# FINAL REPORT

by order of the

Austrian Marshall Plan Foundation

presented by:

**Verena Venek, BSc**

Los Angeles, August 2014

# Automatic Suicidal Risk Assessment based on Clinician-Patient Verbal and Nonverbal Behavior

Master Thesis

**Verena Venek**
**(1210310003)**
**10/3/2014**

Performed at the

**University of Southern California Institute for Creative Technologies**

and at the

**Fachhochschule Kärnten/Carinthia University of Applied Sciences**

School of Medical Information Technology

Supervised by

Dipl.-Ing. Dr. techn. Johannes Oberzaucher | Prof. Dr. Stefan Scherer

(Verena Venek)

# Abstract

*Suicide is the third leading cause of death in the United States of America for ages 13 to 18 and in some countries youth suicide is even the second leading cause of death for ages 10-24 according to Centers for Disease Control and Prevention and WHO. Hence, there is a need for a valid, rapid and accurate way to identify suicidal risk; it is especially important to be able to predict potential repeat occurrences of attempted suicides. In this thesis the conversation dynamics, verbal information and acoustic features of 30 suicidal and 30 non-suicidal patients of ages 13 to 18 interviewed at the Cincinnati Children's Hospital Medical Center Emergency Department and the single clinician were statistically evaluated. Beside the identification of non-suicidal vs. suicidal patients, data of 17 suicidal non-repeaters and 13 repeaters was investigated. A hierarchical classification was used to confirm the discriminative faculty of the features to characterize the psychological states of the adolescents. Ensemble classifiers and SVM classifiers were used and tested with data from entire interviews and with data from the first five open-ended ubiquitous questions and answers. I was able to show that the analysis of verbal and nonverbal interactions between patient and clinician can identify the suicidal risk of adolescents. The investigation of the first five questions of the interview yielded similar promising results in comparison to the information observed in the entire interview. The backchannel of the clinician which lasted less than 700ms was useful to identify the suicidal risk including the patients' features. The hierarchical SVM-ensemble classification yielded an accuracy of 78.3% using data from entire interviews. Thus, the suicidal risk assessment of non-suicidal adolescents, suicidal non-repeaters and suicidal repeaters using clinician's and patient's verbal and nonverbal behaviors yielded promising results.*

**Keywords:** *Youth suicide, adolescent, interaction, repeater, hierarchical classifier.*

# Kurzfassung

*In den Vereinigten Staaten von Amerika ist Selbstmord von 13- bis 18-Jährigen die dritthäufigste Todesursache und in manchen Ländern sogar die zweithäufigste von Jugendlichen von 10 bis 24 Jahren laut CDC und WHO. Daher ist eine genaue, rasche und vor allem gültige Identifizierung des Selbstmordrisikos notwendig. Außerdem ist die Erkennung eines potenziellen wiederholten Selbstmordversuchs wichtig. In dieser Arbeit wird die statistische Evaluierung der Eigenschaften von 30 selbstmordgefährdeten Patienten und 30 Nicht-Selbstmordgefährdeten durchgeführt, die im Cincinnati Children's Hospital Medical Center in der Notfallaufnahme befragt wurden. Dabei wurden die Konversationsdynamik, die verbalen und die akustischen Eigenschaften der 13- bis 18-jährigen Patienten und des Klinikers betrachtet. Bei der statistischen Evaluierung wurden die Eigenschaften der Selbstmordgefährdeten und Nicht-Selbstmordgefährdeten untersucht. Außerdem wurde die Unterscheidung der Selbstmordgefährdeten mit mindestens einem erneuten Versuch und derjenigen, die es kein weiteres Mal versucht hatten, evaluiert. Eine hierarchische Klassifikation wurde verwendet um die diskriminative Fähigkeit der Eigenschaften zu bestätigen, dass sie den psychologischen Status von Jugendlichen charakterisieren können. Dafür wurden Ensemble und SVM Algorithmen zur Klassifikation verwendet. Neben der Untersuchung der gesamten Interviews wurden auch die ersten fünf Fragen und Antworten der 'ubiquitous questions' eingehender betrachtet. Durch die Analyse der verbalen und nicht verbalen Interaktionen zwischen Patient und Kliniker war es möglich auf das Selbstmordrisiko der Jugendlichen rückzuschließen. Das Selbstmordrisiko konnte auch bei Betrachtung der ersten fünf Fragen und Antworten ermittelt werden. Der 'Backchannel' des Klinikers beschreibt Abschnitte des Interviews, die weniger als 700ms dauern. Die daraus gewonnenen Informationen halfen ebenfalls das Selbstmordrisiko zu bestimmen. Die hierarchische SVM-Ensemble Klassifikation ergab eine Genauigkeit der Unterscheidung von 78.3% unter Verwendung des gesamten Interviewmaterials. Es konnten vielversprechende Ergebnisse erreicht werden, wenn es darum ging das Selbstmordrisiko zwischen nicht-selbstmordgefährdet, selbstmordgefährdet aber ohne wiederholtem Selbstmordversuch und selbstmordgefährdet mit wiederholtem Selbstmordversuch anhand von verbalen und nicht verbalen Verhalten von Patienten und Klinikern zu bewerten.*

**Suchbegriffe:** *Selbstmordrisiko, Jugendliche, Interaktion, Selbstmordversuch, hierarchische Klassifikation.*

# Dedication

I dedicate this thesis to my family; my parents, Robert and Sabine, and my "little" sister Vanja. Thank you so much for your amazing love and support throughout my life.
I love you.

# Acknowledgements

I would like to thank my supervisor at USC ICT Stefan Scherer for his support and instructions. I am also grateful to Mr. Oberzaucher for spending time providing useful suggestions about this thesis.

Special thanks are also given to the Institute for Creative Technologies. I have learned many things during my time there, thanks to my amazing colleagues. I really appreciate your support.

# Content

# 1. Introduction

More than 800,000 people die annually from suicide according to the World Health Organization (WHO). Especially the amount of suicide among youth is alerting. In some countries, suicide is the second leading cause of death for 10-24 years old people [WHO14]. In the United States, suicide is the third leading cause of death for ages 13 to 18 [CDC14].

Unfortunately, many never seek professional care previously. Thus, suicidal behavior often remains undetected and undiagnosed. The National Youth Risk Behavior Surveillance [KKS+14] stated that 17% of high school students (i.e. ages 14-18) seriously considered attempting suicide and 13.6% made a suicide plan, while 8% actually attempted suicide one or more times. These statistics are illustrated in Figure 1-1. The bar graph shows that 17% of the interviewed high school students between the ages 14 to 18 considered to attempt suicide, 13.6% of them already made a suicide plan and 8% actually attempted suicide once or more times at the time of the survey. Hence, the task is to detect suicide risk in adolescents before a suicide can be attempted or re-attempted.

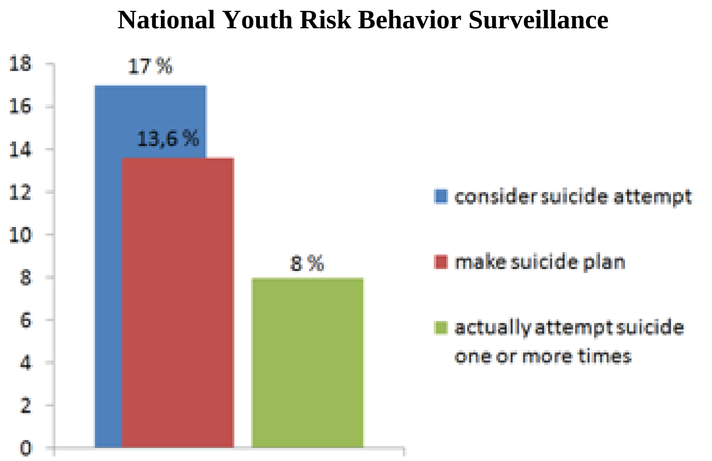


Figure 1-1: Percentages of National Youth Risk Behavior Surveillance results regarding the suicidal behavior survey [KKS+14]

Nevertheless, the number of suicidal attempts is highest among adolescents and young adults, as shown in Figure 1-2 [JBB+06]. In this smoothed hazard function the relationships between onset of lifetime suicidal ideation, plan and attempt among black people is illustrated related to reported ages of onset. The curve increases with age until it peaks in the adolescence and young adulthood. Afterwards, it slightly declines and rapidly increase again in the mid-fifties. To sum it up, there is the highest risk of suicidal ideation, plan and attempt approximately between ages of 15 and 25.

However, they tend less frequently to kill themselves. Hence, suicidal ideation is more common than actually attempting suicide. Parellada et al. [PSM+08] found that 10% of their 104 participants ages 12 to 17 are likely to repeat a suicidal attempt. As the authors mention, this percentage of the repetitive attempts would be higher when they considered suicide attempts followed by any medical health service not only those followed by a psychiatric admission.

**Suicidal hazard function**

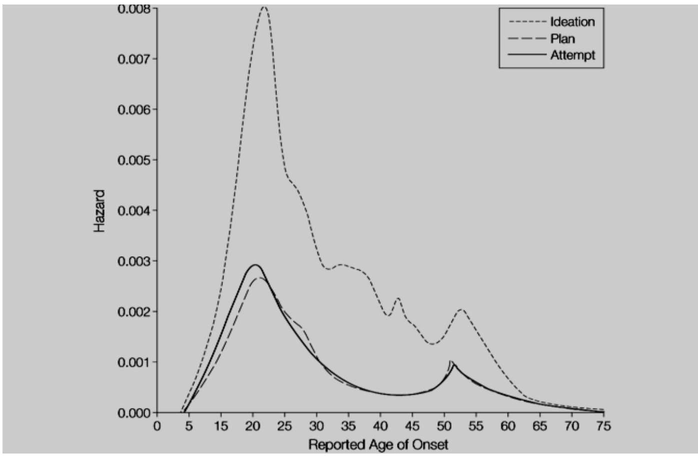


Figure 1-2: Hazard function curve for suicidal ideation, plan and attempt corresponding to ages up to 75 years [JBB+06]

The ability to prevent suicide depends on an early recognition of the suicidal risk (i.e. suicidal or non-suicidal patient) of individuals. Detecting suicidal risk in the primary care settings, for example in the Emergency Departments (EDs) in hospitals, is advantageous. Ahmedani et al. [ASS+14] investigated the health contacts in the year before suicide death. They found that almost all of their participants received health services in the year before their lethal suicide (83%) but half of them were not diagnosed suffering from a mental health disorder. They stated that a mental health and suicide risk assessment has to be performed in general medical settings more thoroughly to begin treatment earlier and prevent re-attempts or suicide. There, an interaction between patient and clinician can yield to a suicidal risk assessment.

This interaction is mainly based on written questionnaires and scales, especially developed to investigate suicidal behavior in adolescents [PBS+11, RM99]. Hence, mainly the patients' verbal behavior is investigated by the clinicians to decide the level of suicidal risk or even to prevent another suicide attempt [HJM05]. Especially the lack of time plays a crucial role for the clinicians in the ED [BHW+13]. This and the fact that many cases remain undiagnosed or untreated lead to the need of improvement in the reliability of suicide certification and reporting [WHO14, ASS+14, Dav04]. Furthermore, there is a need for a valid, rapid and accurate way to identify suicidal risk, especially to predict a potential re-attempt of the suicidal patients [HBP09].

The main task of this master thesis is to investigate if suicidal behavior can be characterized by observing the audio-based verbal and nonverbal interaction between patient and clinician. Hence, the suicidal risk of individuals should be investigated by observing the clinician-patient conversation including the communicated verbal information and voice characteristics.

There are several studies investigating the verbal and nonverbal behavior of suicidal patients and, therefore, assessing the suicidal risk of patients [Pen11, SPM13, OSW+04]. Even the importance of the clinician-patient interaction used to relate to psychological states of mind was observed, for example in [HHR95] and [BDS02]. Hence, studies exist which discuss the association of verbal and nonverbal cues with clinical conditions, such as suicidal risk. However, unusually is the combination of both: adding information which has not been used by now to study if the interaction between patient and clinician can yield to a promising suicidal risk assessment. Thus, a support for the clinician could be developed to detect suicidal risk as early as possible. The verbal behavior is described throughout this thesis as *verbal information* which is communicated by patients and clinician with words and content. The nonverbal behavior in the thesis includes voice quality and vocal behaviors, in the further context called *conversational information* and *acoustic information*. This conversational dynamic information describe interview characteristics such as speak times, pause times and interruptions of both, patient and clinician. Acoustic information are associated with tone of voice and other different parameters characterizing the phonation type of individuals. Due to the missing of visual information of the interviews, the nonverbal behavior in this study excludes visual and facial communications like head nodding or appearances of smiles which, among others, [FMH82] and [HHH+01] already investigated to be useful to assess patients' psychological states.

The combination of verbal and nonverbal behavior descriptors, separated by clinician and patient, is aimed to obtain promising suicidal risk assessment results, including especially also the distinction between non-repeaters and repeaters. For this separation, suicidal patients which attempted suicide more than once before the interview are referred as *repeaters*. The *non-repeaters* are suicidal patients with one or none suicidal attempt, but at least showing signs of suicidal gestures or ideation. The aim is to observe if an identification of non-repeaters and repeaters is possible by observing the interviews.

Especially the dynamic between interviewer and interviewee, i.e. using features of patients and clinician and using conversational dynamic features, are hoped to be of promising usefulness in identifying suicidal risk. Statistical analyses are performed to identify the most significant faculty of the descriptors to characterize the states of the patients. The statistically significant features will be employed by a machine learning classification algorithm to confirm their discriminative faculty. Thus, it is investigated if it is possible to support the suicide risk assessment in the ED by objectively quantifying behaviors during the conversation between patients and clinician.

Furthermore, the five questions and answers of the interview, i.e. the ubiquitous question (UQ) section, are analyzed and determined if it would lead to comparative or similar suicidal risk assessment results than using the entire interview. Using just these five open-ended questions would implicate a brief and clear way to assess suicidal risk among adolescents.

As a further information source in the classification task, the backchannel of the clinician is investigated. The backchannel describes speech patches lasting less than 700 milliseconds including voice information. It is aimed to observe how the backchannel differs between non-suicidal adolescents, suicidal non-repeaters and repeaters.

The main hypotheses of the master thesis are summarized in Table 1**.**

Table 1: Summary of the three main hypotheses throughout this thesis

| | |
|---|---|
| **H1** | The suicidal risk of a person can be recognized by analyzing the verbal and nonverbal interaction between patient and clinician. It can be determined if there is a difference in the verbal and nonverbal behavior of the participants as well as of the clinician between non-suicidal patients, suicidal non-repeaters and suicidal repeaters. |
| **H2** | The investigation of the first section of the interview including the ubiquitous questions yields comparative results than analyzing the entire interview. The use of five questions can support the clinician with a rapid and valid way to assess suicidal risk among adolescents. |
| **H3** | The acoustical information of the backchannel of the clinician is useful to determine suicidal risk together with the patient's characteristics. This means that observing speech fragments of the clinician which last less than 700 milliseconds can yield a supportive result to identify the suicidal risk of adolescents. |

Hence, the difference of the present work to others, especially to [SPM13], is that, in addition to the investigation of acoustic features, conversational information and verbal information and additional acoustic information are acquired to characterize suicidal speech of adolescents between the ages of 13 and 18. Moreover, this work focuses on the dynamics between the clinician and patient during an interview setting and analyzes them separately as well as jointly. The verbal and nonverbal behaviors and differences are investigated between the two classification cases, non-suicidal vs. suicidal adolescents and suicidal non-repeaters vs. suicidal repeaters. Furthermore, the ability of a classifier to discriminate these three classes is investigated. A hierarchical ensemble classifier is implemented which first discriminates the suicidal from the non-suicidal adolescents and then classifies non-repeaters and repeaters.

The investigated dataset in this master project includes patients from the Cincinnati Children's Hospital Medical Center (CCHMC) ED with suicidal ideation, gestures or attempts. Suicidal ideation is referred to be "*any self-reported thoughts of engaging in suicide-related behavior*", while suicidal gestures are characterized as a suicide-related behavior of persons "*who have no intention of killing themselves*" [OBM+96]. Attempted suicides are defined as a possible self-injurious behavior with a non-lethal outcome but with the proved intention of the attempter to kill her- or himself. A suicide attempt may or may not lead to injuries [OBM+96].

At CCHMC 60 interviews with 30 suicidal patients and 30 non-suicidal ones were recorded, transcribed and speech features of the patients and the single trained social worker, i.e. the clinician, were extracted. In the method section, the dataset, the investigated features and the statistical evaluation methods including the Analysis of Variance (ANOVA) are introduced. Furthermore, the structures and the evaluation of the hierarchical classification are described. The results section deals with the significant differences of the statistical evaluation considering non-suicidal vs. suicidal patients and suicidal non-repeaters vs. repeaters separately. Moreover, the different classification approaches, including ensemble, Support Vector Machines (SVMs) and a combination of both, are tested. The classification determined whether the patient is suicidal or not and then decided if the suicidal patient attempted suicide before or not. The results are then discussed and separated into the outcomes of the statistical evaluation and of the classification. The final section concludes with a summary of the master thesis' observations and a look to future work.

# 2. Methods

In this section, the used dataset will be described including the applied questionnaires and how many and which adolescents were considered. Moreover, the investigated features will be introduced including conversational, verbal and acoustic information. Furthermore, the used statistical evaluation methods are explained as well as the structure of the hierarchical classification. Finally, the used statistical parameters to evaluate the classification performance are mentioned.

## 2.1. Dataset

Within a controlled trial from March 2011 through October 2011, 60 interviews with 30 non-suicidal and 30 suicidal adolescent patients from the CCHMC ED were recorded. Thirty male and thirty female adolescents were interviewed by one single male clinician and asked to respond to 16 questions comprised of the Columbia Suicide Severity Rating Scale (C-SSRS version 1/14/2009 [PBS+11]), Suicidal Ideation Questionnaire-Junior (SIQ-JR version 1987 [RM99]) and the Ubiquitous Questionnaire (UQ version 2011[Pes10]). For the study, 60 adolescent patients between the ages of 13 and 18 were identified from the hospital's electronic medical records as potential participants (average age of 15.47 years with $\sigma$=1.5). The bar plot in Figure 2-1 illustrates the number of participants separated by ages. The x-axis corresponds to the age groups, the y-axis plots the number of participants. Four 13-year-old, ten 16-year-old, ten 17-year-old and seven 18-year-old adolescents participated in the study. The major age groups were 14-year-old and 15-year-old, fifteen and fourteen, respectively. As potential subjects, 30 patients were chosen that had come to the ED with suicidal ideation, gestures or attempts.

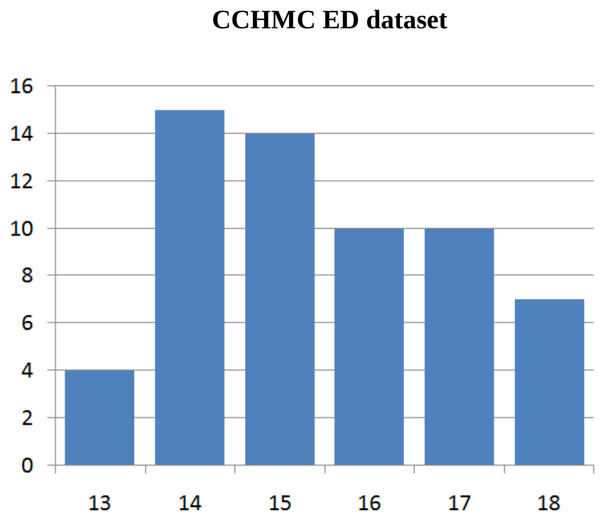


Figure 2-1: Age distribution of the participants of the CCHMC ED dataset

Thirteen *suicidal repeaters* were identified in the CCHMC dataset due to their total number of actual suicidal attempts and their total number of actual attempts in the past six months. If one of these two parameters were larger than one, the subject was categorized as a repeater. Seven of the adolescents were male and six were female adolescents between the age of 14 and 18. The remaining 17 suicidal adolescents were categorized as *non-repeaters*. Their potential controls, also referred as *non-suicidal*, were patients with orthopedic injuries due to the fact that they are seen as having the fewest biological and neurological perturbations of all of the ED patients. Furthermore, they were omitted from the study if they had a history of major mood disorder or if first-degree family members had a history of suicidal behavior. The distribution of the dataset to the three possible statuses non-suicidal, suicidal repeater and non-repeater are illustrated in Figure 2-2. The pie chart depicts the interviews of the provided dataset. Thirty non-suicidal adolescents, seventeen suicidal non-repeaters and thirteen suicidal repeaters participated in the study.

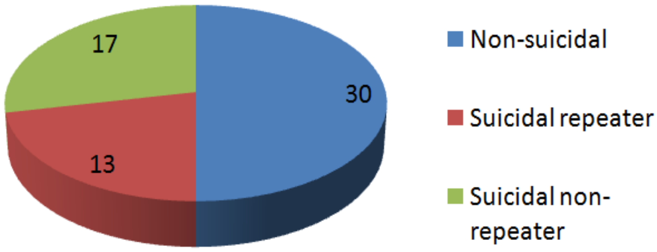


Figure 2-2: Structure of the CCHMC ED dataset labeled with number of interviews

The participation of the patients had to be consent by their parent(s) or legal guardian(s) and him- or herself. Furthermore, they had to be verified as appropriate for the study by the attending physician(s). Each patient received $75USD compensation for participation. The interviews were audio recorded in a private examination room using one single tabletop microphone. Hence, the speech segments including the voice utterances of the clinician and the patient on the single mono channel of the recordings were manually annotated. The average signal-to-noise ratio of the audio sampling was 17.2 dB at 16kHz. Moreover, all interviews were transcribed on a question-response level by using ELAN annotation software[1][SW08].

In general, all the interviews with suicidal patients lasted longer than those with the control ones. The mean duration of the interviews with suicidal patients was 869 seconds (14.5 minutes). In comparison, the average length of the interviews with the controls were almost halved: interviews lasted approximately 491 seconds (8.2 minutes). The mean duration of the interviews with the suicidal repeaters was approximately 856 seconds (14.3 minutes) while the interviews with the suicidal non-repeaters lasted on average 879 seconds (14.7 minutes).

## 2.2. Investigated Features

In this sub-section, investigated features are introduced which were obtained by analyzing the interviews' transcripts and acoustic feature data. The audio-based characteristics were divided into three sub-groups: the conversational information, verbal information and acoustic information features. For the feature extraction Matlab software version R2013a 8.1.0.604 was used to determine the conversational information, Linguistic Inquiry and Word Count (LIWC) 2007[2] [PCI+14] to analyze the transcripts to obtain verbal information and the

[1] http://tla.mpi.nl/tools/tla-tools/elan/
[2] http://www.liwc.net/index.php

Collaborative Voice Analysis Repository (COVAREP)[3] was used to acquire the acoustic features. The transcripts of the recorded interviews were annotated by using ELAN annotation software. They included time references and content of the spoken words of patients and clinician.

### 2.2.1. Conversational Information

The transcripts of the recorded interviews were obtained by using ELAN annotation software which transcribed the interviews including time stamps. These time references marked the start and the end time of the patients' and the clinician's speech patches. Also the sixteen questions including the five open-ended ubiquitous questions [Pes10] were identified by time stamps. The conversation dynamic information features were extracted from the transcribed interviews by using Matlab.

Thus, the speaking times of the patients and clinician were able to be determined as well as many other conversational information characteristics including, for example, respond times to the ubiquitous questions. The time information features with their units extracted from the interviews' transcripts were:

- Total interview duration in seconds
- Total number of words
- Respond times to each ubiquitous question in seconds
- Summed respond times for ubiquitous questions in seconds
- Overlap rate in number per second

Furthermore, there were features determined for clinician and patients separately:

- Speak time in %
- Pause time in %
- Words per second rate
- Interview duration in seconds
- Number of words
- Pause duration in seconds
- Number of breaks
- Number of spoken words
- Clinician-speaks-over-patient rate in number per second
- Patient -speaks-over-clinician rate in number per second

Especially the descriptors with percentage and rate units were considered to be of importance for the classification because they were likely to be more consistent to variability between the interviews with non-suicidal and suicidal patients. In particular the *speak time*, *pause time percentages*, *words per second rate* and the different *overlap rates* were of interest for the classification.

[3] http://covarep.github.io/covarep/

### 2.2.2. Verbal information

The verbal information features were determined by analyzing the transcript data using LIWC2007 software [PCI+14]. This tool is designed to quickly and efficiently analyze written or transcribed verbal text sequentially. Hence, the contents of the answers of the patients as well as of the questions and reactions of the clinician were able to be analyzed separately. The output of this analysis tool are word category scales related to 80 categories. Thus, verbal information was possible to be determined from the interviews between patients and clinician. LIWC2007 simultaneously compares words in the transcripts and words existing in the LIWC2007 Dictionary. The LIWC2007 Dictionary consists of nearly 4,500 words and word stems. If a word is found in the dictionary the particular word category or categories are incremented. Each word or word stem can be related to one or more word categories. One example is given by [PCI+14]: *"the word 'cried' is part of five word categories: sadness, negative emotion, overall affect, verb and past tense verb"*.

Among others, the provided LIWC2007 output features are standard linguistic dimensions like *personal pronouns*, *first person singular pronouns* e.g. 'I, my, mine', *impersonal pronouns* and terms indicating *past tense* and *negation*. Moreover, the word categories related to *positive emotion* and *negative emotion* were investigated. Also *tentative* words like 'maybe, perhaps' or 'guess' were observed. The paralinguistic dimensions *nonfluencies* like 'er, hm, umm' and *assent* words like 'agree, okay, yes' were investigated. An entire list of the word categories and examples can be found in [Pen14]**.**

The used word categories mentioned in [SP01] were especially considered for the statistic evaluation and subsequently for the classification task. Stirman and Pennebaker [SP01] investigated the word use in the poetry of non-suicidal and suicidal poets by performing LIWC analyses on the poets' works. Their social integration theories stated that suicidal poets used more references to themselves and used fewer words related to others. For example, suicidal poets used more self-related words like 'I' or 'my' as well as they spoke less about their families or friends. Due to the proved self-preoccupation of suicidal risk patients [SP01] this was not surprising. Dr. Pennebaker had already given an overview of what the content of speech or rather especially the usage of pronouns could tell about emotional states of individuals in [Pen11].

The hopelessness models were another approach in Stirman's and Pennebaker's work. Suicide can be often related to hopelessness [PB92, Shn98], hence, [SP01] analyzed the usage of negative emotion terms and words related to death. They were able to prove the significant use of terms related to death but there were not any significant results regarding the use of negative emotion terms. Other relevant LIWC features determined by them were a higher use of references to the past in the suicidal group. The use of past tense was able to be referred to physical pain or sadness, both suicidal risk factors [Shn98, SP01].

The list including the investigated LIWC categories and examples is given in Table 2.

Table 2: LIWC features investigated by Stirman and Pennebaker with LIWC examples of the word categories [SP01]

| | Examples |
|---|---|
| Social integration theory | |
| 1st pers singular | I, me, mine |
| 1st pers plural | we, us, our |
| Communication | talk, share |
| Hopelessness theory | |
| Positive emotion | love, nice, sweet |
| Negative emotion | hurt, ugly, nasty |
| Death | bury, coffin, kill |
| Other relevant data | |
| Past tense | went, ran, had |
| Sexual | horny, love, incest |

### 2.2.3. Acoustic information

The acoustic information provides several benefits to investigate voice quality including the determination of phonation types [SAA+14, SHY+14]. As already shown in several studies the phonation type provides a certain ability to distinguish between depressed and suicidal patients [OSS+04, FSS+00].

The freely available COVAREP toolbox was used to obtain the acoustic information features by processing the speech signals. COVAREP is a collaborative speech analysis repository available for Matlab and Octave [DKD+14]. The toolbox offers an extensive selection of open-source robust and tested speech processing algorithms enabling comparative and cooperative research within the speech community [SPM13].

In this study, the voice source and voice quality-related features of each interview were sampled at 100 Hz. Therefore, for each second of the interview acoustic features were provided. To determine the corresponding patches of clinician and patients within each interview the information provided by the time stamps of the transcripts were again used and the logical feature Voice-Uttering-Voice (*VUV*) was used to determine actual speaking.

After separating the data into patches corresponding to patients and clinician, the features were analyzed. Furthermore, the features of the backchannel of the clinician were extracted due to findings in [SPM13]. The backchannel was comprised of speech patches of the interviewer with durations smaller than 700 milliseconds. These patches included words of assent, nonfluencies or also fillers like 'uhm'.

Acoustical measures, which have been useful in several studies [Kan12, SHY+14] to characterize the voice quality from breathy to tense dimension, are described below. The abbreviations in the parentheses following the feature names will be used to refer to them throughout this thesis.

- The *fundamental frequency* ($f_0$) is the base frequency of the speech signal [FSO+12]. It includes the pitch information of individual's speech. The method for a $f_0$ tracking and simultaneous voicing detection based on residual harmonics is introduced in [DA11]. Unvoiced speech segments, i.e. times when no vocal fold vibration appears, were not analyzed for any of the extracted features.

- The *normalized amplitude quotient* (NAQ) and *quasi-open quotient* (QOQ) are both derived from amplitude measurements of the glottal source signal estimated by iterative adaptive inverse filtering (IAIF [ABV92]). The NAQ describes the normalized amplitude quotient of the differentiated glottal flow. The QOQ is measured by detecting the peak in the glottal flow and finding the time points previous to and following this point that descent below 50% of the peak amplitude. The duration between these two time points is divided by the local glottal period to get the QOQ measure. [SHY+14]

- The *parabolic spectral parameter* (PSP) is derived by fitting a parabolic function to the lower frequencies in the glottal flow spectrum. The result of the computation estimates how the spectral decay of an obtained glottal flow behaves with respect to a theoretical limit corresponding to maximal spectral decay. The PSP allows a comparison of glottal flows in terms of their spectral decays, even when fundamental frequencies of voices is different. [ASV97]

- The *maxima dispersion quotient* (MDQ) and the *peak slope* (PS) are generated by using a dyadic wavelet transform employing a cosine-modulated Gaussian pulse as the mother wavelet. Maxima of the speech signal are measured across the scales, on a fixed-frame basis, and a regression line is fit to these maxima which provides the PS measure. The feature is essentially an effective correlate of the spectral slope of the signal. Among others, using the glottal closure instants (GCI) the dispersion of peaks in relation to the GCI position is averaged across different frequency bands and then normalized to the local glottal period which yields the MDQ parameter. For tense voice, where the sharp closing of the glottis is analogous to an impulse excitation, the maxima are tightly aligned to the GCI, whereas for laxer phonation the maxima become highly dispersed. [Kan12]

- The Liljencrants-Fant (LF) model parameter *Rd* is one of the R-parameters of the LF model characterizing the glottal source. *Rd* captures most of the covariation of the LF model parameters [FLL85]. Reference [Obi12] has shown that this feature improved the classification of different levels of vocal effort from expressive speech significantly.

- The tracking of the formants is introduced in detail in [BDdA+04]. The *first* and the *second formants* (*F1, F2)* are the vocal tract resonance frequencies which describe the first two spectral peaks with the lowest frequencies of the speech signal. They identify and characterize primarily vowels [Lad96].

## 2.3. Statistical evaluation

The statistical evaluations of the feature groups were all executed using Matlab. Statistical evaluation methods aim to test statistically significances of data. For the statistical investigations of the features' behaviors between suicidal and non-suicidal adolescents as well as between repeaters and non-repeaters, one-way ANOVA was performed. The variation between two groups and within each group should be observed as well as the faculty of descriptors to characterize non-suicidal patients, suicidal repeaters and suicidal non-repeaters from each other. [Ler11]

In general, ANOVA describes the analysis of variances or rather of the means of several sample groups to measure a difference between these groups. Hence, significance of group differences should be revealed. The one-way ANOVA investigates the significant difference of the dependent variables each at a time among the samples considering one independent variable. The assumptions for the statistical test are normality, equal variance and independence of errors. The description and characteristics of the one-way ANOVA can be read in chapter 3 of [Ler11] and in [SS11].

Regarding the ANOVA null hypothesis, all sample means are meant to be equal:

$$H_0 : \mu_1 = \cdots = \mu_N \tag{1}$$

This null hypothesis is tested by ANOVA, i.e. it decides whether the overall ANOVA null hypothesis will be retained or rejected in the case of two levels. If the ANOVA yields a significance, i.e. the means differ from each other, then the ANOVA null hypothesis can be rejected. However, the ANOVA does not depict which means of the single sample groups differ from each other. This requires post hoc tests which were not used throughout this thesis. [Ler11]

The heart of the ANOVA is the F-test. Therefore, the F-value is computed:

$$F = \frac{Between - group\ variance}{Within - group\ variance} \tag{2}$$

which describes the ratio of *between-group variance* (also called effect variance) and *within-group variance* (also called error variance). The higher the F-value the better to find significant differences within the dataset. The ANOVA observes how unpredictable a F-value is.

After obtaining the F-value, this value can be located in the F distribution which is a probability density function. The F distribution describes a group of distributions with varying degrees of freedom as parameters which determine the shape of the function. The F-values are found on the x-axis and the probability density to observe a certain F-value are illustrated on the y-axis. The F distribution is an asymmetric function with a right exceeding line which asymptotes out to the x-axis. As a probability density function the total area under the F distribution equals one and describes a probability value, the p-value. The p-value is the probability that a located F-value can be observed. Looking at Figure 2-3, it is obvious why larger F-values are better for the significance of the statistical hypothesis test. The larger the F-value, the smaller the p-value due to the asymmetric shape of the F distribution. The smaller the p-value, the more significant is the ratio of the two variances (compare (2)). As already mentioned above, the ANOVA compares the ratio of variances. In the left graph the area under the F distribution corresponds to the p-value of observing a F-value of 1, the right one to the p-value of observing a F-value of 4. The smaller and, hence, more significant p-value of these observations is the one with the F-value of 4. [Pac12]

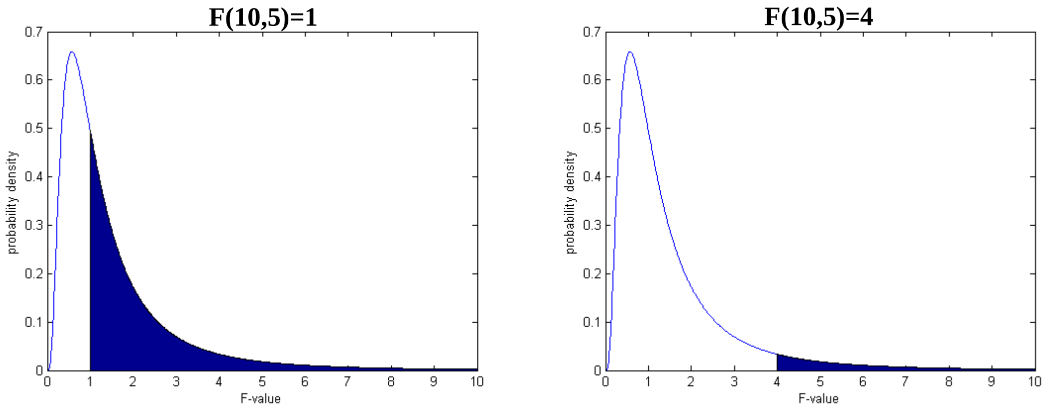


Figure 2-3: F distribution with degrees of freedom (df1=10, df2=5). Left: probability of observing F-value of 1; Right: probability of observing F-value of 4.

The p-value can be compared to the significance level α, i.e. whether an ANOVA result is significant or not. "*The significance level of a hypothesis test, denoted by α, is the probability of rejecting the null hypothesis* $H_0$ *when it is true*" [Dod08]. Commonly, p-values smaller than 5% are called statistically significant due to α (also named Type I error) valued at 0.05 [SS11].

Furthermore, the analysis of the interviews was performed in two ways: First, the features extracted from the entire interview were taken for the ANOVA. This case will be referred in this work as the *complete case*. In this case, interactions with non-suicidal adolescents took on average 491 seconds (8.2 minutes), with suicidal re-attempting patients 856 seconds (14.3 minutes) and with suicidal non-repeaters approximately 879 seconds (14.7 minutes). The participants were asked sixteen questions during the interviews.

Second, the features derived from the section of the interview in which the five ubiquitous questions were asked and answered were analyzed. This was decided especially due to the standardized ubiquitous questions which promised balanced significant analysis results, and due to the reduced duration of the interview. In this thesis, this will be referred as the *UQ case*. Here, the interviews' parts lasted on average with non-suicidal controls 286 seconds (4.8 minutes), with suicidal repeaters 521 seconds (8.7 minutes) and with suicidal non-repeaters 554 seconds (9.2 minutes). In the UQ part of the interviews five open-ended questions were asked. They are mentioned in Table 3.

Table 3: The five open-ended questions which were asked in the first partition of each interview defining the *UQ* case

| | |
|---|---|
| **Q1** | Does it hurt emotionally? |
| **Q2** | Do you have any fear? |
| **Q3** | Are you angry? |
| **Q4** | Do you have any secrets? |
| **Q5** | Do you have hope? |

## 2.4. Hierarchical classification

Subject independent classification was performed to confirm and identify the discriminative faculty of the investigated features. The task of the classification between non-suicidal, suicidal non-repeaters and suicidal repeaters was realized by separating the discrimination into two layers. The used algorithm to develop the hierarchical classifier in each layer and architecture of the classification will be described in section **2.4.1**. In addition, the evaluation of the classification performance will be discussed in section **2.4.2**.

### 2.4.1. Structure of the hierarchical classifier

In the first layer of the hierarchical classifier, a classifier discriminates non-suicidal from suicidal patients. The classifier was trained and tested with a classification matrix containing 41 investigated features of the 60 interviews with the adolescents. The features included conversational information, verbal information and acoustical information.

In the second layer, another classifier was trained and tested by using 18 features. The aim of this classification was to distinguish suicidal non-repeaters and suicidal repeaters. The positively labeled patients from the previous layer were forwarded to the non-repeaters vs. repeaters level. Different to the first layer's training, the classification feature matrix primarily comprised of acoustical features of the clinician and patients.

Three approaches were investigated using the hierarchical classification. First, the *hierarchical ensemble classifier* was trained and tested, the same was performed with a *hierarchical SVM classifier*. Furthermore, a hierarchical classifier was employed which used the SVM algorithm in the first layer to distinguish between non-suicidal and suicidal adolescents. In the second layer, the ensemble classifier was used to classify suicidal non-repeaters and repeaters. The third classifier will be referred as *hierarchical SVM-ensemble classifier*. In Figure 2-4 the general hierarchy of the classification is illustrated. In the upper classification layer the classifier predicts whether the patient is non-suicidal or suicidal. The second classifier in the bottom layer decides then if the suicidal patient is a non-repeater or a repeater.

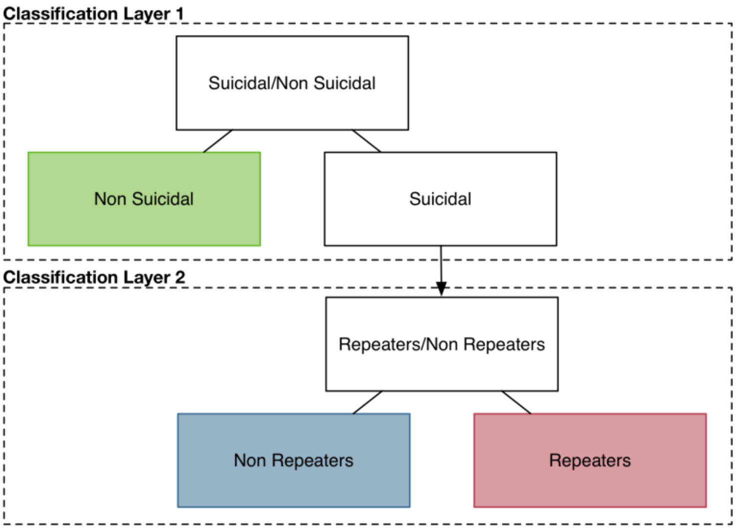


Figure 2-4: Structure of the hierarchical classifier

In the following paragraphs, the two machine learning algorithms, the ensemble boosting and the SVM algorithm, will be explained. Matlab was used to design the three hierarchical classifiers. The testing of the hierarchical classifiers was performed with a leave-one-speaker-out approach.

- ***Ensemble boosting***

Due to the high popularity as boosting algorithm the used ensemble algorithm was decided to be the *AdaBoostM1* algorithm which enhances a set of weak classifiers to a powerful ensemble. The AdaBoostM1 [FS97] algorithm describes a combination of weak learners which are trained by sequentially minimizing a certain loss function. In the binary case, this would be the exponential loss

$$\sum_{n=1}^{N} w_n e^{-y_n f(x_n)} \tag{3}$$

where $w_n$ are the original passed normalized observation weights, $y_n$ is the true class label (i.e. {0, 1}) and $f(x_n)$ is the predicted classification score. For each learner from index *t*=1...40 the weighted classification error is calculated by AdaBoostM1:

$$\varepsilon_t = \sum_{n=1}^{N} d_n^{(t)} I(y_n \neq h_t(x_n)) \tag{4}$$

where $d_n^{(t)}$ is the weight of the $n^{\text{th}}$ interview of the weak classifier *t*. *I* is the indicator function and $y_n$ describes the true class labels. Moreover, $x_n$ is a vector of predictor values for observation n, and $h_t$ describes a hypothesis or rather the predicted output of the learner with index t. The weights are updated iteratively providing each time a different distribution over the training set (i.e. weak learners receive sequentially an updated distribution). This base learning algorithm aims to find the hypothesis $h_t$ with small error $\varepsilon_t$. [SFB+97, HTF08]

The combination of the trained learners provides the prediction function for new data

$$\boldsymbol{f}(\boldsymbol{x}) = \sum_{t=1}^{T} \boldsymbol{\beta_t h_t}(\boldsymbol{x}) \tag{5}$$

with

$$\boldsymbol{\beta_t} = \frac{\mathbf{1}}{\mathbf{2}} \mathbf{log} \frac{\mathbf{1} - \boldsymbol{\varepsilon_t}}{\boldsymbol{\varepsilon_t}} \tag{6}$$

The variable $\beta_t$ describes weak hypotheses in the ensemble. [FS97]

The testing of the classifier was performed with a leave-one-speaker-out approach. Thus, subsets of the existing data were generated by using the $n^{th}$ interview for testing and the *N-1* interviews for training of the classifier. After *N* loops, the confusion matrix was determined for the classifier.

In the hierarchical classification, the ensemble classifier was used to distinguish whether non-suicidal and suicidal patients in the first layer and/or suicidal non-repeaters and suicidal repeaters in the second one. As weak learners or rather subject independent classifiers 40 decision trees were selected for both layers.

- ***Support Vector Machine***

Beside the ensemble distinction approach, SVM algorithms were used. The SVM classification was realized by using the Matlab version of the *libsvm.m* toolbox from the LIBSVM from [CL11]. LIBSVM is a library of SVMs. The SVMs can be employed as a classifier which uses linear algorithms. In the case of a binary classification, the task is to find a hyperplane which separates the two classes with the maximal possible margin.

The SVM algorithm aims to solve the following optimization problem:

$$\min_{\boldsymbol{w},b,\xi} \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w} + C \sum_{i=1}^{l} \xi_i \tag{7}$$

$$\text{subject to } y_i(\boldsymbol{w}^T \Phi(\boldsymbol{x}_i) + b) \geq 1 - \xi_i, \tag{8}$$

$$\xi_i \geq 0$$

in which $(\boldsymbol{x}_i, y_i)$, $i=1,...,l$ describe a training of instance-label pairs where $\boldsymbol{x}_i \in R^n$ and $\mathbf{y} \in \{1,-1\}$. The training vectors $\mathbf{x}_i$ are mapped to a higher dimension by the function $\Phi$. In this dimension the hyperplane with the maximal margin is investigated by the SVM. For the mapping the SVM uses kernel functions, looking like this:

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \Phi(\boldsymbol{x_i})^T \Phi(\boldsymbol{x_j}) \tag{9}$$

Among others, the radial basis function (RBF) is used as kernel function with γ as kernel parameter:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{(-\gamma||\boldsymbol{x}_i - \boldsymbol{x}_j||^2)}, \qquad \gamma > 0 \tag{10}$$

The proposed procedure applying a SVM with the LIBSVM toolbox is given by [HCL03]. First of all the data should be transformed into a SVM package. Then the data should be scaled linearly to a normalized range, for example [0,1]. This normalization has the advantage to weight smaller numeric ranges equally to greater ones which avoids descriptor dominance. Another advantage is that numerical difficulties during the calculation can be avoided. [HCL03]

Afterwards the RBF kernel can be considered with $K(\boldsymbol{x}, \boldsymbol{y}) = e^{-\gamma||\boldsymbol{x}-\boldsymbol{y}||^2}$. Especially if the relation between class labels and features is nonlinear, it is beneficial that this kernel nonlinearly maps samples into a higher dimensional space. In comparison to other possible kernels, e.g. the polynomial kernels, the RBF has fewer numerical difficulties. [HCL03]

The next step is to find the best parameters *C* and *γ* for the kernel. This is realized by using cross-validation. "*The aim is to find satisfying kernel parameters so that the classifier can accurately predict testing data"* [HCL03]. The cross-validation describes a partitioning of the existing data into *k* subsets. Each subset is used for training of the classifier and the residual partition for testing it. Hence, each interview of the entire training set is predicted once so the cross-validation accuracy describes the percentage of data which are correctly labeled. To find the best kernel parameters *C* and *γ*, the grid-search using cross-validation is an approach. Various pairs of (*C*, *y*) values are tested and the one with the best cross-validation accuracy is picked. [HCL03]

After the best kernel parameters were determined, the classifier could be trained with the entire training set using the best *C* and *γ*. The last step was to test the classifier. Due to the leave-one-speaker-out approach the existing data were separated by using the $n^{th}$ interview for testing and the *N-1* interviews for training of the classifier. After *N* loops, the confusion matrix was determined for the hierarchical classifier.

### 2.4.1. Evaluation of the hierarchical classifier

After testing the hierarchical classifiers, measures to assess the classification performance were determined. One of these measures are the accuracies of the hierarchical classifiers which were computed by using the elements of the confusion matrix. The confusion matrix contains the *true positives (TP)*(i.e. suicidal patients or repeater), *true negatives (TN)* (i.e. non-suicidal patient or non-repeater), *false positives (FP)* and *false negatives (FN)* of the classification. The accuracy can be determined by using the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

After training and determining the classification accuracies of the two classification layers separately, an overall accuracy result over the hierarchical classifier was obtained.

For the evaluation of the hierarchical classification results the *recall*, the *precision* and the *F measure* (or *F score*) of the confusion matrices were employed. These ratios assess also the classification performance of a machine learning algorithm.

The recall is described as follows:

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

The true positives of the confusion matrix are divided by the total number of actual positives. The recall also described as true positive rate or sensitivity *"is the fraction of positive examples predicted correctly by a model"* [SW10].

The precision can be computed by using the true positives and the total number of predicted positives:

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

Hence, the precision can be referred to the positive predictive value.

The F measure is a ratio which combines precision and recall. It weights precision and recall evenly:

$$F\ measure = \frac{2 * (Precision * Recall)}{Precision + Recall} \tag{14}$$

According the interpretation of the measures it is similar to the one of the accuracy: the closer to the value one the better. The formula and descriptions of the parameters can be read in chapter 2 of [Ler11].

Due to the distinction of three different classes in the hierarchical classification, the calculations of the statistical measures to evaluate the classification performance were performed by using the multi-class formula.

Furthermore, classification results were computed by using different datasets as input of the hierarchical classifiers. First, all the investigated data including patients' features, clinician's features and clinician's backchannel's features were forwarded to the machine learning algorithm. Moreover, the hierarchical classification accuracy by using only the patients' data was investigated. Afterwards, the outcome of the patients' and the clinician's features was determined. Finally, another classification result was obtained by using the patients' and only the clinician's backchannel's descriptors. Also these results were evaluated by using accuracy, confusion matrix, recall, precision and F measure.

# 3. Results

The outcome of the two main tasks are given in the following sections. First, the results of the statistical evaluation of the non-suicidal vs. suicidal patients will be shown and afterwards the significant differences of the suicidal non-repeater vs. repeater are mentioned. In both investigations the results of the *complete* and the *UQ* case will be given. Secondly, the classification performances using different classification methods in the two layers of the hierarchical classifier will be mentioned. Furthermore, the impact of patients' and clinician's features to the performance will be evaluated.

## 3.1. Statistical evaluation

The results of the statistical evaluation of the features are listed in this section including significances and graphical representations. Two cases were statistically analyzed by using ANOVA: non-suicidal vs. suicidal adolescents with $N=60$ and suicidal non-repeaters vs. repeaters with $N=30$. The significance level was stated to be at least $p < 0.05$. The audio-based features were divided into three sub-groups: the conversational information, verbal information and acoustic information features.

In the given result tables the means, the standard deviations and the p-values are represented. Significant p-values smaller than 0.05 are marked as * and p-values smaller than 0.01 as **. In the boxplots the red line indicates the medians of the distribution of the different features. It also shows the spread and the symmetry of the data. Hence, it is a graphical method to statistically compare different datasets.

The statistical results corresponding to the clinician's speech or patients are specified with subscripts $_C$ and $_P$, respectively. If there is not an obvious affiliation neither to clinician nor to patient no subscript is used. Subscripts are not used for the backchannel's features of the clinician.

### 3.1.1. Non-suicidal vs. suicidal evaluation

This sub-section includes the statistical results of the ANOVAs of the investigated features between suicidal patients and their controls. The two tables represent the statistically significant descriptors of the patients (Table 4) and of the clinician (Table 5) for the *complete* case including means, standard deviations and the p-values of the ANOVAs. For the *UQ* case the significant features are listed with their statistical characteristics in Table 6 corresponding to the patients and Table 7 refers to the clinician's significant descriptors.

#### 3.1.1.1. *Complete* case

***Conversational information***

The words per second rate of the clinician was higher at interviews with non-suicidal adolescents ($\mu_C$= 2.94 words per second) than the one with the suicidal subjects ($\mu_C$= 2.78 words per second, $p_C$<0.01). The speak time given in percentage was significantly different. The suicidal patients spoke 43% on average of the entire interview, while the non-suicidal adolescents occupied 32% of the interview ($p_P$<0.01). The interviewer spoke 31% when interacting with a suicidal adolescent and 46% with a non-suicidal one ($p_C$<0.01), see

Figure 3-1. In the left boxplot the distributions of the patients' feature speak time is illustrated. The suicidal patients spoke more than their controls. The right boxplot illustrates the conversational feature speak time of the clinician. The clinician spoke less during interviews with suicidal patients than with the non-suicidal patients.

Also the pauses between the speech segments were considered: the suicidal adolescents paused on average 14% of the interview, the non-suicidal ones 8% ($p_P$<0.01). In the non-suicidal interviews, the clinician protruded with the mean pause time percentage of 15% ($p_C$<0.05). Patients spoke less over their interlocutor than the clinician did. Thus, the clinician-speaks-over-patient rate showed a significant difference ($p_C$<0.01), as it is represented in Figure 3-2. In this graph, it is shown that the clinician interrupted his interlocutors more often while interacting with non-suicidal patients than with suicidal ones.

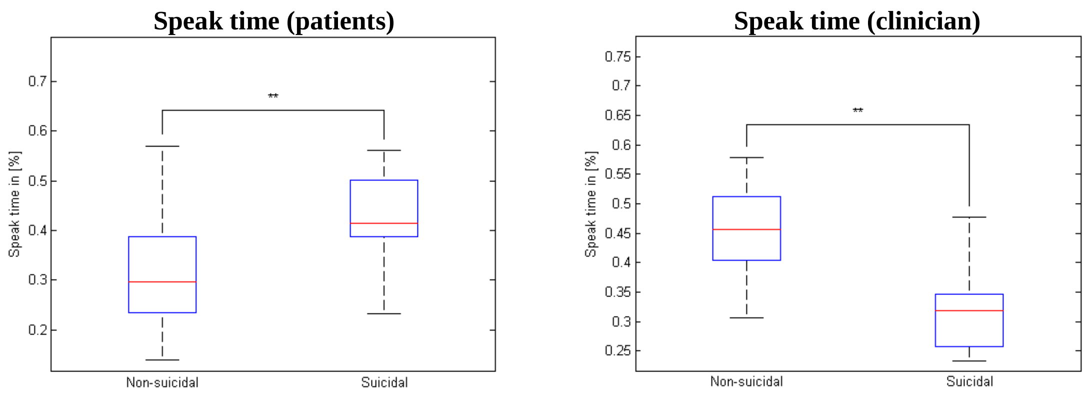


Figure 3-1: Non-suicidal vs. suicidal patients' (left) and clinician's (right) speak time in percentage (*complete* case)

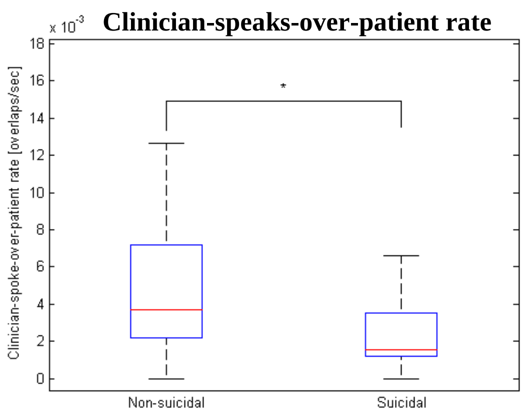


Figure 3-2: Non-suicidal vs. suicidal clinician-speaks-over-patient rate in overlaps per second (*complete* case)

***Verbal information***

For the analyses related to the verbal information, the data was separated into clinician's and patients' feature datasets. Suicidal patients used more often personal pronouns than their controls ($p_P$<0.01), especially self-related first person singular pronouns ($p_P$<0.01). Therefore, see Figure 3-3. This boxplot shows that the suicidal patients referred more often to themselves than the non-suicidal ones did.

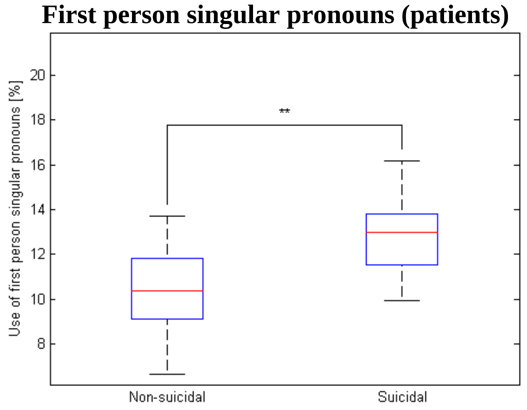


Figure 3-3: Non-suicidal vs. suicidal use of first person singular pronouns of the patients (*complete* case)

Moreover, suicidal patients referred to the past 3.7% of the total interview while non-suicidal adolescents did so 2.1% on average ($p_P$<0.01). Also the clinician referred more often to the past while speaking to suicidal patients ($p_C$<0.01). Adolescent controls used more often assent words than suicidal subjects ($p_P$<0.01). Nonfluencies were observed in controls' interviews ($p_P$<0.01) more often. The clinician used more nonfluencies during interviews with the suicidal patients ($p_C$<0.05), which can be observed in Figure 3-4. The left boxplot shows that nonfluencies were used more often by the non-suicidal patients than by the suicidal ones. The clinician used more words referred as nonfluencies during interviews with the suicidal patients than with their controls as illustrated in the right boxplot.

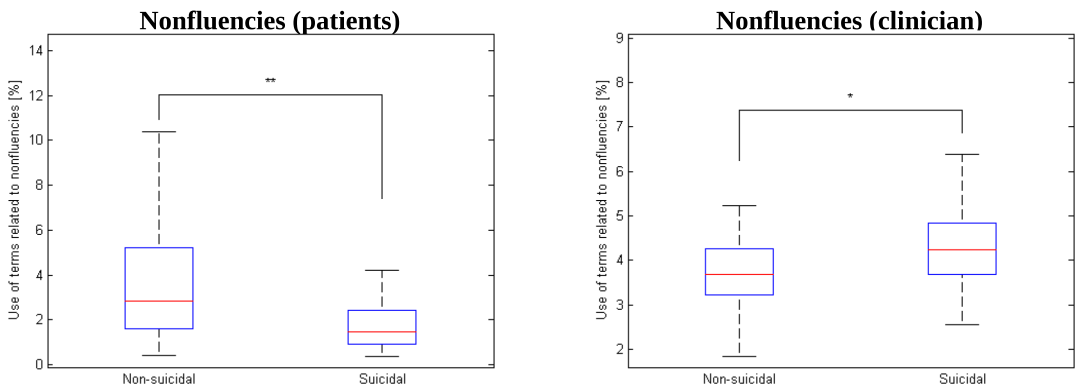


Figure 3-4: Non-suicidal vs. suicidal use of terms related to nonfluencies of the patients (left) and of the clinician (right) (*complete* case)

Non-suicidal patients used more often terms related to negation than the suicidal subjects ($p_P$<0.01). Suicidal adolescents used terms related to negative emotion more often than their controls ($p_P$<0.01) and non-suicidal patients used words correlated to positive emotion more often ($p_P$<0.01), as it is illustrated in Figure 3-5. The left boxplot represents that the suicidal patients used more often negative emotion words than the non-suicidal adolescents. The right one illustrates that the suicidal patients used less often positive emotion words than their controls did.

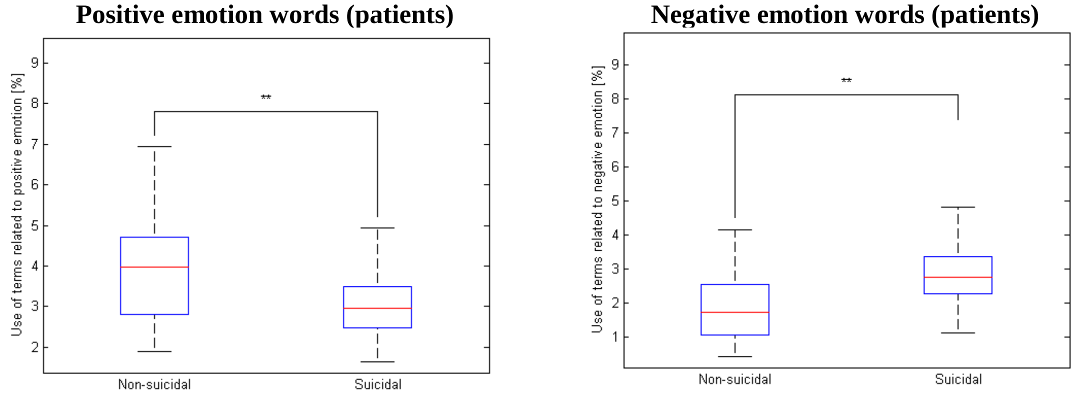


Figure 3-5: Non-suicidal vs. suicidal use of terms related to positive emotion of the patients (left) and use of terms related to negative emotion of the patients (right) (*complete* case)

Tentative terms were used 5.6% on average by non-suicidal adolescents, while suicidal patients used them on average 4.6% of the entire duration ($p_P$<0.01). Also the clinician used tentative words more often in interviews with non-suicidal patients ($p_C$<0.01). The use of tentative words of patients and clinician is depicted in Figure 3-6. The left boxplot shows that the suicidal patients used less often tentative terms than their controls. In the right one it is shown that the clinician used more often tentative words during interviews with non-suicidal adolescents than with suicidal patients.

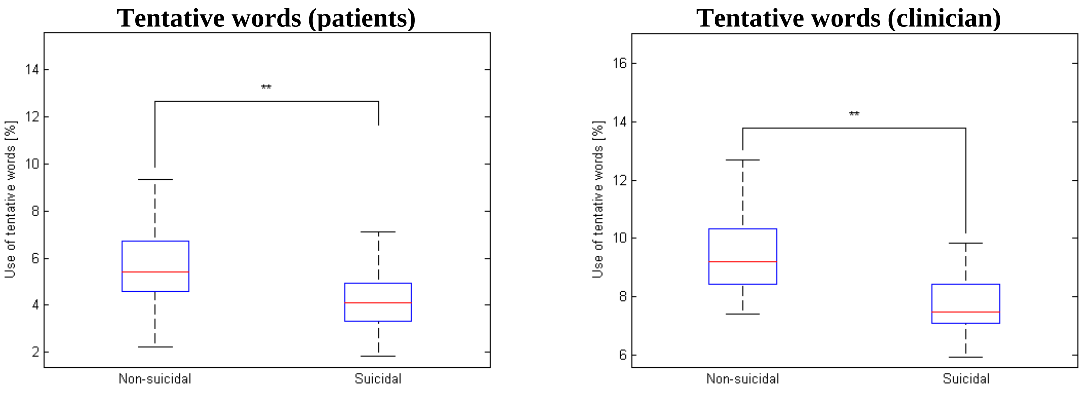


Figure 3-6: Non-suicidal vs. suicidal use of tentative words of the patients (left) and of the clinician (right) (*complete* case)

Furthermore, the use of impersonal pronouns showed a significance for clinicians and patients ($p_C$<0.01 and $p_P$<0.01, respectively). The clinician had a significant use of second person, third person singular and first person plural pronouns. Hence, the feature personal pronoun was also significantly and was selected as proper classification feature ($p_C$<0.01).

***Acoustic information***

The clinician's and patients' acoustic features were investigated separately. Suicidal patients spoke on average with a lower $f_0$ than their controls ($p_P$<0.01). The clinician showed the same pattern ($p_C$<0.01), as it can be observed in Figure 3-7. The left boxplot shows that the suicidal patients used a broader spectrum of $f_0$ than the non-suicidal adolescents. The right boxplot represents the $f_0$ of the clinician. A broader spectrum of $f_0$ of the clinician can be observed while interacting with suicidal patients.

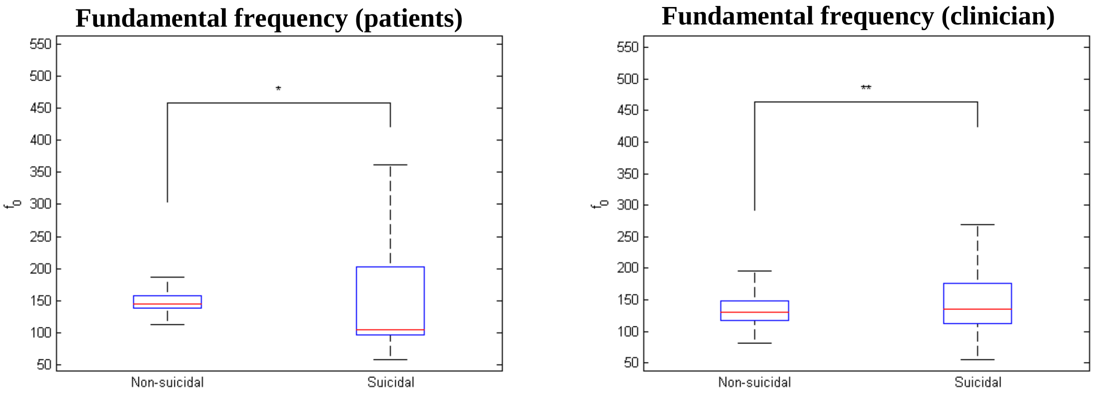


Figure 3-7: Non-suicidal vs. suicidal $f_0$ of the patients (left) and
$f_0$ of the clinician (right) (*complete* case)

The NAQ and QOQ measures were higher, i.e. the patients and clinician spoke with a breathier voice during interviews with suicidal patients. The QOQ distributions within the used dataset is represented in Figure 3-8. The left boxplot represents that the suicidal patients spoke breathier during the interviews than their controls. The right boxplot shows that the clinician spoke with breathier voice while interacting with suicidal patients.

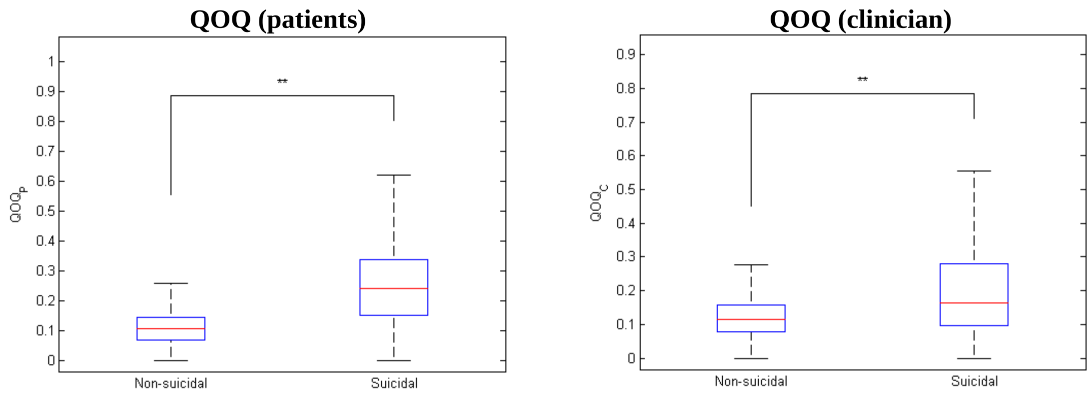


Figure 3-8: Non-suicidal vs. suicidal QOQ of the patient (left) and
QOQ of the clinician (right) (*complete* case)

The impression of a breathier phonation type was also observed for the MDQ parameter ($p_C$<0.01and $p_P$<0.01). Moreover, the PS measure implied a breathier conversation with suicidal patients ($p_C$<0.01, $p_P$<0.01). The PSPs were on average higher during the interviews with the non-suicidal adolescents ($p_C$<0.01, $p_P$<0.01). Similar to the measures already mentioned, the *Rd* parameter was higher during interviews with the suicidal patients ($p_C$<0.01, $p_P$<0.01). Investigating the significance of the formants, only the first formant *F1* was significantly different ($p_C$<0.01, $p_P$<0.01). Regarding the acoustic features of the clinician's backchannel, the significance of the PS result ($p_C$<0.01) showed that the clinician reacted with breathier voice during interviews with suicidal adolescents. As already investigated in the case of analyzing the complete interviewer's acoustic features, the NAQ and the QOQ were respectively higher during interviews with suicidal patients ($p_C$<0.01and $p_C$<0.01, respectively), see Figure 3-9. When comparing these two boxplots, it is obvious that the behavior of the NAQ of the clinician considering the entire interview is very similar to the NAQ distributions of the clinician's backchannel. Hence, differences of the interviews with suicidal and non-suicidal patients can be even investigated by just considering the acoustic information of the clinician's backchannel. Also the PSP was significantly different ($p_C$<0.01).

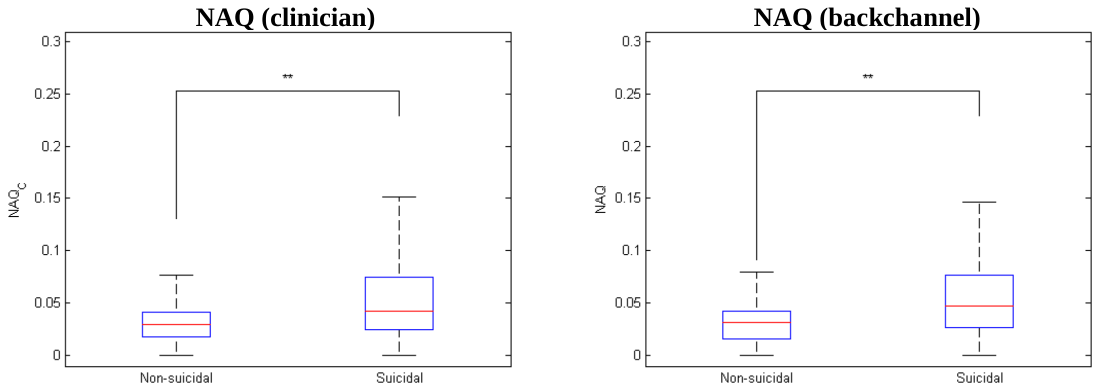


Figure 3-9: Non-suicidal vs. suicidal NAQ of the clinician (left) and NAQ of the clinician's backchannel (right) (*complete* case)

Table 4: Statistical significant results of the patients' features of the non-suicidal vs. suicidal evaluation with *N=60* (*complete* case)

| Feature | Suicidal μ (σ) | Non-suicidal μ (σ) | p-value |
|---|---|---|---|
| Speak time percentage | 0.43 (0.09) | 0.32 (0.1) | ** |
| Pause time percentage | 0.14 (0.04) | 0.08 (0.04) | ** |
| Personal pronouns | 16.72 (1.74) | 13.48 (2.5) | ** |
| 1st person singular pronoun | 12.73 (1.71) | 10.37 (1.9) | ** |
| Impersonal pronouns | 6.93 (1.18) | 5.60 (1.73) | ** |
| Past tense | 3.74 (1.71) | 2.07 (1.42) | ** |
| Negation | 4.05 (1.23) | 5.96 (2.02) | ** |
| Positive emotion | 3.03 (0.80) | 3.95 (1.25) | ** |
| Negative emotion | 2.95 (1.08) | 1.76 (0.88) | ** |
| Tentative | 4.35 (1.72) | 5.64 (1.70) | ** |
| Nonfluencies | 1.90 (1.38) | 3.81 (2.76) | ** |
| Assent | 1.97 (0.99) | 4.56 (3.05) | ** |
| $f_0$ | 220.82(25.10) | 150.62 (11.58) | ** |
| NAQ | 0.08 (0.02) | 0.03 (7.1e$^{-3}$) | ** |
| QOQ | 0.31 (0.07) | 0.11 (0.03) | ** |
| PSP | 0.36 (0.06) | 0.50 (0.09) | ** |
| MDQ | 0.14 (4.9e$^{-3}$) | 0.11 (0.01) | ** |
| PS | -0.20 (0.04) | -0.24 (0.03) | ** |
| *Rd* | 1.63 (0.16) | 1.10 (0.20) | ** |
| *F1* | 620.43(89.94) | 544.66 (121.3) | ** |

Table 5: Statistical significant results of the clinician's features of the non-suicidal vs. suicidal evaluation with *N=60* (*complete* case)

| Feature | Suicidal μ (σ) | Non-suicidal μ (σ) | p-value |
|---|---|---|---|
| Clinician-speaks-over-Patient | 2.49 (2.65) | 4.78 (4.07) | * |
| Words per second rate | 2.78 (0.22) | 2.94 (0.19) | ** |
| Speak time percentage | 0.32 (0.06) | 0.46 (0.07) | ** |
| Pause time percentage | 0.12 (0.04) | 0.15 (0.05) | * |
| Personal pronouns | 11.84 (0.82) | 11.23 (0.66) | ** |
| 1st person singular pronoun | 2.03 (0.44) | 2.30 (0.63) | 0.056 |
| Impersonal pronouns | 7.41 (1.06) | 6.29 (0.92) | ** |
| Past tense | 2.23 (0.73) | 1.19 (0.48) | ** |
| Tentative | 7.63 (0.97) | 9.32 (1.28) | ** |
| Nonfluencies | 4.31 (1.02) | 3.76 (0.89) | * |
| $f_0$ | 194.16(36.59) | 137.07 (24.66) | ** |
| NAQ | 0.07 (0.02) | 0.03 (8.1e$^{-3}$) | ** |
| QOQ | 0.26 (0.05) | 0.11 (0.03) | ** |
| PSP | 0.33 (0.08) | 0.48 (0.11) | ** |
| MDQ | 0.13 (8.1e$^{-3}$) | 0.11 (0.01) | ** |
| PS | -0.23 (0.04) | -0.25 (0.03) | ** |
| *Rd* | 1.62 (0.20) | 1.19 (0.23) | ** |
| *F1* | 622.79(127.43) | 596.09 (155.2) | ** |
| **Backchannel**: | | | |
| NAQ | 0.07 (0.02) | 0.03 (0.01) | ** |
| QOQ | 0.26 (0.07) | 0.10 (0.04) | ** |
| PS | -0.24 (0.04) | -0.25 (0.03) | ** |

#### 3.1.1.2. *UQ* case

***Conversational information***

The speak time of the patients was statistically significant ($p_P<0.01$). Suicidal patients spoke more than their controls. The suicidal patient also paused more than the non-suicidal adolescents ($p_p<0.01$). The patient-speaks-over-clinician rate showed significant difference ($p_P<0.05$) as well as the clinician-speaks-over-patient rate ($p_C<0.05$), see Figure 3-10. In the left boxplot it can be seen that the non-suicidal patients spoke more over their interlocutor than the suicidal patients did. The right boxplot shows that the clinician spoke less over the suicidal patients than over the non-suicidal ones during the interviews.

Summarized, the patients as well as the clinician spoke over each other more often in the non-suicidal case. The clinician spoke less during interviews with suicidal adolescents ($p_C<0.01$) and paused also less ($p_C<0.05$).

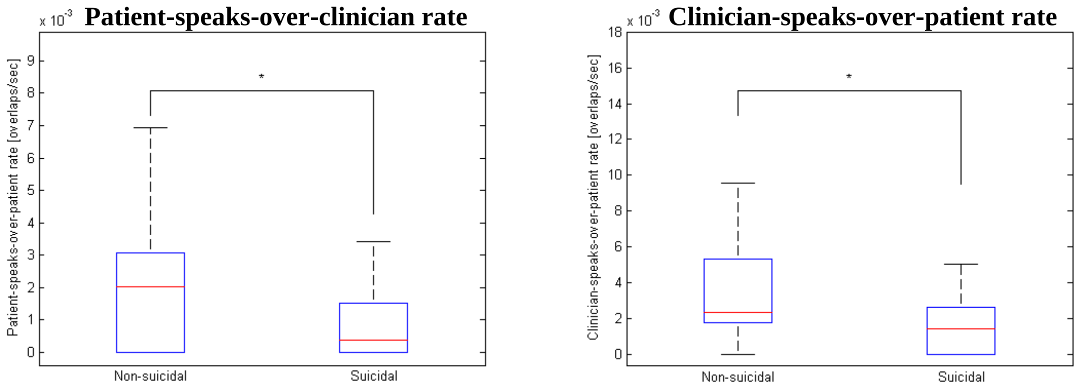


Figure 3-10: Non-suicidal vs. suicidal patient-speaks-over-clinician rate (left) and the clinician-speaks-over-patient rate (right) in overlaps per second (*UQ* case)

***Verbal information***

Suicidal patients used more personal pronouns ($p_P<0.01$), first person singular pronouns ($p_P<0.01$), impersonal pronouns ($p_P<0.01$) and terms indicating the past ($p_p<0.05$). As represented in Figure 3-11, the suicidal patients referred on average more often to themselves by using first person singular pronouns more times than their controls did.

Words corresponding to negation were more often mentioned by the controls. The non-suicidal patients referred on average more often to positive emotions ($p_P<0.01$) than to negative emotions ($p_P<0.01$), see Figure 3-12. Interpreting the left boxplot it shows that the suicidal patients referred less often to positive emotion words than their controls did. The right boxplot shows that the suicidal patients used more often negative emotion words than their non-suicidal controls. Nonfluencies and assent terms were significantly different ($p_P<0.01$).

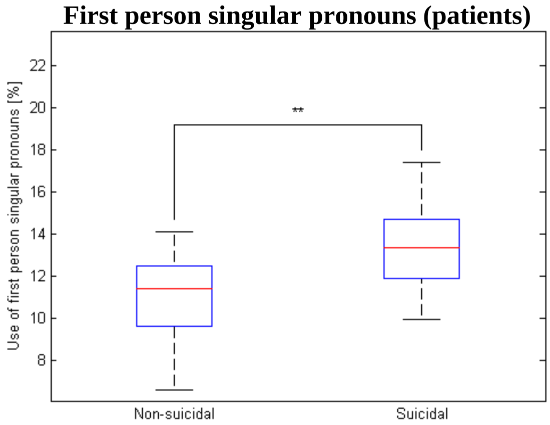


Figure 3-11: Non-suicidal vs. suicidal use of first person singular pronouns of the patients (*UQ* case)

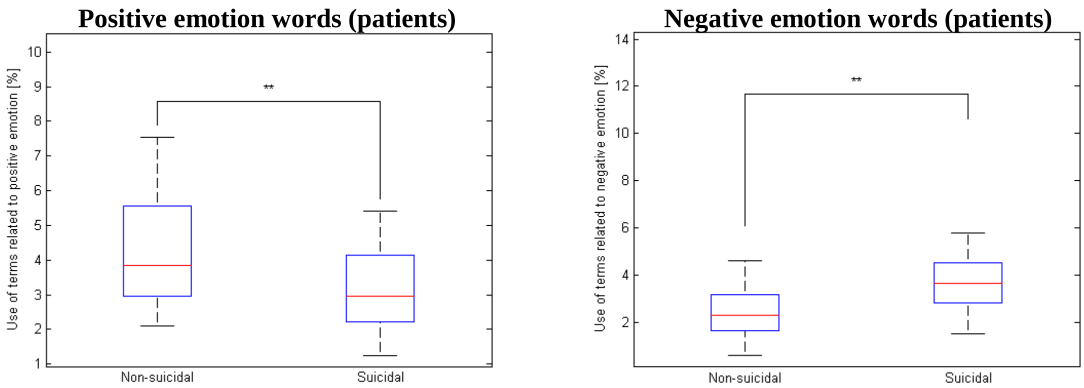


Figure 3-12: Non-suicidal vs. suicidal use of terms related to positive emotion (left) and to negative emotion (right) of the patients (*UQ* case)

The clinician used personal pronouns ($p_C<0.05$) and impersonal pronouns ($p_C<0.01$) more often while interacting with suicidal patients than with their controls. He referred to himself by using first person singular pronouns less during interviews with suicidal patients ($p_C<0.01$). The usage of tentative words and terms indicating the past by the clinician showed significant differences ($p_C<0.01$). Figure 3-13 represents the usage of tentative words by the clinician. The distribution of the use of tentative words shows that the clinician used less tentative words while interacting with suicidal patients than with the non-suicidal ones.

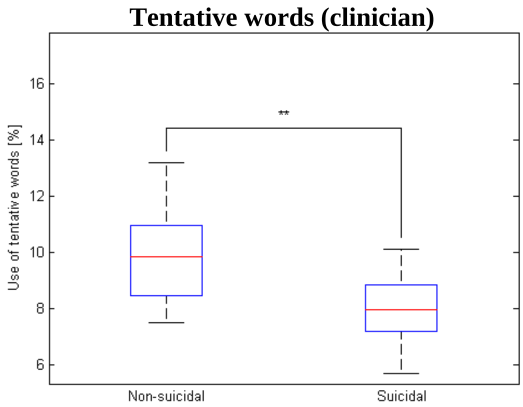


Figure 3-13: Non-suicidal vs. suicidal use of tentative words of the clinician (*UQ* case)

### *Acoustic information*

Suicidal patients spoke on average with higher $f_0$ than the non-suicidal ones ($p_P<0.01$), see left boxplot in Figure 3-14. Non-suicidal patients used a narrower range of frequencies than the suicidal patients did. The higher NAQ, QOQ and MDQ implied that the suicidal patients spoke with breathier voices ($p_P<0.01$). Also the PS indicated that ($p_P<0.01$). The PSP value was higher for the non-suicidal adolescents. The parameter *Rd* and the first formant *F1* were as well significantly different ($p_P<0.01$). The *Rd* distributions of non-suicidal vs. suicidal patients is represented in the right boxplot of Figure 3-14.

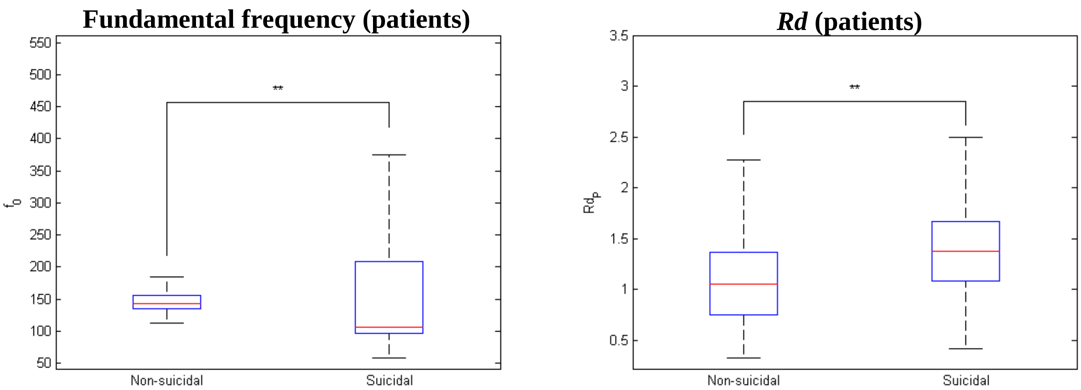


Figure 3-14: Non-suicidal vs. suicidal $f_0$ (left) and *Rd* (right) of the patients (*UQ* case)

The clinician's speech characteristics NAQ, QOQ and PS indicated a breathier conversation during interviews with suicidal patients ($p_C<0.01$). The PSP showed significant difference too ($p_C<0.01$) as well as the *Rd* parameter ($p_C<0.01$). Also *F1* was statistically significant ($p_C<0.01$).

The clinician's backchannel showed significant differences for the NAQ, QOQ and the PS ($p_C$<0.01), as represented in Figure 3-15. The left boxplot represents that the clinician used a breathier voice while interacting with suicidal patients. This behavior could be already observed in the backchannel of the clinician, as illustrated in the right boxplot. All of these three features indicated a breathier conversation while interacting with suicidal patients.

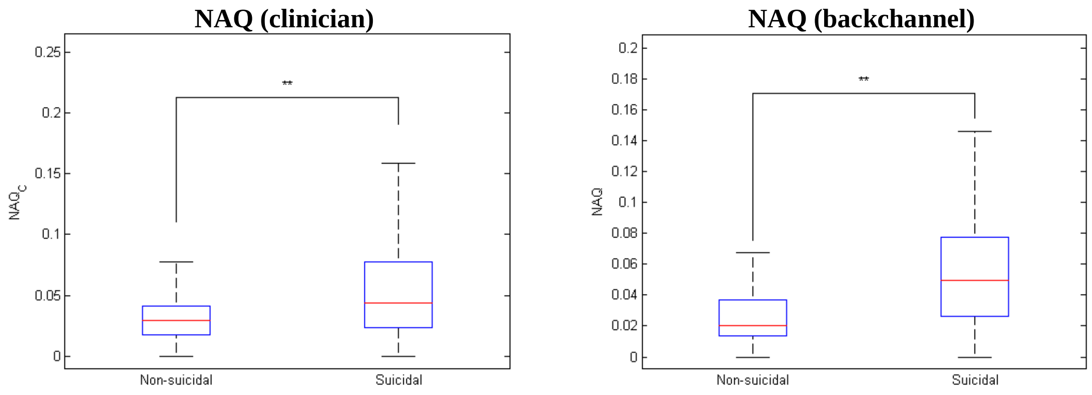


Figure 3-15: Non-suicidal vs. suicidal NAQ of clinician (left) and of the clinician's backchannel (right) (*UQ* case)

Table 6: Statistical significant results of the patients' features of the non-suicidal vs. suicidal evaluation with *N=60* (*UQ* case)

| Feature | Suicidal μ (σ) | Non-suicidal μ (σ) | p-value |
|---|---|---|---|
| Speak time percentage | 0.43 (0.10) | 0.32 (0.11) | ** |
| Pause time percentage | 0.14 (0.04) | 0.08 (0.04) | ** |
| Patient-speaks-over-clinician | 0.93 (1.17) | 2.13 (2.17) | * |
| Personal pronouns | 16.52 (2.38) | 13.36 (2.66) | ** |
| 1st person singular pronoun | 13.31 (1.84) | 11 (1.96) | ** |
| Impersonal pronouns | 7.99 (1.61) | 6.07 (2.32) | ** |
| Past tense | 3.07 (1.91) | 1.89 (1.73) | * |
| Negation | 3.97 (1.38) | 6.54 (2.61) | ** |
| Positive emotion | 3.21 (1.14) | 4.22 (1.46) | ** |
| Negative emotion | 3.84 (1.60) | 2.58 (1.19) | ** |
| Nonfluencies | 1.72 (1.54) | 3.89 (2.97) | ** |
| Assent | 1.98 (1.04) | 4.79 (3.36) | ** |
| $f_0$ | 224.03(26.19) | 149.01 (10.89) | ** |
| NAQ | 0.08 (0.02) | 0.03 (7.3e$^{-3}$) | ** |
| QOQ | 0.31 (0.07) | 0.11 (0.03) | ** |
| PSP | 0.36 (0.06) | 0.52 (0.07) | ** |
| MDQ | 0.13 (4.7e$^{-3}$) | 0.11 (0.01) | ** |
| PS | -0.20 (0.04) | -0.25 (0.04) | ** |
| *Rd* | 1.65 (0.15) | 1.09 (0.18) | ** |
| *F1* | 613.56 (90.76) | 538.75 (104.23) | ** |

Table 7: Statistical significant results of the clinician's features of the non-suicidal vs. suicidal evaluation with *N=60* (*UQ* case)

| Feature | Suicidal μ (σ) | Non-suicidal μ (σ) | p-value |
|---|---|---|---|
| Clinician-speaks-over-patient | 1.84 (2.03) | 3.60 (3.52) | * |
| Speak time percentage | 0.30 (0.06) | 0.44 (0.08) | ** |
| Pause time percentage | 0.13 (0.05) | 0.16 (0.06) | * |
| Personal pronouns | 12.06 (1.10) | 11.45 (0.96) | * |
| 1st person singular pronoun | 2.50 (0.51) | 3.07 (1.01) | ** |
| Impersonal pronouns | 9.00 (1.03) | 7.65 (1.16) | ** |
| Past tense | 1.72 (0.91) | 0.91 (0.70) | ** |
| Tentative | 7.97 (1.13) | 9.81 (1.61) | ** |
| NAQ | 0.07 (0.01) | 0.03 ($8.9e^{-3}$) | ** |
| QOQ | 0.26 (0.05) | 0.11 (0.03) | ** |
| PSP | 0.32 (0.07) | 0.49 (0.11) | ** |
| PS | -0.23 (0.04) | -0.25 (0.03) | ** |
| *Rd* | 1.62 (0.20) | 1.19 (0.22) | ** |
| *F1* | 614.04 (123.65) | 593.97 (144.21) | ** |
| **Backchannel**: | | | |
| NAQ | 0.07 (0.02) | 0.02 (0.01) | ** |
| QOQ | 0.26 (0.06) | 0.09 (0.04) | ** |
| PS | -0.24 (0.04) | -0.24 (0.03) | ** |

### 3.1.2. Suicidal non-repeater vs. repeater evaluation

In this section, the 30 recorded interviews with the suicidal patients were statistically analyzed to determine significant distinctions between suicidal non-repeaters and repeaters. The statistical characteristics of the significant descriptors are listed in Table 8 corresponding to the patients and to the clinician for the *complete* case. Table 9 relates to the patients' and clinician's significantly different features in the *UQ* case.

#### 3.1.2.1. *Complete* case

***Conversational information***

In the interviews with the non-repeaters the overlap rate as well as the clinician-speaks-over-the-patient rate was higher than in those with the repeaters ($p<0.05$ for both features), as it is represented in Figure 3-16. The clinician and patients spoke over each other more often in interviews with non-repeaters (left boxplot). This was also observed in the clinician-speaks-over-patient rate in the right boxplot.

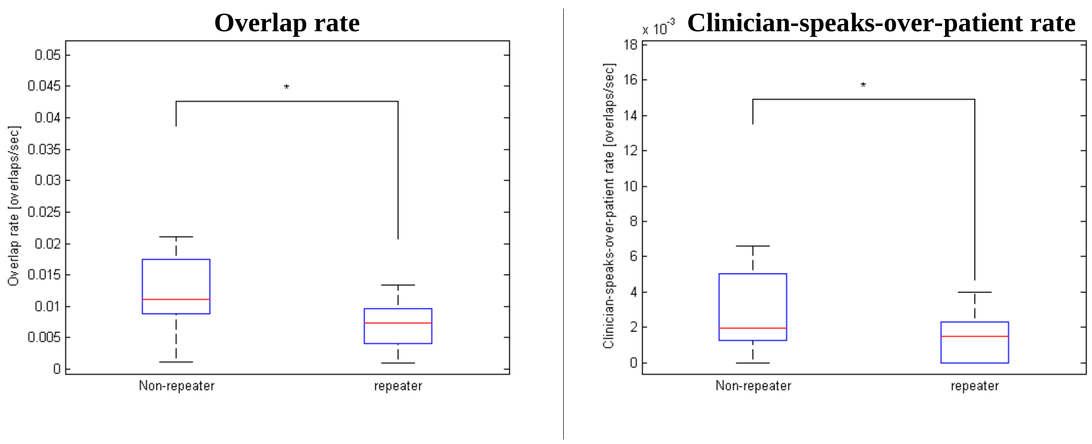


Figure 3-16: Non-repeater vs. repeater overlap rate (left) and clinician-speaks-over-patient rate (right) (*complete* case)

### *Verbal information*

In the non-repeater vs. repeater case, there were not found any relevant significant differences in this feature group. Although the suicidal repeaters used first person singular pronouns slightly less often than the non-repeaters, statistically significant differences were not found for the pronouns, see Figure 3-17. This boxplot represents the significant difference between non-suicidal patients, suicidal repeaters and non-repeaters. Moreover, it shows that there is no statistically significant difference between repeaters and non-repeaters. There are statistically significances between non-suicidal adolescents and the suicidal group including repeaters and non-repeaters.

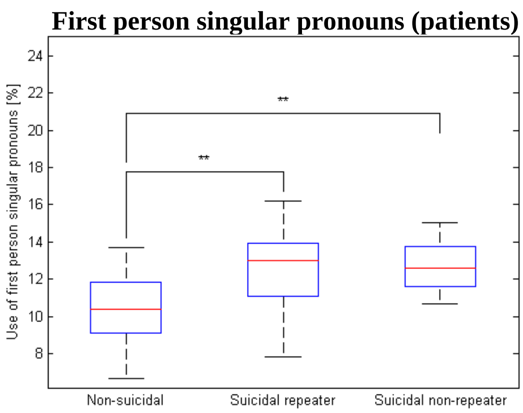


Figure 3-17: Non-suicidal vs. repeater vs. non-repeater use of first person singular pronouns of patients (*complete* case)

### *Acoustic information*

In the interviews with the suicidal repeaters, patients and clinician spoke on average with a lower fundamental frequency while interacting with the non-repeaters, see Figure 3-18. Here, the $f_0$ of the patients (left) and the clinician (right) are compared. It was noticed that the patients' and clinician's distributions of the $f_0$ show similar behavior. An adaptation of the clinician to the patients was identified. The data related to the non-repeaters' interviews showed a greater range of frequencies than the interactions with repeaters.

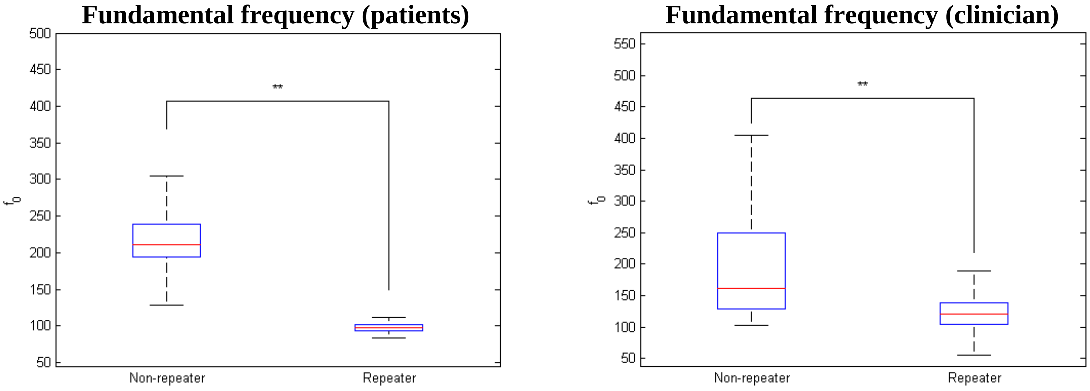


Figure 3-18: Non-repeater vs. repeater $f_0$ of the patients (left) and of the clinician (right) (*complete* case)

The other more significant acoustic features including NAQ, QOQ, MDQ and PS showed results characterizing a breathier voice during interviews with the non-repeaters, as illustrated in Figure 3-19. For patients (left boxplot) and clinician (right boxplot) significant differences were able to detect in the non-suicidal vs. suicidal evaluation as well as in the repeater vs. non-repeater one. On the left side the distribution of the QOQ values of the patients are shown and on the right side the ones corresponding to the clinician.

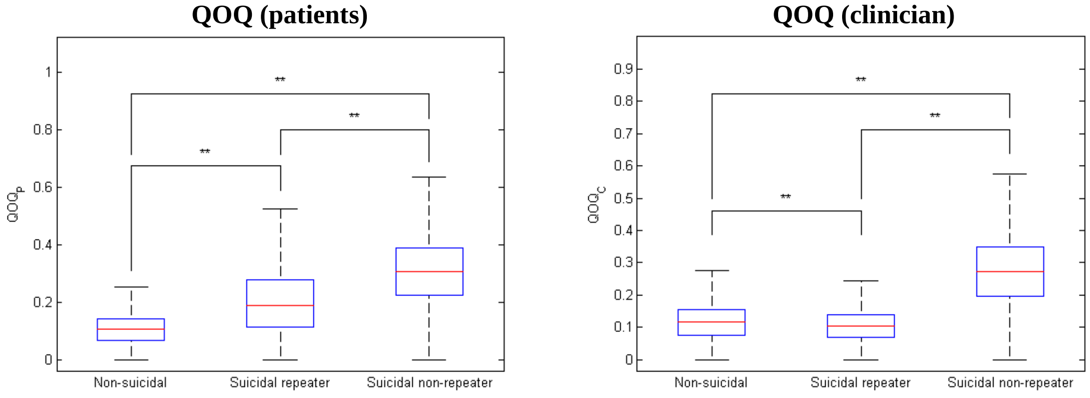


Figure 3-19: Non-suicidal vs. repeater vs. non-repeater QOQ of patients (left) and clinician (right) (*complete* case)

The suicidal non-repeaters had on average higher values for the PSP than the repeaters ($p_P$<0.01). The PSP of the clinician, instead, was higher (i.e. breathier voice) during interviews with suicidal repeaters ($p_C$<0.01). The *Rd* parameter of the clinician during interviews had a higher value while speaking to non-repeaters ($p_C$<0.01).

Moreover, the backchannel of the interviewer was analyzed separately. Similar to the investigation of the acoustic features of the clinician, the parameters correlated to breathier voice were used more often during the interviews with the suicidal non-repeaters. Additionally, the formants *F1* and *F2* were significantly different ($p_C$<0.01 and $p_C$<0.01, respectively).

Table 8: Statistical significant results of the non-repeater vs. repeater evaluation with *N=30* (*complete* case)

| **Feature** | **Repeater μ (σ)** | **Non-repeater μ (σ)** | **p-value** |
|---|---|---|---|
| Overlap rate | 7.75 (4.92) | 13.40 (8.24) | * |
| *Patient* | | | |
| NAQ | 0.03 (0.01) | 0.08 (0.02) | ** |
| QOQ | 0.2 (0.05) | 0.31 (0.07 | ** |
| PSP | 0.24 (0.05) | 0.36 (0.06) | ** |
| MDQ | 0.08 (0.01) | 0.14 ($4.9e^{-3}$) | ** |
| PS | -0.27 (0.03) | -0.2 (0.04) | ** |
| *Clinician* | | | |
| Clinician-speaks-over-patient | 1.38 (1.38) | 3.33 (3.08) | * |
| NAQ | 0.03 ($9.1e^{-3}$) | 0.07 (0.02) | ** |
| QOQ | 0.1 (0.04) | 0.26 (0.05) | ** |
| PSP | 0.39 (0.13) | 0.33 (0.08) | ** |
| MDQ | 0.11 (0.02) | 0.13 ($8e^{-3}$) | ** |
| PS | -0.25 (0.03) | -0.23 (0.04) | ** |
| *Rd* | 1.35 (0.18) | 1.62 (0.2) | ** |
| **Backchannel:** | | | ** |
| NAQ | 0.02 (0.01) | 0.07 (0.02) | |
| QOQ | 0.10 (0.06) | 0.26 (0.07) | ** |
| MDQ | 0.10 (0.02) | 0.13 (0.01) | ** |
| *F1* | 810.8 (213.2) | 701.8 (120.9) | ** |
| *F2* | 1666.2 (330.1) | 1536 (158) | * |

#### 3.1.2.2. *UQ* case

***Conversational information***

The overlap rate of the interviews with the suicidal non-repeaters was on average higher than with the repeaters ($p<0.05$), as represented in Figure 3-20. This conversational information feature was the only one showing statistically significance. There were not investigated any further significant differences in this feature group. The total duration of the interviews with the suicidal repeaters was on average 520.82 seconds (~8.7 minutes). The interviews with the suicidal non-repeaters lasted on average 553.85 seconds (~ 9.2 minutes).

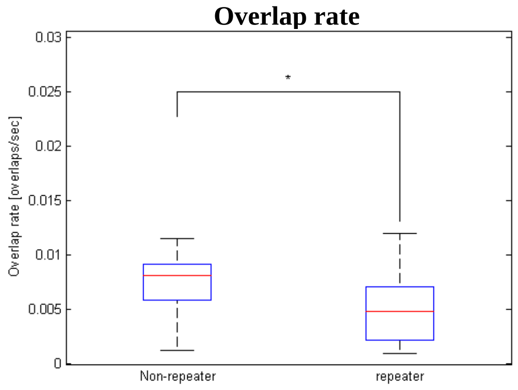


Figure 3-20: Non-repeater vs. repeater overlap rate (*UQ* case)

***Verbal information***

Similar to section **3.1.2.1**, significant differences were not found for the investigated verbal features. Regarding the means of the features, suicidal non-repeaters referred slightly more often to first person singular pronouns ($\mu_P$=12.65 for the repeaters and $\mu_P$=12.78 for the non-repeaters).

***Acoustic information***

The NAQ ($p_P<0.01$), QOQ ($p_P<0.05$), PSP ($p_P<0.01$), MDQ ($p_P<0.01$) and PS ($p_P<0.01$) of the patients indicated a breathier conversation during the interviews with the non-repeaters, as it is represented in Figure 3-21 for the patients' NAQ and PS. On the left side the distributions of the NAQ of the non-repeaters and repeaters are shown. On the right side the significant difference of the PS of the patient is shown. The results indicate that the non-repeaters spoke with a breathier voice than the repeaters. The parameter *Rd* was slightly higher for the non-repeaters than for the repeaters ($p_P<0.01$).

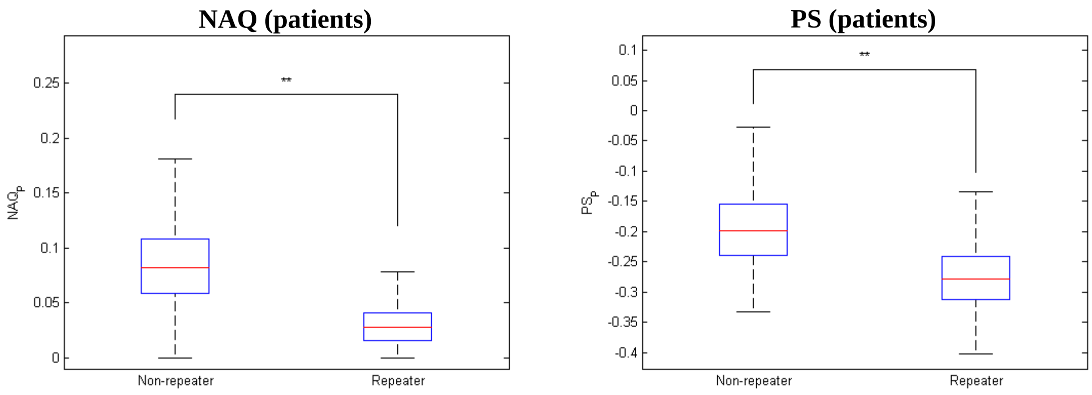


Figure 3-21: Non-repeater vs. repeater NAQ (left) and PS (right) of the patients (*UQ* case)

The values of the clinician's acoustic features showed that in conversation with non-repeaters the clinician used a breathier voice. The NAQ ($p_C$<0.01), QOQ ($p_C$<0.01), MDQ ($p_C$<0.01) and PS ($p_C$<0.01) means were slightly higher while interacting with suicidal non-repeaters, see Figure 3-22. On the left side the NAQ of the clinician is represented, on the right side the QOQ. The distributions show that the clinician spoke with a breathier voice while interacting with suicidal non-repeaters. The PSP ($p_C$<0.01) was slightly higher during interviews with suicidal repeaters as well as *Rd* ($p_C$<0.01).

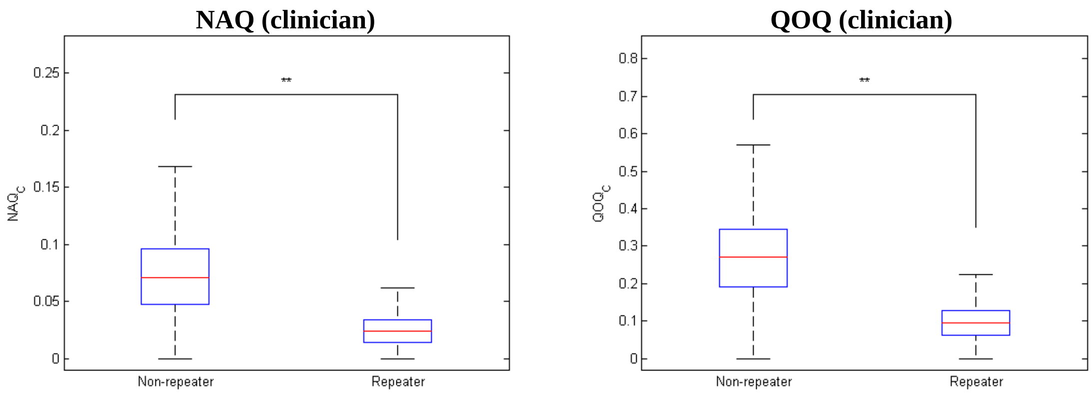


Figure 3-22: Non-repeater vs. repeater NAQ (left) and QOQ (right) of the clinician (*UQ* case)

From the backchannel of the clinician, the NAQ, QOQ and MDQ showed significant differences ($p_C$<0.01), as it is represented in Figure 3-23. The distributions of the NAQ (left boxplot) and the QOQ (right boxplot) of the backchannel of the clinician, which described patches lasting less than 700 milliseconds, indicate that the clinician reacted with a breathier voice during interviews with suicidal non-repeaters. These three parameters characterized the breathier voice of the clinician during interviews with suicidal non-repeaters. *F1* and *F2* were as well significantly different.

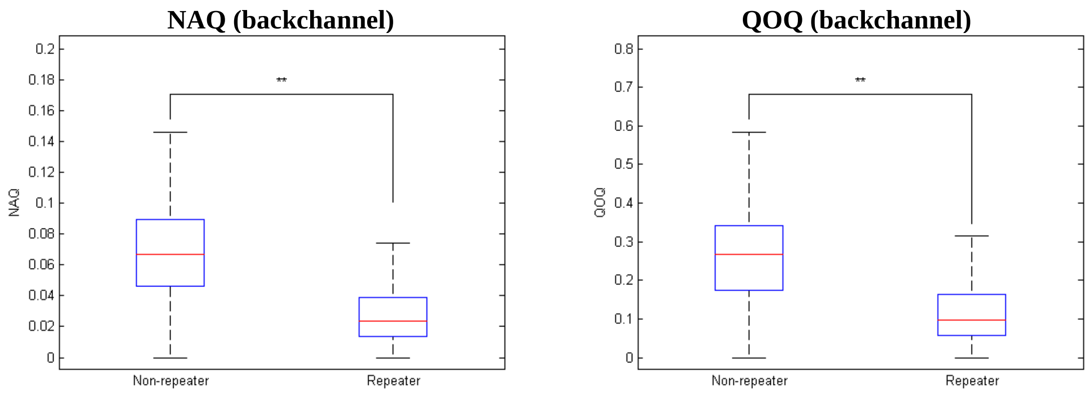


Figure 3-23: Non-repeater vs. repeater NAQ (left) and QOQ (right) of the clinician's backchannel (*UQ* case)

Table 9: Statistical significant results of the non-repeater vs. repeater evaluation with *N=30* (*UQ* case)

| | **Repeater** | **Non-repeater** | |
|---|---|---|---|
| **Feature** | **μ (σ)** | **μ (σ)** | **p-value** |
| Overlap rate | 4.81 (3.23) | 8.06 (4.36) | * |
| *Patient* | | | |
| NAQ | 0.03 (0.01) | 0.08 (0.02) | ** |
| QOQ | 0.20 (0.05) | 0.31 (0.07) | * |
| PSP | 0.24 (0.04) | 0.36 (0.06) | ** |
| MDQ | 0.08 (0.01) | 0.13 (0.005) | ** |
| PS | -0.27 (0.03) | -0.20 (0.04) | ** |
| *Rd* | 1.20 (0.19) | 1.65 (0.15) | ** |
| *Clinician* | | | |
| NAQ | 0.02 (0.01) | 0.07 (0.01) | ** |
| QOQ | 0.09 (0.04) | 0.26 (0.05) | ** |
| PSP | 0.37 (0.13) | 0.32 (0.07) | ** |
| MDQ | 0.10 (0.02) | 0.13 (0.01) | ** |
| PS | -0.25 (0.03) | -0.23 (0.04) | ** |
| *Rd* | 1.35 (0.19) | 1.62 (0.20) | ** |
| **Backchannel:** | | | |
| NAQ | 0.02 (0.01) | 0.07 (0.02) | ** |
| QOQ | 0.09 (0.06) | 0.26 (0.06) | ** |
| MDQ | 0.10 (0.02) | 0.13 (0.01) | ** |
| *F1* | 827.81 (196.29) | 681.62 (112.10) | * |
| *F2* | 1696.19 (334.87) | 1521.12 (140.75) | * |

## 3.2. Classification

In the following sections the classification results of the hierarchical ensemble classification, of the hierarchical SVM classification and of the hierarchical SVM-ensemble classification are given. The *complete* and *UQ* case classification results are shown as well as the results corresponding to the different datasets.

There were four datasets investigated. First, all the obtained features were used as input of the classification. Secondly, only the patients' features were used for training and testing. Moreover, the results obtained by using patients' and clinician's features excluding the backchannel ones were investigated. Eventually, the patients' and the clinician's backchannel features were employed for the training and test phase of the hierarchical classifier.

### 3.2.1. Ensemble Classification

For the *complete* case, an accuracy of 90% was achieved in the non-suicidal vs. suicidal distinction while the non-repeater vs. repeater layer classification delivered an accuracy of 60%. Sixty interviews entered the classification stage, 27 of them were correctly labeled as non-suicidal. The positive labeled one (i.e. suicidal) were forwarded to the non-repeater vs. repeater level, and there 7 true positives (i.e. repeaters) and 10 true negatives (i.e. non-repeaters) were able to be identified correctly. The confusion matrix of the hierarchical classifier is provided in Table 10. The classification over the complete hierarchy yielded an accuracy of 73.3%. The corresponding F scores, precision and recall measures are also given in Table 10. In the approach described above, the features of both or rather of the dynamic between clinician and patients were used.

When the features related to the patients entered the hierarchical classifier, it yielded an accuracy of 61.7%, with an accuracy in the non-suicidal vs. suicidal layer of 76.7% and 60% in the non-repeater vs. repeater layer. The clinician's features without the backchannel ones were added to the patient's features to run through the machine learning algorithm and the accuracy was able to be increased to 68.3% (non-suicidal vs. suicidal: 90%; non-repeater vs. repeater: 46.7%). Moreover, the impact of the backchannel information of the clinician was forwarded together with the patient's features to the hierarchical classifier. This yielded an accuracy of 56.7% (non-suicidal vs. suicidal: 76.7%; non-repeater vs. repeater: 53.3%). The accuracy bars of the single investigations are represented in Figure 3-24. The corresponding confusion matrices, F scores, precision and recall results are shown in Table 12, also for the *UQ* case. The graph in Figure 3-24 represents the accuracies of the two layers and of the hierarchical ensemble classifier in general for each dataset which was used as input for the classification. The blue bars represent the non-suicidal vs. suicidal accuracies, the red ones the non-repeater vs. repeater layer and the green ones the non-suicidal vs. repeaters vs. non-repeaters.

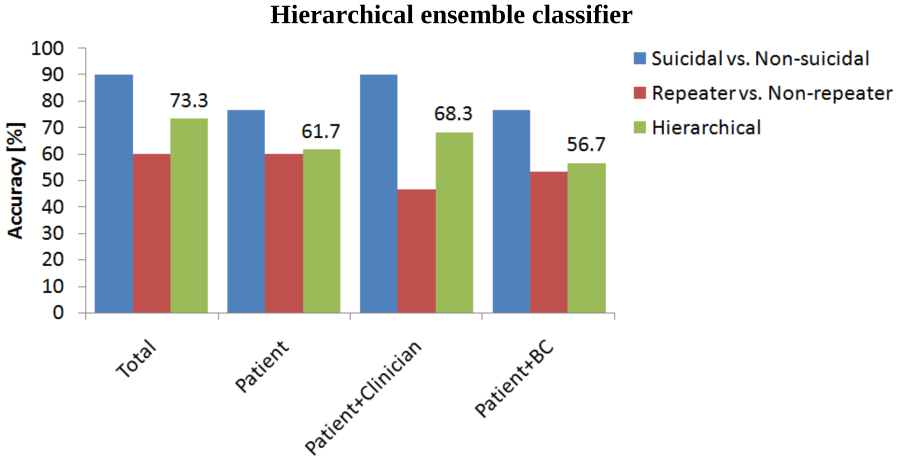


Figure 3-24: Accuracy bars of the hierarchical ensemble classifier (green) including accuracies of the two layers separately (*complete* case)

Table 10: Confusion matrix of the hierarchical ensemble classifier (*complete* case) including accuracy and the parameters recall, precision and F score

| **Prediction** / **Actual** | Non-suicidal | Suicidal non-repeater | Suicidal repeater | **Recall** |
|---|---|---|---|---|
| Non-suicidal | 27 | 1 | 2 | 0.90 |
| Suicidal non-repeater | 2 | 10 | 5 | 0.59 |
| Suicidal repeater | 1 | 5 | 7 | 0.54 |
| **Precision** | 0.90 | 0.63 | 0.50 | **Accuracy** |
| **F score** | 0.90 | 0.61 | 0.52 | 73.3% |

Additionally, the *UQ* case was investigated. The accuracy of the classification over all previous utilized features declined by 10% to 63.3% in comparison to the *complete* case. The classification performance of the non-suicidal vs. suicidal distinction as well as the non-repeater vs. repeater one decreased in the *UQ* case to 86.7% and to 50%, respectively. 26 patients were correctly labeled as non-suicidal, 7 as suicidal non-repeaters and 5 as repeaters. The corresponding accuracies of the two layers of the classifier are listed in Table 11.

For the investigation of the three additional investigated datasets the patients' features yielded an accuracy of 61.7%, adding the clinician's features resulted in an accuracy of 65% and an accuracy of 66.7% was obtained by using patients' and clinician's backchannel's features, see Figure 3-25.

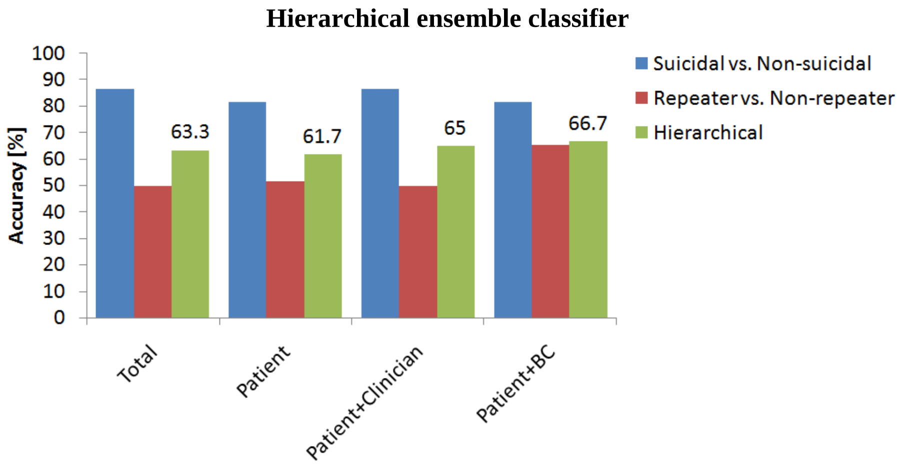


Figure 3-25: Accuracy bars of the hierarchical ensemble classifier (green) including accuracies of the two layers separately (*UQ* case)

Table 11: Confusion matrix of the hierarchical ensemble classifier (*UQ* case) including accuracy and the parameters recall, precision and F score

| **Prediction** / **Actual** | Non-suicidal | Suicidal non-repeater | Suicidal repeater | **Recall** |
|---|---|---|---|---|
| Non-suicidal | 26 | 3 | 1 | 0.87 |
| Suicidal non-repeater | 1 | 7 | 5 | 0.54 |
| Suicidal repeater | 3 | 9 | 5 | 0.29 |
| **Precision** | 0.87 | 0.37 | 0.45 | **Accuracy** |
| **F score** | 0.87 | 0.44 | 0.36 | 63.3% |

Table 12: Accuracies in the two layers and the overall accuracy for each dataset used as input for the hierarchical ensemble classifier (BC=backchannel)

| | Non-suicidal vs. suicidal | Non-repeater vs. repeater | Hierarchical |
|---|---|---|---|
| *Complete* case | | | |
| Patient+Clinician+BC | 90% | 60% | 73.3% |
| Patient | 76.7% | 60% | 61.7% |
| Patient+Clinician | 90% | 46.7% | 68.3% |
| Patient+BC | 76.7% | 53.3% | 56.7% |
| *UQ* case | | | |
| Patient+Clinician+BC | 86.7% | 50% | 63.3% |
| Patient | 81.7% | 51.6% | 61.7% |
| Patient+Clinician | 86.7% | 50% | 65% |
| Patient+BC | 81.7% | 65.5% | 66.7% |

### 3.2.2. SVM classification

The hierarchical SVM classifier performed the distinction between non-suicidal patients, suicidal non-repeaters and repeaters with an accuracy of 76.7% in the *complete* case. The accuracies related to the non-suicidal vs. suicidal layer was 95% and the one referred to the non-repeater vs. repeater distinction was 62.1%. In total 29 patients were truly labeled as non-suicidal, 15 as suicidal non-repeaters and 2 as suicidal repeaters. The corresponding confusion matrix, F measures, precision and recall values can be found in Table 13 and the accuracies of the different used datasets are given in Table 15 and represented in Figure 3-26.

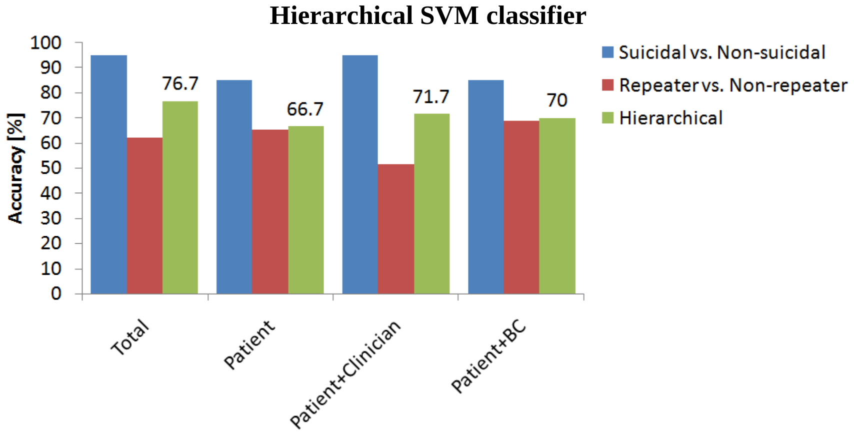


Figure 3-26: Accuracy bars of the hierarchical SVM classifier (green) including accuracies of the two layers separately (*complete* case)

Table 13: Confusion matrix of the hierarchical SVM classifier (*complete* case) including accuracy and the parameters recall, precision and F score

| **Prediction** / **Actual** | Non-suicidal | Suicidal non-repeater | Suicidal repeater | **Recall** |
|---|---|---|---|---|
| Non-suicidal | 29 | 1 | 0 | 0.97 |
| Suicidal non-repeater | 2 | 15 | 9 | 0.63 |
| Suicidal repeater | 0 | 2 | 2 | 0.33 |
| **Precision** | 0.94 | 0.83 | 0.18 | **Accuracy** |
| **F score** | 0.95 | 0.71 | 0.24 | 76.7% |

In the *UQ* case, the classifier yielded an accuracy of 58.3%. As non-suicidal 24 adolescents were correctly labeled, 11 as non-repeaters and none of the repeaters were correctly labeled. In the first layer the accuracy was 80%, while in the second layer 56.7% were achieved. The confusion matrix and the evaluation measures are represented in Table 14. The accuracies of the different datasets are also given in Table 15 and illustrated in Figure 3-27.

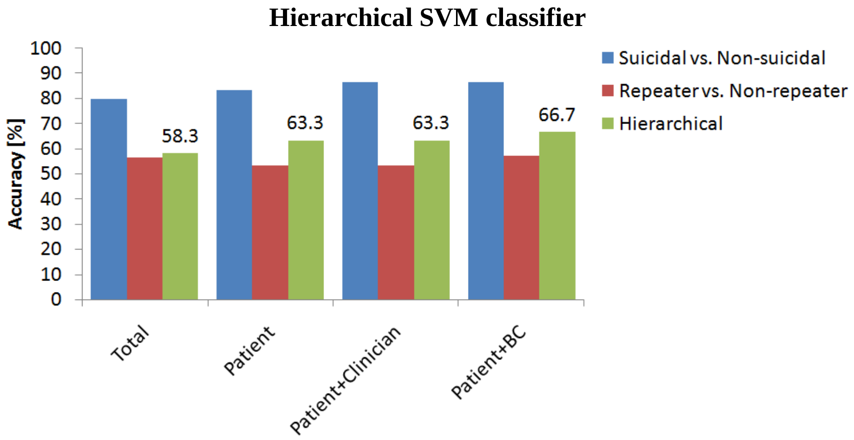


Figure 3-27: Accuracy bars of the hierarchical SVM classifier (green) including accuracies of the two layers separately (*UQ* case)

Table 14: Confusion matrix of the hierarchical SVM classifier (*UQ* case) including accuracy and the parameters recall, precision and F score

| **Prediction** / **Actual** | Non-suicidal | Suicidal non-repeater | Suicidal repeater | **Recall** |
|---|---|---|---|---|
| Non-suicidal | 24 | 6 | 0 | 0.80 |
| Suicidal non-repeater | 3 | 11 | 10 | 0.46 |
| Suicidal repeater | 3 | 3 | 0 | 0 |
| **Precision** | 0.80 | 0.55 | 0 | **Accuracy** |
| **F score** | 0.80 | 0.50 | 0 | 58.3% |

Table 15: Accuracies in the two layers and the overall accuracy for each dataset used as input for the hierarchical SVM classifier (BC=backchannel)

| | Non-suicidal vs. suicidal | Non-repeater vs. repeater | Hierarchical |
|---|---|---|---|
| *Complete* case | | | |
| Patient+Clinician+BC | 95% | 62.1% | 76.7% |
| Patient | 85% | 65.5% | 66.7% |
| Patient+Clinician | 95% | 51.7% | 71.7% |
| Patient+BC | 85% | 69% | 70% |
| *UQ* case | | | |
| Patient+Clinician+BC | 80% | 56.7% | 58.3% |
| Patient | 83.3% | 53.3% | 63.3% |
| Patient+Clinician | 86.7% | 53.3% | 63.3% |
| Patient+BC | 86.7% | 57.1% | 66.7% |

### 3.2.3. SVM-ensemble classification

In the *complete* case, the overall accuracy was valued with 78.3%. In the SVM classification non-suicidal vs. suicidal layer the accuracy resulted in 95%. In the ensemble classification layer labeling between non-repeaters and repeaters the accuracy was 62.7%. 29 patients were labeled correctly as non-suicidal, 12 as non-repeaters and 6 as repeaters. The confusion matrix, the F measures, recall and precision values are listed in Table 16. The accuracy bars of the different classification results for the four different input datasets are shown in Figure 3-28 and listed in Table 18.

In the *UQ* case, the hierarchical SVM-ensemble classification yielded an accuracy of 60%. This meant a 80% classification accuracy in the first layer and 53.3% in the second layer. Of the non-suicidal patients 24 were correctly labeled. Seven non-repeaters and 5 repeaters were correctly classified. Table 17 includes the confusion matrix, F measures, recall and precision values and Figure 3-29 illustrates the different accuracy bars of the four investigated cases.

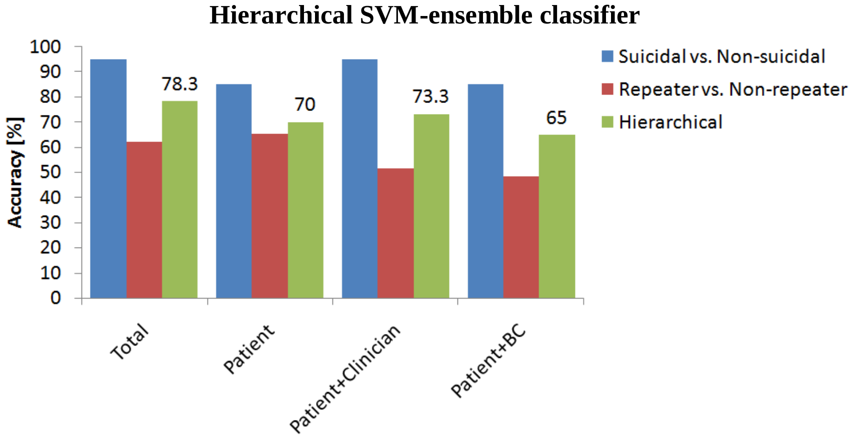


Figure 3-28: Accuracy bars of the hierarchical SVM-ensemble classifier (green) including accuracies of the two layers separately (*complete* case)

Table 16: Confusion matrix of the hierarchical SVM-ensemble classifier (*complete* case) including accuracy and the parameters recall, precision and F score

| **Prediction** / **Actual** | Non-suicidal | Suicidal non-repeater | Suicidal repeater | **Recall** |
|---|---|---|---|---|
| Non-suicidal | 29 | 0 | 1 | 0.97 |
| Suicidal non-repeater | 2 | 12 | 5 | 0.63 |
| Suicidal repeater | 0 | 5 | 6 | 0.55 |
| **Precision** | 0.94 | 0.71 | 0.50 | **Accuracy** |
| **F score** | 0.95 | 0.67 | 0.52 | 78.3% |

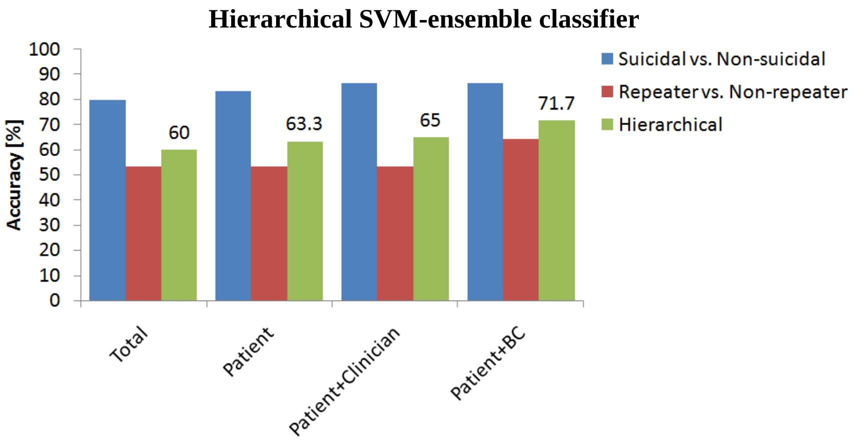


Figure 3-29: Accuracy bars of the hierarchical SVM-ensemble classifier (green) including accuracies of the two layers separately (*UQ* case)

Table 17: Confusion matrix of the hierarchical SVM-ensemble classifier (*UQ* case) including accuracy and the parameters recall, precision and F score

| **Prediction** / **Actual** | Non-suicidal | Suicidal non-repeater | Suicidal repeater | **Recall** |
|---|---|---|---|---|
| Non-suicidal | 24 | 4 | 2 | 0.80 |
| Suicidal non-repeater | 3 | 7 | 5 | 0.47 |
| Suicidal repeater | 3 | 7 | 5 | 0.33 |
| **Precision** | 0.80 | 0.39 | 0.42 | **Accuracy** |
| **F score** | 0.80 | 0.42 | 0.37 | 60% |

Table 18: Accuracies in the two layers and the overall accuracy for each dataset used as input for the hierarchical SVM-ensemble classifier (BC=backchannel)

| | Non-suicidal vs. suicidal | Non-repeater vs. repeater | Hierarchical |
|---|---|---|---|
| *Complete* case | | | |
| Patient+Clinician+BC | 95% | 62.7% | 78.3% |
| Patient | 85% | 65.5% | 70% |
| Patient+Clinician | 95% | 51.7% | 73.3% |
| Patient+BC | 85% | 48.3% | 65% |
| *UQ* case | | | |
| Patient+Clinician+BC | 80% | 53.3% | 60% |
| Patient | 83.3% | 53.3% | 63.3% |
| Patient+Clinician | 86.7% | 53.3% | 65% |
| Patient+BC | 86.7% | 64.3% | 71.7% |

# 4. Discussion

In this section the results of the statistical evaluation as well as of the classification will be discussed and findings of the thesis project will be introduced. Moreover, the most promising classification result will be presented and explained.

## 4.1. Statistical evaluation

In general, the statistical evaluations of the *complete* and the *UQ* case show that the investigated features in the *complete* case are likely to be observed in the evaluation of the *UQ* case, as stated within the thesis' second hypothesis. This means that already the short partition of each interview including the UQ is sufficient to characterize non-suicidal patients, suicidal non-repeaters and repeaters from each other. In the following paragraphs, the three feature groups' results will be discussed and the differences in the significances between the *complete* and *UQ* case will be mentioned.

Regarding the **conversational features**, the interviews with the suicidal patients last longer than the others, while the patients speak on average longer as the clinician but also pause longer as their controls. The clinician interrupts the suicidal patients less than the non-suicidal ones. Considering the non-repeater vs. repeater investigation, the clinician interrupts the repeaters less often than the non-repeaters. Also the suicidal patients do not interrupt the interviewer's speech as often as their controls do.

In the *UQ* case the patient-speaks-over-clinician rate extends the set of significant features in the non-suicidal vs. suicidal investigation. There is no significant difference for the clinician's words per second rate. The non-repeater vs. repeater evaluation of the *UQ* case does not result in a significant difference for the clinician-speaks-over-patient rate.

The investigation of the **verbal features** shows that they are useful for the non-suicidal vs. suicidal classification but not for the distinction of non-repeaters vs. repeaters. Any relevant verbal features, neither from the patients nor from the clinician, show significant differences for non-repeaters vs. repeaters. Hence, in the following paragraphs only the non-suicidal vs. suicidal investigations will be discussed in detail.

First, the social integration theory of [SP01] is verified. In all cases, the mean percentages of self-related pronouns (e.g. 'I, my, mine') of the suicidal group are higher than the ones of the controls. Moreover, the p-values show in the *UQ* and the *complete* case significant differences. A reduced use of terms related to family or friends is not measureable in this study. Suicidal patients use on average more references to others than their controls, especially second and third person pronouns. However, the p-values are not statistically significant. Moreover, a lower rate of using first person plural references is observable but shows no significant differences.

Although [SP01] showed significances for the use of terms related to death, an increased use of terms related to death is not observed in the suicidal group within this dataset. This could be due to the fact that the type of data [SP01] analyzed was written poetry and not interviews. However, with the provided dataset it is possible to show significance for the use of negative and positive emotion words. Suicidal patients use less terms related to positive emotion and more terms related to negative emotion than their non-suicidal controls. As in [SP01] suicidal patients refer more often to the past.

As already investigated in [SPM13], **acoustic features** of patients can help to determine suicidal risk. However, in this thesis the non-repeater vs. repeater are additionally compared and also the clinician's backchannel is investigated. The statistical evaluation of this feature group shows that the acoustic features of the patients and clinician are especially useful to discriminate non-repeaters and repeaters, but also non-suicidal and suicidal patients from each other. There are just minor differences in the significances of the features between *complete* and *UQ* case. The acoustic features of the clinician's backchannel show useful significances. The evaluation deduces that the clinician speaks breathier with the suicidal adolescents, especially with the non-repeaters, and this is already determined by just investigating the backchannel instead of the clinician's acoustic features of the entire interview. Moreover, an adaptation of the clinician's voice features to the ones of the patient is possible to be observed. This can be already confirmed by just looking at the features of the backchannel instead of the entire clinician's information.

Summarized, the statistical evaluation of the feature groups show that the discrimination of the non-suicidal vs. suicidal patients is easier to perform due to high significant differences within the three feature groups. Hence, the thesis' first hypothesis is able to be confirmed. Verbal and nonverbal behavior of the patient as well as of the clinician can characterize the suicidal risk of the patients. The non-repeater vs. repeater investigation reveals that especially acoustic features and, therefore, nonverbal information should be considered to distinguish non-repeaters from repeaters.

## 4.2. Classification

The hierarchical structure of the classifier allows to distinguish non-suicidal and suicidal patients in the first layer and then the suicidal non-repeaters and repeaters are classified out of the suicidal patients in the second layer. The best classification result using this hierarchical structure is obtained by employing a SVM classifier in the first layer and an ensemble classifier in the second layer. Thus, the SVM-ensemble classification achieves an accuracy of 78.3%, which is the highest one compared to the other classifiers' performances as it is illustrated in Figure 4-1. In the *UQ* case the hierarchical SVM classifier results in a higher accuracy (63.3%) than the SVM-ensemble classifier (60%), see also Figure 4-1. The bars of this plot correspond to the classification accuracies used in the hierarchical structure. The blue bars correspond to the ensemble classifier, the red ones to the SVM classifier and the green ones to the SVM-ensemble classifier, which uses SVMs in the first layer and an ensemble boosting algorithm in the second layer. In the *complete* case the SVM-ensemble classification achieves the highest accuracy. In the *UQ* case the SVM classification is most promising.

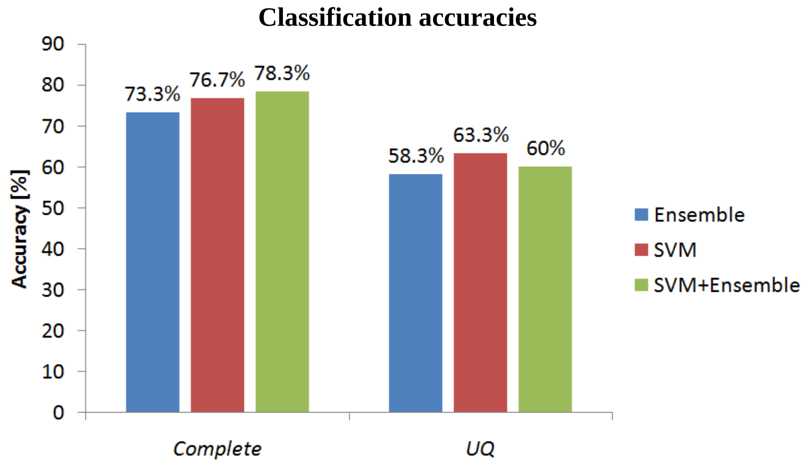


Figure 4-1: Comparison of the classification results (*complete* and *UQ* case)

The classification between suicidal non-repeaters and repeaters in the second layer is more difficult than the distinction between non-suicidal and suicidal patients. As already mentioned in section **4.1**, the differences between non-suicidal and suicidal patients are higher than the non-repeater vs. repeater investigation. In the end, 41 features are used for the hierarchical classification in the first layer. Table 4 in section **3.1.1.1** on page 25 shows the used features of the patients. On the same page Table 5 lists the used features of the clinician. In the second layer, 18 features are used to train and test the classifier (see Table 8 in section **3.1.2.1** on page 33).

As it is illustrated in Figure 4-2, in the first layer 6 conversational, 16 verbal and 19 acoustical information features are used to classify non-suicidal and suicidal patients. In the second layer, the classification is performed by considering 2 conversational features and 16 acoustical features of patients and clinician. The blue bars show the number of the conversational information features, the red one shows the number of the verbal information features and the green ones of the acoustical information features. In the suicidal vs. non-suicidal distinction all feature groups are considered for the classification. Comparing repeaters and non-repeaters less features were significantly different and, therefore, no verbal characteristics are relevant for the classification but the nonverbal ones are. Hence, the lack of the verbal information differences between non-repeaters and repeaters leads to a high weighting of the nonverbal behavior of patients and clinician.

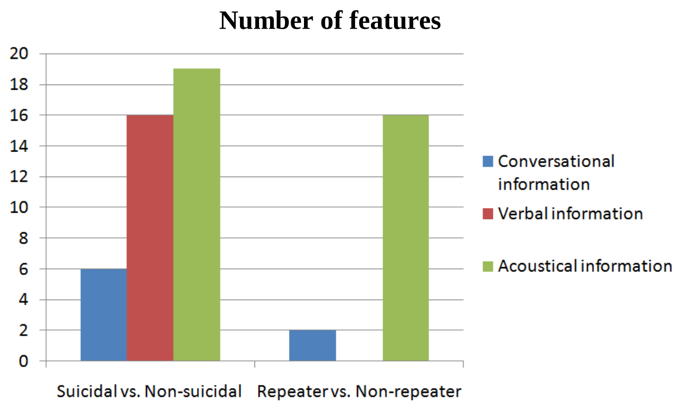


Figure 4-2: Distribution of the feature groups corresponding to the two layers of the hierarchical classifier

Furthermore, the influence of the clinician's features is investigated by using subsets of the features corresponding to the patients' and the clinician's ones for the SVM-ensemble classification. Hence, it is possible to show that the clinician's features increase the accuracy of the classification. Especially in the *complete* case, the performance of the hierarchical classifier is improved by using the patients' and the clinician's conversational, verbal and acoustic features for the classification task. By only using the patients' and the clinician's backchannel's features, the accuracy of the hierarchical classifier decreases from 70% (only patients' features) to 65%. In the end, by considering all the obtained features, including clinician's, patients' and backchannel's ones, the classification results in the best hierarchical classifier performance. Hence, the positive effect of considering clinician's features is possible to be shown.

Although the positive influence of the backchannel's features is not possible to be shown in the *complete* case, observing the *UQ* case leads to interesting results: they implicate that the performance of the ensemble classifier is possible to be increased by adding patients', clinician's and backchannel's features sequentially. It is worth mentioning that the ensemble classification using the patients' and the clinician's backchannel features obtained the best performance (accuracy of 66.7%) within this investigation. Thus, only by adding the backchannel's features the non-suicidal patients, suicidal non-repeaters and repeaters are possible to be identified. As a result, the thesis' third hypothesis is possible to be verified. So the acoustical information of the backchannel of the clinician is useful to determine suicidal risk together with the patient's characteristics.

To sum it up, the hierarchical classification yields a possibility to distinguish between non-suicidal adolescents, suicidal non-repeater and repeaters. In the *complete* case the hierarchical SVM-ensemble classification structure leads to the highest performance comparing to hierarchical SVM and ensemble classifier. In the *UQ* case the best classification performance is obtained by using the hierarchical ensemble classifier in both layers. The positive effect by considering the clinician's features of the interview is observed. Hence, it is possible to achieve the evidence that a classification between non-suicidal patients, suicidal non-repeaters and repeaters is possible.

# 5. Conclusion

In this thesis, the ability to classify non-suicidal patients, suicidal non-repeaters and suicidal repeaters between the ages of 13 and 18 is investigated. Statistical analyses reveal significant features of the interaction between patients and clinician. These were expected to characterize verbal and nonverbal properties of suicidal adolescents and interviewer behavior. Verbal information features are confirmed to be useful to discriminate non-suicidal vs. suicidal adolescents. For the discrimination of suicidal non-repeaters and repeaters nonverbal acoustic information is shown to be most useful. The classification between non-suicidal adolescents, suicidal non-repeaters and suicidal repeaters is possible. A hierarchical classification structure using SVMs and/or ensemble classifier obtains promising classification results.

Conversational, verbal and acoustic information are shown to characterize suicidal speech of adolescents. In the first case, the statistical analysis of non-suicidal vs. suicidal patients reveals significant differences in each investigated feature group. All three feature groups are advantageous to discriminate between non-suicidal and suicidal patients.

In the second case, suicidal non-repeaters vs. suicidal repeaters, not as many significant differences are identified than in the investigations of non-suicidal vs. suicidal patients. Significances are found for the conversational features of the interviews. Unlike the evaluation regarding the analysis between suicidal subjects and their controls, the verbal information features are lacking significant differences which complicates the classification task. Nevertheless, the acoustic features, especially known for distinguishing breathy to tense voices, show statistical significances. Thus, it can be argued that written or verbal questionnaires just addressing the patients' verbal information might not be enough to identify a suicidal repeater, because clinicians could miss the revealing information of the patients' nonverbal information. Furthermore, the need of computer-aided support and the assessment of nonverbal conversational content is crucial to identify suicidal repeaters.

The discriminative faculty of the identified features is possible to be confirmed by the hierarchical SVM-ensemble classification which yields an accuracy of 73.3% for the *complete* case. In the *UQ* case the SVM classification achieves the most promising performance with an accuracy of 63.3%.

It is easier to differentiate between non-suicidal and suicidal patients than between suicidal non-repeaters and repeaters because more information of the conversation can be used for the classification. The hierarchical structure of two subsequent classifiers achieves promising results discriminating non-suicidal adolescents, suicidal non-repeaters and suicidal repeaters. In the first layer conversational, verbal and acoustic information features or rather patients' and clinician's verbal and nonverbal behaviors are used to characterize non-suicidal and suicidal adolescents. However, the discrimination between suicidal non-repeaters and repeaters in the second layer requires especially nonverbal acoustic information, no verbal information and just two conversational features of the interviews.

Regarding the backchannel of the clinician it is found that almost each significant acoustic feature of the clinician's complete dataset is observed in the backchannel information too.

Although the clinician's behavior is predictive because he knew about the state of the patient, the adaptation of the clinician's voice to the patient's one is already observed by just looking at the speech fragments that lasted less than 700 milliseconds. The clinician spoke with a breathier voice to the suicidal repeater patients than to the non-suicidal ones. In a future project, it would be interesting to investigate the adaptation of a 'clueless' clinician. This could be realized by not letting a clinician know about the psychological states of the patients while interacting with them.

While this study shows promising results it does not consider the future but rather the past of the suicidal non-repeaters and repeaters. Further, the performance of the non-repeater vs. repeater classification stage or the prediction of suicidal repeaters could be improved by extending the classification features, especially the nonverbal ones. A prospective study including an extended multimodal approach analyzing visual information is planned for future work.

Nevertheless, this study shows that a classification between non-suicidal adolescents, suicidal non-repeaters and suicidal repeaters is possible by considering the audio-based verbal and nonverbal behaviors of patients and clinician as well as the dynamic between interviewer and interviewee. However, the verbal behavior of the patients and the clinician are unfortunately not useful to distinguish non-repeaters from repeaters. Nevertheless, the identification of the suicidal risk considering non-repeaters and repeaters requires in the first layer of a hierarchical classification structure conversational, verbal and nonverbal information of the interlocutors. In the second and last layer especially nonverbal characteristics of patients and clinician are useful.

The discriminative faculty of the conversational, verbal and acoustic features was confirmed. Hence, a progress in the additional support of suicidal risk assessment of adolescents was possible to be identified.

# Table of Abbreviations

| | | |
|---|---|---|
| ANOVA | … | Analysis of Variances |
| CCMHC | … | Cincinnati Children's Hospital Medical Center |
| COVAREP | … | Collaborative Voice Analysis Repository for speech technologies |
| C-SSRS | … | Columbia Suicide Severity Rating Scale |
| ED | … | Emergency Department |
| ELAN | … | annotation software |
| *F1* | … | first formant |
| *F2* | … | second formant |
| FN | … | False Negative |
| FP | … | False Positive |
| GCI | … | Glottal Closure Instants |
| LF | … | Liljencrants-Fant |
| LIBSVM | … | Library for Support Vector Machines |
| LIWC | … | Linguistic Inquiry and Word Count |
| MDQ | … | Maximal Dispersion Quotient |
| NAQ | … | Normalized Amplitude Quotient |
| PS | … | Peak Slope |
| PSP | … | Parabolic Spectral Parameter |
| QOQ | … | Quasi-Open Quotient |
| RBF | … | Radial Basis Function |
| SIQ-JR | … | Suicidal Ideation Questionnaire-Junior |
| SVM | … | Support Vector Machine |
| TN | … | True negative |
| TP | … | True positive |
| UQ | … | Ubiquitous Question |
| VUV | … | Voice-Uttering-Voice |
| WHO | … | World Health Organization |

# Table of Figures

# Table of Tables

# Bibliography


[ASS+14] B. Ahmedani, G. Simon, C. Stewart, A. Beck, B. Waitzfelder, R. Rossom, F. Lynch, A. Owen-Smith, E.M. Hunkeler, U. Whiteside, B.H. Operskalski, M.J. Coffey, L.I. Solberg, "Health Care Contacts in the Year Before Suicide Death", in *Journal of General Internal Medicine*, vol. 29 (6), 2014, Jun., pp. 870-877.

[ABV92] P. Alku, T. Bäckström and E. Vilkman, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering", in *Speech Communication*, 11 (2-3), 1992, pp. 109-118.

[ASV97] P. Alku, H. Strik, and E. Vilkman, "Parabolic spectral parameter - A new method for quantification of the glottal flow", in *Speech Communication,* vol. 22, 1997, pp. 67-79.

[BDS02] R. Beck, R. Daughtridge and P. Sloane, "Physician-patient communication in the primary care office: a systematic review", in *American Board of Family Practice,* vol. 15 (1), 2002, Jan.-Feb., pp. 25-38.

[BHW+13] L. Block, R. Habicht, A.W. Wu, S.V. Desai, K. Wang, K.N. Silva, T. Niessen, N. Oliver and L. Feldman, "In the Wake of the 2003 and 2011 Duty Hours Regulations, How Do Internal Medicine Interns Spend Their Time?", in *Journal of General Internal Medicine,* vol. 28, no. 8, 2013, pp. 1042-1047.

[BDdA04] B. Bozkurt, B. Doval, C. d'Alessandro and T. Dutoit, "Improved differential phase spectrum processing for formant tracking", in *International Conference on Spoken Language Processing (ICSLP),* Jeju Island, KR, 2004, pp. 2421-2424.

[CDC14] Centers for Disease Control and Prevention (2014, July 28). Retrieved from *WISQUARS - Leading Causes of Death Reports, National and Regional, 1999-2011* [Online]. Available: http://webappa.cdc.gov/sasweb/ncipc/leadcaus10_us.html

[CL11] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines" in *ACM Transactions on Intelligent Systems and Technology,* vol. 2, 2011, pp. 27:1-27:27.

[Dav04] K. Davis, "Detecting Suicide Risk in Adolescents and Adults in an Emergency Department: a Pilot Study", in *Honors Projects,* Paper 11, 2004.

[DKD+14] G. Degottex, J. Kane, T. Drugman, T. Raitio and S. Scherer, "Covarep- a collaborative voice analysis repository for speech technologies", in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014),* Florence, IT, 2014, pp. 960-964.

[DA11] T. Drugman, and A. Abeer, "Joint robust voicing detection and pitch estimation based on residual harmonics", in *Proceedings of Interspeech 2011 (ISCA)*, San Jose, CA, USA, 2011, pp. 1973-1976.

[Dod08] Y. Dodge, "Significance Level", in *The Concise Encyclopedia of Statistics*, New York, NY, USA: Springer, 2008, pp. 488-490.

[FMH82] L. Fairbanks, M. McGuire, and C. Harris, "Nonverbal Interaction of Patients and Therapists During Psychiatric Interviews", in *Journal of Abnormal Psychology*, vol. 91, no. 2, 1982, pp. 109-119.

[FLL85] G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow", in *STL-QPSR*, vol. 26, no. 4, 1985, pp. 1-13.

[FSO+12] S. Finkelstein, S. Scherer, A. Ogan, L.-P. Morency, and J. Cassell, "Investigating the Influence of Virtual Peers as Dialect Models on Student's Prosodic Inventory", in *ISCA Workshop on Child-Computer Interfaces (WOCCI)*, Portland, OR, USA, 2012.

[FSS+00] D. France, R. Shiavi, S. Silverman, M. Silverman and D. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk", in *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, 2000, July, pp. 829-837.

[FS97] Y. Freund, and R.E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting", in *Journal of Computer and System Sciences*, vol. 55, 1997, December, pp. 119-139.

[HHR95] J.A. Hall, J.A. Harrigan and R. Rosenthal, "Nonverbal behavior in clinician-patient interaction", in *Applied & Preventive Psychology*, vol. 4,1995, pp. 21-37.

[HTF08] T. Hastie, R. Tibshirani and J. Friedman, "Ensemble Learning", in *The Elements of Statistical Learning*, 2nd ed., New York: Springer, 2008, ch. 16, pp. 605-625.

[HJM05] V. Haynal-Reymond, G. Jonsson and M. Magnusson, "Non-verbal Communication in Doctor-Suicidal Patient Interview", in *The Hidden Structure of Interaction: From Neurons to Culture Patterns*, Amsterdam, NL: IOS Press, 2005, ch. 9.

[HHH+01] M. Heller, V. Haynal-Reymond, A. Haynal and M. Archinard, "Can Faces Reveal Suicide Attempt Risks?", in *The Flesh of the Soul:The Body We Work with*, Bern, CH: Peter Lang, 2001, pp. 243-269.

[HBP09] L.M. Horowitz, E.D. Ballard and M. Pao, "Suicide screening in schools, primary care and emergency departments", in *Current Opinion in Pediatrics*, vol. 21, no. 5, 2009, Oct., pp. 620-627.

[HCL03] C.-W. Hsu, C.-C. Chang and C.-J. Lin. (2014, Aug. 4). *A Practical Guide to Support Vector Classification* (Initial version 2003; Last updated 2010) [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

[JBB+06] S. Joe, R.E. Baser, G. Breeden, H.W. Neighbors and J.S. Jackson, "Prevalence of and Risk Factors for Lifetime Suicide Attempts Among Blacks in the United States", in *JAMA* , vol. 296, no. 17, 2006, Nov., pp. 2112-2123.

[Kan12] J. Kane, "Tools for analysing the voice - Developments in glottal source and voice quality analysis", Ph.D. dissertation, Phonetics and Speech Laboratory, Trinity College Dublin, Dublin, Ireland, 2012.

[KKS+14] L. Kann, S. Kinchen, S. Shanklin, K. Flint, J. Hawkins, W. Harris, R. Lowry, E. O'Malley Olsen, T. McManus, D. Chyen, L. Whittle, E. Taylor, Z. Demissie, N. Brener, J. Thornton, J. Moore, S. Zaza, "Youth Risk Behavior Surveillance - United States, 2013" in *Centers for Diease Control and Prevention MMWR Surveillance Summaries*, vol. 63, no. 4, 2014, Jun., pp. 1-168.

[Ler11] G. Leroy, *Designing User Studies*, London, UK: Springer, 2011.

[Obi12] N. Obin, "Cries and Whispers: Classification of vocal effort in expressive speech", in *Proceedings of Interspeech*, Portland, OR, USA, 2012.

[OBM+96] P.W. O'Carroll, A.L. Berman, R.W. Maris, E.K. Moscicki, B.L. Tanney, M.M. Silverman, "Beyond the tower of Babel: A nomenclature for suicidology", in *Suicide & Life-Threatening Behavior,* vol. 26, 1996, pp. 237-252.

[OSS+04] A. Ozdas, R.G. Shiavi, S.E. Silverman, M.K. Silverman and D.M. Wilkes, "Investigation of Vocal Jitter and Glottal Flow Spectrum as Possible Cues for Depression and Near-term Suicidal Risk", in *IEEE Transactions on Biomedical Engineering,* vol. 51, no. 9, 2004, pp. 1530-1540.

[OSW+04] A. Ozdas, R.G. Shiavi, D.M. Wilkes, M.K. Silverman and S.E. Silverman, "Analysis of Vocal Tract Characteristics for Near-Term Suicidal Risk Assessment", in *Methods of Information in Medicine*, vol. 43, no. 1, 2004, pp. 36-38.

[Pac12] L. Pace, "One-Way Analysis of Variance", in *Beginning R: An Introduction to Statistical Programming*: Apress, 2012, ch. 10, pp. 139-147.

[PSM+08] M. Parellada, P. Saiz, D. Moreno, J. Vidal, C. Llorente, M. Álvarez, P. García-Portilla, A. Ruiz-Sancho, C. Arango, J. Bobes, "Is attempted suicide different in adolescent and adults?", in *Psychiatric Research*, vol. 157, 2008, pp. 131-137.

[Pen14] Pennebaker Conglomerates, Inc. (2014, Jun. 29). *Table 1: LIWC2007 Output Variable Information* [Online]. Available: http://liwc.net/descriptiontable1.php#

[Pen11] J.W. Pennebaker, *The Secret Life of Pronouns: What Our Words Say About Us*, 1st ed.: Bloomsbury Press, 2011.

[SP01] S.W. Stirman and J.W. Pennebaker, "Word Use in the Poetry of Suicidal and Nonsuicidal Poets", in *Psychosomatic Medicine,* vol. 63, 2001, pp. 517-522.

[PCI+14] J.W. Pennebaker, C.K. Chung, M. Ireland, A. Gonzales and R.J. Booth. (2014, May 2). *The Development and Psychometric Properties of LIWC2007* [Online]. Available: http://www.liwc.net/LIWC2007LanguageManual.pdf

[Pes10] Pestian, J., "A conversation with Edwin Shneidman", in *Suicide and Life-Threatening Behavior,*vol. 40, no. 5, 2010, pp. 516-523.

[PBS+11] K. Posner, G.K. Brown, B. Stanley, D.A. Brent, K.V. Yershova, M.A. Oquendo, G.W. Currier, G.A. Melvin, L. Greenhill, S. Shen and J.J. Mann, "The Columbia-Suicide Severity Rating Scale: Initial Validity and Internal Consistency Findings From Three Multisite Studies With Adolescents and Adults", in *Am J Psychiatry*, vol. 168, no. 12, 2011, Dec., pp. 1266-1277.

[RM99] W.M. Reynolds and J.J. Mazza, "Assessment of Suicidal Ideation in Inner-City Children and Young Adolescents: Reliability and Validity of the Suicidal Ideation Questionnaire-JR", in *School Psychology Review,* vol. 28, no. 1, 1999, pp. 17-30.

[SFB+97] R.E. Schapire, Y. Freund, P. Bartlett and W.S. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods", in *Machine Learning: Proceedings of the Fourteenth International Conference*, San Francisco, CA, USA, 1997, pp. 322-330.

[SHY+14] S. Scherer, Z. Hammal, Y. Yang, L.-P. Morency and J. Cohn, "Dyadic Behavior Analysis in Depression Severity Assessment Interviews", in *Proceedings of the ACM International Conference on Multimodal Interaction (ICMI)*, Istanbul, TUR, 2014 (accepted for publication).

[SPM13] S. Scherer, J. Pestian and L.-P. Morency, "Investigating the speech characteristics of suicidal adolescents", in *IEEE Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* Vancouver, BC, CAN, 2013, pp. 709-713.

[SW08] H. Sloetjes and P. Wittenburg, "Annotation by category – ELAN and ISO DCR", in *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, MAR, 2008.

[SS11] M. Spiegel and L. Stephens, "Analysis of Variance", in *Statistics*, 4th ed., USA: McGraw-Hill Companies, Inc., 2011, ch. 16, pp. 403-407.

[SW10] C. Sammut and G.I. Webb, "Sensitivity", in *Encyclopedia of Machine Learning*, US: Springer, 2010, p. 901.

[WHO14] World Health Organization. (2014, March 15). *Mental Helath-Suicide Prevention (SUPRE)* [Online]. Available: http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/