Adolescent Suicidal Risk Assessment in Clinician-Patient Interaction

Verena Venek, Stefan Scherer, *Member, IEEE*, Louis-Philippe Morency, *Member, IEEE*, Albert "Skip" Rizzo, *Member, IEEE*, and John Pestian

Abstract—Youth suicide is a major public health problem. It is the third leading cause of death in the United States for ages 13 through 18. Many adolescents that face suicidal thoughts or make a suicide plan never seek professional care or help. Within this work, we evaluate both verbal and nonverbal responses to a five-item ubiquitous questionnaire to identify and assess suicidal risk of adolescents. We utilize a machine learning approach to identify suicidal from non-suicidal speech as well as characterize adolescents that repeatedly attempted suicide in the past. Our findings investigate both verbal and nonverbal behavior information of the face-to-face clinician-patient interaction. We investigate 60 audio-recorded dyadic clinician-patient interviews of 30 suicidal (13 repeaters and 17 non-repeaters) and 30 non-suicidal adolescents. The interaction between clinician and adolescents is statistically analyzed to reveal differences between suicidal versus non-suicidal adolescents and to investigate suicidal repeaters' behaviors in comparison to suicidal non-repeaters. By using a hierarchical classifier we were able to show that the verbal responses to the ubiquitous questions sections of the interviews were useful to discriminate suicidal and non-suicidal patients. However, to additionally classify suicidal repeaters and suicidal non-repeaters more information especially nonverbal information is required.

Index Terms—Behavior analytics, clinician-patient interaction, hierarchical classifiers, ubiquitous questions, youth suicide

1 Introduction

In the United States of America suicide is the third leading cause of death for ages 13 to 18 according to the Centers for Disease Control and Prevention.[1] Adolescents and young adults have the highest number of suicide attempts and the number unreported incidents of suicidal ideation or gestures can only be speculated about. One problem of suicide prevention is that many who suffer never seek professional care or help. This makes it difficult to assess suicidal risk of young patients. Overall, suicidal behavior often remains untreated or undetected. Automatic suicidal risk assessment based on behavioral analyses in short interviews could help to identify suicidal behavior effectively and pervasively. To improve the success of suicidal risk classification and assessment we need to understand and investigate verbal and nonverbal behavior of suicidality [1], [2], [3], [4].

In this work, we investigate interviews of one clinician with 60 young patients 30 suicidal and 30 non-suicidal adolescents. Within this work we utilize the ubiquitous questionnaire (UQ) [1] that consists of five open-ended ubiquitous questions. These questions address patients' emotional states and aim to allow the patient to freely disclose their feelings rather than require responses of many more specific questions, e.g., in a written questionnaire that often result in basic yes/no responses. Here, we investigate both verbal and nonverbal behaviors of the entire interviews and objectively quantify the interaction of clinician and patient. In particular, two distinct segments of the interviews are investigated: (1) the UQ segment containing only the ubiquitous questions (see Table 1) and (2) the non-UQ segment, comprising questions about the patients' sleeping patterns, friends and social-networking habits (see Table 2). Furthermore, our investigations focus both on suicidal versus non-suicidal adolescents' behaviors as well as repeater versus non-repeater behaviors. Within this context, it is important to identify repeaters as their suicidal risk is considerably higher. Within this work, adolescents with repeated suicidal attempts in the past are referred as repeaters throughout this paper. The group of non-repeaters consists of suicidal adolescents with one or no suicidal attempt, but at least showing signs of suicidal gestures or ideation. Please note that the present work is a retrospective study, which could lead to a number of confounding behavior on both the clinician's as well as the participants' side. However, we believe that our evaluations are an important stepping-stone towards technologies that are capable of predicting suicidal behaviors in the future. In particular, due to the focus of our investigations on nonverbal and verbal responses to more indirect questions rather than questions that heavily rely on self-disclosure. As acknowledged in related work (in particular with respect to military culture)

1. http://www.cdc.gov/violenceprevention/suicide/youth_suicide.html

- V. Venek is with the University of California, Institute for Creative Technologies, Playa Vista, CA, USA, and with the Carinthia University of Applied Sciences, Klagenfurt, Austria, Europe. E-mail: verena.venek@gmx.at.
- S. Scherer and A. Rizzo are with the University of California, Institute for Creative Technologies, Playa Vista, CA, USA. E-mail: {scherer, rizzo}@ict.usc.edu.
- L.-P. Morency is with the Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA. E-mail: morency@cs.cmu.edu.
- J. Pestian is with the Department of Pediatrics, Biomedical Informatics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, OH, USA. E-mail: John.Pestian@cchmc.org.

Manuscript received 13 Aug. 2015; revised 8 Jan. 2016; accepted 13 Jan. 2016. Date of publication 17 Jan. 2016; date of current version 6 June 2017. Recommended for acceptance by R.A. Calvo.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TAFFC.2016.2518665

TABLE 1
Ubiquitous Questions

ID	Question
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?

The five open-ended questions which were asked in the first segment of each interview defining the UQ case.

TABLE 2
Non-Ubiquitous Questions

ID	Question
Q6	Have you ever had anybody, either in your family or a friend that you know of that may have tried suicide or committed suicide?
Q7	How many hours in a day do you think you spend on the internet?
Q8	How close or connected do you feel to your family?
Q9	How close or connected do you feel to your peers?
Q10	Do you attend a religious service on a regular basis?
Q11	Do you have any past history of like physical abuse or neglect or anything like that?
Q12	Do you have any access to firearms?
Q13	On a school night how much do you think you sleep?
Q14	And what about on a weekend when you don't have school?
Q15	Do you feel you currently have insomnia where you have problems sleeping?
Q16	Have there been any recent changes in your sleep habits?

The 11 open-ended questions which were asked in the second segment of each interview defining the non-UQ case.

the willingness to disclose such information to healthcare providers is lacking due to a number of reasons, including fear of possible repercussions and stigmatization [5]. In addition, a study investigating a very similar population to the one investigated in the present work (i.e., adolescents) have identified that in-fact low self-disclosure is closely associated and significantly correlated with suicidal thinking as well as suicidal attempts [6]. The present work establishes indicators that do not require the subject to disclose specific information.

In order to distinguish suicidal versus non-suicidal adolescents and repeaters versus non-repeaters, we perform an automatic hierarchical classification experiment. With the classification experiment we investigate the discriminative faculty of the identified nonverbal and verbal features. In addition to the investigations of specific behaviors we seek to identify the discriminative faculty of the five ubiquitous questions to assess suicidal risk among adolescents within clinician-patient interactions. To achieve this, the statistical results as well as the classification results of the UQ case are compared to these of the non-UQ case.

The remainder of this paper is organized as follows: Section 1.1 states our main research questions. In Section 2 we introduce related work and provide an overview of the used dataset and the investigated features in Sections 3 and 4. The statistical results of the feature analysis are provided for the suicidal versus non-suicidal patients and for the suicidal repeaters versus non-repeaters in Section 5. The classification results and training datasets are described and shown in Section 6. The most important and promising findings of this study are summarized in the discussion and conclusions Sections 7 and 8.

1.1 Research Questions

Within this work, we address three main research questions:

- **RQ1:** Is it possible to recognize the suicidal risk of a person by analyzing the verbal and nonverbal behavior patterns within structured clinician-patient interactions? In particular, we seek to identify characteristics of suicidal speech in the interviews and distinguish between suicidal and non-suicidal adolescents' behavior.
- **RQ2:** Is it possible to recognize if a patient has already attempted suicide more than once by analyzing the verbal and nonverbal behavior patterns? We want to find characteristics of repeaters' speech patterns and distinguish repeaters from non-repeaters in the classification task.
- **RQ3:** Is it possible to recognize the suicidal risk of a person by using the verbal and nonverbal interaction of clinician and young patients within the ubiquitous question context only? In particular, we want to identify which part of the interaction produces the most distinguishable behavioral patterns. Specifically, we split the interview into two segments: the UQ segment and non-UQ segment of the interview.

2 Related Work

Several researchers have investigated the correlates between severe depression, suicide, and the characteristics of voice and speech [2], [3], [4], [7], [8], [9], [10], [11], [12]. This work was motivated by investigations of [2], which analyzed 16 interviews with adolescents from the dataset of the Cincinnati Children's Hospital Medical Center (CCHMC) Emergency Department (ED). The researchers analyzed 16 interviews including eight suicidal and eight 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 percent by using the Hidden Markov Model and 75 percent 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, [2] stated that the clinician's voice adapted to the patient's one.

In [3] 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 500 ms). 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 versus suicidal 85 percent correct, depressed versus suicidal 75 percent correct, control versus depressed 90 percent correct). Lachenbruch's holdout validation was employed, in which a single data point is left out of the training of the classifier. This held out data point is then later utilized in the test phase of the approach. 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 [7] 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-attempt 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 could be detected. Furthermore, the researchers stated that the nonverbal behavior of the interviewer accurately reflects which patients were repeaters and which were not.

In [13], 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.

Gunn and Lester also utilized LIWC in their work and investigated language changes in Twitter posts of a female adolescent 24 hours prior to her suicide [10]. They found that for example contrary to the observations within the present work and the Stirman and Pennebaker study positive emotional words increased towards the last hours prior to the suicide. In addition, they observed a trend of decreasing use of self-related words towards the suicide in their work. While these findings are of significant interest the researchers have stated some concern about the authenticity of the postings as well as the brevity of the posts.

Zhang et al. further investigated the predictability of suicidal risk using latent Dirichlet allocation as a feature extraction procedure and linear regression as an automatic approximation approach [12]. The utilized data was obtained from microblog entries on a Chinese blogging platform. The researchers achieved a minimal root mean square error with this approach of around 11 for an approximation on a 1-32 scale for the suicide probability scale. In related work by O'Dea et al. around 14,000 Tweets potentially containing suicide related material were collected and manually coded for level of "concern". Human raters agreed moderately with a $\kappa = 0.55$ and an automatic machine learning approach based on support vector machines approached human performance with an overall accuracy of 76 percent [11].

In a similar study to the present work [14] the ability to classify non-suicidal patients, suicidal repeaters and suicidal non-repeaters between the ages of 13 and 18 was investigated. The behaviors of the entire interviews were objectively quantified during the interaction of clinician and patient. Verbal information features were proven to be useful to discriminate non-suicidal versus suicidal adolescents. For the discrimination of suicidal repeaters and non-repeaters nonverbal acoustic information was shown to be most useful. The discriminative faculty of the identified features could be confirmed by the hierarchical ensemble classification, which yielded an accuracy of 73.3 percent.

The difference of the present work to others, especially to [14], is that the interviews are separated into the UQ case, containing verbal and nonverbal behavior features of the segments including the five ubiquitous questions, and the non-UQ case. The non-UQ case consisting of the residual eleven questions of the interview is investigated to offer a comparison. The ability of the verbal and nonverbal information of the five open-ended ubiquitous questions to assess suicidal risk is observed. Furthermore, a different approach of the hierarchical classification is introduced. The classification is separated into two layers to label non-suicidal adolescents, suicidal repeaters and suicidal non-repeaters. Support vector machine classification is used to distinguish suicidal and non-suicidal patients. In the second layer the suicidal patients are classified into repeaters and non-repeaters by using an ensemble classifier.

3 Dataset

The investigated dataset is the same as in [14] but has been extended since the conduction of the study. 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 Emergency Department have been recorded under institutional review board approval (IRB #2008-1421).

Thirty male and 30 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 [15]), Suicidal Ideation Questionnaire-Junior (SIQ-JR version 1987 [16]) and the Ubiquitous Questionnaire (UQ version 2011 [1]). 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 of 15.47 years with $SD = 1.5$). 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 16 kHz. Moreover, all interviews were transcribed on a question-response level by using ELAN annotation software. In general, all the interviews with suicidal patients last longer than those with the control ones. The mean duration of the interviews with suicidal patients is 869 seconds. In comparison, the average lengths of the interviews with the controls are almost halved: interviews last approximately 490 seconds. In the UQ case, the interviews' parts last 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 are asked.

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. The features are separated into three groups: conversation dynamic features, verbal information and acoustic information features. Each is extracted from clinician's and patients' interview sections.

Conversation dynamic features are extracted by analyzing the interviews' transcripts by using Matlab. This feature group includes speech and pausetime percentages of clinician and patients as well as words per second rates and overlap rates. Overlaps are defined as interrupting the speech of the interlocutor. For example, the clinician or patient does not allow his or her interlocutor to finish a sentence by interrupting with words of agreement or of incentive. 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 [17]. 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 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 (v.1.2.0), a collaborative speech analysis repository available for Matlab and Octave [18] is used. 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 700 ms. These patches include words of assent, non-fluencies, or fillers like "uhm". Acoustical measures, which have been useful in several studies 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 paper.

- *Fundamental frequency (f_0):* This parameter includes the pitch information of individuals' speech. The method for f_0 tracking and simultaneous voicing detection based on residual harmonics is introduced in [19]. 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 [20].
- *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 percent of the peak amplitude. The duration between these two time points is divided by the local glottal period to get the QOQ measure [21].
- *Parabolic spectral parameter (PSP):* This measure is derived by fitting a parabolic function to the lower frequencies of 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 [22].
- *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 [23].
- *Peak slope (PS):* The feature is essentially an effective correlate of the spectral slope of the speech signal [23].
- *Liljencrants-fant model parameter Rd:* This measure is one of the R-parameters of the Liljencrants-Fant (LF [24]) model characterizing the glottal source. Rd captures most of the covariation of the LF model parameters. Reference [25] 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 [26]. 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

The interview segments containing the five ubiquitous questions and the answers to them are investigated separately against the residual verbal and nonverbal content of

TABLE 3
Suicidal versus Non-Suicidal Patients' Mean Values and Statistical Significant Results

	UQ		Non-UQ	
	Suicidal	**Non-S.**	**Suicidal**	**Non-S.**
Overlap rate	0.93	0.21*	1.31	1.32
Speech time	0.42	0.32**	0.42	0.33**
Pause time	0.14	0.08**	0.12	0.07**
Personal Pron.	16.52	13.36**	16.74	13.41**
1st person Pron.	13.31	11**	11.92	9.53**
Impersonal Pron.	7.99	6.07**	5.21	4.88
Past Tense	3.07	1.89**	4.31	2.35**
Negation	3.97	6.54**	4.56	5.45
Positive emotion	3.21	4.22**	2.88	3.55
Negative emotion	3.84	2.58**	1.42	0.65**
Tentative	4.76	5.26	3.88	6.04**
Death	0.4	0.33	0.34	0.14*
Nonfluencies	1.72	3.89**	2.29	3.87*
Assent	1.98	4.79**	2.14	4.44**
f_0	224.03	149.01**	210.96	152.02**
NAQ	0.08	0.03**	0.08	0.03**
QOQ	0.31	0.11**	0.29	0.11**
PSP	0.36	0.52**	0.34	0.48**
MDQ	0.13	0.11**	0.14	0.1**
PS	−0.19	−0.25**	−0.22	−0.24**
Rd	1.65	1.09**	1.56	1.11**
F1	613.56	538.75**	641.48	549.77**

Suicidal versus Non-Suicidal (i.e., Non-S.) nonverbal and verbal patients' behavior observations. Features are abbreviated as introduced in Section 4 and pronouns are abbreviated with Pron.

TABLE 4
Suicidal versus Non-Suicidal Clinician's Mean Values and Statistical Significant Results

	UQ		Non-UQ	
	Suicidal	**Non-S.**	**Suicidal**	**Non-S.**
Overlap rate	1.84	3.59*	0.98	2.43**
Words per second	2.74	2.83	2.89	3.14**
Speech time	0.3	0.44**	0.35	0.48**
Pause time	0.13	0.16*	0.11	0.12
Personal Pron.	12.06	11.45*	11.44	10.88
1st person Pron.	2.49	3.07**	1.26	1.28
Impersonal Pron.	9.01	7.65**	4.89	4.39
Past Tense	1.72	0.91**	2.79	1.54**
Tentative	7.97	9.81**	7.24	8.66**
Death	0.11	0.03	0.44	0.4
Nonfluencies	4.58	4.34	3.89	3.04**
f_0	192.29	138.12	198.28	135.37**
NAQ	0.07	0.03**	0.07	0.03**
QOQ	0.26	0.11**	0.27	0.12**
PSP	0.32	0.49**	0.35	0.46**
MDQ	0.13	0.11	0.13	0.11**
PS	−0.23	−0.24**	−0.23	−0.25**
Rd	1.62	1.19**	1.61	1.19**
F1	614.04	593.97**	642.06	599.5**
Backchannels				
NAQ	0.07	0.02**	0.06	0.04**
QOQ	0.26	0.09**	0.27	0.13**
PS	−0.24	−0.24**	−0.22	−0.25

Suicidal versus Non-Suicidal (i.e., Non-S.) nonverbal and verbal clinician's behavior observations. Features are abbreviated as introduced in Section 4 and pronouns are abbreviated with Pron.

the interviews. Hence, there are two cases investigated: the UQ case and the non-UQ case. Statistical analyses are performed to determine significant differences of the features to distinguish between suicidal and non-suicidal adolescents. Two cases are statistically analyzed by using ANOVA: suicidal versus non-suicidal adolescents and suicidal repeaters versus non-repeaters. The significance level is stated to be at least $p < 0.05$. 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 machine learning experiments (see Section 6). Statistical results corresponding to the clinician's speech or patients are specified with subscripts C and P, respectively.

5.1 Suicidal versus Non-Suicidal Evaluation

This section introduces the statistical results of the ANOVAs of the investigated features between suicidal patients and their controls (**RQ1**). The p-values of the statistically significant features corresponding to the patients are listed in Table 3 and the ones corresponding to the clinician in Table 4. The corresponding mean values of the samples are as well given in Table 3 for the patients and in Table 4 for the clinician respectively. The results are separated into UQ case and non-UQ case (**RQ3**).

Conversation. The patient-speaks-over-clinician rate is significant solely in the UQ case. Suicidal patients interrupt more often their clinician than the non-suicidal ones. The clinician-speaks-over-patient rate is significantly different in the non-UQ case and in the UQ one. In both cases the clinician speaks less over their interlocutors while interacting with suicidal patients. The speech time and pause time percentages of the patients are significantly different in both cases. The suicidal patients speak more than the non-suicidal adolescents. Considering the clinician, the speech time percentage is significantly different in both cases but the pause times are only significant in the UQ case. The clinician talks more during the interviews with the non-suicidal patients. He also pauses more often in the UQ case. However, in the non-UQ case the significantly different words per second rate of the clinician is added. The speech of the clinician shows on average a higher words per second rate during the conversation with non-suicidal adolescents.

Verbal information. The following significances are found regarding the patient's verbal features: In the UQ case as well as in the non-UQ case the use of personal pronouns and first person singular pronouns are significantly different. On average the suicidal patients use more often personal pronouns and references to themselves (see Fig. 1). The use of impersonal pronouns is just significant in the UQ case as well as the use of terms related to negation and positive emotion. The non-suicidal adolescents use more often negation terms and references to positive emotion. On the other hand, the use of tentative words and of terms related to death are significantly different only in the non-UQ case. Suicidal patients refer more often to death. The use of terms related to negative emotion (see Fig. 2), of assent words and nonfluencies (e.g., "um", "uh"), as well as references to the past are significantly different for both cases. Suicidal patients refer more often to negative emotion. Nonfluencies and assent terms are more often used by non-suicidal adolescents.

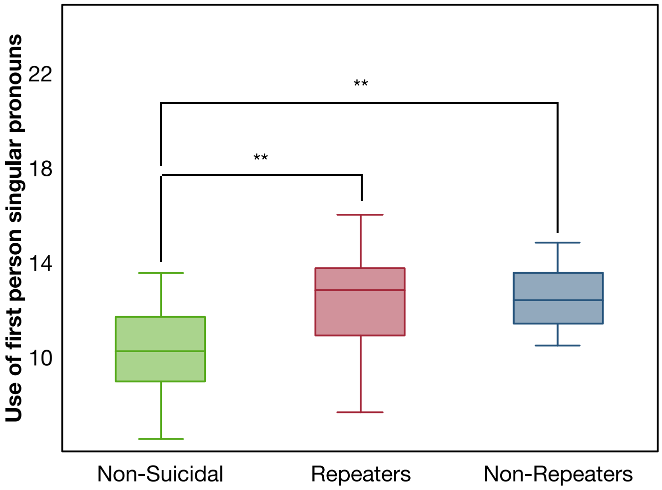

Fig. 1. Use of first person singular pronouns of non-suicidal and suicidal patients in the UQ case.

Different to the patients' verbal features, there are not as many similarities in the clinician's content between the two cases. The use of personal pronouns, first person singular pronouns and impersonal pronouns are significantly different just in the UQ case. The clinician uses more often personal pronouns in interviews with suicidal patients and less first person singular pronouns. The use of nonfluencies is only significant in the non-UQ case. During interviews with suicidal patients the clinician uses more terms indicating nonfluencies. The references to the past and to tentative terms are significantly different for both cases. As well as for the patients, the clinician uses on average more often terms referred to the past while interactions with suicidal patients.

Acoustic information. All the investigated acoustic features of the patients show significant differences in both cases. These include f_0, NAQ, QOQ, PSP, MDQ, PS, Rd and F1. The suicidal patients speak on average with a higher f_0. Regarding the clinician's acoustic features not as many comparable statistical results are obtained. The NAQ, QOQ, PSP, PS, Rd and F1 show significant difference in both cases. The NAQ, QOQ, MDQ and PS are indicating a breathier

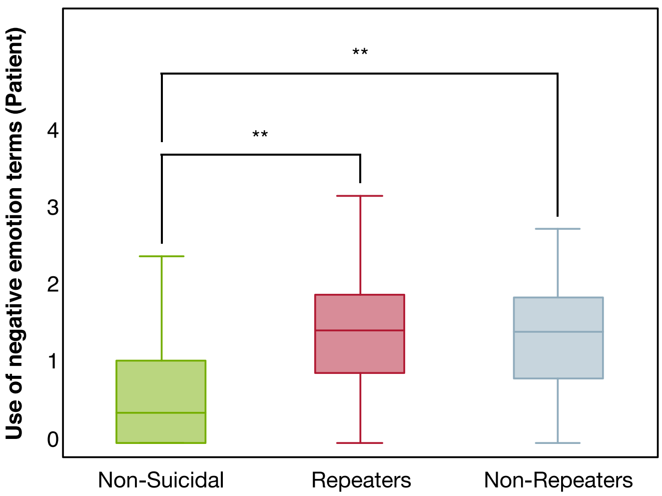

Fig. 2. References to negative emotion of non-suicidal and suicidal patients in the non-UQ case.

TABLE 5
Repeater versus Non-Repeater Mean Values

		UQ		Non-UQ	
		Repeater	Non-rep.	Repeater	Non-rep.
	Overlap rate	4.8	8.1*	5.6	7
Patient	$\mathbf{f_0}$	96.65	224.03	100.62	210.96**
	NAQ	0.03	0.08**	0.03	0.08**
	QOQ	0.2	0.31*	0.2	0.29**
	PSP	0.24	0.36**	0.25	0.34**
	MDQ	0.08	0.13**	0.08	0.14*
	Rd	1.2	1.65**	1.13	1.56**
	PS	−0.27	−0.2**	−0.26	−0.22**
	F1	502.81	613.56	494.48	641.48**
	F2	1,462.87	1,618.99**	1,401.35	1,621.12**
Clinician	$\mathbf{f_0}$	118.85	192.29**	124.36	198.28
	NAQ	0.02	0.07**	0.03	0.07**
	QOQ	0.09	0.26**	0.12	0.27**
	PSP	0.37	0.32**	0.42	0.35**
	MDQ	0.1	0.13**	0.11	0.13
	PS	−0.25	−0.23**	−0.24	−0.23
	Rd	1.35	1.62**	1.36	1.61*
	F1	692.62	614.04	627.94	642.06**
	F2	1,678.94	1,580.66	1,589.29	1,583.47*
Back-Ch.	**NAQ**	0.02	0.07**	0.03	0.06**
	QOQ	0.09	0.26*	0.12	0.27**
	MDQ	0.09	0.13**	0.11	0.13
	F1	827.81	681.62*	773.2	766.92*
	F2	1,696.19	1,521.12*	1,600.15	1,583.82

conversation with suicidal patients. Only in the non-UQ case the f_0 and the MDQ are significantly different. When considering the backchannel of the clinician, the NAQ, QOQ and PSP are significantly different in both cases. Also within the backchannel it is shown that the conversation is breathier with suicidal patients. The Rd and the PS are added to the significant features of the UQ case.

5.2 Suicidal Repeater versus Non-Repeater Evaluation

In this section, the 30 recorded interviews of the suicidal patients are statistically analyzed to determine significant features and distinctions between suicidal repeaters and non-repeaters (**RQ2**). The p-values of the statistically significant patients' and clinician's features are summarized in Table 5 along the corresponding mean values. The results are also separated into the UQ case and the non-UQ case (**RQ3**).

Conversation. In the interviews with the non-repeaters and repeaters, solely in the UQ case the conversational feature overlap rate is barely significantly different. More overlaps are detected in the interviews with the non-repeaters. All the other conversational information show no significances.

Verbal information. In the repeater versus non-repeater case, there are not observed any relevant statistically significances of the verbal information features, neither in the UQ case nor in the non-UQ case.

Acoustic information. The acoustic information features show more significant features in the non-UQ case. The f_0 and the F1 of the patients are only significantly different in the non-UQ case. They are on average higher during interviews with the non-repeaters. The NAQ, QOQ, PSP, MDQ,

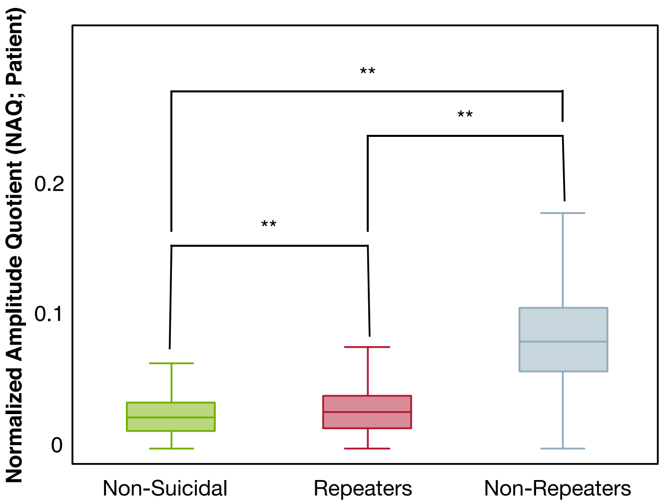

Fig. 3. Acoustic feature NAQ of non-suicidal, non-repeaters and repeaters in the UQ case.

Rd, PS and F2 are statistically relevant in both cases indicating breathier voices of the non-repeaters. In the UQ case these clinician's acoustic features are exclusively significant: the f_0, MDQ and PS. Only in the non-UQ case the two formants F1 and F2 are statistically different. In both cases the NAQ, QOQ, PSP and Rd show significant differences (Figs. 3 and 4). It also indicates breathier conversations with non-repeaters. Moreover, the acoustic information of the backchannel of the clinician is considered. In both cases the NAQ, QOQ and F1 are significantly different. The MDQ and F2 are only significantly different in the UQ case. Even the backchannel information refer to a breathier conversation with non-repeaters.

6 Hierarchical Classification

We developed a hierarchical classification with two layers (see Fig. 5). In the first layer the suicidal patients and the non-suicidal patients are distinguished using SVM classification with radial basis function kernel. The suicidal labeled patients are forwarded to the second layer. There, the suicidal repeaters and the non-repeaters are classified using the ensemble algorithm AdaBoostM1, as it is described in [27].

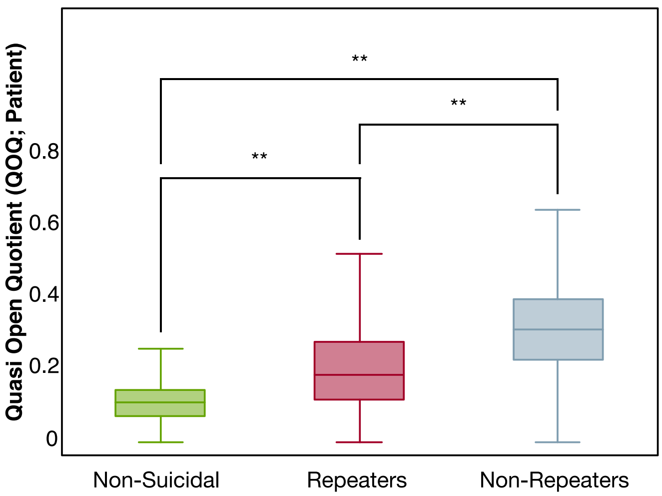

Fig. 4. Acoustic feature QOQ of non-suicidal, non-repeaters and repeaters in the UQ case.

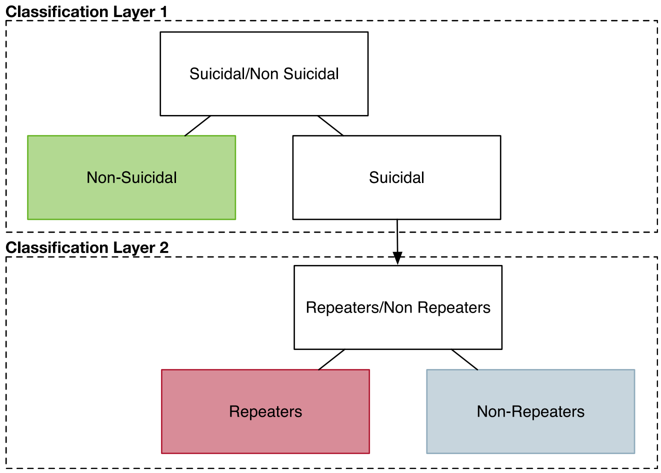

Fig. 5. Hierarchical structure of the classification.

The classification is performed subject-independently to confirm and identify the discriminative faculty of the statistically significant features in the UQ and in the non-UQ case. The testing of the hierarchical classification is performed with a leave-one-interview-out approach. In Fig. 5 the general hierarchical structure of the classification is illustrated.

In the UQ case 37 statistically significant features of patients and clinician are used to distinguish between suicidal and non-suicidal adolescents (see first columns of Tables 3 and 4). The features include 6 conversational, 14 verbal and 17 acoustic features. To classify non-repeaters from repeaters 20 features are used including one conversational feature and 19 acoustic features (see first column of Table 5). The distribution of the features corresponding to the classification layer is illustrated in Fig. 6A. Using only the patients' features the hierarchical classification yields an accuracy of 56.7 percent. The classification of suicidal versus non-suicidal patients yields an accuracy of 85 percent. The distinction between repeaters and non-repeaters achieves an accuracy of only 34.5 percent. By adding the clinician's features the accuracy in the first layer is improved (90 percent) but in the second layer it even decreases (33.3 percent). The last tested training set contains patients', clinician's and backchannel features. The classification of suicidal and non-suicidal patients achieves an accuracy of 88.3 percent. The accuracy of the classification of repeaters and non-repeaters is valued with 51.6 percent. This means a total accuracy of the hierarchical classification of 66.7 percent. The results are represented in Table 6. In addition, we provide the

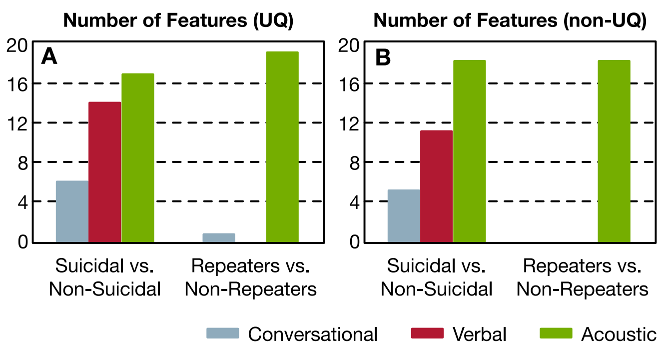

Fig. 6. Number of features used for the classification in the *UQ case (A)* and *non-UQ (B)* of the respective feature groups.

TABLE 6
Classification Accuracies of the Leave-One-Speaker-Out Datasets (UQ) in %

Feature Set	Suicidal versus non-Suicidal	Repeater versus non-Repeater	Hierarchical Classification
Patient Only	85	34.5	56.7
Patient+Clinician	90	33.3	58.3
All Features	88.3	51.6	66.7

TABLE 8
Classification Accuracies of the Leave-One-Speaker-Out Datasets (Non-UQ) in %

Feature Set	Suicidal versus non-Suicidal	Repeater versus non-Repeater	Hierarchical Classification
Patient Only	76.7	50	56.7
Patient+Clinician	85	39.4	51.7
All Features	86.7	67.7	71.7

confusion matrix, recall, precision, and F-scores for the best performing hierarchical classification using UQ only with 66.7 percent accuracy in Table 7.

In the non-UQ case, 34 features are used in the first layer (see second column of Tables 3 and 5). The set consists of 5 conversational, 11 verbal and 18 acoustic features of patients and clinician. In the repeater versus non-repeater 20 features are used for training and testing including solely 18 acoustic features (see second column of Table 5). Fig. 6B represents the distribution of the features over the two classification layers. The hierarchical classification yields an accuracy of 55 percent when using just the patients' features. The accuracy decreases to 51.7 percent when the clinician's features are added, although the accuracy of the suicidal versus non-suicidal layer increases from 76.7 to 85 percent. However, the accuracy in the second layer decreases from 50 to 39.4 percent. When the features of the patients, clinician and clinician's backchannel are used to test the hierarchical classification, an accuracy of 71.7 percent is achieved. The accuracy in the first layer increases to 86.7 percent and in the second layer it even increases to 67.7 percent. The corresponding accuracies are listed in Table 8. In addition, we provide the confusion matrix, recall, precision, and F-scores for the best performing hierarchical classification using non-UQ only with 71.1 percent accuracy in Table 9.

We compare the provided accuracies of the classification experiments to the ones achieved in [14] by using ensemble classifiers in both layers. For the classification the complete interview information was used. An accuracy of 90 percent was achieved for the suicidal versus non-suicidal distinction while the suicidal repeaters versus suicidal non-repeaters layer classification delivered an accuracy of 60 percent. The classification over the complete hierarchy yielded an accuracy of 73.3 percent. When using all the information from the clinician-patient interaction, the accuracies are highest. However, in the UQ case an accuracy of 90 percent is achieved when using patients' and clinician's features in the suicidal versus non-suicidal classification. In the non-UQ case, the repeater versus non-repeater classification yields a higher accuracy than in the complete case of [14]. Thus, the performance of the present classification experiments is comparable to the results in [14]. Nevertheless, the highest hierarchical classification accuracy was performed when using the complete information of the clinician-patient interactions [14].

7 Discussion

In the following discussion section we want to answer the research questions we introduced in Section 1.1. In particular, the ability to characterize suicidal speech (**RQ1**) is discussed, as well as the characterization of repeaters' and non-repeaters' verbal and nonverbal behaviors in the interviews (**RQ2**). In the paragraph labeled as **RQ3**, the suicidal risk classification utilizing the UQ and non-UQ interaction contexts separately will be discussed.

RQ1: We reveal a large set of characteristic features using statistical analyses of the 60 interviews of 30 suicidal and 30 non-suicidal patients. In total, we find 22 statistically significant patients' features and 21 that characterize the clinician. The statistical evaluation shows that each feature group comprises statistically significant features when distinguishing between suicidal and non-suicidal adolescents, in both investigated cases. In particular, conversational, verbal, and acoustic information characterize suicidal speech of adolescents. Nonverbal as well as verbal information are highly supportive to identify suicidal risk of adolescents. Below we discuss each feature group separately in detail.

Conversational features of the interviews show that the clinician interrupted suicidal patients less often than the non-suicidal ones. Suicidal patients' turns overlap the clinician on average more often than the non-suicidal adolescents. The

TABLE 7
Confusion Matrix of the Hierarchical Classifier (Non-UQ Case) Including Recall, Precision and F-Score (All Features; Accuracy 66.7 percent)

	Non-suicidal	Suicidal non-repeater	Suicidal repeater	Recall
Non-suicidal	26	2	2	0.87
Suicidal non-repeater	1	9	6	0.56
Suicidal repeater	2	7	5	0.36
Precision	0.90	0.50	0.38	
F-score	0.88	0.53	0.37	

TABLE 9
Confusion Matrix of the Hierarchical Classifier (Non-UQ Case) Including Recall, Precision And F-Score (All Features; Accuracy 71.1 percent)

	Non-suicidal	Suicidal non-repeater	Suicidal repeater	Recall
Non-suicidal	24	3	4	0.74
Suicidal non-repeater	1	13	5	0.68
Suicidal repeater	1	3	6	0.6
Precision	0.92	0.68	0.4	
F-score	0.84	0.68	0.48	

duration of the interviews depend on the state of the patient: Interviews with suicidal patients lasted about 15 minutes while the interviews with non-suicidal ones took only about 8 minutes. This finding is also reflected in the speech time of the patients and clinician. During interviews with suicidal patients the clinician pauses longer and speaks less, while the patients pause less and speak longer. The self-preoccupation associated with suicidal risk can be a possible explanation of this finding [13]. The mere length of the interview, however, could be an artifact of the fact that the clinician knew of the state of the patient.

Verbal information. The social integration theories of Stirman and Pennebaker [13] are confirmed for suicidal adolescent patients by our study: the suicidal adolescents referred to themselves using first person singular pronouns more often than the non-suicidal patients. The increased use of references to one-selves is a further indication of the self-preoccupation of suicidal risk patients. In addition, suicidal risk patients refer more often to the past as Stirman and Pennebaker showed in their study. Moreover, we find that suicidal patients use less terms related to positive emotions but refer more to negative emotions. This complements the work of Stirman and Pennebaker who hypothesized that suicidal poets would use negative emotional terms more frequently but could not find them [13]. We find that terms related to negation and assent are more frequently used by non-suicidal patients. Perhaps, this finding can be explained by more frequent short assent answers of the non-suicidal patients, such as "yes" or "no". As Stirman and Pennebaker we also find the use of terms related to death in the speech of suicidal patients more often than of the non-suicidal ones. The references to death with terms like "kill, die, death" are of high importance when assessing the suicidal risk, as hopelessness is often related to suicide [28], [29]. The clinician's verbal behavior differs in one category: Although, the clinician uses more personal pronouns while interacting with suicidal patients, he refers less often to himself. This is definitely caused by the interaction with the patient using words like "you, your", etc. in the questions and incentive interruptions like "You can go ahead and just briefly explain it". The use of second person singular pronouns is not significantly different. We identify the higher use of terms related to death and past tense also in the speech of the clinician with suicidal patients.

Acoustic information. During interviews with suicidal patients the clinician as well as the suicidal adolescents show breathier voice characteristics. The use of a breathier voice was already referred to suicidal speech by [7]. This also can be related to hopelessness and depressive disorders, which are two symptoms that characterize the suicidal risk of a person [2]. On average the clinician as well as the suicidal patients spoke with a higher fundamental frequency during these interviews. We observe that the clinician's speech aligns to the patients' ones by lowering or raising speech fundamental frequency. Here, it is important to emphasize that the clinician was the same for all 60 interviews. Also the backchannel utterances' (e.g., "yeh", "uh huh", "okay") features of the clinician alone show breathier voice characteristics while interacting with suicidal patients. 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 700 ms [2].

RQ2: When evaluating the significant differences and changes of the features between repeaters and non-repeaters, mainly acoustic information are statistically significant. Only in the UQ case one conversational feature (i.e., overlap rate) can be identified as discriminative. It can be argued that written or verbal questionnaires 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. While we can not rule out the possibility that the utilized questions were not suitable to elicit responses that would allow discrimination between repeaters and non-repeaters, there is considerable evidence that disclosure or the lack thereof is associated with suicidal risk and ideation [5], [6]. Hence, it could be argued that the need of computer-aided support and the assessment of nonverbal conversational content are of considerable importance to identify suicidal repeaters.

Acoustic information. The interviews with non-repeaters are characterized by breathier voices of both clinician and patients. Similarly, Öjehagen et al. [30] stated that repeaters in their study acted more often impulsively and that their suicidal intent was less severe than that of the non-repeaters. In our analysis, non-repeaters as well show more characteristic behaviors that are found for suicidal adolescents overall. Hence, one could argue, as in Öjehagen et al. [30], non-repeaters show more severe behaviors of suicidal risk than repeaters (see Figs. 3 and 4). Otherwise, the effects difference between repeaters and non-repeaters could be explained by the interpersonal theory of Van Orden et al. [31]. The human fear system can be altered by suicidal behavior particularly by reducing the fear of death because of repeating attempts and increasing the physical pain tolerance. Those two factors can result from suicidal behavior, e.g., non-lethal attempts. This habituation process may reflect the greater acquired capability for suicidal behavior in repeaters and results in being more similar to non-suicidal adolescents than non-repeaters (Fig. 3).

Due to the lack of the significant differences of verbal and conversational features, it would be a benefit to assess further nonverbal behavior of repeaters. Some studies already focused on the facial expressions of clinicians and repeaters as well [7], [30]. Unfortunately, we do not have access to video recordings for the present work. These studies found that the clinicians again showed significantly different behavior while interacting with repeaters without being aware of it. We also find the adaptation of the clinician's voice characteristics to those of the repeaters [2].

RQ3: For both interaction contexts, namely UQ and non-UQ segments of the interviews, significantly different features are identified when analyzing suicidal versus non-suicidal adolescents. In other words, the statistical evaluation of suicidal versus non-suicidal patients reveals more discriminative features than the one of repeaters versus non-repeaters. We identify that in particular nonverbal behavior comprising voice characteristics describe repeaters and non-repeaters [6]. When considering the number of significant features, more features were significantly different in the UQ case than in the non-UQ case. This finding indicates that the UQ indeed elicit more emotional responses than the more fact based questions in the non-UQ segment.

To further investigate the discriminative power of the segments, we performed an automatic leave-one-speaker-out hierarchical classification comparing UQ segments with non-UQ segments. In the first layer a SVM classifier is trained and tested to classify repeaters and non-repeaters. In the second layer an ensemble classifier is used. In the UQ case as well as in the non-UQ case conversational, verbal, and acoustic features are used to discriminate suicidal from non-suicidal adolescents. In the UQ case conversational and acoustic information are used to distinguish repeaters from non-repeaters in the second layer of the hierarchical classifier. However, in the non-UQ case solely acoustic features are used for the classification. The most promising hierarchical classification accuracy is achieved with 71.7 percent in the non-UQ case. In the second layer, in which the distinction between repeaters and non-repeaters is performed, the highest accuracy at 67.7 percent is achieved. Regarding the classification of suicidal and non-suicidal patients the UQ case reveals promising results with an accuracy of 88.3 percent. Moreover, the accuracy of the classification of suicidal and non-suicidal adolescents is increased to even 90 percent when considering the patients' and clinician's features without the backchannel's information.

The UQ are clearly the most discriminative for the classification of suicidal and non-suicidal patients. However, to additionally classify suicidal repeaters and suicidal non-repeaters more information is required. This can be improved by further nonverbal information of clinician and patients, e.g., analyzing facial expressions [7], [30]. In summary, for the hierarchical classification in both cases, the classification of suicidal and non-suicidal patients requires conversational, verbal, and acoustic information. However, to distinguish repeaters and non-repeaters almost solely acoustic features are needed. We characterized suicidal speech considering repeater versus non-repeater with verbal and nonverbal audio-based features. We also show that a hierarchical classification can determine the three classes: non-suicidal adolescents, repeaters, and non-repeaters. In the future, a multimodal study including video-based features of patients and clinician could improve the modest accuracies of the hierarchical classification. In particular, the classification of repeaters and non-repeaters could be improved as our investigations reveal that especially nonverbal features are important for the discrimination. Most notably, we discovered that the repeater identification depends on both the nonverbal behaviors of clinician and patient. With respect to the differences of the UQ and the non-UQ segments of the interviews: The classification of suicidal and non-suicidal adolescents can be satisfactorily performed by using just the five open-ended questions included in the UQ segments. Concerning repeater versus non-repeater we needed the non-UQ segments to gather more information of the repeaters' behavior.

8 Conclusions

In this study, two cases in suicidal risk assessment are investigated and compared. First, the interview segments containing five open-ended ubiquitous questions and answers are tested if they have the potential to classify non-suicidal patients, suicidal repeaters, and suicidal non-repeaters between the ages of 13 and 18. Moreover, the residual interview segments, called non-UQ case, are investigated to set a comparison to the UQ segments. After significant differences in the feature groups of conversational, verbal, and acoustic information are determined statistically, they are used in an automatic classification experiment. Verbal and nonverbal behavior of clinician-patient interactions are useful to distinguish between suicidal and non-suicidal patients. To discriminate between repeaters and non-repeaters especially acoustic information features are useful. The highest accuracy can be achieved using the non-UQ case. However, the most promising accuracy regarding the classification between suicidal and non-suicidal patients is achieved using the UQ segments. Information acquired in the UQ segments is enough to distinguish between suicidal and non-suicidal adolescents. To additionally characterize repeaters more information is required.

Overall, the study showed that the ubiquitous questions part of the interviews can help assess the suicidal risk of adolescents, i.e., distinguish suicidal and non-suicidal patients. We could show that verbal behavior of young patients and clinicians is important to assess suicidal risk within an interview, but assessing the potential of suicidal re-attempt nonverbal behavior information, in particular acoustic features, is important. Therefore, we believe that the investigated approach holds the potential to additionally support clinicians when assessing suicidal risk of adolescents in the ED.

Acknowledgments

The work depicted here is sponsored by the US Army Research Laboratory (ARL) under contract number W911NF-14-D-0005. Statements and opinions expressed and content included do not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred. Verena Venek and Stefan Scherer contributed equally to the paper. John Pestian is the corresponding author.

References

[1] J. P. Pestian, "A conversation with Edwin Shneidman," *Suicide Life-Threatening Behavior*, vol. 40, no. 5, pp. 516–523, 2010.
[2] S. Scherer, J. P. Pestian, and L.-P. Morency, "Investigating the speech characteristics of suicidal adolescents," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 709–713.
[3] 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 Trans. Biomed. Eng.*, vol. 51, no. 9, pp. 1530–1540, Sep. 2004.
[4] D. J. France, R. G. Shiavi, S. E. Silverman, M. K. Silverman, and D. M. Wilkes, "Acoustical properties of speech as indicators of depression and suicidal risk," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 829–837, Jul. 2000.
[5] M. D. Anestis and B. A. Green, "The impact of varying levels of confidentiality on disclosure of suicidal thoughts in a sample of United States National Guard personnel," *J. Clin. Psychol.*, vol. 70, no. 10, pp. 1023–1030, 2015.
[6] N. Horesh, G. Zalsman, and A. Apter, "Suicidal behavior and self-disclosure in adolescent psychiatric inpatients," *The J. Nervous Mental Disease*, vol. 192, no. 12, pp. 837–842, 2004.

[7] V. Haynal-Reymond, G. K. Jonsson, and M. S. Magnusson, *Non-Verbal Communication in Doctor-Suicidal Patient Interview* (series The Hidden Structure of Interaction). Amsterdam, The Netherlands: IOS Press, 2005, ch. 9.

[8] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. Quatieri, "A review of depression and suicide risk assessment using speech analysis," *Speech Commun.*, vol. 17, pp. 10–49, 2015.

[9] S. Scherer, G. Lucas, J. Gratch, A. Rizzo, and L.-P. Morency, "Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews," *IEEE Trans. Affective Comput.*, 2015, Doi: 10.1109/TAFFC.2015.2440264

[10] J. F. Gunn and D. Lester, "Twitter postings and suicide: An analysis of the postings of a fatal suicide in the 24 hours prior to death," *Suicidology*, vol. 17, no. 3, pp. 28–30, 2012.

[11] B. O'Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, "Detecting suicidality on Twitter," *Internet Interventions*, vol. 2, pp. 183–188, 2015.

[12] L. Zhang, X. Huang, T. Liu, A. Li, Z. Chen, and T. Zhu, *Using Linguistic Features to Estimate Suicide Probability of Chinese Microblog Users* (series Human Centered Computing). New York, NY, USA: Springer, 2015, pp. 549–559.

[13] S. W. Stirman and J. W. Pennebaker, "Word use in the poetry of suicidal and nonsuicidal poets," *Psychosomatic Med.*, vol. 63, no. 4, pp. 517–522, 2001.

[14] V. Venek, S. Scherer, L.-P. Morency, A. Rizzo, and J. P. Pestian, "Adolescent suicidal risk assessment in clinician-patient interaction: A study of verbal and acoustic behaviors," in *Proc. Spoken Lang. Technol.*, 2014, pp. 277–282.

[15] K. Posner, G. K. Brown, B. Stanley, D. Brent, K. V. Yershova, M. A. Oquendo, G. W. Currier, G. A. Melvin, L. Greenhill, S. Shen, and J. J. Mann, "The ColumbiaSuicide severity rating scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults," *The Am. J. Psychiatry*, vol. 168, no. 12, pp. 1266–1277, Dec. 2011.

[16] 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 Psychol. Rev.*, vol. 28, no. 1, pp. 17–30, 1999.

[17] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *J. Lang. Social Psychol.*, vol. 29, no. 1, pp. 24–54, 2010.

[18] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, "Covarep - A collaborative voice analysis repository for speech technologies," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2014, pp. 960–964.

[19] T. Drugman and A. Abeer, "Joint robust voicing detection and pitch estimation based on residual harmonics," in *Proc. Interspeech*, 2011, pp. 1973–1976.

[20] P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parameterization of the glottal flow," *J. Acoust. Soc. Am.*, vol. 112, no. 2, pp. 701–710, 2002.

[21] T. Hacki, "Klassifizierung von glottisdysfunktionen mit hilfe der elektroglottographie," *Folia Phoniatrica*, pp. 43–48, vol. 41, no. 1, 1989.

[22] P. Alku, H. Strik, and E. Vilkman, "Parabolic spectral parameter—a new method for quantification of the glottal flow," *Speech Commun.*, vol. 22, no. 1, pp. 67–79, 1997.

[23] J. Kane, "Tools for analysing the voice - developments in glottal source and voice quality analysis," Ph.D. dissertation, Phonetics and Speech Lab., School of Linguistic, Speech and Commun. Sci., Trinity College Dublin, Ireland, 2012.

[24] G. Fant, J. Liljencrants, and Q. Lin, "A four parameter model of glottal flow," KTH, Speech Transmission Laboratory, Quarterly Rep. Stockholm, Sweden, vol. 4, pp. 1–13, 1985.

[25] N. Obin, "Cries and whispers: Classification of vocal effort in expressive speech," in *Proc. Interspeech*, 2012, pp. 2234–2237.

[26] B. Bozkurt, B. Doval, C. d'Alessandro, and T. Dutoit, "Improved differential phase spectrum processing for formant tracking," in *Proc. Interspeech*, 2004, pp. 2421–2424.

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

[28] K. Petrie and R. Brook, "Sense of coherence, self-esteem, depression and hopelessness as correlates of reattempting suicide," *Brit. J. Clinical Psychol.*, vol. 31, no. 3, pp. 293–300, 1992.

[29] E. S. Shneidman, *The Suicidal Mind*. London, U.K.: Oxford Univ. Press, 1998.

[30] A. Öjehagen, G. Regnell, and L. TräskmanBendz, "Deliberate self-poisoning: Repeaters and nonrepeaters admitted to an intensive care unit," *Acta Psychiatrica Scandinavica*, vol. 84, no. 3, pp. 266–271, 1991.

[31] K. A. Van Orden, T. K. Witte, K. C. Cukrowicz, S. Braithwaite, E. A. Selby, and T. Joiner, "The interpersonal theory of suicide," *Psychol. Rev.*, vol. 117, no. 2, pp. 575–600, 2010.

Verena Venek received the Dipl.-Ing. in health care IT at the Carinthia University of Applied Sciences, Austria, in 2014, and the bachelor of science in engineering at the same university in 2012. She performed her bachelor thesis project at Siemens AG Healthcare in Erlangen, Germany (2012), and she worked on her master thesis at the USC Institute of Creative Technologies in Playa Vista, Los Angeles, CA (2014). During this period, she conducted research in the field of multimodal verbal and nonverbal assessment of suicidal risk in adolescents. Her studies on this topic are published at the IEEE SLT Workshop 2014. In 2015, she joined as a junior researcher the research Department Mobile and Web-based Information Systems at Salzburg Research in Austria. In current projects, she is responsible for spatio-temporal data analysis of movement data. In the research group eHealth, she deals with assessment of user requirements and assistive systems for active and assisted living, particularly for elderlies.

Stefan Scherer (https://schererstefan.net) received the degree of Dr. rer. nat. from the faculty of Engineering and Computer Science at Ulm University in Germany with the grade summa cum laude (i.e., with distinction). He is a research assistant professor in the Department of Computer Science at the University of Southern California (USC) and a research associate at the USC Institute for Creative Technologies where he leads research projects funded by US National Science Foundation (NSF) and ARL. His research aims to automatically identify, characterize, model, and synthesize individuals' multimodal nonverbal behavior within both human-machine as well as machine-mediated human-human interaction. He is a member of the IEEE.

Louis-Philippe Morency received the master and PhD degrees from the MIT Computer Science and Artificial Intelligence Laboratory. He is an assistant professor in the Language Technology Institute at Carnegie Mellon University where he leads the Multimodal Communication and Machine Learning Laboratory (MultiComp Lab). His research focuses on creating algorithms and probabilistic models to analyze, recognize and predict human subtle communicative behaviors during social interactions. In 2008, he was selected as one of "AI's 10 to Watch" by *IEEE Intelligent Systems*. He has received seven best paper awards in multiple ACM- and IEEE-sponsored conferences for his work on context-based gesture recognition, multimodal probabilistic fusion and computational models of human communication dynamics. He is a member of the IEEE.

Albert "Skip" Rizzo is a clinical psychologist and the director of Medical Virtual Reality at the University of Southern California the Institute for Creative Technologies. He is also a research professor with the USC Department of Psychiatry and at the USC Davis School of Gerontology. He conducts research on the design, development and evaluation of virtual reality systems targeting the areas of clinical assessment, treatment and rehabilitation across the domains of psychological, cognitive and motor functioning in both healthy and clinical populations. This work has focused on PTSD, TBI, Autism, ADHD, Alzheimer's disease, stroke, and other clinical conditions. In spite of the diversity of these clinical R&D areas, the common thread that drives all of his work with digital technologies involves the study of how interactive and immersive virtual reality simulations can be usefully applied to address human healthcare needs beyond what is possible with traditional tools and methods. In 2010, he received the American Psychological Association Award for Outstanding Contributions to the Practice of Trauma Psychology for his R&D work on VR exposure therapy and in 2015 he received the Society for Brain Mapping and Therapeutics *Pioneer in Medicine* award. He is a member of the IEEE.

John Pestian received the PhD and MBA degrees. He is a professor of pediatrics, psychiatry, and biomedical informatics at Cincinnati Children's Hospital Medical Center and the University of Cincinnati. The mission of the Pestian Lab is to develop advanced multimodal technology for the care of neuropsychiatric illness. His research into identifying the optimal neuropsychiatric medications has been used to help more than 230,000 patients and the lab's research using machine-learning for suicide classification is now being tested in the school setting. More about the research team can be found at pestianlab.cchmc.org.