their opinions on different topics, products and services online. With the help of these emotions, psychologists can analyse the mental state of the person as well as neurological illness of the person. IBM is conducting research on the statements written by patients to extract the information about how patients feel by looking on to the words and the language used in the sentences. This method is useful in detecting which patients are severely affected and need what kind of psychological treatment. Text emotion detection can be used to prevent suicides by providing appropriate treatment to the psychologically affected patients and later on tracking their mental state for affective treatment.

2.2.2.4 Text emotion detection in Finance

Social communication platforms are becoming more and more popular for the people to express their opinions. These platforms attract the attention of financial investors to examine and analyze the opinions of the individual users. There are some difficulties while performing text emotion detection in finance. These can be identifying useful information content, representing unstructured text in a structured format under scalable framework and quantifying this structured data. With the help of more robust categorization systems based on new classifiers and features and through the

incorporation of more news sources and authors these problems can be solved as well as the accuracy can be improved.

2.2.2.5 Text emotion detection on non-academic services at HSRW

The research "Sentiment Analysis and Machine Learning Techniques: Case Study of Rhein-Waal University" aimed to identify and analyze the sentiments of the students of the Rhein-Waal University towards available non-academic services and leisure activities. This analysis helped Rhein-Waal University of Applied Sciences to explore these areas and improve student's life and experience. However, there were some limitations to this project such as drafting of questionnaire, platform limitations and data accuracy. The accuracy of this system can be improved using voice emotion detection as it analyses on voices not on the words (Saha, 2018).

2.2.3 Facial Emotion Detection and Multimodal Detection

Multimodal emotion detection is the combination of different emotion detection techniques into one unified package. Such emotion detection techniques include voice emotion detection, facial emotion detection, posture/gesture detection and more. There have been a lot of areas where multimodal emotion detection is applied.

2.2.3.1 Media Retrieval and Indexing

Emotion detection techniques can be applied to video clips and movies to greatly aid the process of indexing (Paleari et al., 2010). For instance, movies can be indexed into genres using both visual- and auditory-based detection methods on actors' and actresses' facial expressions and voices. Emotion detection techniques can also further be used to provide better summarizations of movies, assist the trailer creation process and help predict viewers' emotions while watching movies (Baveye, 2015).

2.2.3.2 Video Game Testing

Facial and/or multimodal emotion detection can be used for testing video games' reception (Bahreini et al., 2014). During the testing of a video game, gamers' feedbacks are a valuable asset for perfecting the beta video game into a final product. Facial emotion detection can thus be applied in real-time to detect the gamers' facial expressions, helping video game developers better understand gamers' emotions throughout different parts of the video game. As emotion detection in such an area is inherently less effective and more intrusive when for instance written-based feedback forms are used, facial/multimodal emotion detection is faster, more efficient and reliable for understanding user experience.

2.2.3.3 Software Usability Testing

Facial and/or multimodal emotion detection techniques can be applied in extended software usability testing. Using cameras, biometric sensors and keystroke analysis tools, test users' emotional

expressions when using the software can be recorded and analyzed (Kołakowska et al., 2014). With the help of multimodal emotion detection techniques, extended usability tests can be designed and carried out to acquire and understand users' experiences and thus help improve the product. Examples of the types of software that can benefit from multimodal emotion detection techniques include web pages, integrated development environment (IDE), and generally those with a graphical user interface(GUI) to interact with the end-users.

3 Methodology

The project consists of the following nine stages illustrated below and which are laid out in detail in this chapter.

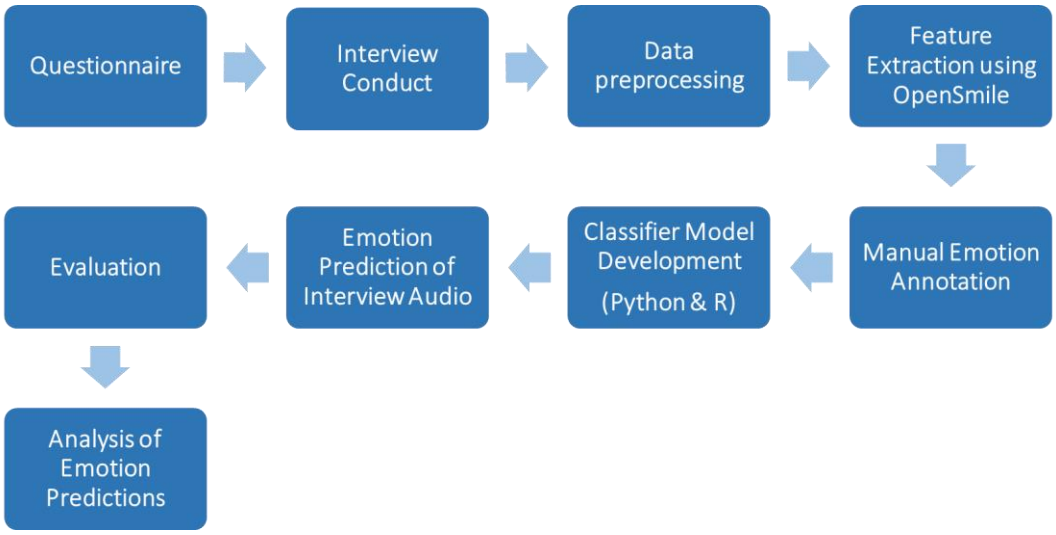


Figure 2 Project Stages

3.1 Research, Questionnaire and Interview

3.1.1 Psychological Research

One approach to analyzing emotions from survey participants may be to ask for their emotions directly and to analyze the answers based on the content of responses. This research's intent in using voice emotion detection opposed to this approach is to detect more accurate information regarding the survey participant's emotions and opinions. The causes leading to this hypothesis are explained in the following.

Truth of information: In surveys, participants may have a restriction in explaining their emotions openly and without omissions, since they may fear reactions to their opinions. Even in anonymized surveys there may be doubt on the adequate anonymization, careless storage or sharing of data as well as the reliability of interviewers. Not only protection of oneself may be aimed for, but also protection of lecturers who may be associated with a certain critic. For example, if a certain lecturer

may be known for poorly prepared material but the teaching style would be perfectly suited for one of his students, this student may refrain from criticizing poorly prepared material in general. With voice emotion detection, it is very difficult for a non-professional actor to fake emotions, as emotions are recognizable in a natural and subtle form.

Level of politeness: Closely related to the truth of information, students may be used to filtering their true emotions under a filter of politeness, which may prevent them from directly criticizing and exposing their emotions. Dependent on academic and work experience as well as culture and education, critic for others, especially of people in higher positions, may be valued high or may be regarded as inappropriate. As for untruthful statements, voice emotion detection may be able to detect emotions despite this distortion through politeness.

Verbalization of opinions: Given the same opinion, based on knowledge of the spoken language, people may express their opinions differently. If certain terms are not well known or seem foreign to a person, there is a chance of missing precision in sentences or even accidental expression of opinions different to those, which are intended to be conveyed. Words may be connoted differently according to a person's experiences and culture. As voice emotion detection is not based on words, it may correctly indicate emotions and hence opinions for a topic.

Missing time or interest in survey: In a lack of time or interest, a person may answer questions with lack of detail or without thinking much about it. Asking questions that trigger emotions, survey participants may tend to engage more in the questions and emotions may be received from fewer words then from elaborate explanations.

Subconscious emotions: Depending on the level of self-reflectiveness, students may be able to consciously detect their emotions in different levels. Emotions may be subconscious and may be not even known to the students feeling these emotions. In this case, voice emotion detection may help to find out the unintentionally hidden emotions.

Consequently, there are many aspects, which lead to a hypothesis that voice emotion detection may add additional value to surveys by analyzing emotions despite distortions in language.

3.1.2 Questionnaire Design and Interview

To receive feedback from the students of HSRW, a questionnaire which focuses on five main areas of teaching services was designed. These areas include:

3.1.2.1 Questionnaire Design

- Teaching methodology and style
- Course curriculum

- Quality of teaching
- Timing of classes
- Evaluation patterns.

The questionnaire was reviewed and adjusted with suggestions from the head of psychology department of the University of Applied Sciences Rhein-Waal. In the discussion with the head of the psychology department, the major challenge to stimulate strong and honest emotions from students was discussed leading to the suggestion to show different pictures, ask triggering questions and to describe two scenarios (one positive and another negative scenario related to a fictitious student's experience at HSRW).

The suggestion to create two scenarios for one question was implemented to extend the length of the answers responded by the interviewee. As suggested, both positive as well as negative scenarios were presented to not influence the answer to one direction. As the statement of scenarios was estimated the most viable implementable solution under a balanced narration of differently connotated emotions and experiences. To lead interviews to an emotional atmosphere, emotionally connotated words such as "frustrated", "tired" or "optimistic" were used. Additionally, stories were written and told in an informal way on purpose, as a formal tone impedes emotional reactions. As one example, the question related to the timing of the classes is as follows:

Timing of classes

What is your opinion on the start times and duration of classes in your studies?

- Student 1: The student feels uncomfortable with the start times of his classes. He has the habit of working late in the night, so classes starting at 8:00 are too early for him. In his prior experience, everything started at 9:30, so he feels stressed and half asleep trying to attend a very early class. Then, if the class is continuing for multiple hours and continues until the evening, he feels very tired not able to concentrate. With this, he feels frustrated and angry. Especially block courses result in big disappointment.
- Student 2: Student 2 feels that block courses are very effective, since they are very condensed. Some courses are only on few days then running for the whole day. After that, they leave a long time until the next lecture. She feels motivated about the fact that she can manage her time on her own and therefore she also feels confident and optimistic about the grades in these subjects. She also feels comfortable with the start times.

In this question the scenarios are explained as a negative experience of the student 1 and the positive experience of the student 2. In similar way, scenarios were created for each question. A reinforcement of the creation of a more emotional atmosphere through experiences was achieved through the

interview of multiple students at once. At the end of each interview, one question was added that was asking the interviewed person, if she/he was in an emotional state to verify whether the person participating in the interview was in an emotional state.

3.1.2.2 Interview Conduct

In total, 30 interviews were taken, out of which 13 interviews were given by master students and 18 interviews by bachelor students. In total, the interviewed students were from 9 countries on 4 continents. It was observed that students were sharing their honest experiences related to the teaching services in HSRW. A separate room was booked for conducting interviews as the noise factor would be minimum in this room.

3.2 Implementation

An overview of the implementation can be seen from the figure below.

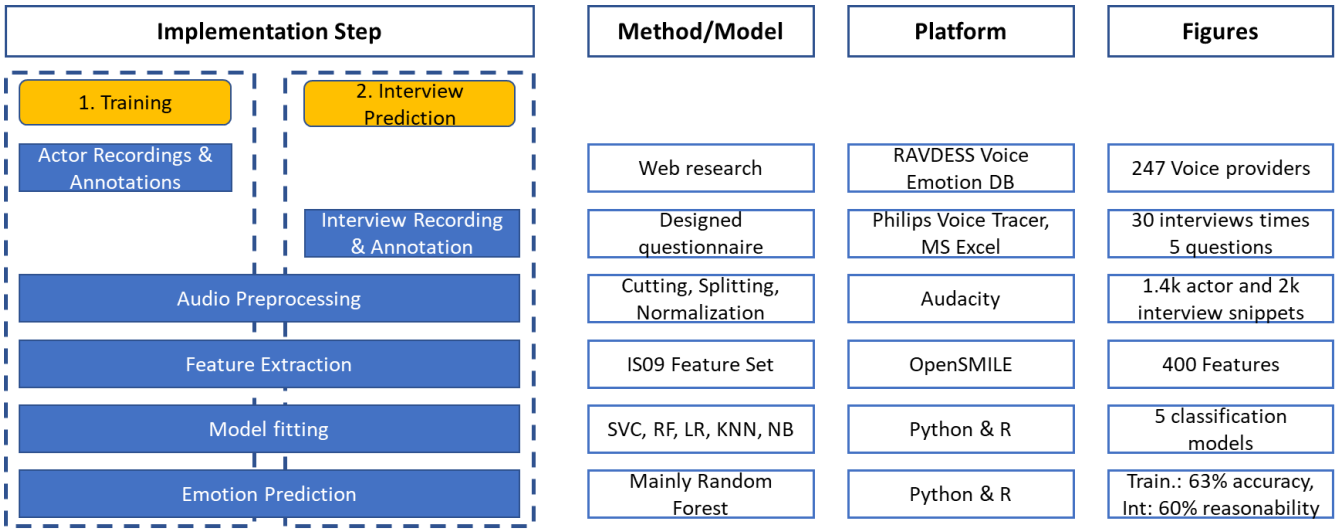


Figure 3 Implementation Overview

3.2.1 Data pre-processing of interview audio files

During the interviews, audio files were recorded which could afterwards be used for voice emotion detection. However, the waveform audio files from these interviews had to be prepared to be in an appropriate format prior to classification.

In the first step, all interview files were split according to interviewee and answer so that one waveform audio file would represent one answer from one interviewee. Questions from the interviewers were cut out to prevent using these voices in the classification. The cutting of these audios was done with the help of a software named Audacity. Audacity is a free, user-friendly audio editor and recorder for Windows, Mac OS and other operating systems.

The recording conditions and specifications of the interviews should be as similar as possible to those of the actor recordings, as these influence the MFCC values in the feature vectors. Since the audial

influences of the environment of the interviews as well as the microphone specifications may be different to the analogous situation during the recording of the actor files, the influence of these factors was reduced by normalizing the audio recording tracks of both interviews as well as actor data. The normalization as well as other audio adjustment was conducted also with Audacity®. With normalization, the peak amplitude of audio tracks can be configured to assimilate the balance of audio tracks. The maximum amplitude was normalized to -1,0 dB and the DC offset was centered on 0.0. The removal of DC offsets may prevent the deteriorating effects of noise and can permit extended volume of the recording.

Additionally, since the actor recordings usually have a length of ca. six seconds, this length should as well be used for the recordings in the interviews. Therefore, all interview answers have been split into parts of six seconds, of which each would later be used to predict the emotions from the respective interviewed people. Lastly, the sampling rates were equalized between actor data and interview data by setting the sampling rate to the one detected in the actor data: 48 kHz.

3.2.2 Feature Extraction

For feature extraction, an open-source real time capable feature extraction tool called OpenSMILE was used. It was initially developed at TU München and it is currently supported by audEERING. This tool is used for feature extraction as well as pattern recognition and it enables the user to extract large sets of audio features in real-time. OpenSMILE toolkit is modular and provides flexible feature extraction for signal processing and machine learning applications. OpenSMILE does not expose an API for developers, instead the source code which was written in C++ has to be manually compiled. It has a fast, efficient and flexible architecture and it can run on various platforms such as Mac OS, Windows and Linux. This tool can be used offline in batch mode for processing large data sets.

OpenSMILE provides config files for the frequently used feature sets. The most frequently used feature sets in OpenSMILE are as follows:

- Chroma features for key and chord recognition
- MFCC for speech recognition
- PLP for speech recognition
- Prosody (Pitch and loudness)
- The INTERSPEECH 2009 Emotion Challenge feature set
- The INTERSPEECH 2010 Paralinguistic Challenge feature set
- The INTERSPEECH 2011 Speaker State Challenge feature set
- The INTERSPEECH 2012 Speaker Trait Challenge feature set
- The INTERSPEECH 2013 ComParE feature set

- The MediaEval 2012 TUM feature set for violent scenes detection.
- Three reference sets of features for emotion recognition (older sets, obsoleted by the new INTERSPEECH challenge sets)
- Audio-visual features based on INTERSPEECH 2010 Paralinguistic Challenge audio features.

To extract the features, the INTERSPEECH 2009 emotion challenge feature set was used. It is represented by the config file config/emo_ISO9.conf. It contains in total 384 features and the features are saved in ARFF file where new features were appended to an existing file, which process is referred to as batch processing. With the help of *config/emo_ISO9* command and changing the directory path to the drive where all audio files were stored.

3.2.3 Classification Predictor Training

The first step in the implementation of a predictor was to train this predictor with the voices from the actor database and then validate, if the accuracy of the predictor was high enough to deliver a good prediction of the actor emotion. The minimum accuracy of a predictor to deliver value would need to be higher than random guessing. Random guessing would, with endless data points, amount to the value of one divided by the number of predictable classes. As we classify into eight different emotions, the value of random guessing is: 1 / 8 = 0.125 = 12.5%. Consequently, 12.5% accuracy at least had to be surpassed for any classifier to deliver value a higher value than random guessing.

In literature, SVM classifiers have been found to perform best for voice emotion detection. Therefore, SVM classification was used to predict emotions also in our case. However, results of SVM were also compared to other classification algorithms, i.e. Random Forest Classification, Logistic Regression, K-Nearest Neighbor and Naïve Bayes. The machine learning models implemented in the Python package *scikit-learn* (short: sklearn) were used for emotion predictions.

The training data consists of 1,440 feature vectors from actor audio snippets. In every snippet, one emotion was spoken out. Afterwards, the features from the audio snippets have been extracted with OpenSMILE with the IS09_Emotion configuration and stored into an ARFF file. The feature vectors were then scaled to standard values. Additionally, according to the location of the audio snippets in the file structure that was created based on their emotions, a csv file has been created beforehand, stating the emotion of every record in the feature vector. This emotion represents the dependent variable in every machine learning model. In the final step of data preprocessing, the data was split into a training set with 85% of the records and a test set with 15% of the training data, with which the accuracy of classification can be validated.

The process of model fitting is not deterministic, since on the one side the random choice of training and test set and as well the model fitting with this data can lead to different models. To compare a distribution of accuracies, the process of train-and-test-split, standard scaling, model fitting and prediction of test values was repeated 100 times. The results of these accuracies were saved into lists and displayed with a seaborn boxplot, as presented below.

In the boxplot, the distribution of classification accuracies of SVM Classification (SVC), Random Forest Classification (RF), Logistic Regression (LR), K-Nearest Neighbor (KNN) and Naïve Bayes (NB) can be seen. SVC, as already stated in literature, seems to perform best on the training data among the classifiers with a median accuracy of 63% and accuracies from 53% to 69%. Second best, Logistic Regression presents itself with a median of 57% and a range from 49% to 66%. Random Forest Classification follows Logistic Regression with a median of 55% and a range from 45% to 66%. KNN and NB perform equally well with median accuracies of 53% and values from 45% to 59%. All accuracies surpass the minimum requirement described above and show acceptable accuracy rates. Since, however, these accuracy levels may be due to overfitting and may perform differently on a different test set, all classifiers were also used in the next step of interview classification.

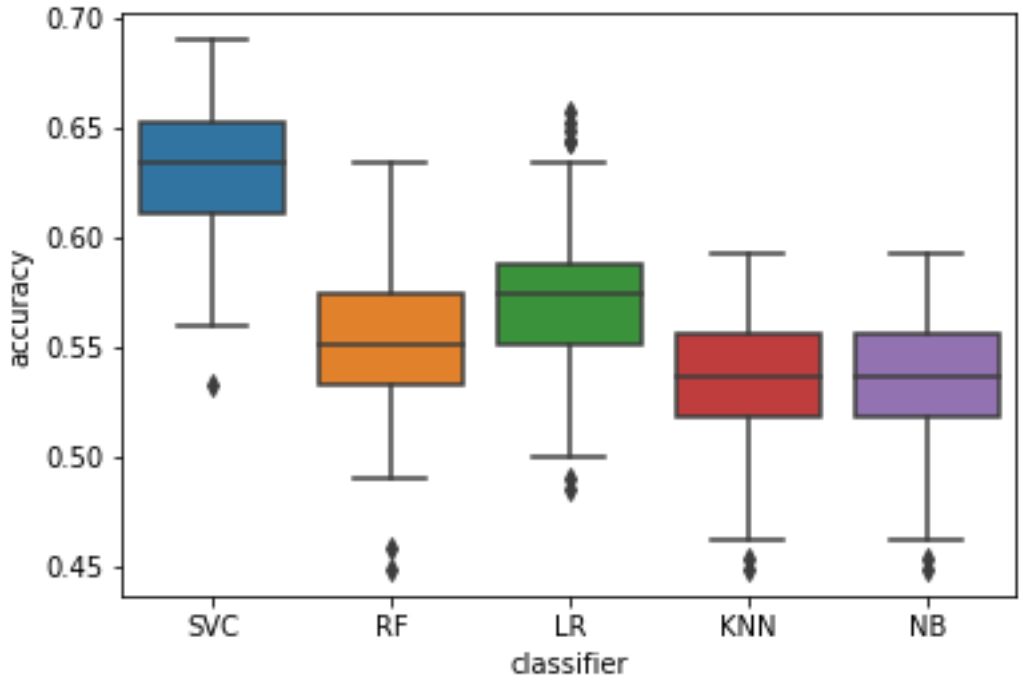


Figure 4 Accuracy Comparison from Different Classifiers

3.2.4 Classification of Interview Audios

In the interview predictor training, 100% of the actor data may be used for training the emotion prediction classifiers. Therefore, instead of creating a train-test-split like in section 3.2.3 “Predictor training”, in this script, the entire annotations and feature vectors were used. All emotions were

encoded with a label encoder and the continuous feature set was standardized, equally as in the prediction of the training data. Moreover, all machine learning algorithms tested with the training data, i.e. Support Vector Classification, Logistic Regression, K-Nearest Neighbor, Naïve Bayes and Random Forest Classification were used for prediction of the interview data.

As test data, the interview snippets of six seconds length were used. All answers were annotated with both an emotion perceived based on the uttered content of an answer and the mere voice of an answer. As multiple audio snippets per interviewed person and answer were loaded into Python, each of these audio snippets was to be predicted and the group of predictions per person-answer was to be compared against the annotations. The measurement of strength of an emotion in an answer was conducted by counting the times that a certain emotion has been predicted for a person-answer and dividing it by the number of snippets for the respective answer. However, as model fitting may produce different models even despite fitting with the same data, to attain more stable results, each model was fitted 50 times and all files have been predicted with each of these models. In consequence, the final emotion indication per person-answer has been identified by comparing the counts of predictions from all 50 models.

The results from this prediction were saved into five spreadsheets, one per answer, which list the single person-answers together with their predictions, all the five prediction algorithms gave and the strength of these algorithms.

4 Evaluation

After the recording of interviews was done, the audio files were split based on the interviewee's response to each category of question. A further splitting was done by normalizing the files using Audacity and annotating the files manually. The annotation of emotion was done by choosing emotion from a set of 8 types of emotions – "happy", "sad", "disgust", "surprised", "angry", "fearful", "calm" and "neutral". Two types of manual annotation, voice-based annotation and content-based annotation were carried out.

- Content-based manual annotation: This process involves the understanding of the content of the interviewee from their response as comprehended by human intelligence.
- Voice-based manual annotation: This process involves detecting the frequency and waveform of the vocal quality. For annotating manually, humans can only focus on the physical qualities like pitch, pace, roughness and tone of the students when being stimulated with the survey questions.

The machine prediction outputs were cross-validated with human assessment by manual listening of the audio file for each question. Each machine prediction was then assessed to be either reasonable or non-reasonable based on human observation. For each audio file per question, additional comments on why the prediction was (non-)reasonable were also made. Based on the statistics of the validation process, the accuracy of the classifier model was calculated. As seen in Figure 5, the accuracy of the classifier was 60%.

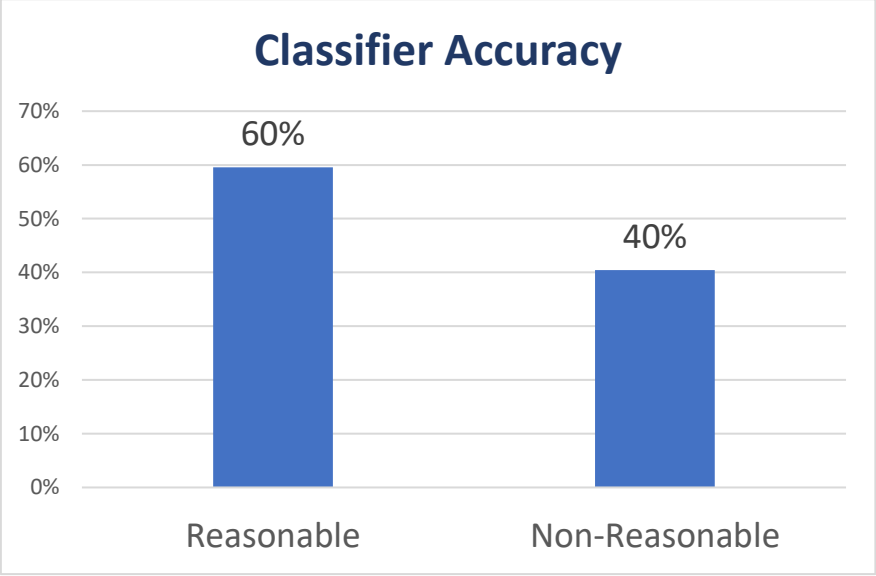


Figure 5 Classifier Accuracy

5 Results, Analysis and Discussion

5.1 Predictor Analysis

Dissimilarities between training and test data (covariate shift)

Some of the Machine learning algorithms could not predict the interview data effectively, while they could predict training data with a high accuracy. For example, the support vector classification reached a median accuracy of 63% in the training data, but predicted the emotion “fearful” for almost every audio snippet from the interview data. Due to this reason, the suspicion arose that the test data deviates significantly from the training data. To verify this, the covariate shift was analyzed with a method described in an article posted by (Gupta, 2018). This method merges a common dataset of both training data and test data with a labelling of whether a record belongs to either of these origins. A random forest classifier is then fitted to the merged dataset to analyze whether a certain record belongs to training or test data.

When this methodology was applied to the actor and interview data, a created random forest classifier was able to predict the type of a record with a 100% accuracy. This means that there is a high certainty of the existence of a covariate shift between training and test data. This covariate shift is detrimental to the success of the predictor, as the training data should be as similar to the test data as possible.

To understand, how an algorithm can decide whether a vector of features belongs to the training data or test data, a decision tree classifier as a simpler machine learning than random forest classification has been applied. The decision tree fitted to the same data of both training and test sets can predict the origin of data with an accuracy of 99.94%. The basis for this decision is taken from the features visualized in Figure 6. With attribute 17 (“pc_fftMag_mfcc_sma[1]_amean”) and attribute 200 (“pcm_RMSenergy_sma_de_linregerrQ”), the two data origins can be differentiated without annotation. Even when dropping both columns, other features are similarly valuable for the decision tree classifier and bring about an almost equally high accuracy of 99.6%.

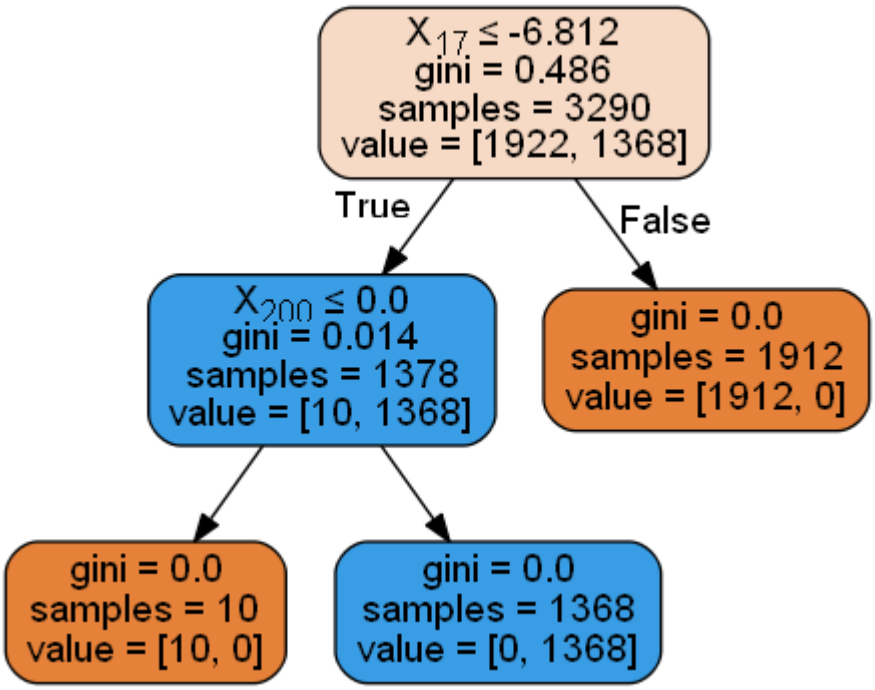


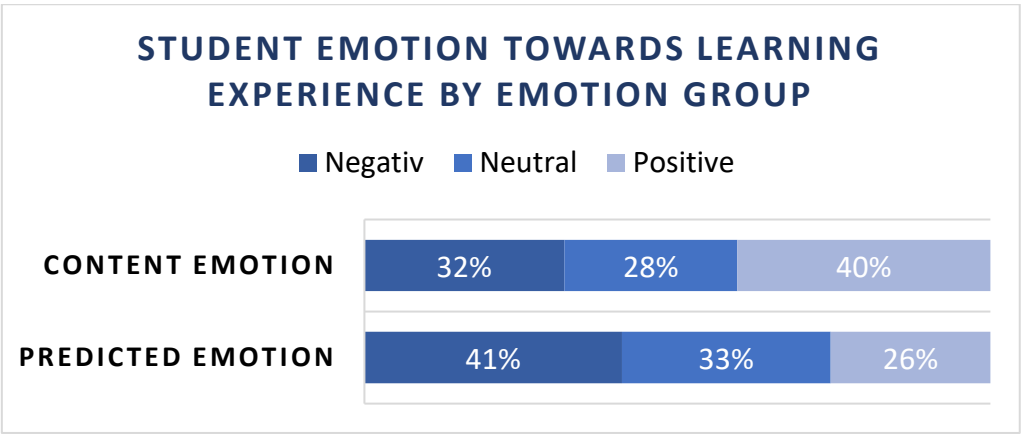
Figure 6 Decision Tree Classification

5.2 Academic Results, Analysis and Discussion

This section describes the results and findings after the implementation of voice emotion detection techniques on the review of students collected through the process of interview at HSRW. The analysis was done by comparing the emotions as per the voice pattern and the content spoken in the interview. In all the analysis, the content-based emotion indicates the emotions as per the content spoken in interview such as happy, good, interesting, etc. and the predicted emotion shows the emotion which classifier predicted as per waveform, pitch, frequency, etc.

5.2.1 Overall learning experience (Python classifier)

5.2.1.1 As per general emotion group



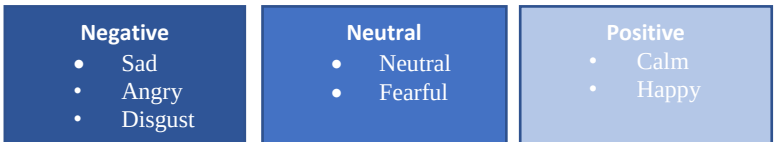


Figure 7 General Emotion Group, Predicted vs. Content-based

In this section, the analysis was done for three general emotions - negative, neutral and positive. From the figure above, it can be seen that the emotions of students were quite spread out as per both, voice pattern and content. The percentage of positive and negative emotions for predicted emotions were 26% & 41%. However, the percentage of positive and negative emotions for content-based emotions were reversed i.e. 40% & 32%. One reason could be that the interviewees did not reveal their true feelings within the content. Moreover, the percentages of students that showed negative emotion towards learning experience based on voice pattern (predicted) and content were 41% & 32% respectively, which is a matter of concern as these percentages are significant.

5.2.1.2 As per specific individual emotion

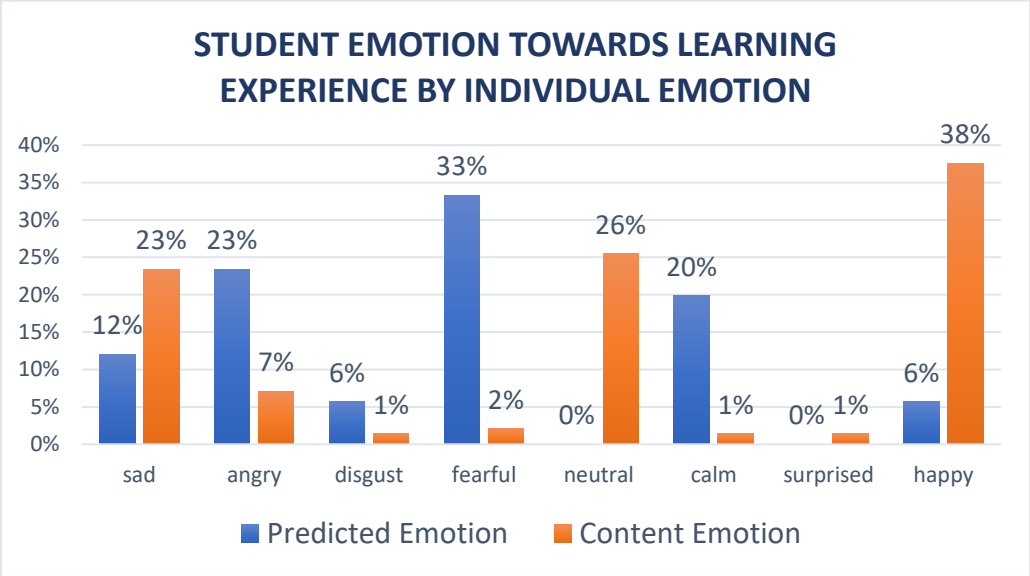


Figure 8 Individual Emotion, Predicted vs. Content-based

In this section, the analysis was done for 8 different types of emotions. From the figure above, it can be seen that the emotion '*fearful'* was predicted for 1/3 of the total students interviewed, which could be due to the students' nervousness and uncomfortableness in giving reviews via interview. As per prediction, the fraction of students feeling *angry* and *sad* were greater than those feeling *happy* and *calm*, 1/3 & 1/4 of total respectively. Moreover, the percentage of students *sad* and *angry*, based on content is also significant (30 %). This could also be a point of concern for the university management.

5.2.2 Learning experience as per 5 different areas of teaching service (Python classifier)

5.2.2.1 Teaching methodology and style

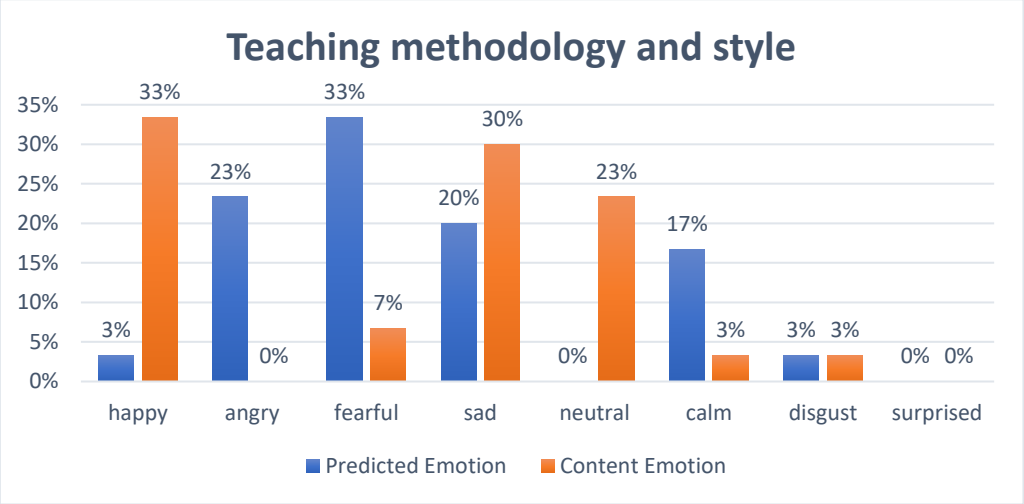


Figure 9 Emotions on Teaching Methodology and Style

From Figure 9, it can be observed that the percentage of student feeling *sad* and *angry* with the 'teaching methodology and style' as per voice pattern (predicted) is quite significant (43%). Also, the percentage as per content is 30%. This is a point of concern for the university management.

5.2.2.2 Course Curriculum

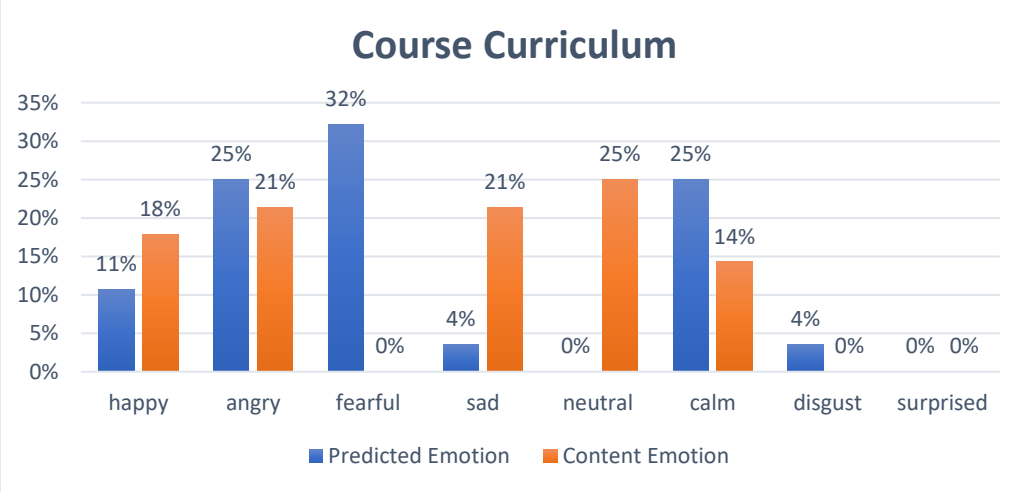


Figure 10 Emotions on Course Curriculum

From Figure 10, it can be seen that the emotions of student with the 'course curriculum' was wide spread, based on voice pattern as well as content. This means the course curriculum was quite satisfactory for the students.

5.2.2.3 Timing of classes

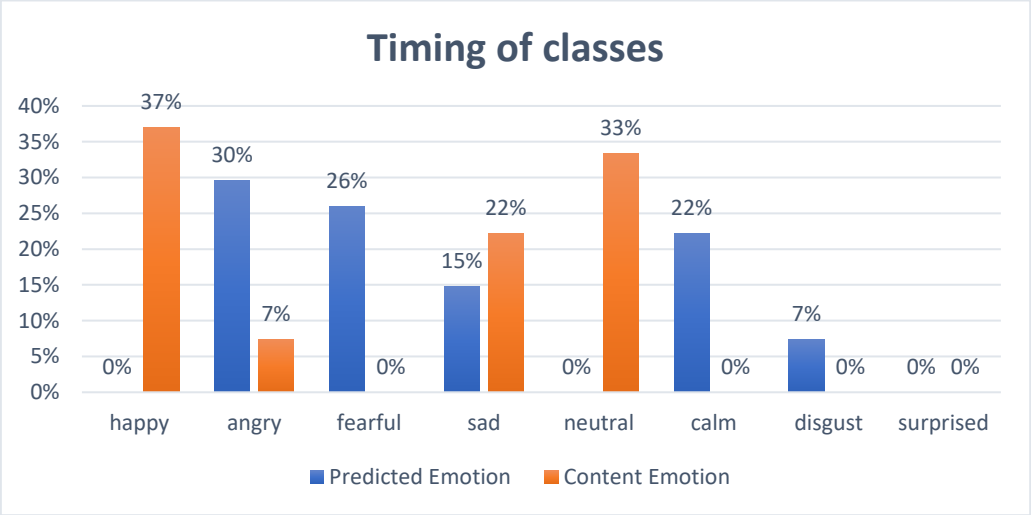


Figure 11 Emotions on Timing of Classes

From Figure 11, it can be observed that the 'timing of classes' was also one of the major areas of concern. 45% of students were *sad* and *angry* as per voice pattern (predicted) and 29% as per content.

5.2.2.4 Quality of teaching

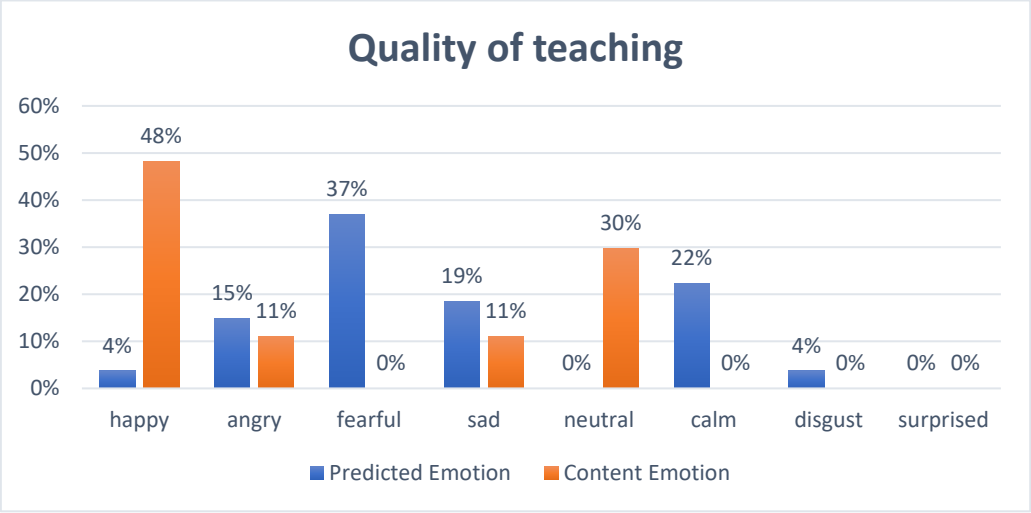


Figure 12 Emotions on Quality of Teaching

From Figure 12, it can be seen that the percentage of student 'happy' with the 'quality of teaching' as per content is significantly high (48%). However, the percentage of happy and calm is only 26%. So, it can be concluded that the students are not revealing their true emotion in content and not feeling comfortable in giving review on it. This leads to prediction of 'fearful' emotion to 37%.

5.2.2.5 Evaluation patterns

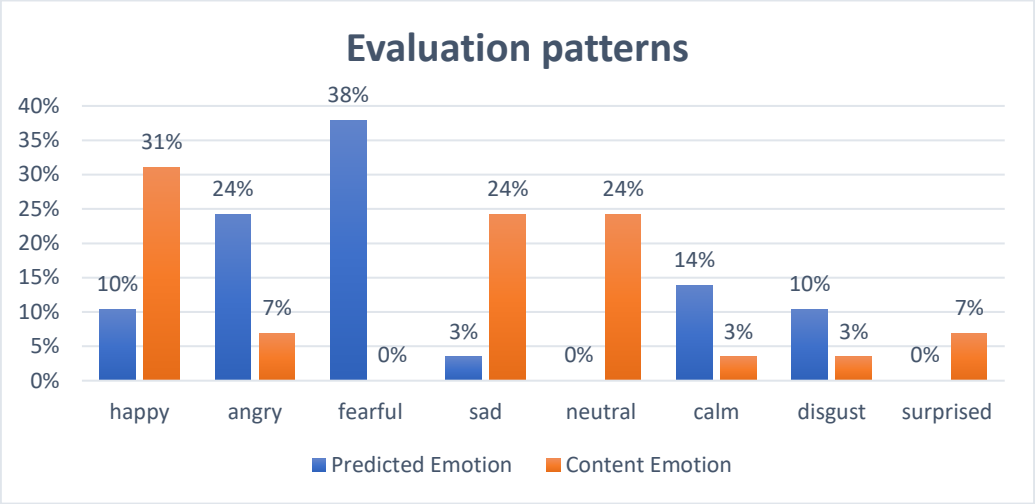


Figure 13 Emotions on Evaluation Patterns

From Figure 13, it can be seen that the student emotions are wide spread out with the 'evaluation patterns' based on content as well as voice pattern. This is not a point of concern for the management.

5.2.3 Comparison of overall learning experience in Python and R

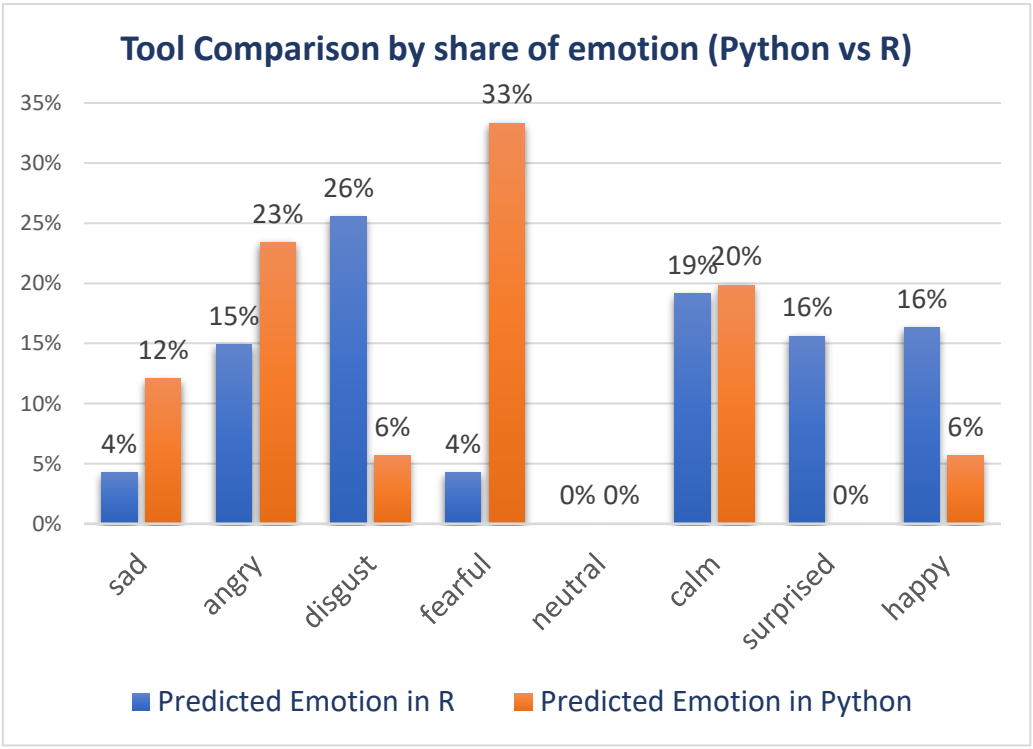


Figure 14 Tool Comparison by share of emotion (Python vs R)

In this section, students' emotions towards overall learning experience using 'Python' and 'R' classifier are analyzed. From Figure 14, it is observed that the prediction varies by changing the tool. However, the percentage of students with negative emotions in both the tool is significant i.e. R (45%)

& Python (41%). And it is a matter of concern for the university management. Also, there is high variation for emotion ‘fearful’ and ‘surprised’ between the tool, which needs further research.

6 Challenges and Limitations

6.1 Challenges in Methodology

- Detection of emotions from voice requires vocal expert, so the manual annotation might not be accurate.
- The audios were split into 6 seconds and emotions from each 6-second audio file were different from each other. Furthermore, the final emotion of each respective question was the emotion with the highest occurrence among the 6-second audio files whereas manual annotation is done for each question. Therefore, there was a mismatch between manual annotation and machine prediction. This mismatch could not be prevented due to unrealistic time constraint in annotating: 1,991 single files diligently on a 6-second basis.
- In some cases, there was a difference in content-based emotion and voice emotion, which means the interviewees were trying to change their emotions during the interview.
- The questionnaire and interview style still may need to be improved in order to fetch stronger and yet not directed, manipulated emotions.

6.2 Mismatch Between Feelings Towards a Topic and Shown Emotions

- Some interviewees might have been nervous and uncomfortable while giving interviews, thus their emotions were predicted to be fearful irrespective of content.
- Some interviewees were taking short pauses in between to think while giving answers which broke the continuity, thus emotions were changing with respect to speed because the training dataset consisted of actor audios. Even though taking pauses in between is a normal situation in real-life interviews, this does not exist in training dataset with actor voice snippets in which only one sentence is spoken at a time.
- For most of the interviewees, English was the second language in which some were not so fluent and/or found it comfortable talking. This resulted in more breaks or interjections, which the machine may have interpreted as fear in many cases.
- In some cases, interviewees were either comparing their experiences at Rhein-Waal University to their prior studies or spending a big part of their answer to talk about their prior studies, thus the emotion, which was predicted, was not entirely based on their experiences at Rhein-Waal University.

6.3 Model complexity

- There were 400 features in the feature set used in this project which made it difficult to understand how the calculation of the prediction was conducted in detail and how each individual feature influenced the prediction.
- The MFCC features were results from transformations of wave files and the values were by themselves very difficult to interpret by humans.
- The contribution weight of features was highly balanced across the 400 features and there was no major influence of a single feature. This statement can be proved with the help of Figure 15.

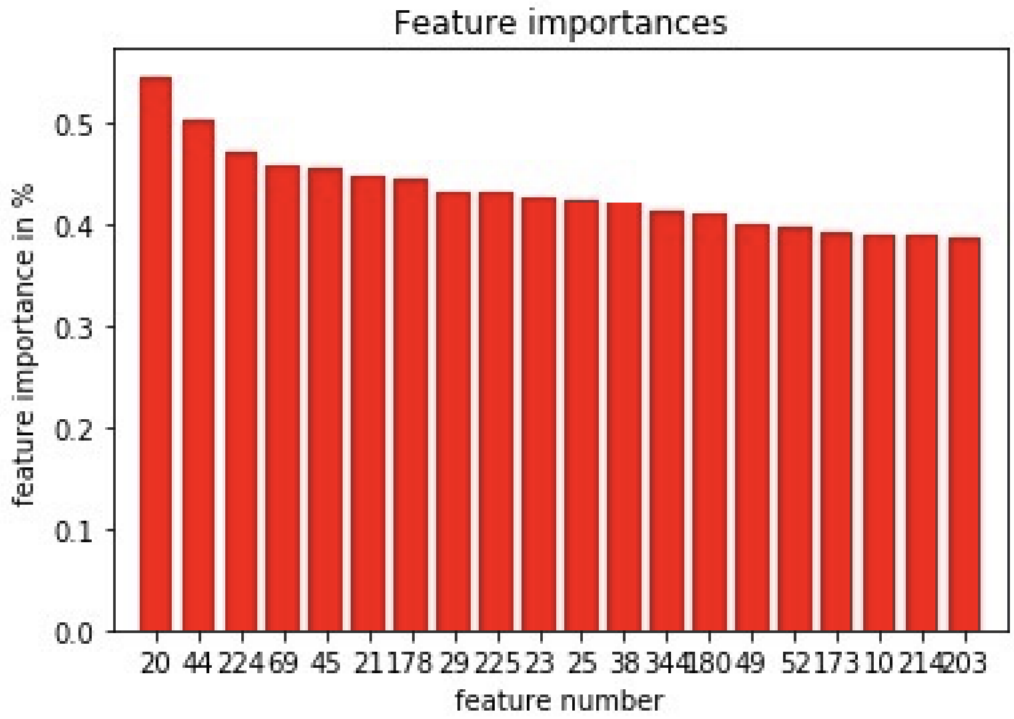


Figure 15 Feature Importances

Figure 15 shows the 20 most contributing features for random forest prediction. It is evident from the figure that the contribution of these 20 features is almost the same and even the most important features do not reach 1 percent contribution.

6.4 Machine weaknesses

- For interviews with a lot of interjections (such as "Errm", "Ahm"), the machine misinterpreted the emotion as *disgust* in many cases.
- "*Calm*" and "*Sad*" are inherently very close to each other in audial emotion expression, since all react to few changes in the voices.

- The emotions "*happy*" and "*angry*" are inherently very close to each other in audial emotion expression, as both react to agitation.
- Answers with cynical laughter were in many cases predicted to be "*happy"*.

7 Conclusions and Future Works

In this project paper, the technology of voice emotion detection has been considered and applied in the use case of interviews, specifically finding out the emotions of students at the University of Applied Sciences Rhein-Waal. To achieve this, the fundamentals of emotion analysis as well as technical building blocks of voice emotion detection have been presented together with the application areas, in which voice emotion detection is already used or seen as a potential tool in the future. On grounds of this knowledge, interviews have been conducted with 30 students on their learning experience in the University of Applied Sciences Rhein-Waal. With the use of an emotion classifier that we have created, the detected emotions of the interviews have been analyzed on their prediction success and on the emotional status among students.

One goal of this research project is the derivation of statements about the emotions that students from the University of Applied Sciences Rhein-Waal feel towards the five areas of concern "Teaching Methodology and Style", "Timing of Classes", "Quality of teaching", "Course curriculum" and "Evaluation patterns". Statements could be created and were presented in chapter 5.2 "Academic Results, Analysis and Discussion" .As a general finding across all interview areas it has been found that the emotions across different students are wide-spread and that there is not one unison emotional reaction. This means that the students experience varies and that they may even react differently to the similar experiences. Hence, as a further research it is recommended to evaluate the factors that lead to certain emotions and for which observation group that is true. Both strengths expressed with a positively connotated emotion, such as happiness or calmness, as well as weaknesses expressed by negatively connotated emotions, such as anger or sadness, may be analyze. A further conclusion is that not all students are satisfied with their learning experience and that it may be possible to optimize the learning experience in the future.

A second goal of this paper was to use specifically the technology of voice emotion detection to classify the emotions of the students. Through a training with a voice emotion database, an emotion classifier could be implemented that could classify any given audio snippet into one of eight emotions, with the support vector classification in training even reaching an accuracy of 63% in median. The application of this classifier to the conducted interviews posed additional challenges, however. The manual annotation of the eight emotions to the interviews could not be conducted with a satisfactory degree of certainty. This hindered the verification of emotion predictions and made the use of another

method, namely the review of reasonability, necessary. Out of this challenge, a study on the correct annotation of emotions is recommended. The annotation as well as prediction was further faced with an obstacle of answers with a low emotional state. To trigger an emotional state is difficult in interviews, since no specific emotions may be triggered, and the questions have to be neutral. Therefore, further research in triggering emotions in a neutral way is proposed. Mispredictions also have been found to be the consequence of misinterpretations of certain sounds or the differentiation between two emotions that are audially similar. As machines may learn to integrate new scenarios it is estimated that misinterpretations may be resolved with more training data. This training data could for example integrate cases of interjections and thinking parts as noise that can be neglected.

Further research also needs to be conducted on machine-learning platforms designed for Python and R which caused variation in the classification of emotions, especially for "*fearful*" and "*surprised*". Parameters from functions calls in Python and R packages should be examined and tested in consideration with the underlying algorithms and feature sets.

In summary, returns of this research paper include the general emotional state of students of the University of Applied Sciences Rhein-Waal, an emotional classifier model and evaluation as well as a revelation of weaknesses of voice emotion classifiers and the challenges of their use in interviews. Directions of further research are laid open, of which positive results are expected to strongly improve outcomes of interview analysis with voice emotion detection and render this a valid method in the future.

8 References

BAHREINI, K., NADOLSKI, R. & WESTERA, W. Improved multimodal emotion recognition for better game-based learning. International Conference on Games and Learning Alliance, 2014. Springer, 107-120.

BAVEYE, Y. 2015. *Automatic prediction of emotions induced by movies.* Ecole Centrale de Lyon.

BIERSACK, S. & KEMPE, V. 2005. Exploring the influence of vocal emotion expression on communicative effectiveness. *Phonetica,* 62**,** 106-119.

BOERSMA, P. 2002. Praat, a system for doing phonetics by computer. *Glot international,* 5.

BRAVE, S. & NASS, C. 2003. Emotion in human-computer interaction. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications***,** 81-96.

BREAZEAL, C. 2004. Social interactions in HRI: the robot view. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* 34**,** 181-186.

BURKHARDT, F., PAESCHKE, A., ROLFES, M., SENDLMEIER, W. F. & WEISS, B. A database of German emotional speech. Ninth European Conference on Speech Communication and Technology, 2005.

BUSSO, C., BULUT, M., LEE, C.-C., KAZEMZADEH, A., MOWER, E., KIM, S., CHANG, J. N., LEE, S. & NARAYANAN, S. S. 2008. IEMOCAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation,* 42**,** 335.

CAI, C., XU, Y., KE, D. & SU, K. 2015. A fast learning method for multilayer perceptrons in automatic speech recognition systems. *Journal of Robotics,* 2015.

CHAN, S. W. & CHONG, M. W. 2017. Sentiment analysis in financial texts. *Decision Support Systems,* 94**,** 53-64.

CHAVAN, V. M. & GOHOKAR, V. 2012. Speech emotion recognition by using SVM-classifier. *International Journal of Engineering and Advanced Technology (IJEAT),* 1**,** 11-15.

DAUBECHIES, I. 1992. Ten lectures on wavelets, vol. 61 of CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, Pa, USA.

DOUGLAS-COWIE, E., COWIE, R. & SCHRÖDER, M. A new emotion database: considerations, sources and scope. ISCA tutorial and research workshop (ITRW) on speech and emotion, 2000.

EKMAN, P. 1992. An argument for basic emotions. *Cognition and Emotion,* 6**,** 169-200.

EKMAN, P., O'SULLIVAN, M., FRIESEN, W. V. & SCHERER, K. R. 1991. Invited article: Face, voice, and body in detecting deceit. *Journal of nonverbal behavior,* 15**,** 125-135.

EYBEN, F., WÖLLMER, M. & SCHULLER, B. Opensmile: the munich versatile and fast open-source audio feature extractor. Proceedings of the 18th ACM international conference on Multimedia, 2010. ACM, 1459-1462.

GUPTA, S. 2018. *How (dis)similar are my train and test data?* [Online]. Available: https://towardsdatascience.com/how-dis-similar-are-my-train-and-test-data-56af3923de9b [Accessed 08 December 2018].

HANSEN, J. H. & BOU-GHAZALE, S. E. Getting started with SUSAS: A speech under simulated and actual stress database. Fifth European Conference on Speech Communication and Technology, 1997.

KHARA, S., SINGH, S. & VIR, D. A Comparative Study of the Techniques for Feature Extraction and Classification in Stuttering. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018. IEEE, 887-893.

KOŁAKOWSKA, A., LANDOWSKA, A., SZWOCH, M., SZWOCH, W. & WROBEL, M. R. 2014. Emotion recognition and its applications. *Human-Computer Systems Interaction: Backgrounds and Applications 3.* Springer.

LARTILLOT, O. 2011. Mirtoolbox user's manual. *Finnish Centre of Excelence in Interdisciplinary Music Research*.

LITMAN, T. 2017. *Autonomous vehicle implementation predictions*, Victoria Transport Policy Institute Victoria, Canada.
LIVINGSTONE, S. R. & RUSSO, F. A. 2018. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PloS one,* 13**,** e0196391.
LUGOVIĆ, S., DUNĐER, I. & HORVAT, M. Techniques and applications of emotion recognition in speech. Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2016 39th International Convention on, 2016. IEEE, 1278-1283.
MAGLIO, P. P., SRINIVASAN, S., KREULEN, J. T. & SPOHRER, J. 2006. Service systems, service scientists, SSME, and innovation. *Communications of the ACM,* 49**,** 81-85.
MATHAI, P. P. A Review on Sentiment Analysis and Text-To-Speech.
OATLEY, K. & JOHNSON-LAIRD, P. N. 1987. Towards a Cognitive Theory of Emotions. *Cognition and Emotion,* 1**,** 29-50.
PALEARI, M., HUET, B. & CHELLALI, R. Towards multimodal emotion recognition: a new approach. Proceedings of the ACM international conference on image and video retrieval, 2010. ACM, 174-181.
PRAKASH, C., GAIKWAD, V., SINGH, R. R. & PRAKASH, O. 2015. Analysis of Emotion Recognition System through Speech Signal Using KNN & GMM Classifier. *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE),* 10.
RHEIN-WAAL, H. 2017. Enrollment increases at Rhine-Waal University of Applied Sciences (press release).
RONIT SAHA, S., MOUNISHA JULURU, DAMIAN ANAMELECHI ANYAMELE, ARUN KUMAR HOLLA BOMMANABILLU RAGHAVENDRA 2018. Sentiment Analysis and Machine Learning Techniques: Case Study of Rhein-Waal University. Hochschule Rhein-Waal: Hochschule Rhein-Waal.
SAKSAMUDRE, S. K., SHRISHRIMAL, P. & DESHMUKH, R. 2015. A review on different approaches for speech recognition system. *International Journal of Computer Applications,* 115.
SCHERER, K. R. A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology. Sixth International Conference on Spoken Language Processing, 2000.
SCHERER, K. R. & CESCHI, G. 1997. Lost luggage: a field study of emotion–antecedent appraisal. *Motivation and emotion,* 21**,** 211-235.
SCHERER, K. R. & CESCHI, G. 1997. Lost luggage: a field study of emotion–antecedent appraisal. *Motivation and emotion,* 21**,** 211-235.
SCHULLER, B. W. 2018. Speech emotion recognition: two decades in a nutshell, benchmarks, and ongoing trends. *Commun. ACM,* 61**,** 90-99.
SHIVHARE, S. N. & KHETHAWAT, S. 2012. Emotion detection from text. *arXiv preprint arXiv:1205.4944*.
TAN, P.-N. 2007. *Introduction to data mining*, Pearson Education India.
VOGT, T. 2010. Real-time automatic emotion recognition from speech.
VOGT, T., ANDRÉ, E. & BEE, N. EmoVoice—A framework for online recognition of emotions from voice. International Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems, 2008. Springer, 188-199.
WAARAMAA, T., ALKU, P. & LAUKKANEN, A.-M. 2006. The role of F3 in the vocal expression of emotions. *Logopedics Phoniatrics Vocology,* 31**,** 153-156.

Appendices

Questionnaire

We will not collect your name or any information that could give an indication about your identity. Please also do not reveal the identity and names of other students, faculties or teachers.

Demographics

- Which country are you from?
- Are you a bachelor's or a master's student?

Teaching methodology and style

How would you comment on your satisfaction with teaching methodology and style? What is done well, would could be improved? To illustrate this question, we would like to tell you two scenarios:

- [explanation]: Stating teaching methodology we refer to different styles of teaching. These may be influenced for example by a teacher's personality, culture, beliefs and mood. Some examples may be that the style differs in formal vs. informal, practical vs. theoretical, engaging vs. monologue based, difficult vs. easy, fast vs. slow. Methods may also differ in practical vs. theoretical style, homework and assignments vs. longer lectures.
- Student 1: He thinks that his education system is delivery based. Delivery based means, everything is taught in class. Additionally, assignments are given to build expertise on a topic, but now he is in Germany and the teaching method is totally different, where it is more inclined to self-learning. For him, it is difficult to manage his time, because he is in the habit of learning delivery-based, which is making him feel stressed, pressured and overwhelmed. This is, because he is used to receiving information pre-processed and expressed in natural language by the teacher, therefore understanding now is lower and he feels sad.
- Student 2: In an American way taught class, all points from the content are taught in class and many assignments are given that have to be delivered each week. Although this is known back from school, this is not a common lecture style the German student knows from university. Currently he has mixed feelings about this lecture. On the one hand, he feels that he really digs deep into the lecture content and has the opportunity to apply everything very practically, which makes him feel confident about the exam. On the other hand, the assignments require a lot of time and not every assignment he feels to be necessary, when he long has got the concept of it, which makes him feel stressed and partly angry due to the plethora of classes and hours he puts into university and still feels overwhelmed handling all of it.

Course Curriculum

How do you feel about the course curriculum? Are changes required or is it fine?

- [explanation]: With course curriculum we refer to the content and area of curriculum, topicality and inclusion of current trends, number subjects, offering of electives, depth of the courses and applicability.
- Student 1: A student feels there are too many courses and there is too little time left for it. To her, too many courses are fit into one semester. As all subjects from one semester have to be completed in that same semester, it is difficult to decide on which and how many subjects to take. This feels like a burden, when just starting with a course. This also adds up to the time expenditure of daily life such as travelling to university, cooking and so on. She feels that she cannot completely grasp the subject in one semester, which is unsatisfying and makes her feel less confident regarding the understanding of the topic. She would like to give her best in every subject, but cannot do due to the time schedule.
- Student 2: A student feels grateful for having taken obligatory courses that he would not have taken, if he would have to make the choice. In those courses, he got to know the importance of the courses after attending them. Especially, due to soft skill courses that may seem unimportant, he now feels more confident and conscious about his actions. Apart from that he feels the course curriculum to be very good designed for his future work and not as generalized and broad as other masters. He can see the direction and feels happy about taking that direction.

Timing of classes

What is your opinion on the start times and duration of classes in your studies?

- Student 1: The student feels uncomfortable with the start times of the classes. He has the habit of working late in the night, so classes starting at 8:00 are too early for him. In his prior experience, everything started at 9:30, so he feels stressed and half asleep trying to attend a very early class. Then, if the class continues for multiple hours and continues until the evening, he feels very tired not being able to concentrate. With this, he feels frustrated and angry. Especially block courses result in big disappointment.
- Student 2: Student 2 feels that block courses are very effective, since they are very condensed. Some courses are only on few days then running for the whole day. After that, they leave a long time until the next lecture. She feels motivated about the fact that she can manage her time on her own and therefore she also feels confident and optimistic about the grades in these subjects. She also feels comfortable with the start times.

Quality of teaching

What do you think about the preparation of course and material for your subjects?

How knowledgeable do you see your professors in the classes they teach?

- There is one teacher who is very active and very intense in his classes. He has much experience and is knowledgeable about the lecture. Also, he is demanding and needs perfection in everything. Student 1 feels happy and motivated about this, as this knowledge and confidence reflects on him and engages him. Student 2 sees him walk in the class with so much energy and feels pressurized with all the demands that he states and scared to match the criteria.
- There is a second teacher who is not well prepared, casual, tired and not very present. Student 1 feels bored and his interest gone in this subject. Although the subject that is taught appeals to him, he does not feel like attending classes and unmotivated. Student 2 feels positive, as demands are lower and she is able to do something creative in this class. She can study what she feels is important.

Evaluation patterns

What are your thoughts about the setup and conduct of exams and other methods of grading? Are they designed properly in your field?

- Student 1 feels hectic about the design of evaluation. He has ca. eight subjects parallely and he needs to submit weekly parts that may be graded and are taken into account to the full grade. In past, he could better manage time for evaluation, as he could concentrate a few weeks before the exams on evaluation. He could handle different subjects as required, some continuously prepared and others prepared before the exams. The fact that exams can only be taken once per year also pressurizes him and frightens him.
- Student 2 is happy about the fact that for some subjects there are only presentations to be prepared and no written exams or reports to be submitted. This way, student 2 can dedicate her time to other subjects. With having to continuously hand in work she also feels motivated to work continuously and stretch energy across the semester.

Emotional State

Do you currently feel to be in an emotional state?

Code Repository

https://github.com/Vincentxyz/ARP_Emotions

Declaration of Authenticity

We hereby declare that the work presented herein is our own work completed without the use of any aids other than those listed. Any material from other sources or works done by others has been given due acknowledgement and listed in the reference section. Sentences or parts of sentences quoted literally are marked as quotations; identification of other references with regard to the statement and scope of the work is quoted.

The work presented herein has not been published or submitted elsewhere for assessment in the same or a similar form. We will retain a copy of this assignment until after the Board of Examiners has published the results, which we will make available on request.





Ronit Saha (24939)	Viet-Hung Vu (26078)	Sachin Kumar (24914)	Aishwarya Narkar (24854)	Vincent Finn Alexander Meyer zu Wickern (25142)