ADOLESCENT SUICIDAL RISK ASSESSMENT IN CLINICIAN-PATIENT INTERACTION: A STUDY OF VERBAL AND ACOUSTIC BEHAVIORS

Verena Venek [1,2], Stefan Scherer[1], Louis-Philippe Morency[1], Albert (Skip) Rizzo[1], John Pestian[3]

[1]University of Southern California, Institute for Creative Technologies, Playa Vista, CA
[2]Carinthia University of Applied Sciences, Klagenfurt, Austria, Europe
[3]Department of Pediatrics, Biomedical Informatics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, OH

ABSTRACT

Suicide among adolescents is a major public health problem: it is the third leading cause of death in the US for ages 13-18. Up to now, there is no objective way to assess the suicidal risk, i.e. whether a patient is non-suicidal, suicidal re-attempter (i.e. repeater) or suicidal non-repeater (i.e. individuals with one suicide attempt or showing signs of suicidal gestures or ideation). Therefore, features of the conversation including verbal information and nonverbal acoustic information were investigated from 60 audio-recorded interviews of 30 suicidal (13 repeaters and 17 non-repeaters) and 30 non-suicidal adolescents interviewed by a social worker. The interaction between clinician and patients was statistically analyzed to reveal differences between suicidal vs. non-suicidal adolescents and to investigate suicidal repeaters' behaviors in comparison to suicidal non-repeaters. By using a hierarchical ensemble classifier we were able to successfully discriminate non-suicidal patients, suicidal repeaters and suicidal non-repeaters.

Index Terms— Suicide, adolescent, interaction, re-attempt, hierarchical classifier

1. INTRODUCTION

Youth suicide is the third leading cause of death for ages 13-18 in the United States of America. Adolescents and young adults have the highest number of suicide attempts and many never seek professional care. Thus, suicidal behavior often remains untreated or undetected. The National Youth Risk Behavior Surveillance [1] stated that 17% of high school students (i.e. ages 14-18) seriously considered attempting suicide, while 13.6% made a suicide plan and 8% actually attempted one or more times. Detecting suicidal risk (i.e. suicidal or non-suicidal patient) in the primary care settings, for example in the Emergency Departments (ED) in hospitals, is advantageous. There, an interaction between patient and clinician can yield to a risk assessment. This interaction is mainly based on written questionnaires and scales, especially developed to investigate suicidal behavior in adolescents [2][3]. Hence, mainly the patients' verbal behavior is investigated by the clinicians to decide the level of risk or even to prevent another suicide attempt [4][5]. Especially the lack of time plays a crucial role for the clinicians in the ED. There is a need for a valid, brief and accurate way to identify suicidal risk, especially to predict a potential re-attempt of the suicidal patients [6].

The aim of this project is to investigate if suicidal risk can be determined by observing the patient-clinician conversation including the communicated verbal information and voice characteristics. Hence, the research aims to investigate if it is possible to support the suicide risk assessment by objectively quantifying behaviors during the interaction of clinician and patient during an interview. Furthermore, the study also considers non-repeaters and repeaters and observes if there is a possibility to identify them by observing the dynamic between interviewer and interviewee. In this study, adolescent suicidal patients which attempted suicide more than once are referred as repeaters. The group of non-repeaters includes suicidal adolescents with one or none suicidal attempt, but at least showing signs of suicidal gestures or ideation.

The following section presents related work in suicidal risk assessment. Section 3 and 4 give an overview of the used dataset and the investigated features. In section 5 the statistical results of the feature analysis are given and in section 6 the hierarchical classification is described with results. This work concludes with a discussion and a conclusion of the performed study.

2. RELATED WORK

Several researchers have investigated the correlates between severe depression, suicide, and the characteristics of speech [5][7][8]. This work was motivated by investigations of [7], which analyzed 16 interviews with adolescents from the dataset of the Cincinnati Children's Hospital Medical Center (CCHMC) ED. The researchers analyzed 16 interviews including 8 suicidal and 8 non-suicidal patients with ages between 13 and 17. One of the research aims was the extraction of discriminative speech features for a standard machine learning classification algorithm. They were able to achieve classification accuracies of 81.25% by using the Hidden Markov Model and 75% using the Support Vector Machine algorithm (SVM). They were also able to identify features related to the patients' speech which enhanced the classification results. They revealed that suicidal adolescents spoke usually with breathier voices than the non-suicidal patients. Moreover, [7] stated that the clinician's voice adapted to the patient's one.

In [8], the speech of 10 male suicidal, 10 male depressed, and 10 male control subjects, ages 25 to 65, was analyzed in great detail. The data for the suicidal subjects were obtained from a large spectrum of recording setups comprising, for example, suicide notes recorded on tape. The other two groups were recorded under more controlled conditions at Vanderbilt University. For each subject the researchers concatenated speech to clips of 30 seconds

of uninterrupted speech (i.e., removing pauses larger than 500ms). Then they analyzed jitter in the voiced parts of the signal as well as glottal flow spectral slope estimates. Both features helped to discern the classes in binary problems with high above-chance accuracies by utilizing simple Gaussian mixture model-based classifiers (e.g., control vs. suicidal 85% correct, depressed vs. suicidal 75% correct, control vs. depressed 90% correct). A holdout validation was employed. However, the fact that the recordings were done over such a large variance of recording setups, as acknowledged by the authors themselves, makes it difficult to assess "the accuracy about the extracted speech features and, therefore, the meaningfulness of the classification results." Nevertheless, the fact that the researchers have analyzed real-world data with speech recorded from subjects shortly before they attempted suicide is remarkable and needs to be acknowledged.

Further, in [4] the nonverbal communication in interviews between doctors and suicidal patients was investigated to classify between repeaters (i.e. patients who re-attempted suicide within the next 24 months) and non-repeaters (i.e. patients who *did not* re-attempted within the next 24 months) by using coded facial behavior of the interlocutors. In the repeater's group, a broader less frequently occurring variation of patterns were able to be detected. Furthermore, the researchers stated that the nonverbal behavior of the interviewer accurately reflects which patients were repeaters and which were not. In [5]**,** Stirman and Pennebaker investigated the word use in the poetry of suicidal and non-suicidal poets by performing Linguistic Inquiry and Word Count (LIWC) analyses of the poets' works. Their social integration theories stated that suicidal poets used more references to themselves and showed a reduced use of words related to others, e.g. suicidal poets used more self-related words like "I" or "my" as well as they spoke less about their families or friends.

The difference of the present work to others, especially to [7], is that, in addition to the investigation of acoustic features, conversational information and verbal 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, suicidal vs. non-suicidal adolescents and suicidal repeaters vs. suicidal non-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 repeaters and non-repeaters.

3. DATASET

Within a controlled trial from March 2011 through October 2011, 60 interviews with 30 suicidal and 30 non-suicidal adolescent patients from the Cincinnati Children's Hospital Medical Center (CCHMC) Emergency Department (ED) have been recorded. Thirty male and thirty female adolescents were interviewed by one single trained social worker and asked to respond to 16 questions comprised of the Columbia Suicide Severity Rating Scale (C-SSRS version 1/14/2009 [2]), Suicidal Ideation Questionnaire-Junior (SIQ-JR version 1987 [3]) and the Ubiquitous Questionnaire (UQ version 2011 [9]). 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 σ=1.5). As potential subjects, 30 patients were chosen that had come to the ED with suicidal ideation, gestures or attempts. 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 > 1, 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 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 participation of the patients had to be consent by their parent(s) or legal guardian(s) and him- or herself. Furthermore, he or she 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. 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. In comparison, the average length of the interviews with the controls were almost halved: interviews lasted approximately 490 seconds.

4. INVESTIGATED FEATURES

In this section, the investigated audio-based features which are obtained by analyzing the interviews' transcripts and acoustic feature data are introduced.

Conversation dynamic features are extracted by analyzing the interviews' transcripts by using Matlab. This feature group includes *speak* and *pause time* percentages of clinician and patients as well as *words per second rates* and *overlap rates*. If clinician or patient do not allow their interlocutors finish speaking, i.e. interrupt each other in the middle of the sentence, including words of agreement and of incentive, this is considered as overlap. The rate of the overlaps is maintained by dividing the number of overlaps by the duration of the interview.

Verbal information features are gathered by analyzing the transcript data of the interviews using LIWC software [10]. The features are word category scales related to 80 categories provided by the LIWC analysis separated for patients and clinician. The utilized verbal features are among others standard linguistic dimensions like *personal pronouns*, *1st person singular pronoun* like 'I, my, mine', *impersonal pronouns* and terms indicating *past tense* and *negation*. Moreover, the word categories related to *positive emotion* and *negative emotion* are used as well as *tentative* words like 'maybe, perhaps' or 'guess'. Also the paralinguistic dimensions *nonfluencies* like 'er, hm, umm' and *assent* words like 'agree, okay, yes' are investigated.

Acoustic information. For the processing of the speech signals, the freely available COVAREP toolbox, a collaborative speech analysis repository available for Matlab and Octave [11][1] is used.

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

COVAREP provides an extensive selection of open-source robust and tested speech processing algorithms enabling comparative and cooperative research within the speech community. Furthermore, the acoustic features of the clinician's backchannel are analyzed. The backchannel is defined as speech segments of the interviewer with durations smaller than 700ms. These patches include words of assent, non-fluencies, or fillers like 'uhm'.

Below the utilized acoustic features, including the ones which characterize voice qualities from breathy to tense dimension, are introduced in more detail.

- *Fundamental frequency (f_0)*

This parameter includes the pitch information of individuals' speech. The method for a f_0 tracking and simultaneous voicing detection based on residual harmonics is introduced in [12]. Unvoiced speech segments, i.e. times when no vocal fold vibration appears, were not analyzed for any of the extracted features.

- *Normalized Amplitude Quotient (NAQ)*

The NAQ describes the normalized amplitude quotient of the differentiated glottal flow. [13]

- *Quasi-Open Quotient (QOQ)*

The QOQ is measured by detecting the peak in the glottal flow and finding the time points previous to and following this point that descend 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. [13]

- *Parabolic Spectral Parameter (PSP)*

This measure 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 f_0 of voices is different. [14]

- *Maxima Dispersion Quotient (MDQ)*

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. [15]

- *Peak Slope (PS)*

The feature is essentially an effective correlate of the spectral slope of the speech signal. [15]

- *Liljencrants-Fant model parameter Rd*

This measure is one of the *R*-parameters of the Liljencrants-Fant (LF [16]) model characterizing the glottal source. *Rd* captures most of the covariation of the LF model parameters. Reference [17] has shown that this feature improved the classification of different levels of vocal effort from expressive speech significantly.

- *Formants (F1, F2)*

The tracking of the formants is introduced in detail in [18]. The first and the second formants F1 and 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.

5. STATISTICAL ANALYSIS

Two cases are statistically analyzed by using ANOVA: suicidal vs. non-suicidal adolescents and suicidal repeaters vs. non-repeaters. The significance level is stated to be at least $p < 0.05$. In addition, the mean and the standard deviations of the individual features are calculated.

The significant features are expected to characterize verbal and acoustic properties of suicidal adolescents and interviewer behavior. The discriminative faculty of the identified features is then confirmed by the machine learning classification (see section 6). Statistical results corresponding to the clinician's speech or patients are specified with subscripts $_C$ and $_P$, respectively.

5.1. Suicidal vs. non-suicidal evaluation

This sub-section introduces the statistical results of the ANOVAs of the investigated features between suicidal patients and their controls. The statistical results are listed in Table 1 and Table 2.

Table 1: Statistical significant results of the patients' features of the suicidal vs. non-suicidal evaluation

	Suicidal	Non-suicidal	
Feature	**μ (σ)**	**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)	**
Non-fluencies	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)	**

Conversation. The words per second rate of the clinician is higher at interviews with non-suicidal adolescents (μ_C= 2.94 words per second) than the one with the suicidal subjects (μ_C= 2.78 words per second, $p_C<0.01$). The speaking time given in percentage is significantly different. The suicidal patients speak 43% on average of the whole interview, while the non-suicidal adolescents occupy 32% of the interview ($p_P<0.01$). The interviewer speaks 31% when interacting with a suicidal adolescent and 46% with a non-suicidal one ($p_C<0.01$). Also the pauses between the speech segments are 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 protrudes with the mean pause time percentage of 15% ($p_C<0.05$). Patients speak less over their interlocutor than the clinician does. Thus, the clinician-speaks-over-patient rate shows a significant difference ($p<0.01$).

Verbal information. For the analyses related to the verbal information, the data is separated into clinician's and patients' feature datasets. Suicidal patients use more often personal pronouns than their controls ($p_P<0.01$), especially self-related first personal singular pronouns ($p_P<0.01$). Moreover, suicidal patients refer to the past 3.7% of the total interview while non-suicidal adolescents do so 2.1% on average ($p_P<0.01$). Also the clinician refers more often to the past while speaking to suicidal patients ($p_C<0.01$). Adolescent controls use more often assent words than suicidal subjects ($p_P<0.01$). More often non-fluencies are observed in controls' interviews ($p_P<0.01$). The clinician uses more non-

fluencies during interviews with the suicidal patients (p_C<0.05). Non-suicidal patients use more often terms related to negation than the suicidal subjects (p_P<0.01). Suicidal adolescents use terms related to negative emotion more often than their controls (p_P<0.01) and non-suicidal patients use words correlated to positive emotion more often (p_P<0.01). Tentative terms are used 5.6% on average by non-suicidal adolescents, while suicidal patients use them on average 4.6% of the whole duration (p_P<0.01). Also the clinician uses tentative words more often in interviews with non-suicidal patients (p_C <0.01). Furthermore, the use of impersonal pronouns shows a significance for clinicians and patients (p_C<0.01 and p_P <0.01, respectively). The clinician has a significant use of second person, third person singular and first person plural pronouns. Hence, the feature personal pronoun is also significantly and is selected as proper classification feature (p_C <0.01).

Table 2: Statistical significant results of the clinician's features of the suicidal vs. non-suicidal evaluation

	Suicidal	Non-suicidal	
Feature	**μ (σ)**	**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)	0.05
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)	**
Non-fluencies	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)	**

Acoustic information. The clinician's and patients' acoustic features are investigated separately. Suicidal patients speak on average with a lower f_0 than their controls (p_P<0.01). The clinician shows the same pattern (p_C<0.01). The NAQ and QOQ measures are higher, i.e. the patients and clinician speak with a breathier voice during interviews with suicidal patients. This is also valid for the MDQ parameter (p_C<0.01and p_P<0.01). Moreover, the PS measure implies a breathier conversation with suicidal patients (p_C <0.01, p_P <0.01). The PSPs are 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 is higher during interviews with the suicidal patients (p_C<0.01, p_P<0.01). From the formants only the F1 measure is 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) shows that the clinician reacts 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 are respectively higher during interviews with suicidal patients (p_C<0.01and p_C<0.01 respectively). Also the PSP is significantly different (p_C<0.01).

5.2. Suicidal repeater vs. non-repeater evaluation

In this section, the 30 recorded interviews with the suicidal patients are statistically analyzed to determine significant features and distinctions between suicidal repeaters and non-repeaters. The results are summarized in Table 3.

Conversation. 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).

Table 3: Statistical significant results of the repeater vs. non-repeater evaluation

	Repeater	Non-repeater	
Feature	**μ (σ)**	**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)	*

Verbal information. In the repeater vs. non-repeater case, there are not found significant differences in this feature group. Although the suicidal repeaters use first person singular pronouns slightly less often than the non-repeaters, statistically significant differences are not found for the pronouns, see Fig. 1.

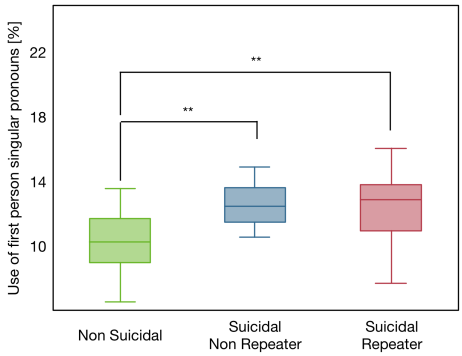

Fig. 1: Boxplot of patients' use of first singular pronouns during the interviews. This feature is significantly different between suicidal and non-suicidal patients but not significant enough to discriminate between repeaters and non-repeaters.

Acoustic information. In the interviews with the suicidal repeaters, patients and clinician speak on average with a lower

fundamental frequency while interacting with the non-repeaters. The other more significant acoustic features including NAQ, QOQ, MDQ and PS show results characterizing a breathier voice during interviews with the non-repeaters (see Fig.2). Except the PSP: the suicidal non-repeaters have on average higher values than the repeaters (p_P<0.01). The PSP of the clinician, instead, is higher (i.e. breathier voice) during interviews with suicidal repeaters (p_C<0.01). The Rd parameter of the clinician during interviews has a higher value while speaking to non-repeaters (p_C<0.01).

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

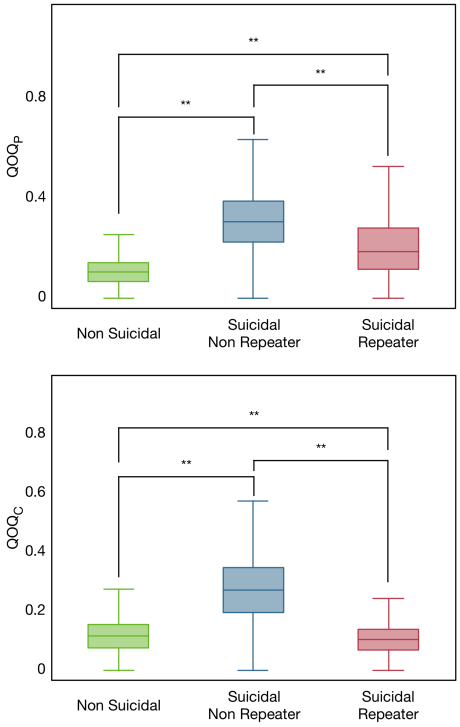

Fig. 2: Boxplot comparison of patients' and clinician's QOQ during the interviews. Interactions with suicidal patients show a breathier conversation than with the non-suicidal ones.

6. HIERARCHICAL CLASSIFICATION

Subject independent classification is performed to confirm and identify the discriminative faculty of the investigated features. The *AdaBoostM1* algorithm, as described in [19], is used to create the classifier. The testing of the ensemble classification is performed with a leave-one-speaker-out approach. In Fig. 3 the hierarchical structure of the classification is illustrated. The task of the classification between non-suicidal, suicidal repeaters and suicidal non-repeaters is realized by separating the discrimination into two layers: first, an ensemble classifier which discriminates suicidal from non-suicidal adolescents is trained and tested with a classification matrix containing the investigated features of the 60 interviews. In the end, 41 features (see Table 1 and Table 2) are used for this classification. In the second layer, another ensemble classifier is trained, however, by using 18 (see Table 3 without verbal information features) almost only acoustical features and discriminating the suicidal classified interviews by the previous layer. An accuracy of 90% is achieved for the suicidal vs. non-suicidal distinction while the suicidal repeaters vs. suicidal non-repeaters layer classification delivers an accuracy of 60%. 60 interviews enter the classification stage, 27 of them are correctly labeled as non-suicidal. The positively labeled ones are forwarded to the repeaters vs. non-repeaters level, and there 7 true positives (i.e. repeaters) and 10 true negatives (i.e. non-repeaters) are able to be identified correctly. The confusion matrix of the hierarchical classifier is provided in Table 4. The classification over the complete hierarchy yields an accuracy of 73.3%. The corresponding F1 scores, precision and recall measures are given in Table 5.

Table 4: Confusion matrix of hierarchical classification

Actual \ Prediction	Non-suicidal	Suicidal non-repeater	Suicidal repeater
Non-suicidal	27	1	2
Suicidal non-repeater	2	10	5
Suicidal repeater	1	5	7

Table 5: F1-scores, precision and recalls of the hierarchical classification

Condition	F1-score	Precision	Recall
Non-suicidal	0.90	0.90	0.90
Suicidal non-repeater	0.61	0.63	0.59
Suicidal repeater	0.52	0.50	0.54

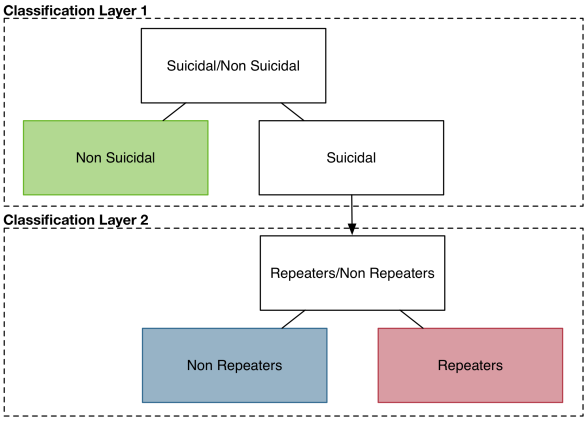

Fig.3: Hierarchical structure of the ensemble classification

7. DISCUSSION

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

In the second case, suicidal repeaters vs. suicidal non-repeaters, are not as many significant differences identified. 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.

Regarding the backchannel of the clinician we found that almost each significant acoustic feature of the clinician's complete dataset is observed only in the backchannel information too. Although the clinician knows about the state of the patient, the adaptation of the clinician's voice to the patient's one could be already observed by just observing the speech fragments that lasted less than 700ms. The clinician speaks with a breathier voice to the suicidal repeater patients than to the non-suicidal ones.

Between suicidal and non-suicidal patients it is easier to differentiate than between suicidal repeaters and non-repeaters because more information of the conversation can be used for the classification. The hierarchical structure of two subsequent ensemble classifiers reveals satisfying results discriminating non-suicidal adolescents, suicidal repeaters, and suicidal non-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 between non-suicidal and suicidal adolescents. However, the discrimination between suicidal repeaters and non-repeaters in the second layer requires especially nonverbal acoustic information, none verbal information and just two conversational features of the interviews.

8. CONCLUSION

In this study, the ability to classify non-suicidal patients, suicidal repeaters and suicidal non-repeaters between the ages of 13 and 18 is investigated. Therefore, statistical analyses reveal significant features of the interaction between patients and clinician. These are expected to characterize verbal and acoustic properties of suicidal adolescents and interviewer behavior. Verbal information features are proven to be useful to discriminate non-suicidal vs. suicidal adolescents. For the discrimination of suicidal repeaters and non-repeaters nonverbal acoustic information is shown to be most useful. The discriminative faculty of the identified features is able to be confirmed by the hierarchical ensemble classification which yielded an accuracy of 73.3%.

While our study shows promising results it does not consider the future but rather the past of the suicidal repeaters and non-repeaters. Further, the performance of the repeater vs. non-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 repeaters, and suicidal non-repeaters is possible by considering the verbal and nonverbal behaviors of patients and clinician as well as the dynamics between interviewer and interviewee. Hence, a progress in the additional support of suicidal risk assessment of adolescents could be identified.

9 . REFERENCES

[1] L. Kann et al., "Youth Risk Behavior Surveillance - United States, 2013", *Centers for Disease Control and Prevention MMWR Surveillance Summaries*, vol. 63, no. 4, June 2014.

[2] K. Posner et al., "The Columbia-Suicide Severity Rating Scale: Initial Validity and Internal Consistency Findings From Three Multisite Studies With Adolescents and Adults", *Am J Psychiatry*, vol. 168, no. 12, pp. 1266-1277, December 2011.

[3] 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", *School Psychology Review*, vol. 28, no. 1, pp. 17-30, 1999.

[4] V. Haynal-Reymond, G.K. Jonsson and M.S. Magnusson, "Nonverbal Communication in Doctor-Suicidal Patient Interview", *The Hidden Structure of Interaction*, IOS Press, ch.9, 2005.

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

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

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

[8] 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", *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1530-1540, 2004.

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

[10] J.W. Pennebaker, C.K. Chung, M. Ireland, A. Gonzales, R.J. Booth. *The Development and Psychometric Properties of LIWC2007* [Online]. Available: http://www.liwc.net/LIWC2007LanguageManual.pdf

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

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

[13] S. Scherer, Z. Hammal, Y. Yang, L.-P. Morency and J.F. Cohn, "Dyadic Behavior Analysis in Depression Severity Assessment Interviews," *ICMI '14*, 2014.

[14] P. Alku, H. Strik and E. Vilkman, "Parabolic spectral parameter - A new method for quantification of the glottal flow", *Speech Communication*, vol. 22, pp. 67-79, 1977.

[15] J. Kane, "Tools for analysing the voice - Developments in glottal source and voice quality analysis", PhD dissertation, Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Ireland, 2012.

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

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

[18] B. Bozkurt, B. Doval, C. d'Alessandro and T. Dutoit, "Improved differential phase spectrum processing for formant tracking", *Proceedings of Interspeech ICSLP*, pp. 2421-2424, 2004.

[19] R.E. Schapire, Y. Freund, P. Bartlett and W.S. Lee, "Boosting the margin: A new explanation for the effectiveness of voting methods", *Annals of Statistics*, vol. 26, no. 5, pp. 1651-1686, 1998.

.