First Aid Stress Deduction

Timothy Dingeman 5993709
June 27, 2014

Bachelor Thesis
Credits: 18 EC

Bachelor Opleiding Kunstmatige Intelligentie

UNIVERSITEIT VAN AMSTERDAM

Faculty of Science
Science Park 904
1098 XH Amsterdam

Supervisors

dr. B. Bredeweg	dr. D.J.M. Weenink
Informatics Institute	Department of Linguistics
Faculty of Science	Faculty of Humanities
University of Amsterdam	University of Amsterdam
Science Park 904	Spuistraat 210
1098 XH Amsterdam	1012 VT Amsterdam

Abstract

Emergency workers are likely to experience a fair amount of stress in their work which revolves around the provision of first aid. Due to stress, fatal errors can be made. A first aid speech system can be of help when dealing with stress by reducing the chance of inaccuracies. This speech system should recognize the stress level in the emergency worker's voice, and provide assistance through different protocols depending on the stress level.

To designate a stress level, the acoustic properties of emergency workers voices are analysed. The most valuable acoustic properties will be the foundation for a stress classifier to be able to classify new speech signals according to the three aforementioned stress levels. The most valuable acoustic properties to classify upon a male test set are the fundamental frequency, the first formant frequency and the difference between two intensity objects within a speech signal. The stress classifier has an accuracy of 82% predicting unseen data as "neutral" and "stressed".

Contents

1 Introduction

While providing first aid, emergency workers can experience an increase of their stress level, which can have adverse effects on their productivity level [1]. When an emergency worker malfunctions because of an overwhelming and stress-inducing emergency situation, the victim can be harmed unnecessarily due to inaccurate decision making. Furthermore, stress can decelerate aspiring emergency workers' learning processes [2]. The introduction of an automated first aid speech system that assists an emergency worker while providing first aid can provide him or her with a practical and safe approach to respond to stress. The aforementioned first aid system differentiates between separate first aid procedures, depending on the emergency workers' stress level. This system cannot only be utilized by experienced emergency workers, but also by emergency workers in training. A crucial advantage of this speech system is that it is non-invasive. In addition, its user can keep both hands free to complete his or her tasks.

To derive a stress level measurement from the acoustic signal of the emergency worker's voice, research in phonetics provides an analysis of the physiological aspects of stress and their effects on speech. In a stressful situation, an individual's respiration increases. Respiration, then, leads to an increase in subglottal pressure, which affects different acoustic properties of the speech system [3]. These properties give prospect to classifying stress and subsequently assigning different stress levels to the emergency worker's verbal communication.

Earlier research describes detecting stress upon acoustic properties is feasible, although a robust stress classifier has not been realized so far [4, 5, 6, 7]. A voice based stress detector called StressSense [7] has been developed to detect stress with a mobile device. However, the most robust stress classifier regarding StressSense has an accuracy of 71.30% to allocate stress to unseen data. The accuracy of their training data has been allocated 67.60%. The accuracy is considered contemptible regarding a decent stress diagnosis using a first aid speech system. A solid built stress classifier with a high accuracy is necessary regarding a proper functionality of the first aid speech system. Therefore different classifiers are considered to find the most accurate and accelerated approach. In stressSense, the classification of stress levels is done in a simplistic division between a "stressed"- and an "unstressed" state of the subject. The current research focuses on a more nuanced classification that distinguishes three levels of stress: "neutral", "stressed", and "severely stressed". The research question of this inquiry will thus be:

> Can the acoustic signal of an emergency workers' voice be utilized as an indicator for his or her stress level, in order for a first aid speech system to differentiate in procedures depending on that stress level?

The organisation of this thesis is as follows. A conceptual model of the first aid speech system will be presented in section 2. The groundwork for this research is described in section 3. In this section, three different types of research are delineated to clarify the fundamentals of this analysis: research in phonetics, psychological research and military research. The method and approach taken with regards to the analysis is presented in

section 4 and the approach taken according to the classifiers is presented in section 5. Section 6 illustrates the results obtained concerning the methodology. In section 7 the evaluation will be presented. The conclusion regarding the research question is disclosed in section 8. The last section, section 9, will deal with limitations and work in prospect.

2 First Aid Speech System

An emergency worker's stress can, in many ways, be compared to the pressure a soldier in the army is under. Three different types of stress reactions could be perceived by from the behaviour of the soldier [8]:

1. "Mild" stress causes the soldier to show odd behaviour in comparison to his or her usual conduct, such as disorientation. The soldier is still capable of communicating once he or she undergoes the "Mild" stress level.

2. "Serious" stress results in an incapability of the soldier to communicate properly using verbal communication. The soldier will stutter, mumble or will deliver incomprehensible speech.

3. "Very serious" stress entails a complete inability to communicate in any way.

These stress reactions provide us with a useful division that could also be applied to emergency workers' responses to stressful situations. This research will maintain a focus on the first type of stress reaction, namely "mild" stress. When experiencing this particular type of stress, the emergency worker is still capable of communicating. The "mild" stress could be subdivided into three different stress levels which are detectable through speech:

1. The "neutral" stress level is indicated when no stress is measured within the acoustic signal of a person's voice.

2. The "stressed" stress level should be indicated between the "neutral" and the "severely stressed" stress level.

3. The "severely stressed" stress level is indicated when the emergency worker has the highest amount of stress according to the features of his or her voice.

Once a stress level is derived from the acoustic signal, it has to be decided on which values the above mentioned stress levels should occur.

When the emergency worker is in a neutral state, there is no necessity of providing additional information. Supplying additional information could even be considered superfluous and thus uncalled for once an emergency worker is not stressed. When the emergency worker is stressed, however, the additional information will consist of explaining the first aid techniques in more detail. The amount of details should vary upon the amount of stress of the emergency worker; an higher stress level will lead to a more detailed description of the first aid techniques [8].

Once an emergency worker suffers from stress, he or she is more likely to develop a tunnel vision and premature closure: a tendency to stop considering other possible diagnoses after a diagnosis is reached [9, 10]. To prevent the emergency worker from suffering from tunnel vision and premature closure [10], he or she should be accommodated with additional information of first aid techniques [8]. The first aid protocol consists of the following main steps:

1. Secure personal safety.
2. Secure victim's safety.
3. Report situation.
4. Verify and control fatal bleedings of the victim.
5. Verify and clear the airway of the victim.
6. Verify the respiration rate of the victim.
7. Verify the heart rate of the victim.
8. Verify the consciousness of the victim.

Supplementary to the additional information, the emergency worker should be advised to use optional calming techniques [8], to control oneself and proceed with the protocol. Due to timing, which is critical, only two fast variances of calming techniques are of importance in a first aid scenario:

- "Muscle Relaxation": The emergency worker should tighten all of his or her muscles at the same time for fifteen minutes or longer. Then he or she should relax the muscles to relief the tension.
- "Stomach Breathing": The emergency worker should slowly breathe in, hold his or her breath for two to five seconds and slowly breathe out. This procedure can be repeated five times.

Each of the aforementioned main steps is divided in three procedures P_1, P_2 and P_3 regarding the first aid speech system; the aforementioned step 3 "report the situation" is used as an example:

1. The procedure P_1 is based on an emergency worker in a unstressed condition, therefore the first aid system does not provide additional information regarding the protocol. The emergency worker is expected to report the situation without help of the first aid system.
2. The procedure P_2 is based on an emergency worker in a "mild stressed" condition. In this case the first aid system provides additional information regarding the protocol.

For example, the emergency worker is expected to acquire aid: He or she has to answer questions about the corresponding situation, in order to report the situation successfully with regards to the first aid system.

3. The procedure P_3 is based on an emergency worker in a "severely stressed" condition, therefore the first aid system provides the aforementioned calming techniques, furthermore providing additional information regarding the protocol. The emergency worker is expected to acquire aid to report the situation considering the first aid system and is advised to utilize one of the aforementioned muscle relaxation techniques, such as "Muscle Relaxation".

According to the above-mentioned procedures, a first aid speech system is set up in figure 1, which discriminates on the basis of the three aforementioned levels of stress detected by speech analysis.

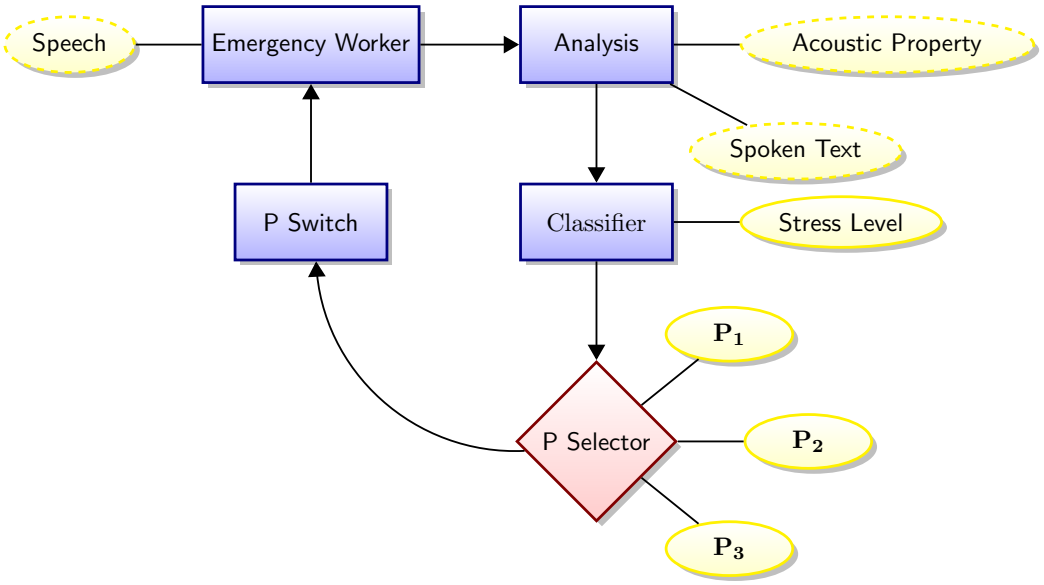

Figure 1: Representation of the first aid speech system: The emergency worker's speech is analysed once the first aid speech system is activated by the emergency worker. The acoustic properties are extracted from the speech signal once the analysis is conducted. A stress level is indicated according to the acoustic properties. The P selector discriminates between the three types of procedures depending on the aforementioned stress level introduced in section 2.

3 Groundwork

3.1 State of the Art

Stress among individuals can be categorized into different levels [11]. The Social Readjustment Rating Scale (SRRS) scores [11, 12, 13] and the Perceived Stress Scale (PSS) [14] offer useful measurements of different degrees of stress. These scales show that, the more change one is undergoing, the higher the increase in stress.

The stress levels as indicated in the quantifications of the aforementioned scales are based on experiences of individuals that have taken place in the past. However, the measurement of stress *while* experiencing it could be a highly useful addition to the current theories that exist concerning stress, especially with regards to the functioning of a first aid speech system.

The application StressSense is developed to monitor and identify stress in real-life situations. This application detects stress in acoustic environments with the use of microphones in smartphones [7]. In the evaluation regarding this voice based stress detector, different audio fragments were examined on outdoor and indoor scenarios. The audio fragments were labelled with the acoustic features of these recordings to create a dataset. The same type of identification of vocal features is examined to detect emotion [15].

Regarding StressSense, relevance of the different features is measured by the method of information gain based feature ranking [7]. The ranking of features proved the standard deviation of the fundamental frequency to be the most important vocal indicator to assign stress. The second important feature is the speaking rate. The third important feature is the range of the fundamental frequency. These notions will be elaborated upon in section 3.2.

However, the most robust stress classifier regarding StressSense, also claimed to be the universal model of StressSense, has an accuracy of 71.30% upon the test data and an accuracy of 67.60% upon the training data. A first aid speech system acquires a higher accuracy in emergency situations. The alternative classifiers StressSense provides, require input from the user for adaptation and therefore not beneficial to the productivity of a first aid speech system. StressSense measures multiple acoustic features which causes delay upon execution. Therefore reducing the features to classify Stress is a contribution towards the productivity of the first aid speech system.

3.2 The Acoustic Signal Under Stress

A recorded speech signal is represented in a waveform to analyse the signal [16]. A waveform, also called spectrum, is a representation of the air pressure in decibel measured by the parameter time in seconds. The human auditory system possesses a distribution that performs spectral analysis to resolve incoming speech signals into sine waves. In order to analyse the speech signal, the speech signal is cut up in different segments. The segments are compared by overlapping them and analysing their differences. The segments are specified as analysis intervals. Figure 2 illustrates the aforementioned procedure.

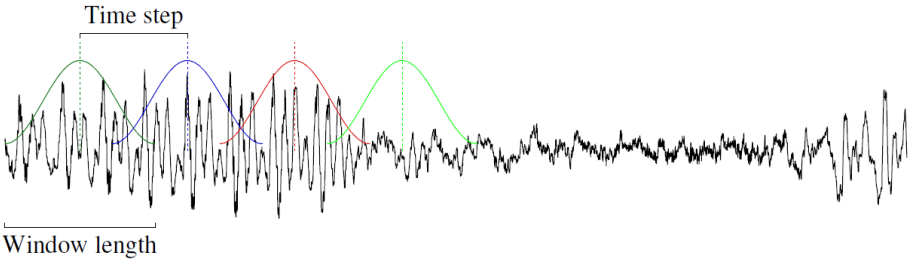

Figure 2: The representation of the speech signal. "Window length" is the duration of the segment subject to analysis. The duration of the segment can differ due to the speaker's properties; for example, male speakers have a lower frequency range than female speakers to analyse the fundamental frequency which will be explained later in this section. "Time Step" is the amount of overlap between two succeeding frames [16].

The fundamental frequency refers to the periodicity in the speech sound [17]. A high fundamental frequency corresponds to a high frequency sound wave, just as a low fundamental frequency corresponds to a low frequency sound wave. Intensity is defined as the average power in decibel weighed by time in seconds. The intensity of a speech signal can be an indicator for stress, since speakers can express more power in speech upon when they are stressed. The human basilar membrane, which is located in the inner ear, constantly resolves frequency. The spectrogram is a representation of the aforementioned membrane, which represents the frequency in Hertz measured by the time in seconds[18].

For a speech system to detect stress, it is necessary to analyse the physiological aspects of stress and their effects on speech. Once an individual is subject to a stressful situation, his or her respiration increases. Respiration will then most likely intensify subglottal pressure during speech. This increase in subglottal pressure is evidently a rudimentary factor of increasing fundamental frequency [3].

Aside from fundamental frequency (F_0), speech signals can also contain formant frequencies which are crucial features of phonemes, perceptible linguistic units that constitute the spoken word in a given language.

In phonetics, a formant is defined as an acoustic resonance of the human vocal tract. Different physiological aspects, comparable to the abovementioned increase in respiration, exist in relation to the formant frequencies. The formant's frequencies differ from each other: the first formant, the one with the lowest frequency, is called F_1, the second F_2, et cetera. Articulation is influenced by stress [3], thus the formant frequencies (in particular F_1, F_2) can indicate a certain stress level.

Two noise features, namely jitter and shimmer, will be extracted from the sound signal, since noise measurement can indicate peak shifts in the fundamental frequency: noise can indicate stress [19]. Jitter is the average difference between the constant periods in the sound signal. Jitter arises when the speaker is not able to provide a constant duration to consecutive periods in the speech signal. Shimmer is the average difference between the

amplitudes of constant periods divided by the mean value of the amplitude. Shimmer arises when the speaker is not able to provide a constant amplitude to consecutive periods in the speech signal.

Another measurement of high-frequency noise is the band energy difference in decibel. Band energy difference can, comparable to jitter and shimmer, indicate peak shifts. The mean intensities of two sound objects within the signal are calculated. Consequently the mean intensity is computed for both objects and the difference of the aforementioned mean values is calculated [20].

3.3 Tools

Preprocessing

To assemble data, the sound is retrieved from a movie file of the film *The Call* [21]. Audacity[22], an audio editing application, offers a technique to derive only the centre channel from a movie file to reduce background noise. Praat[17] a system for doing phonetics offers an approach to label spoken words in the sound file in combination with a textgrid file. A textgrid file is an annotation file in which the textual information about the interval is presented. The labelling is performed by evaluating the original film fragments. Audio measurement cannot be conducted with along audio file: therefore the script save-small-files.praat [23] is used to divide different small audio files according to the labels of the textgrid file. This script generates small audio files: the titles are the spoken words spoken in different states and by different genders combined with the time of recording.

Analysis Tools

Praat [17], offers techniques to analyse the fundamental frequency, formants, intensity and the noise features. Different functions acquire the acoustic properties described in section 3.2. For all the functions except *To_Spectrum*, a time range needs to be specified. The aforementioned functions are introduced in table 1. For a full explanation of the algorithms which measure the fundamental frequency, intensity, jitter and shimmer, please refer to [24, 18, 17].

Function	**Function Description**
To_Spectrum()	Creates a spectrum object.
To_Intensity()	Creates an intensity object consisting ofmeasurements regarding to a maximum frequency according to a specified time step.
To_Pitch()	Creates a pitch object consisting of measurements in a specified range regarding to a specified time step.
To_Formant()	Creates a formant object consisting of measurements with a specified maximum frequency and a specified number of formants
Get_band_energy_difference()	Calculates the intensity of two objects specified in ranges and calculates the difference between the objects.
Voice_report()	Creates a voice report consisting of noise features, such as jitter and shimmer.
extractNumber()	Extracts a number from the voice report.
Get_minimum()	Returns the minimum of an object's measurements for a given unit using a given method
Get_maximum()	Returns the maximum of an object's measurements for a given unit using a given method
Get_mean()	Returns the mean value of an object's measurements for a given unit.
Get_quantile()	Calculates the quantile value of an object's measurements that is twice the median
Get_standard_deviation()	Calculates the standard deviation of an object's measurements.

Table 1: The functions of Praat which are utilized to analyse the speech signal

Logistic Regression

Stanford machine learning courses [25] contribute a method to implement logistic regression in MATLAB for which the ensuing formulas are used. The output of the classifier is retrieved from a hypothesis function ($h_\theta(x)$) is defined as:

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} = P(y = 1|x; \theta)$$

This function is often referred to as the "sigmoid function" or logistic funtion. $h_\theta(x)$ is the predicted likelihood that $y = 1$, given input x, parameterized by θ. The output of $h_\theta(X)$ should inhere between the range of 0 and 1. The following statement is made out since the output can only be 0 or 1,:

$$P(y = 0|x; \theta) = 1 - P(y = 1|x; \theta)$$

Based on the aforementioned facts, the likelihood of y=1 is estimated when $h_\theta(x) \geq \frac{1}{2}$ and $\theta^T x \geq 0$. The likelihood of $y = 0$ is estimated when $h_\theta(x) \prec \frac{1}{2}$ and $\theta^T x \prec 0$.
The cost function is utilized to optimize the parameters θ in the hypothesis function. θ is optimized by the following function with m number of training examples and feature column i:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} Cost(h(x^i), y^i)$$

The cost function is defined as follows, with feature column i:

$$Cost(h(x^i), y^i) = \begin{cases} -log(h_\theta(x)) & if = 1 \\ -log(1 - h_\theta(x)) & if = 0 \end{cases}$$

The gradient descent function is utilized to fit the parameters θ and thereby to minimize the cost function. The gradient descent function will through all parameters θ frequently and is defined as follows, with m number of training examples, feature column i and parameter row number j:

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} [(h_\theta(x^i) - y^i) x_j^i]$$

Considering logistic regression is prone to overfitting, a regularization term is added to the gradient descent and cost function to exclude outliers by increasing regularization parameter λ. Given m number of training examples, feature column i, regularization parameter λ and parameter row number j; $\frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2$ is added to the cost function and $\frac{\lambda}{m}\theta_j$ is added to gradient descent to perform regularization [26].
The optimalization function "function minimalization unconstrained" ($fminunc$) is employed to find the minimum of the cost function given the gradient [27].

Feature Evaluation

WEKA[28] offers various feature selection techniques that are designed to find the most relevant features. Two search methods in combination with an evaluator are chosen, namely:

1. The search method BestFirst makes use of "hill climbing" to search within the features. This search method is used in combination with the evaluator CfsSubsetEval, which assesses the prediction of each feature and compares the degrees of irrelevance of the features[29].

2. The second search method, Ranker, employs evaluations of the individual features. This search method is used in combination with the evaluator InfoGainAttributeEval, which evaluates the information gain of each feature to the extent of the entire feature set. Information gain is defined as the predicted reduction in entropy when sorting different features. The entropy is a degree of uncertainty; a certain aspect has an entropy value of 0. An entropy of a discrete random variable X with values x is measured by [30]:
$$H(X) = -\sum_i p(x) log_2 p(x)$$
The information Gain is therefore given by:
$$InfoGain = H(Y) - H(Y|X)$$
where X is a feature with value x and Y is a feature with value y,
$$H(Y) = -\sum_{y \in Y} p(y) log_2(p(y)),$$
$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) log_2(p(y|x))$$

WEKA classifiers

Logistic regression in WEKA is used to compare results with the MATLAB implementation. Besides logistic regression, WEKA offers a decision tree learning approach, called J48. The J48 algorithm is applied to construct Univariate Decision Trees by checking whether all cases belong to the same class. The aforementioned information gain is calculated for each feature and the feature with the highest information gain is selected to reproduce a classification [31]. K-fold cross validation of n set observations is introduced to evaluate random constructed subsets. K-fold cross validation separates the set of n observations into K single subsets. Different generated subsets are adressed randomly to evaluate as a training set and the last subset is used for evaluation [30].

Experts

Praat offers a method to implement a Multiple Forced Choice listening experiment, therefore the template file of *"ExperimentMFC 6"* is utilized to construct an experiment to evaluate experts [17].

Statistics

The evaluation of the results of the aforementioned logistic classifier and the results of the listening experiment depend on the statistical "T-test" [32]. The following fundamental formulas are defined to perform a T-test:

An assumption of a specific outcome is stated as follows to perform a T-test: $H_0 : \mu$. The opposite of the assumption is stated as follows: $H_1 :< \mu$ The n number of experts is defined within the solution space:

$$\Omega = \{\omega_1, ..., \omega_n\}$$

Expert ω_i is assigned to result $X(\omega_i)$, which is now defined as stochastic variable.
$\bar{X}$ is decribed as the mean outcome of the samples in Ω.
The next formula is utilized to compute the expectation of the result X according to the solution space Ω:

$$E(X) = \sum_{w \in \Omega} X(\omega) P(\{\omega\})$$

The variation is calculated using the aforementioned formula:

$$VAR(X) = E(X^2) - (E(X))^2$$

The standard deviation σ is the square root of the variation:

$$\sigma = \sqrt{VAR(X)}$$

The calculation of the T value can be performed regarding the aforementioned formulas. The T-test is defined as follows, given n number of experts.

$$T = \frac{\bar{X} - (H_0 : \mu)}{\sigma} \sqrt{n}$$

A statistical significance need to chosen to find the critical value regarding the outcome of t. The critical value c is retrieved from the t-distribution table presented in appendix A. H_0 will be rejected when $t \leq -c$ or when $t \geq c$, and H_0 will not be rejected when $-c < T < c$.

4 Methodology: Data Analysis

4.1 Data Acquisition

To assemble data, the sound is retrieved from a movie file of the film *The Call* [21]. This film contains fragments of emergency workers, victims and dispatchers of the American emergency telephone number in a "stressed" or "neutral" state.

Since this research is solely focused on speech the centre channel, which reduces background noise and emphasized speech, needs to be extracted from the audio file. Audacity[22], an audio editing application, offers a technique to derive only the centre channel from a movie file. The optional FFmpeg library[33] needs to be installed to obtain the sound from the mkv movie file. After the sound channels are made visible in the application, a 16 bit pulse-code modulation centre channel audio file with a frequency of 48000 Hertz is extracted from the file. The extracted sound file is mono, since all DTS channels are mono.

Data acquisition is carried out by assigning intervals to spoken words on this audio file. A lexicon of spoken words is labelled with the acoustic features of the voices of emergency workers, dispatchers and victims. Praat[17] offers an approach to label spoken words in the sound file in combination with a textgrid file. A textgrid file is an annotation file in which the textual information about the interval is presented. Subsequently, it is possible to label the sound in the Praat working environment.

The measurements of the acoustic properties of voices of male speakers differ from those of female speakers; another label ("male" or "female") is thus added in the lexicon, to assign the emergency worker's gender. A label is added to distinguish between recordings made in a "stressed" or "neutral" state, to classify a stress level by use of the approach of supervised learning. Audio measurement cannot be conducted with along audio file: therefore the script save-small-files.praat [23] is used to divide different small audio files according to the labels of the textgrid file. This script generates small audio files: the titles are the spoken words spoken in different states and by different genders combined with the time of recording.

Important to mention is that the labelling of the aforementioned lexicon of spoken words is based constructed an evaluation of an individual expert. The evaluation is based on watching and listening to the original fragments of the film, in addition to solely from listening to the audiofile. The evaluation is based on the perception of stress by an individual and not generally validated by multiple experts. Therefore multiple experts are subject to evaluate the results of the stress classification in section 5.4.

4.2 Analysis

Praat [17] offers analysis tools to derive the acoustic properties of all aforementioned audio files, which are extracted using the textgrid file. A script has been written to extract the meaningful acoustic properties of all files recursively. Two output files are created for each gender: in these files, we will thus see the results of the male and female recordings. Furthermore, a list is manufactured with all of the audio files in the script directory. These

files are of the datatype *"string"*. The audio file of the list item is selected recursively to carry out an analysis for each audio file.

The boundaries of the pitch or formant measurements depend on the gender of the speaker. Hence, the variable *gender*$ is initialized to discriminate between male or female; when the name of the file starts with an f, the string *"F"* is linked to *gender*$. When the file starts with an *m*, the string *"M"* is assigned to *gender*$. the variable *state* is used to discriminate between neutral and stressed recordings by supervised learning algorithms. When the filename starts with *"f.n"* or *"m.n"*, the value of the variable state is set to 0. When the filename starts with *"f.s"* or *"m.s"*, the value of state is set to 1. After the aforesaid steps are taken, based on the list of strings, the sound is read from the directory to create an object to perform measurements. The sound object needs to be selected before every measurement. Different settings are of importance to create a pitch object to conduct pitch analysis:

1. *Time step in seconds* is set to its standard value, namely 0. Praat will measure about 100 pitch values a second with male sound objects. It is not necessary to change this value, because it is sufficiently accurate for speech measurements.

2. *Pitch floor* is set to 75 Hertz for male sound objects, and set to 100 Hertz for female sound objects. Measurements lower than these frequencies are redundant and thus neglected.

3. *Pitch ceiling* is set to 300 Hertz for male sound objects and set to 500 Hertz for female sound objects. Measurements higher than these frequencies are redundant and thus neglected.

Female speakers have a higher pitch level than male speakers. Therefore *pitch ceiling* and *pitch floor* are dependent on gender, to make sure the measurements for each of the pitch objects are accurate. The features extracted from the pitch object are illustrated in table 2.
Similar to the pitch measurements, the settings to create a formant object are also dependent on the speaker's gender. The following settings are relevant to perform formant analysis:

1. *Time step in seconds* is set to its standard value, namely 0. This value is sufficiently accurate for speech measurements.

2. *maximum number of formants* is set to 5, this is the maximum of number of formants that humans can produce during speech.

3. *maximum formant frequency* is set to 5500 Hertz for female sound objects and 5000 for male sound objects.

4. The effective *duration of the analysis window* is set to 0.025 seconds.

5. *Pre-emphasis* is set from 50 Hertz, frequencies below 50 Hertz will be neglected.

#	Feature	Setting Description
1	State	The state of the speaker is 0 or 1.
2	f0minimum	The pitch minimum in Hertz using parabolic interpolation in Hertz.
3	f0maximum	The pitch maximum in Hertz using parabolic interpolation in Hertz.
4	f0range	The pitch maximum minus the pitch minimum in Hertz.
5	f0mean	The mean value of the pitch measurements in Hertz.
6	f0median	$\frac{1}{2}$ quantile of the pitch measurements in Hertz.
7	f0std	The standard deviation of the pitch measurements in Hertz.

Table 2: The State feature is set to true or false according to the speakers stress level. The resulting features are the measurements and settings regarding the pitch object.

#	Feature	Setting Description
8	f1median	$\frac{1}{2}$ quantile of the total amount of first formant measurements in Hertz.
9	f2median	$\frac{1}{2}$ quantile of the total amount of second formant measurements in Hertz.

Table 3: The measurements and settings regarding the formant object.

The formant measurements are obtained from the formant object. Table 3 illustrates the settings of the formant measurements.
The creation of an intensity object is done with the default settings, namely:

1. *Maximum pitch* is set to 100 Hertz.
2. The *time step* of the resulting intensity contour is set to 0.
3. *Subtract mean pressure* is set to "yes": this option will exclude the intensity values of the part without speech.

The intensity measurements are performed once the intensity object is created. The features extracted from the intensity object are summarized in table 4
A spectrum object is created with default value *"yes"* to retrieve the spectral energy difference, . The spectrum object is expressed in decibel. The settings do obtain the spectral difference between two intensity objects are summarized in table 5.

#	Feature	Setting Description
10	stdIntensity	the standard deviation of the total range of intensity measurements in Hertz.
11	medianIntensity	$\frac{1}{2}$ quantile of the total range of intensity measurements in Hertz.

Table 4: The measurements and settings regarding the intensity object.

#	Feature	Setting Description
12	spectrum-Difference	A first object is created with a range between 200 decibel and 400 decibel. A second object is created with a range between 3200 decibel and 6400 decibel. The mean intensity is computed for both objects and the difference of the aforementioned mean values is calculated.

Table 5: The measurements and settings regarding the spectrum object.

Noise values can also be signs of pitch peaks, which can be indicators of stress. Therefore the noise measurements are added to the script. The noise values can be extracted from the voice report option of Praat using the function *extractNumber()*. Table 6 represents the noise values derived from the voice report.

#	Feature	Description
13	mean_jitter	the average difference between the constant periods times 100, which illustrates the jitter in percentages.
14	mean_shimmer	The average difference between the amplitudes of constant periods divided by the mean value of the amplitude times 100, which illustrates the shimmer in percentages.

Table 6: The noise measurements and settings regarding the voice report.

The results of the above-stated process can be seen in two lexicons (male and female) with spoken words in combination with the acoustic properties of the verbal expressions. At the end of the session, the objects created in the process are removed to make sure new sessions will not make use of the same data.

4.3 Data

A test set (unseen data) is created for the classifier to predict the state value of the unseen data entries, based on learning with the aid of a training set (seen data). A male and female test set is extracted from the dataset, consisting of 25 "neutral" entries and 25 "stressed" entries. The remaining data will be set as training set. A test set (unseen data) is created for the classifier to predict the state value of the unseen data entries, based on learning with a training set (seen data). A male and female test set is extracted from the dataset, consisting of 25 "neutral" entries and 25 "stressed" entries. The remaining data will be set as training set.

5 Methodology: Classifiers

5.1 Implementation Stress Classifier

Logistic regression is used as a classifier because of the variety in measurements within the features. Logistic regression is a convex function which means that logistic regression always converges to the global optimum instead of a local optimum. The output of the classifier is retrieved from a hypothesis function ($h_\theta(x)$), which is defined in section 3.3.

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

The Input

The classifiers should train with the variable *Featurevector* consisting of the columns of features. If a test set is loaded, the feature vector also contains the columns of features to classify with. A variable x is initialized which contains the aforementioned features and a variable y is loaded which contains the state, namely 1 ("stressed") or 0 ("neutral").

Cost function

The cost function calculates the optimal parameters of the hypothesis function as described in section 3.3. To implement the cost function in MATLAB, the cost function is rewritten in the following way, with m number of training examples and feature column i [25]:

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^i log(h_\theta(x^i)) + (1 - y^i) log(1 - h_\theta(x^i))]$$

Gradient Descent

Gradient descent fits the parameters calculated by the cost function and earlier described in section 3.3 [25].

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} [(h_\theta(x^i) - y^i) x_j^i]$$

Reduce Overfitting

It is most likely overfitting occurs; a random error can be classified as a correct instance. The addition of regularization provides an approach to reduce the amount of outliers. Two changes to the logistic regression algorithm are necessary to perform regularization with a scalar parameter λ as mentioned in section 3.3:

1. The regularization term to the cost function is defined as follows, with m number of training examples, feature column i, regularization parameter λ and parameter row number j [25]:

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^i log(h_\theta(x^i)) + (1-y^i)log(1-h_\theta(x^i))] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

2. The addition of the regularization term to the gradient descent function is defined as follows, with m number of training examples, feature column i, regularization parameter λ and parameter row number j[25]:

$$\theta_j := \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}[(h_\theta(x^i) - y^i)x_j^i + \frac{\lambda}{m}\theta_j]$$

The optimization function ($fminunc$) is operated to find the minimum of the cost function given the gradient. The amount of iterations for $fminunc$ is set to 200.

Third Stress Level

Because the likelihood of unseen data is set to 0 when $h_\theta(x) < \frac{1}{2}$ and set to 1 when $h_\theta(x) > \frac{1}{2}$ an if-statement is written, which assigns a prediction value of the data entries in the vector $Ytest$ in accordance with the aforementioned statement. The following adjustment is made within the if statement to supply the classifier with three stress levels (*"neutral", "stressed" and "severely stressed"*); after the prediction is made, the median value of the data entries with the property $h_\theta(x) > \frac{1}{2}$ will be set as boundary between "stressed" and "severely stressed". Therefore a third if-statement is written to set a third prediction value (2) to the data entries with values above the aforementioned boundary.

Real Time Stress Classification

Praat is executed within the MATLAB file to record and extract features from speech in real time. The user is asked to enter gender, because the dataset differs depending on gender. The Praat script for females is initialized together with the female dataset and the Praat script for males is initialized together with the male dataset. Praat extracts the features from the recording and the prediction will be made using the model of the dataset and the aforementioned if-statements.

5.2 Feature Evaluation

Redundant acoustic features will be excluded to reduce dimensionality to consequently remove noise features. When the most crucial discriminating features are detected, searching through the combinations of these relevant features is naturally more effective than searching through all of the features in the database. The feature evaluation is based on the two search methods and evaluators presented in section 3.3, namely BestFirst in combination with the evaluator CfsSubsetEval and the search method Ranker in combination with the evaluator InfoGainAttributeEval. A selection of features will be constructed considering the outcome of the aforementioned search methods. The most optimal set of features will be chosen by experimenting with the logistic regression classifier introduced in section 5.

5.3 Other Classifiers

The aforementioned classifier is evaluated using the logistic regression algorithm and decision tree learning in WEKA, considering the set of features chosen in section 5.2. The K-fold cross validation technique proposed in section 3.3 is also utilized to analyse additional data. The amount of folds is set to 10. The dataset file requires the following additional arguments for processing in WEKA:

- @*relation* stress-neutral; defines the name of the relation.
- @*State*$\{0, 1\}$; defines the state to be 1 or 0.
- @*attibute* FeatureName *numeric*; defines the numerical data.
- @*data*; defines the data area.

After adding these arguments, the file extension needs to be changed into an .arff file. The classification results will be summarized in section 6.

5.4 Experts

A listening experiment is set up to evaluate the classification which is constructed by the regularized logistic regression classifier (introduced in section 5.1), utilizing the set of features that is chosen in section 5.2. The experiment is based on the assumption that humans are experts in detecting stress. Twenty experts are asked for their opinion on the question in which state of stress the speakers in the audio files seem to be. The randomized selection of audio files is constructed regarding the test set utilized in section 5.2. The experts have to perform the following task in the listening experiment: the experts have to listen to audio files belonging to the spoken words in the lexicon of the test set. After listening to an audio file, the expert is asked to choose between “neutral”, “stressed” or “severely stressed”.

The evaluation of the aforementioned listening experiment depends on the T-test proposed in section 3.3 utilizing a significance value of 5%.

6 Results

6.1 Data

Every data entry contains 14 features, which are summarized in table 2, 3, 4. The male database is a $203X14$ dimensional matrix and the female database is a $229X14$ dimensional matrix. The amount of data entries are presented in table 7.

Dataset	**Neutral**	**Stressed**
Total Male	103	100
Total Female	115	114
Test Set Male	25	25
Test Set Female	25	25
Training Set Male	78	89
Training set Female	90	75

Table 7: The amount of data entries in each data set.

6.2 Feature Evaluation

The feature evaluation regarding the search method BestFirst in combination with the evaluator CfsSubsetEval is illustrated in table 8.The feature evaluation regarding the search method Ranker in combination with the evaluator InfoGainAttributeEval is illustrated in table 9.

Rank	**Female**	**Male**
1	f0mean	f0mean
2	f0median	f0median
3	f1median	f1median
4	f2median	f0minimum
5	medianIntensity	f0maximum
6	spectrumDifference	mean_jitter
7	f0minimum	mean_shimmer

Table 8: Male and Female ranking results on whole training set using BestFirst

Rank	Female	Gain	Male	Gain
1	f0median	0.2276	f0median	0.4233
2	f0minimum	0.2245	f0mean	0.4223
3	f0mean	0.2179	f0minimum	0.3773
4	f0maximum	0.1458	f0maximum	0.3399
5	spectrumDifference	0.0843	f1median	0.1475
6	medianIntensity	0.0794	mean_jitter	0.0988
7	f2median	0.0638	mean_shimmer	0.0877
8	f1median	0.0637	f0range	0.0681
9	f0range	0	f0std	0.0681
10	f0std	0	spectrumDifference	0.0618
11	mean_jitter	0	stdIntensity	0
12	stdIntensity	0	f2median	0
13	mean_shimmer	0	medianIntensity	0

Table 9: Male and Female ranking results on whole training set using Information Gain Feature Ranking

6.3 Stress Classifier

The results of experimenting on the male test set utilizing the aforementioned rankings are listed in table 10. The most accurate feature set regarding males contains the combination of $f0median$, $f1median$ and $spectrumDifference$ which is illustrated in figure 3. Increasing regression parameter with value $\lambda = 4$ four times on the male test set consisting of features $f0median$, $f1median$ and $mean\_jitter$ increases the training accuracy towards 81.70% and test accuracy towards 82%. Increasing λ with regards to the resulting combinations of male features causes the corresponding test sets accuracy to decrease and the corresponding training sets accuracy to increase.

Male Feature Set	Train %	Test %
f0median	79.08	76
f0median, f1median	81.04	80
f0median, f1median, f0minimum	80.39	80
f0median, f1median, f0minimum,spectrumDifference	80.39	80
f0median, f1median, spectrumDifference	81,05	82
f0median, spectrumDifference, mean_shimmer	79.74	82
f0median, f1median, mean_jitter, spectrumDifference	81.05	82
f0median, f1median,mean_jitter	81.05	82

Table 10: The results of experimenting concerning different combinations of features predicting the male test set based on the male training set, with the training accuracy and the test set accuracy in percentages.

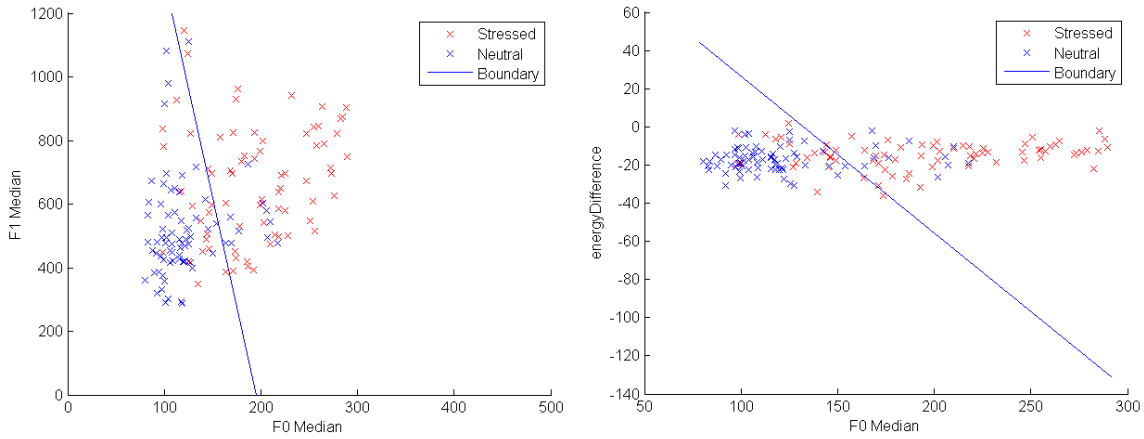

Figure 3: The result of combining the features $f0median$, $f1median$ and $spectrumDifference$ concerning the male training set. The red and blue crosses represent the original state of the data entries in the training set ("neutral" or "stressed") and the boundary is set for the classification of new data.

The results of experimenting with the female test set utilizing the aforementioned rankings are listed in table 11. The most accurate feature set regarding females contains the combination of $f0median$, $f1median$ and $medianIntensity$ which is illustrated in figure 4. Increasing regression parameter with value $\lambda = 2$ upon the female test set consisting of features $f0median$, $f1median$ and $mean\_jitter$ decreases the training accuracy towards 75.98% and increases the test accuracy towards 78%. Increasing λ concerning the resulting combinations of female features causes the corresponding test sets accuracy to decrease and the corresponding training sets accuracy to decrease.

Female Feature Set	**Train %**	**Test %**
f0mean	77.09	74
f0mean, f1median	78.78	68
f0mean, f1median, f0minimum	78.21	72
f0median, f1median, f0minimum	77.65	74
f0median, f1median, f0minimum, medianIntensity	78.21	74
f0median, f1median, medianIntensity	77.09	78
f0median, medianIntensity	78.21	76
f0median, f1median, mean_jitter	77.65	76

Table 11: The results of experimenting with different combinations of features predicting the female test set based on the female training set, with the training accuracy and the test set accuracy in percentages.

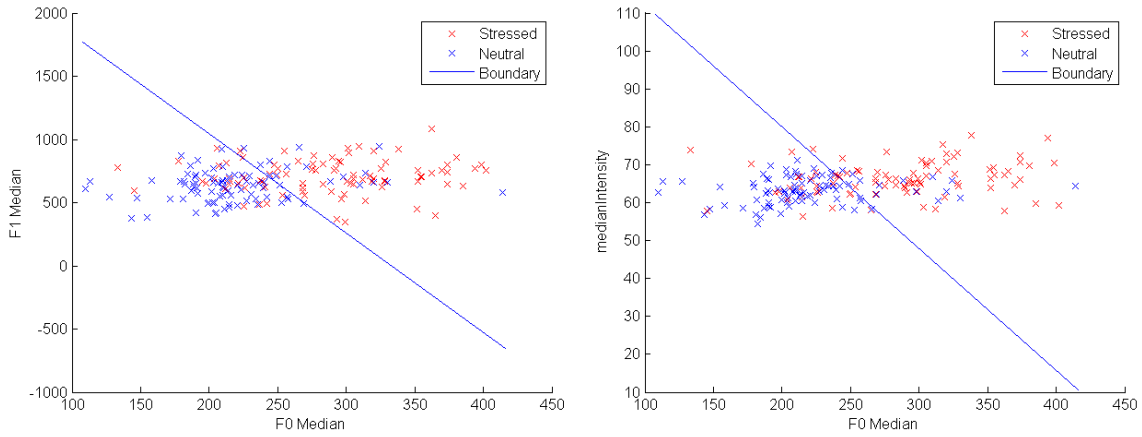

Figure 4: The result of combinating the features $f0median$, $f1median$ and $medianIntensity$ considering the female test set. The red and blue crosses represent the original state of the data entries in the training set ("neutral" or "stressed") and the boundary is set for the classification of new data.

Third Stress Level

The third stress level according to the hypothesis function $h_\theta(x)$ is presented in section 3.3 and section 5. The interval and median value of $h(\theta)$ within range $\theta > 0.5$ regarding the male and female test set and the corresponding features is presented in table 12.

θ	**Min**	**Max**	**Median**
Male	0.5173	0.9935	0.7593
Female	0.5052	0.9877	0.6717

Table 12: The minimum, maximum and median value of $h(\theta)$ regarding the male and female test set and the corresponding features.

6.4 Other Classifiers

Table 13 illustrates the results of the J48 decision tree learning algorithm and the logistic regression algorithm evaluating the male and female test set in WEKA. The logistic regression is evaluated using WEKA to acknowledge the functioning of the stress classifier implemented in MATLAB.

Gender	Data Set	Classifier	Accuracy (%)
Male	Training	Logistic Regression	80.79
Male	Training	J48	86.70
Male	Test	Logistic Regression	80.00
Male	Test	J48	86.00
Male	10 Fold	Logistic Regression	79.31
Male	10 Fold	J48	81.28
Female	Training	Logistic Regression	78.17
Female	Training	J48	80.35
Female	Test	Logistic Regression	78.00
Female	Test	J48	84.00
Female	10 Fold	Logistic Regression	76.92
Female	10 Fold	J48	73.36

Table 13: The accuracy of the male and female test set, training set and 10-fold validation utilizing the J48 pruned tree classifier and the logistic classifier in WEKA.

6.5 Experts

The evaluation of the listening experiment introduced in section 5.4 is performed utilizing the T-test proposed in section 3.3 employing a significance value of 5%. The results of the listening experiment are disclosed in table 14.

Ω	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
Results (%)	60	54	46	52	42	60	48	54	58	52
Ω	ω_{11}	ω_{12}	ω_{13}	ω_{14}	ω_{15}	ω_{16}	ω_{17}	ω_{18}	ω_{19}	ω_{20}
Results (%)	58	42	70	46	58	62	58	54	60	48

Table 14: The results of the listening experiments with regards to the experts Ω. The listening experiment contains the audio fragments concerning the male test set.

The following calculations are performed with regards to the T-test:

The assumption $H_0 : \mu$ is the accuracy of the male classifier according to the features $f0median$, $f1median$ and $spectrumDifference$:

$$(H_0 : \mu) \approx \frac{82}{100}$$

The probability distribution is characterized by the amount of experts and the results of the listening experiment presented in table 14:

$$X : \begin{pmatrix} 0.42 & 0.46 & 0.48 & 0.52 & 0.54 & 0.58 & 0.60 & 0.62 & 0.70 \\ \frac{2}{20} & \frac{2}{20} & \frac{2}{20} & \frac{2}{20} & \frac{3}{20} & \frac{4}{20} & \frac{3}{20} & \frac{1}{20} & \frac{1}{20} \end{pmatrix}$$

The expectation $E(X)$ is calculated with regards to the probability distribution and the formula described in section 3.3:

$$E(X) = \sum_{w \in \Omega} X(\omega) P(\{\omega\}) = 60.20\%$$

The variance is determined to estimate the standard deviation σ:

$$VAR(X) = E(X^2) - (E(X))^2 = E(X^2 - X)^2 = -\frac{61}{1000}$$

The standard deviation σ is determined regarding the variance:

$$\sigma = \sqrt{VAR(X)} \approx \frac{247}{1000}$$

$\bar{X}$ is the average of the results:

$$\bar{X} = \frac{541}{1000}$$

n is the number of experts Ω. $\Omega = w_1, ..., w_n$, therefore the chance of selection of one expert is set to $\frac{1}{20}$. The T-test is determined using the aforementioned values:

$$T = \frac{\bar{X} - (H_0 : \mu)}{\sigma} \sqrt{n} = -5.05$$

Since the statistical significance is set to $\frac{5}{100}$, the critical value c regarding the amount of experts $n = 20$ is 1.724718 regarding the t-distribution table presented in section A. The assumption H_0 will be rejected since $T \leq -c$, since H_0 will not be rejected while $-c < T < c$ with regards to section 3.3.

7 Evaluation

Two training models are set up, since different settings apply to the analyses of the voices of male speakers and female speakers. To retrieve the formant measurements of female voices, the maximum formant frequency is set to 5500 Hertz. It is set to 5000 Hertz to carry out measurements regarding male voices. The fundamental frequency is measured in a range from 75 Hertz to 300 Hertz regarding male voices and a range from 300 Hertz to 500 Hertz regarding female voices.

The pitch measurements and the measurements regarding the first formant are considered most useful when detecting stress, according to the feature evaluations illustrated in table 8 and 9. Therefore different feature sets are evaluated regarding the test results. The principal feature combinations are weighed, applying the stress classifier established in section 5.1. The most optimal set of features to detect stress for male speakers making use

of the stress classifier, consists of $f0median$, $f1median$ and $spectrumDifference$, with a training set accuracy of $81,05\%$ and a test set accuracy of 82% according to table 10.

The most optimal set of features to detect stress for female speakers is $f0median$, $f1median$ and $medianIntensity$ with a training set accuracy of $77,09\%$ and a test set accuracy of 78% according to table 11. The male test set encounters limited underfitting; the female test set encounters limited overfitting. The amount of overfitting and underfitting regarding the female- and male test sets are relatively small and hence neglected. When increasing regression parameter λ to value $\lambda = 4$, the accuracy of the test set improves, as is mentioned in section 6.3. The aforementioned feature set will not be considered as the most optimal feature set, due to the aforementioned results of feature combination $f0median$, $f1median$ and $spectrumDifference$ which have the same results without regularization, and because of the fact that jitter can be caused by noise and by peak shifts mentioned in 3.2. Regarding the rest of the feature combinations, increasing regularization parameter λ causes underfitting regarding both the male and female test set. Therefore increasing regularization parameter λ does not have a positive effect on the other test sets or the training sets.

The estimation of the three stress levels clarified in 3.3 makes use of the values in table 12 to predict unseen data. New predictions of θ within the interval $0.5173 > |H_\theta(X)| < 0.7593$ are classified as "stressed" and predictions within the interval $0.7593 < |H_\theta(X)| > 0.9935$ are classified as "Very stressed" regarding the male training data. Predictions of θ within the interval $0.5052 > |\theta| < 0.6717$ are classified as "stressed" and predictions within the interval $0.6717 < |\theta| > 0.9877$ are classified as "Very stressed" regarding the female training data.

With reference to the results of the classifiers in WEKA (see section 6.4), J48 is a proper evaluator for the current dataset, since the dataset is relatively small. A larger database is necessary to make a justified classification with regards to stress detection. A larger dataset will cause J48 to acquire a large increase of leaves. Therefore J48 is not suitable for detecting stress in a first aid speech system. However, the results of the logistic regression classifier in WEKA are identical to the results the stress classifier mentioned in section 6.3. Therefore it is valid to make use of the aforementioned stress classifier as a tool for stress detection.

It has become clear that H_0 - the assumption that experts classify with the same accuracy as the stress classifier - is proved invalid by the results of the T-test (see section 6.5). The expectation is calculated for experts to $60,20\%$ to detect stress in male speech, according to the listening experiment performed in section 6.5. It can be feasible to detect three levels in stress, considering the expectation of the expert is higher than 50%. Furthermore, the error rate of the stress classification in section 6.3 results in 19% misclassified data entries regarding the male test data, which leads to a higher error rate of the experts. Experts can misclassify their perception of stress, which causes the aforementioned error rate to decline. The data set has been labelled manually while taking into account the visual context of the film "The Call" mentioned in section 4.1. More contextual information was thus provided to construct a more valid data set. The experts have not had this contextual

information to their advantage.

8 Conclusion

Is is stated above, the main concept of the first aid speech system is explained in section 2. A conceptual model of the first aid speech system is illustrated in figure 1. Three general procedures P_1, P_2 and P_3, depending of the emergency workers' stress level are introduced in section 2. The first aid speech system is able to differentiate between the aforementioned procedures in accordance with a "mild", "stressed", and "severely stressed" stress level.

As mentioned in the introductory part of this work, the universal model of StressSense has an accuracy of 71.30%, which is 10% less than the accuracy of the stress classifier presented in section 3. A proper stress detector distinguishes between gender in agreement with the different sound measurements regarding male or female voices, as is mentioned in section 7. Hence the universal model is not suitable as stress detector. The top three features regarding stressSense are the speaking rate, standard deviation and the range of the pitch. The pitch range strongly correlates with the standard deviation and this causes overfitting concerning the parameters.

Articulation and respiration are influenced by stress and are the cause of an higher formant frequency and an higher fundamental frequency as stated in section 3.2. This statement is validated by the results of the most appropriate stress classifier, containing the fundamental frequency- and the formant frequencies features. The principal set of features concerning males is the median of the fundamental frequency, the median of the first formant measurements and the mean spectral difference between two intensity objects. The last feature can be called contemporary regarding the state of the art presented in section 3. Unlike StressSense, the most optimal measurements consist of different types of measurements which are crucial to detect stress.

The optimal tool for detecting stress is logistic regression: stress detection using the J48 decision tree learning algorithm does not function better. J48 has a high accuracy on the small test set presented in section 6.1. However, J48 will not be not suitable for detecting stress in a critical first aid situation due to complexitys mentioned in section 7 and therefore time consuming.

The results of the feature evaluation methods in section 6.2 have not been proven sufficient to retrieve an optimal feature set. Experimenting with the stress classifier presented in section 5.1 has been necessary to retrieve the optimal feature set. The most beneficial acoustic properties of the speech signal to detect stress for males are as follows:

1. The median of the fundamental frequency measurements.
2. The median of the first formant frequency measurements.
3. The band energy difference of two intensity objects regarding the same corresponding speech signal, which is not expected to be valuable regarding stress detection.

The most beneficial acoustic properties of the speech signal to detect stress for females are as follows:

1. The median of the fundamental frequency measurements.
2. The median of the first formant frequency measurements.
3. The median of the intensity measurements.

It is feasible to detect stress when the aforementioned set of features is established. Logical regression reaches an optimum using the features as mentioned in 3.3. Therefore the analysis of the acoustic signal of an emergency workers' voice when the emergency worker is under stress can be used as an indicator for his or her stress level.

A solidly built stress classifier with a high accuracy is necessary with regards to a proper functionality of the first aid speech system as stated in section 1. Since such a stress classifier is not yet validated, the stress detection is not accurate enough in the speech system to apply it in critical situations when providing first aid. However, for emergency workers in training it can be a useful tool to manage stress. In these situations, the first aid speech system will still be a positive accompaniment to the performance of an emergency worker.

9 Discussion

There are a few remarks to make with regards to this research on a general level, as well on a smaller scale.

First of all, the database illustrated in table 7 is considered small. A larger database would yield a better performance. The database only contains words spoken in English: other languages can contain different characteristics. A database should then be validated for every language. The database is labelled by an individual, and thus dependent on a certain level of subjectivity; a ground truth such as Galvanic Skin Response measurements could be a valuable addition to this research.

The results of the T-test in section 6.5, show that experts could not validate the hypothesis. The expert is not able to differentiate the output of the classifier in three levels at the same accuracy as the stress classifier, namely 82%. The experts are expected to classify $60,20\%$ of three differentiating stress levels correctly when taking into account the male test set. As a consequence, the approach that differentiates between three stress levels is not rejected, but does still need more proof.

In section 3.1, one of the top three features regarding StressSense is said to be the speaking rate. The speaking rate has not been measured; experimenting with the speaking rate and the training model would be a useful addition to the validation of the most optimal features.

On a microlevel, the analysis of the speech signal can differ because of a large amount of variable measurements. It would be useful to find the most suitable values regarding the measurements, such as the window length illustrated in figure 2. Even though this point of discussion may seem rather small, it is however of great relevance.

The development of a first aid system and further research considering the procedures of the first aid system would be a great contribution to the profession of emergency workers, who have to make crucial decisions in their line of work.

A Appendix

Ω/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000.000	3.077.684	6.313.752	1.270.620	3.182.052	6.365.674	6.366.192
2	0.288675	0.816497	1.885.618	2.919.986	430.265	696.456	992.484	315.991
3	0.276671	0.764892	1.637.744	2.353.363	318.245	454.070	584.091	129.240
4	0.270722	0.740697	1.533.206	2.131.847	277.645	374.695	460.409	86.103
5	0.267181	0.726687	1.475.884	2.015.048	257.058	336.493	403.214	68.688
6	0.264835	0.717558	1.439.756	1.943.180	244.691	314.267	370.743	59.588
7	0.263167	0.711142	1.414.924	1.894.579	236.462	299.795	349.948	54.079
8	0.261921	0.706387	1.396.815	1.859.548	230.600	289.646	335.539	50.413
9	0.260955	0.702722	1.383.029	1.833.113	226.216	282.144	324.984	47.809
10	0.260185	0.699812	1.372.184	1.812.461	222.814	276.377	316.927	45.869
11	0.259556	0.697445	1.363.430	1.795.885	220.099	271.808	310.581	44.370
12	0.259033	0.695483	1.356.217	1.782.288	217.881	268.100	305.454	43.178
13	0.258591	0.693829	1.350.171	1.770.933	216.037	265.031	301.228	42.208
14	0.258213	0.692417	1.345.030	1.761.310	214.479	262.449	297.684	41.405
15	0.257885	0.691197	1.340.606	1.753.050	213.145	260.248	294.671	40.728
16	0.257599	0.690132	1.336.757	1.745.884	211.991	258.349	292.078	40.150
17	0.257347	0.689195	1.333.379	1.739.607	210.982	256.693	289.823	39.651
18	0.257123	0.688364	1.330.391	1.734.064	210.092	255.238	287.844	39.216
19	0.256923	0.687621	1.327.728	1.729.133	209.302	253.948	286.093	38.834
20	0.256743	0.686954	1.325.341	1.724.718	208.596	252.798	284.534	38.495
21	0.256580	0.686352	1.323.188	1.720.743	207.961	251.765	283.136	38.193
22	0.256432	0.685805	1.321.237	1.717.144	207.387	250.832	281.876	37.921
23	0.256297	0.685306	1.319.460	1.713.872	206.866	249.987	280.734	37.676
24	0.256173	0.684850	1.317.836	1.710.882	206.390	249.216	279.694	37.454
25	0.256060	0.684430	1.316.345	1.708.141	205.954	248.511	278.744	37.251
26	0.255955	0.684043	1.314.972	1.705.618	205.553	247.863	277.871	37.066
27	0.255858	0.683685	1.313.703	1.703.288	205.183	247.266	277.068	36.896
28	0.255768	0.683353	1.312.527	1.701.131	204.841	246.714	276.326	36.739
29	0.255684	0.683044	1.311.434	1.699.127	204.523	246.202	275.639	36.594
30	0.255605	0.682756	1.310.415	1.697.261	204.227	245.726	275.000	36.460
inf	0.253347	0.674490	1.281.552	1.644.854	195.996	232.635	257.583	32.905

Table 15: The critical t values regarding the T-test in section 3.3, given the amount of experts Ω and the significance. The assumption H_0 will be rejected when $t \leq -c$ or when $t \geq c$, and H_0 will not be rejected when $-c < t < c$ [32].

References

[1] M. Tarafdar, Q. Tu, B. S. Ragu-Nathan, and T. Ragu-Nathan, "The impact of technostress on role stress and productivity," *Journal of Management Information Systems*, vol. 24, no. 1, pp. 301–328, 2007.

[2] S.-Å. Christianson, "Emotional stress and eyewitness memory: a critical review.," *Psychological bulletin*, vol. 112, no. 2, p. 284, 1992.

[3] S. A. Patil and J. H. Hansen, "Speech under stress: Analysis, modeling and recognition," 2007.

[4] G. Zhou, J. H. Hansen, and J. F. Kaiser, "Nonlinear feature based classification of speech under stress," *Speech and Audio Processing, IEEE Transactions on*, vol. 9, no. 3, pp. 201–216, 2001.

[5] D. A. Cairns and J. H. Hansen, "Nonlinear analysis and classification of speech under stressed conditions," *The Journal of the Acoustical Society of America*, vol. 96, no. 6, pp. 3392–3400, 1994.

[6] P. Rajasekaran, G. Doddington, and J. Picone, "Recognition of speech under stress and in noise," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86.*, vol. 11, pp. 733–736, IEEE, 1986.

[7] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury, "Stresssense: Detecting stress in unconstrained acoustic environments using smartphones," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pp. 351–360, ACM, 2012.

[8] Koninklijke Landmacht, "Handboek Koninklijke Landmacht Militair, soldiers manual regarding the Royal Dutch Army," 2010.

[9] G. Keinan, "Decision making under stress: scanning of alternatives under controllable and uncontrollable threats.," *Journal of personality and social psychology*, vol. 52, no. 3, p. 639, 1987.

[10] V. R. LeBlanc, "The effects of acute stress on performance: implications for health professions education," *Academic Medicine*, vol. 84, no. 10, pp. S25–S33, 2009.

[11] T. J. Tan and C. Winkelman, "The contribution of stress level, coping styles and personality traits to international students academic performance," *Australian Catholic University, Locked Bag*, vol. 4115, 2010.

[12] T. H. Holmes and R. H. Rahe, "The social readjustment rating scale," *Journal of psychosomatic research*, vol. 11, no. 2, pp. 213–218, 1967.

[13] S. T. Bramwell, M. Masuda, N. N. Wagner, and T. H. Holmes, "Psychosocial factors in athletic injuries: Development and application of the social and athletic readjustment rating scale (sarrs)," *Journal of Human Stress*, vol. 1, no. 2, pp. 6–20, 1975.

[14] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *Journal of health and social behavior*, pp. 385–396, 1983.

[15] K. R. Scherer, "Vocal markers of emotion: Comparing induction and acting elicitation," *Computer Speech & Language*, vol. 27, no. 1, pp. 40 – 58, 2013. Special issue on Paralinguistics in Naturalistic Speech and Language.

[16] David Weenink, "Speech Signal Processing with Praat, lecture notes distributed in Speech Recognition and Synthesis at the University of Amsterdam," 2014.

[17] Boersma, Paul & Weenink, David, "Praat: doing phonetics by computer [Computer program]. Version 5.3.78." `http://www.praat.org/`, 2014.

[18] Paul Boersma, "Chapter 17: Acoustic analysis," in *Research methods in linguistics*, Cambridge University Press, 2014.

[19] P. Boersma, "Stemmen meten met praat," *Stem-, spraak-, en taalpathologie*, vol. 12, pp. 237–251, 2004.

[20] C. J. van As, "Chapter 5: Acoustic measures and signal typing of voice quality in tracheoesophageal speech, and their relations to perceptual evaluations," in *Tracheoesophageal speech. A multidimensional assessment of voice quality*, Budde-Elinkwijk Grafische Producties, 2001.

[21] Sony Pictures Home Entertainment, "The call [Film]," 2013.

[22] "About Audacity. [Computer program]." `http://audacity.sourceforge.net`, Retrieved 2014-05-27.

[23] "Praat Info [Scripts]." `http://aune.lpl-aix.fr`, retrieved on 2014-25-05.

[24] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in *Proceedings of the institute of phonetic sciences*, vol. 17, pp. 97–110, Amsterdam, 1993.

[25] Andrew Ng, "Logistic Regression (Week 3) [Video Lecture]." `https://class.coursera.org/ml-004/`, retrieved on 2014-30-05.

[26] S.-I. Lee, H. Lee, P. Abbeel, and A. Y. Ng, "Efficient l˜ 1 regularized logistic regression," in *Proceedings of the National Conference on Artificial Intelligence*, vol. 21, p. 401, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[27] "fminunc [Tool]." `http://www.mathworks.nl/help/optim/ug/fminunc.html`, retrieved on 2014-30-05.

[28] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[29] M. I. Devi, R. Rajaram, and K. Selvakuberan, "Generating best features for web page classification," *Webology*, vol. 5, no. 1, 2008.

[30] S. Russell, P. Norvig, and A. Intelligence, "A modern approach," *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, vol. 25, 1995.

[31] M. Mathuria, "Decision tree analysis on j48 algorithm for data mining," *Intrenational Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 6, 2013.

[32] B. Van Es, "Kansberekening en Statistiek, lecture notes distributed in Probability and Statistics at the University of Amsterdam," 2012.

[33] "Audacity [Features]." `http://audacity.sourceforge.net`, 2013-03-22. Retrieved 2014-27-05.