# Investigating Voice Quality as a Speaker-Independent Indicator of Depression and PTSD

*Stefan Scherer, Giota Stratou, Jonathan Gratch, Louis-Philippe Morency*

University of Southern California,
Institute for Creative Technologies, Los Angeles
scherer@ict.usc.edu

## Abstract

We seek to investigate voice quality characteristics, in particular on a breathy to tense dimension, as an indicator for psychological distress, i.e. depression and post-traumatic stress disorder (PTSD), within semi-structured virtual human interviews. Our evaluation identifies significant differences between the voice quality of psychologically distressed participants and not-distressed participants within this limited corpus. We investigate the capability of automatic algorithms to classify psychologically distressed speech in speaker-independent experiments. Additionally, we examine the impact of the posed questions' affective polarity, as motivated by findings in the literature on positive stimulus attenuation and negative stimulus potentiation in emotional reactivity of psychologically distressed participants. The experiments yield promising results using standard machine learning algorithms and solely four distinct features capturing the tenseness of the speaker's voice.

**Index Terms**: Voice quality, psychological distress classification, virtual human interaction

## 1. Introduction

Investigations on the nonverbal manifestations of psychological disorders typically rely on labor-intense and tedious manual annotations of behaviors such as facial expressions, gaze patterns, gestures and vocal patterns by trained experts. Perhaps as a result, findings on the link between nonverbal behavior and clinical illness is sparse. Some disorders, such as depression have been heavily studied [1, 2, 3, 4, 5, 6], whereas important societal problems, such as post-traumatic stress disorder (PTSD), remain understudied. This paper explores the promise of automated techniques for early screening of mental illness.

State-of-the-art screening technology for psychological disorders largely builds upon filling out questionnaires, which yield crude scores to assess a person's psychological state. Such screenings usually discard or do not take into account quantitative or qualitative information on nonverbal behaviors. Virtual human and automated perception technology could possibly allow for a combination of structured questionnaires and the quantitative analysis of nonverbal behaviors within automated human-virtual human interaction scenarios. The gathered information on the interviewees' responses to the questions posed by the virtual human along with their automatically analyzed and quantified behavior, could then be used to inform professional healthcare providers with valuable information in both online as well as post-interactional manners, which could in turn render the diagnostic process more effective as well as more holistic.

The speech and voice patterns of participants suffering from depression have been characterized in related work by lowered intensity, increased monotonicity, reduced articulation rate, and varied switching pause duration [7, 3, 6]. In contrast, speech characteristics of PTSD are not well-understood. Here we investigate the potential of a small number of parameters to serve as indicators of psychological disorders within semi-structured virtual human-led interview scenarios. We are interested in the characterization of the participants' voice quality on a breathy to tense dimension and the association to psychological disorders, i.e. depression and PTSD (Research Goal **R1**). We seek robust features that are independent of speaker or gender and demonstrate their discriminative power in leave-one-speaker-out classification experiments (Research Goal **R2**). We evaluate over 40 unique wizard-of-Oz interactions.

A novel aspect of this work is that the virtual human is designed to evoke participant-behaviors that are likely to be diagnostic of psychological disorders. Recent research suggests that clinical illness such as depression may be better discriminated by examining participants reactions to positive and negative emotional stimuli. For example, in a recent large meta-analysis [5], participants suffering from depression show reduced affective reactions to positive emotional stimuli (*positive attenuation*) and increased affective reactions to negative emotional stimuli (*negative potentiation*) in contrast to normal control participants. Based on these findings, we investigate within this work the emotional reactivity of the participants towards varying question polarities and affective stimuli, i.e. the wizard triggers - in a structured manner - questions with neutral, positive, or negative polarity. Based on this contextualization we investigate possible discriminative reactions, namely participants' positive attenuation or negative potentiation [8], and investigate potential probing strategies to enrich basic questionnaires (Research Goal **R3**).

The remainder is organized as follows: Section 2 summarizes related work on voice analysis in the context of psychological disorders. In Section 3, we introduce the investigated dataset and the experimental setup of this work. We describe and motivate the investigated speech features in Section 4. These features are statistically evaluated in Section 5. In Section 6 we summarize the classification results of the question polarity dependent and speaker-independent experiments. Lastly, Sections 7 and 8 discuss the results and conclude.

## 2. Related Work

In [9], for example, the speech of 10 suicidal, 10 depressed, and 10 control participants was analyzed in great detail (male only). The data for the suicidal participants were obtained from a large spectrum of recording setups comprising, for example, suicide notes recorded on tape. The authors 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. depressed 90% correct). The large variance of recording conditions reduces the generalizability of these results.

The study in [10] involved the analysis of glottal flow features as well as prosodic features for the discrimination of depressed read speech of 15 male and 18 female speakers. The extracted glottal flow features comprised instances such as the minimal point in glottal derivative, maximum glottal opening, start point of glottal opening, and start point of glottal closing. The prosodic features extracted consist of fundamental frequency ($f_0$), energy, and speaking rate. The classification was performed on a leave-one-observation-out paradigm, which renders the analysis highly speaker-dependent. Hence, strong classification results were observed, well above 85% accuracy for male speakers and above 90% for female speakers. The authors identified glottal flow features to be chosen by the feature selection algorithm for the majority of the classifiers as well as energy-based features for female speakers.

In [11], facial as well as acoustic measures are used to classify speech of depressed and non-depressed speakers within a face to face interview corpus. Basic measures, such as fundamental frequency ($f_0$) and switching pauses (i.e. pause length after the interviewer poses a question), are used for classification experiments and the classification of speakers that responded to their treatment (i.e. their Hamilton depression score reduced by 50%). Responders and non-responders could be classified with about 80% accuracy in a speaker-dependent experiment.

Within the present work, we address some of the shortcomings of previous work: we characterize the speech of participants suffering from depression or PTSD using a set of speaker-independent voice quality parameters, which robustly discriminate speech on a breathy to tense dimension. In particular, speech and voice characteristics of participants suffering from PTSD have been widely understudied in the past. As mentioned above, previous results might be biased due to speaker dependent classification or varying recording conditions. Hence, we emphasize the investigation of speaker-independence in leave-one-speaker-out classification experiments in this work using our novel dataset that has been recorded in constant high-fidelity conditions, overcoming some disadvantages of the evaluation in [9]. Lastly, we show results for the implications of question polarity on the discrimination capabilities of the investigated voice quality features in human-virtual human interaction.

# 3. Virtual Human Distress Assessment Interview Corpus

In this section, we introduce the analyzed dataset which is similarly structured as the large human-human Distress Assessment Interview Corpus (DAIC), that was described in [12]. The corpus, investigated in the present work, is recorded in a wizard of Oz controlled scenario where a virtual human interacts verbally and nonverbally in a semi-structured manner with a participant[1].

## 3.1. Procedure

The participants were recruited via Craigslist and were recorded at the USC Institute for Creative Technologies. In total 45 participants interacted with the virtual human. Unfortunately, only 43 interactions could be used in this study as for two of the interactions the logging of the virtual human's behavior failed to record. All participants who met requirements (i.e. age greater than 18, and adequate eyesight) were accepted. Their mean age was 41.2 years ($\sigma$ = 11.6; 27 male and 16 female).

For the recording of the dataset we adhered to the following procedure: After a short explanation of the study and giving consent, participants complete a series of questionnaires. These questionnaires included amongst others the PTSD Checklist-Civilian version (PCL-C) and the Patient Health Questionnaire, depression module (PHQ-9) [12].

Two wizards in a separate room control the nonverbal behavior (e.g. head nods, smiles, back-channels) and the verbal utterances including questions and verbal back-channels of the virtual human by selecting pre-recorded behaviors from a menu interface. This wizard-of-Oz setup is the first step towards a fully automatic interaction. The interaction between the participants and the wizard-of-Oz controlled virtual human was designed as follows: The virtual human explains the purpose of the interaction and that it will ask a series of questions. It further, tries to build rapport with the participant in the beginning of the interaction with a series of shallow questions about Los Angeles. Then three phases of positive/negative/positive polarity follow. The positive phases include questions like: "What would you say are some of your best qualities?" or "What are some things that usually put you in a good mood?". The negative phase includes questions such as: "Do you have disturbing thoughts?" or "What are some things that make you really mad?". The questions were pre-recorded and animated using the SmartBody architecture [13].

## 3.2. Condition and Question Polarity Assessment

The PCL-C and PHQ-9 scales provide researchers with guidelines on how to assess the participant's condition based on the responses. Our participant-pool got split into 18 participants that scored above 10 on the PHQ-9, which corresponds to moderate depression and above (cf. [14]), and 18 participants scored below 8. The second condition got split following the PCL-C classification resulting in 23 participants scoring positively for PTSD and 20 negatively. The two positively scoring groups (i.e. depressed and PTSD categories) overlap strongly as the evaluated measurements PHQ-9 and PCL-C correlate strongly (Pearson's $r > 0.8$), as reported in [12].

The questions' affective polarity was assessed based on a 5 point Likert scale, ranging from strongly negative (-2) to strongly positive (+2), by two independent coders. The coder-agreement resulted in a Krippendorff $\alpha$ of 0.86, which indicates to a strong agreement. Questions with a mean rating $r < -1$ were considered negative, $r > +1$ positive, and $-1 > r > +1$ neutral respectively. In total there were 10 unique positive and 10 negative questions, as well as 45 unique neutral questions. Not every participant was asked exactly the same questions, and the wizard adapted the selected questions for each phase based on the conversation flow. Interviews lasted between 5 and 15 minutes.

# 4. Investigated Features

The automatically extracted features were chosen based on previous encouraging results in classifying voice patterns of suicidal adolescents [15] as well as the features' relevance for characterizing voice qualities on a breathy to tense dimension [16, 17].

[1] Sample interaction between the virtual agent and a human actor can be seen here: http://www.youtube.com/watch?v=ejczMs6b1Q4

| | Depression | No-Depression | p-value |
|---|---|---|---|
| **NAQ** | 0.067 (0.034) | 0.100 (0.027) | 0.003 |
| **peakSlope** | -0.362 (0.049) | -0.403 (0.036) | 0.007 |
| **QOQ** | 0.292 (0.103) | 0.367 (0.072) | 0.016 |
| **OQ**$_{NN}$ | 0.609 (0.015) | 0.621 (0.012) | 0.015 |

Table 1: Statistically significant acoustic measures discerning participants with moderate or severe depression (N = 18) and participants without (N = 18). The mean and standard deviation (in parentheses) values as well as the corresponding p-values derived from independent t-tests.

The first three features are derived from the glottal source signal estimated by iterative adaptive inverse filtering (IAIF, [18]). The output is the differentiated glottal flow. The normalized amplitude quotient (**NAQ**, [19]) is calculated using:

$$\text{NAQ} = \frac{f_{ac}}{d_{peak} \cdot T_0} \tag{1}$$

where $d_{peak}$ is the negative amplitude of the main excitation in the differentiated glottal flow pulse, while $f_{ac}$ is the peak amplitude of the glottal flow pulse and $T_0$ the length of the glottal pulse period. The quasi-open quotient (**QOQ**, [20]) is also derived from amplitude measurements of the glottal flow pulse. The quasi-open period 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 parameter. Further, we extract **OQ**$_{NN}$ a novel parameter estimating the open quotient using standard Mel frequency cepstral coefficients and a trained neural network for open quotient approximation [21].

The final feature involves a dyadic wavelet transform using $g(t)$, a cosine-modulated Gaussian pulse similar to that used in [22] as the mother wavelet:

$$g(t) = -cos(2\pi f_n t) \cdot exp\left(-\frac{t^2}{2\tau^2}\right), \tag{2}$$

where the sampling frequency $f_s$ = 16 kHz, $f_n = \frac{f_s}{2}$, $\tau = \frac{1}{2f_n}$ and $t$ is time. The wavelet transform, $y_i(t)$, of the input signal, $x(t)$, at the $i^{th}$ scale, $s_i$, is calculated by:

$$y_i(t) = x(t) * g\left(\frac{t}{s_i}\right), \tag{3}$$

where $*$ denotes the convolution operator and $s_i = 2^i$. This functions essentially as an octave band zero-phase filter bank.

For the **peakSlope** feature [23], the speech signal is used as $x(t)$ in Eq. (3). Maxima are measured across the scales, on a fixed-frame basis, and a regression line is fit to $log_{10}$ of these maxima. The slope of the regression line for each frame provides the peakSlope value. The feature is essentially an effective correlate of the spectral slope of the signal.

## 5. Statistical Feature Evaluation

Here, we report the results of our statistical analysis of voice quality (VQ) features. The results of this analysis are summarized in Tables 1 and 2. In these tables we provide additionally to mean and standard deviation, p-values as derived from independent t-tests.

All investigated features indicate that the voice of participants with moderate to severe depression is more tense than the voice of participants with no-depression, as indicated by the

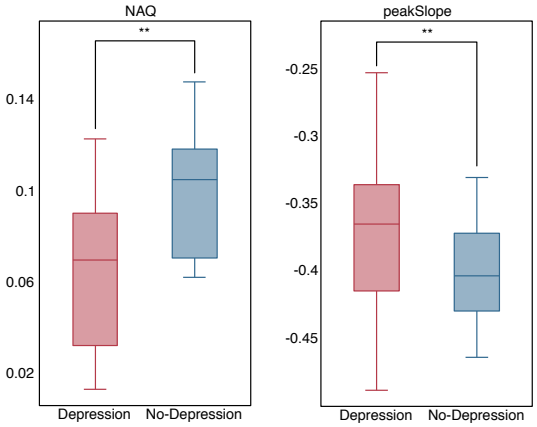


Figure 1: Exemplary visualization of the observed normalized amplitude quotient (NAQ) and the peak slope parameter (peakSlope) as observed for the conditions depression and no-depression within the entire interviews.

| | PTSD | No-PTSD | p-value |
|---|---|---|---|
| **NAQ** | 0.070 (0.035) | 0.097 (0.027) | 0.007 |
| **peakSlope** | -0.371 (0.054) | -0.399 (0.037) | 0.051 |
| **QOQ** | 0.296 (0.100) | 0.358 (0.070) | 0.021 |
| **OQ**$_{NN}$ | 0.611 (0.016) | 0.621 (0.011) | 0.024 |

Table 2: Statistically significant acoustic measures discerning participants with PTSD (N = 20) and without (N = 23).

PHQ-9 measure. Further, all observed values are significantly different for the two groups (i.e. depression and no-depression) as observed with independent t-tests, with NAQ showing the strongest observable difference (Hedges' g = -1.063; p = 0.003). This observation promises a strong separability between the groups and is further illustrated in Figure 1. Hedges' g value is a measure of the effect size found in the data, i.e. the g-value denotes the required shift of the mean of one set to match the mean of the other in magnitudes of standard deviations [24].

Similarly, these effects are observable for the PTSD vs. no-PTSD conditions. However, the effects are not as strong as in the depression condition (e.g. peakSlope does not score significantly different in a t-test with p = 0.051; Hedges' g = 0.603). Due to reduced effect sizes, we expect the classification performance to be below the one observed for the depression condition. Further, it can be seen that the observed feature scores are quite similar between the conditions, as it was mentioned earlier the groups of PTSD and depression highly overlap.

## 6. Classification Experiments

We investigate the discriminative power of the four voice quality features in several classification experiments using standard support vector machines (SVM). The conducted experiments for the classification of participants with and without depression and PTSD are performed multiple times for the different affective question polarities (i.e. positive, negative, and neutral, as well as all available data). All experiments are conducted using a leave-one-speaker-out strategy: for the training of the classifiers in each fold, we leave out the speech samples of one speaker entirely from the training and test the classifiers on the speech of the left-out speaker. We employ SVM with radial basis function kernels. The SVM are trained on the median of the features over the single utterances.

| | Question Polarity | Accuracy (%) | $F_1$ pos/neg |
|---|---|---|---|
| Depression | **All** | **75.00** | 0.769/0.727 |
| | **Neutral** | 63.89 | 0.667/0.606 |
| | **Positive** | 69.44 | 0.718/0.667 |
| | **Negative** | 58.33 | 0.615/0.545 |
| PTSD | **All** | 69.77 | 0.697/0.697 |
| | **Neutral** | **72.09** | 0.714/0.727 |
| | **Positive** | 69.77 | 0.697/0.697 |
| | **Negative** | 52.38 | 0.500/0.545 |

Table 3: Classification results for depression and PTSD classification in leave-one-speaker-out validation experiments. Baseline classification at 50% accuracy for the depression condition and 53.5% for PTSD condition respectively.

For the decision we integrate the classifications for each single speech segment falling into the specific polarity category, depending on the polarity assessment specified in Section 3.2, and form an overall temporally integrated decision for each interaction. The results for the two types of experiments are reported below and summarized in Tables 3.

**Depression Classification:** The performance for the depression classification experiment ranges between 75.00% ($F_1$ score for the conditions depression 0.769 and no-depression 0.727) when utilizing all the data and 58.33% ($F_1$ score for the conditions depression 0.615 and no-depression 0.545) in the negative polarity.

**PTSD Classification:** The performance for the PTSD classification experiment ranges between 72.09% ($F_1$ score for the conditions PTSD 0.714 and no-PTSD 0.727) when utilizing the neutral polarity and 52.38% ($F_1$ score for the conditions PTSD 0.501 and no-PTSD 0.545) in the negative polarity.

## 7. Discussion

In the following we discuss the findings in more detail following the defined research goals R1-R3 in Section 1.

**R1:** The observed participants show more tense voice features in the case of the conditions depression and PTSD when compared to participants in the negative conditions. These findings enrich the commonly investigated state-of-the-art characteristics and could provide clinicians with additional information when assessing the psychological state of a patient. In contrast to more standard speech features, such as fundamental frequency ($f_0$), these investigated voice quality features are more gender independent than $f_0$, which adds to the indicators' robustness. The gender independence is underlined by the following observations: We find no significant difference between genders within the voice quality parameters (except for peakSlope with $\mu$ = -0.338 ($\sigma$ = 0.039) for female depressed participants and $\mu$ = -0.399 ($\sigma$ = 0.041) for male depressed participants; p = 0.006; Hedges' g = 1.451), however, for $f_0$ the groups are highly significantly different ($\mu$ = 181.318 ($\sigma$ = 20.291) for female depressed participants and $\mu$ = 99.032 ($\sigma$ = 17.400) for male depressed participants; $p < 0.001$; Hedges' g = 4.069). The same holds for the other three conditions (i.e. no-depression, PTSD, and no-PTSD)[2].

**R2:** Within the speaker independent classification experiments we could observe promising results using a basic SVM with just four features as input. The SVM could distinguish depression from no-depression with an accuracy of up to 75% and participants with PTSD from no-PTSD with 72.09% on an interview level. These findings underline the discriminative power of the utilized features and confirm the observed large statistical differences. The similar performances between the classification of the two psychological disorders was expected as the groups largely overlap due to the strong correlation between the two measures PHQ-9 and PCL-C [12]. We expect this performance to increase when utilizing additional features as presented in related work (cf. Section 2) as well as more sophisticated classifiers taking advantage of the dynamic nature of speech. Additionally, visual cues such as facial expressions, gaze, gestures, and posture can further improve the achieved performance [12, 25].

**R3:** We could confirm that classification results depend on the question polarity (i.e. the virtual human asks questions with neutral, positive, or negative polarities). Previous work suggests that positive attenuation is common in psychologically distressed participants [5]. Similarly, in our experiments we could distinguish participants with psychological disorders from those without with a higher precision using the responses to the positive and neutral questions. We argue that participants without psychological disorders adapt to the more positive stimuli, whereas participants with psychological disorders remain on at the same level of engagement and do not adapt towards the positive stimuli. With respect to the neutral questions, we believe that the improved performance of the automatic classifiers could be due to their increased faculty of discrimination[3]. Unfortunately, we could not confirm negative potentiation within our experiments, which might be due to the lack of negativity in the investigated stimuli within our experiments. Hence, we plan to further increase the strength of stimuli in the future and investigate this effect further.

## 8. Conclusion

Based on the research goals stated in Section 1, we can identify three main contributions of this work: (1) we observed significant differences in the speaker's voice quality with respect to depression and PTSD when compared to control participants, in particular participants with psychological disorders exhibit more tense voice qualities; (2) the limited number of features allows for good classification performance with about 70% accuracy in a leave-one-speaker-out classification; and (3) question polarity influences the separability of conditions: in the presented experiments, positive and neutral questions allowed for a better disambiguation of the groups than negative ones, which is conform with the discussed positive attenuation in related work [5]. These results are quite promising and we believe that in the future healthcare providers could benefit from such automatic information derived from the patients' voice in order to assess potential psychological disorders more effectively and potentially with higher accuracy.

## 9. Acknowledgements

This work is supported by DARPA under contract (W911NF-04-D-0005) and U.S. Army Research, Development, and Engineering Command and. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

[2] At this point, we would like to note that we are aware of possible pharmaceutical influences on the speaker's voice, however, no medication records were collected from the participants due to IRB limitations.

[3] As an analogy for this we would like to use a yellow traffic light: the more aggressive driver might still cross the light, whereas the careful driver will stop more likely. However, both drivers would pass or stop at a green and red light respectively.

# 10. References


[1] P. Ekman and W. V. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," *Semiotica*, vol. 1, pp. 49–98, 1969.

[2] P. Waxer, "Nonverbal cues for depression," *Journal of Abnormal Psychology*, vol. 83, no. 3, pp. 319–322, 1974.

[3] C. Sobin and H. A. Sackheim, "Psychomotor symptoms of depression," *American Journal of Psychiatry*, vol. 154, no. 1, pp. 4–17, 1997.

[4] J. E. Perez and R. E. Riggio, *Nonverbal social skills and psychopathology*, ser. Nonverbal behavior in clinical settings. Oxford University Press, 2003, pp. 17–44.

[5] L. M. Bylsam, B. H. Morris, and J. Rottenberg, "A meta-analysis of emotional reactivity in major depressive disorder," *Clinical Psychology Review*, vol. 28, pp. 676–691, 2008.

[6] Y. Yang, C. Fairbairn, and J. F. Cohn, "Detecting depression severity from vocal prosody," *IEEE Transactions on Affective Computing*, vol. 99, no. PrePrints, 2012.

[7] J. A. Hall, J. A. Harrigan, and R. Rosenthal, "Nonverbal behavior in clinician-patient interaction," *Applied and Preventive Psychology*, vol. 4, no. 1, pp. 21–37, 1995.

[8] J. Rottenberg, J. J. Gross, and I. H. Gotlib, "Emotion context insensitivity in major depressive disorder," *Journal of Abnormal Psychology*, vol. 114, no. 4, pp. 627–639, 2005.

[9] 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, no. 9, pp. 1530–1540, 2004.

[10] M. Elliott, M. A. Clements, J. W. Peifer, and L. Weisser, "Critical analysis of the impact of glottal features in the classification of clinical depression in speech," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 96–107, 2008.

[11] J. F. Cohn, T. S. Kruez, I. Matthews, Y. Ying, M. H. Nguyen, M. T. Padilla, F. Zhou, and F. De la Torre, "Detecting depression from facial actions and vocal prosody," in *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009, pp. 1–7.

[12] S. Scherer, G. Stratou, M. Mahmoud, J. Boberg, J. Gratch, A. Rizzo, and L.-P. Morency, "Automatic behavior descriptors for psychological disorder analysis," in *accepted for publication at IEEE Conference on Automatic Face and Gesture Recognition*, 2013.

[13] M. Thiebaux, S. Marsella, A. N. Marshall, and M. Kallmann, "Smartbody: behavior realization for embodied conversational agents," in *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 1*, ser. AAMAS '08. International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 151–158.

[14] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The phq-9," *Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, 2001.

[15] S. Scherer, J. P. Pestian, and L.-P. Morency, "Investigating the speech characteristics of suicidal adolescents," in *accepted for publication at International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

[16] S. Scherer, J. Kane, C. Gobl, and F. Schwenker, "Investigating fuzzy-input fuzzy-output support vector machines for robust voice quality classification," *Computer Speech and Language*, vol. 27, no. 1, pp. 263–287, 2013.

[17] J. Kane, S. Scherer, M. Aylett, L.-P. Morency, and C. Gobl, "Speaker and language independent voice quality classification applied to unlabelled corpora of expressive speech," in *accepted for publication at International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

[18] P. Alku, T. Bäckström, and E. Vilkman, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech Communication*, vol. 11, no. 2-3, pp. 109–118, 1992.

[19] P. Alku, T. Bäckström, and E. Vilkman, "Normalized amplitude quotient for parameterization of the glottal flow," *Journal of the Acoustical Society of America*, vol. 112, no. 2, pp. 701–710, 2002.

[20] T. Hacki, "Klassifizierung von glottisdysfunktionen mit hilfe der elektroglottographie," *Folia Phoniatrica*, pp. 43–48, 1989.

[21] J. Kane, S. Scherer, L.-P. Morency, and C. Gobl, "A comparative study of glottal open quotient estimation techniques," in *to appear in Proceedings of Interspeech 2013*. ISCA, 2013.

[22] C. d'Alessandro and N. Sturmel, "Glottal closure instant and voice source analysis using time-scale lines of maximum amplitude," *Sadhana*, vol. 36, no. 5, pp. 601–622, 2011.

[23] J. Kane and C. Gobl, "Identifying regions of non-modal phonation using features of the wavelet transform," *Proceedings of Interspeech, Florence, Italy*, pp. 177–180, 2011.

[24] L. V. Hedges, "Distribution theory for glass's estimator of effect size and related estimators," *Journal of Educational Statistics*, vol. 6, no. 2, pp. 107–128, 1981.

[25] J. Joshi, R. Goecke, G. Parker, and M. Breakspear, "Can body expressions contribute to automatic depression analysis?" in *accepted for publication at IEEE Conference on Automatic Face and Gesture Recognition*, 2013.