Dyadic Behavior Analysis in Depression Severity Assessment Interviews

Stefan Scherer
USC Institute for Creative Technologies
12015 Waterfront Dr.
Playa Vista, CA
scherer@ict.usc.edu

Zakia Hammal
Carnegie Mellon University
5000 Fifth Avenue
Pittsburgh, PA
zakia_hammal@yahoo.fr

Ying Yang
University of Pittsburgh
Biomedical Science Tower
Pittsburgh, PA 15213
yiy17@pitt.edu

Louis-Philippe Morency
USC Institute for Creative Technologies
12015 Waterfront Dr.
Playa Vista, CA
morency@ict.usc.edu

Jeffrey F. Cohn
University of Pittsburgh
210 S. Bouquet St.
Pittsburgh, PA 15260
jeffcohn@cs.cmu.edu

ABSTRACT

Previous literature suggests that depression impacts vocal timing of both participants and clinical interviewers but is mixed with respect to acoustic features. To investigate further, 57 middle-aged adults (men and women) with Major Depression Disorder and their clinical interviewers (all women) were studied. Participants were interviewed for depression severity on up to four occasions over a 21 week period using the Hamilton Rating Scale for Depression (HRSD), which is a criterion measure for depression severity in clinical trials. Acoustic features were extracted for both participants and interviewers using COVAREP Toolbox. Missing data occurred due to missed appointments, technical problems, or insufficient vocal samples. Data from 36 participants and their interviewers met criteria and were included for analysis to compare between high and low depression severity. Acoustic features for participants varied between men and women as expected, and failed to vary with depression severity for participants. For interviewers, acoustic characteristics strongly varied with severity of the interviewee's depression. Accommodation - the tendency of interactants to adapt their communicative behavior to each other - between interviewers and interviewees was inversely related to depression severity. These findings suggest that interviewers modify their acoustic features in response to depression severity, and depression severity strongly impacts interpersonal accommodation.

ICMI'14, November 12 - 16 2014, Turkey
Copyright 2014 ACM 978-1-4503-2885-2/14/11 ...$15.00.
http://dx.doi.org/10.1145/2663204.2663238.

General Terms

H.1.2 User/Machine Systems; H.5.m Miscellaneous; J.4 Social and Behavioral Sciences

Keywords

Depression; Dyadic Interaction; Accommodation; Voice Quality

1. INTRODUCTION

Diagnosis and assessment of depression is almost entirely informed by what patients, their families, or caregivers report. Standardized procedures for incorporating nonverbal behavior and voice characteristics, in particular, are lacking. Their absence is especially salient for depression, a mood disorder for which disruption in emotion experience, communication, and self-regulation are key features [1, 69, 14, 20]. Within the past decade, significant progress has been made in linking voice characteristics to emotion [26, 39, 61, 2, 68], turn-taking, reciprocity [19, 55], and a broad range of interpersonal outcomes [38, 54]. There is strong reason to believe that automatic analysis of voice characteristics could provide a powerful tool to assist in detection and assessment of depression over the course of treatment and recovery. Improved measurement and understanding of the relation between depression and voice characteristics could aid early detection, and lead to better understanding of mechanisms and improved interventions. The lower vocal tract is innervated by the vagal nerve, and thus provides important information about the peripheral physiology of depression [51, 56]. Because depression is one of the most prevalent mental health disorders [44] and a leading cause of disability worldwide [52], the potential contribution of improved measurement is great.

The timing, amplitude, and acoustic features of speech have been investigated [63, 60]. Several investigations have revealed reduced speech variability and monotonicity in loudness and pitch [50, 13, 47, 25], reduced speech [33], reduced articulation rate [11], and increased pause duration [63, 6] when compared to non-depressed comparison participants.

Varied switching pause duration was found to be correlated with depression severity in a within-subject analysis [66]. Findings for acoustic features in general have found depression effects, as well. These effects include increased tension in the vocal tract and the vocal folds in between-subject studies [13, 59], and speech characteristics related to psychomotor retardation [22, 62]. Yang et al. [66], however, found no differences in fundamental frequency (f_0) in relation to depression severity.

With two notable exceptions, previous work has focused on depressed participants and ignored the interview context and their interviewers in particular. Scherer et al. [58] studied suicidal adolescents who were depressed and their interviewers. Suicidal adolescents had more breathy voice qualities than non-suicidal comparison participants, interviewers' acoustic characteristics were correlated with these changes. Backchannels provided by the interviewer were more breathy when they were interviewing suicidal adolescents. Yang et al. [66] found that interviewer, but not interviewee, f_0 mean and variability showed a strong relationship with severity of depression. Interviewers used lower and more variable f_0 when speaking with participants who were more depressed than they did when speaking with participants who were less depressed.

A number of factors might account for the discrepancy with respect to acoustic features between the different studies. The participants in Yang et al., study all met criteria for Major Depressive Disorder and were studied with respect to change in severity over time. Other studies have compared depressed and non-depressed participants. Perhaps most important, the analysis of acoustic characteristics by Yang et al. was limited to f_0. Other work has more comprehensively analyzed acoustic features. To investigate whether the reduced feature set in Yang et al. may have accounted for this discrepancy, we analyze their recordings using more comprehensive, state of the art procedures to assess voice characteristics related to both prosody as well as voice quality, collected in a freely available speech signal processing toolbox [16]. Further, we address the dyadic analysis of acoustic and nonverbal accommodation, which represents the major novelty of this study.

2. METHODS AND MATERIALS

2.1 Participants

Fifty-seven depressed participants (34 women, 23 men) were recruited from a clinical trial for treatment of depression. They ranged in age from 19 to 65 years (mean = 39.65) and were Euro- or African-American (46 and 11, respectively). At the time of study intake, all met DSM-IV [1] criteria [21] for Major Depressive Disorder (MDD). Although not a focus of this report, participants were randomized to either anti-depressant treatment with a selective serotonin re-uptake inhibitor (SSRI) or Interpersonal Psychotherapy (IPT). Both treatments are empirically validated for treatment of depression [37].

2.2 Observational Procedures

Symptom severity was evaluated on up to four occasions at 1, 7, 13, and 21 weeks by ten clinical interviewers (all female). Interviewers were not assigned to specific participants, and they varied in the number of interviews they conducted. Four interviewers were responsible for the bulk of the interviews. The median number of interviews per interviewer was 14.5; four conducted six or fewer. Interviews were conducted using the Hamilton Rating Scale for Depression (HRSD) [34], which is a criterion measure for assessing severity of depression. Interviewers all were expert in the HRSD and reliability was maintained above 0.90. HRSD scores of 15 or higher are generally considered to indicate moderate to severe depression; and scores of 7 or lower to indicate a return to normal [24]. We used these cut-off scores to define the two conditions of high depression and low depression in this study. Subjects scoring between cut-off scores 7 and 15 were excluded from the analysis.

Interviews were recorded using four hardware synchronized analogue cameras and two unidirectional microphones. Two cameras were positioned approximately 15° to the participant's left and right to record their shoulders and face. A third camera recorded a full body view while a fourth recorded the interviewer's shoulders and face from approximately 15° to their right. Audio was digitized at 48 kHz and later down-sampled to 16 kHz for speech processing. Missing data occurred due to missed appointments, or technical problems. Technical problems included failure to record audio or video, audio or video artifacts, and insufficient amount of data. To be included for analysis, we required a minimum of 20 speaker turns with at least 3 seconds in length and at least 50 seconds of vocalization in total. Thus, the final sample was 61 sessions from 36 participants; 47 score high on HRSD and 14 low.

2.3 Preprocessing

To attenuate noise as well as to equalize intensity, Adobe Audition II [54] was used to reduce noise level and equalize intensity. An intermediate level of 40% noise reduction was used to achieve the desired signal-to-noise ratio without distorting the original signal. Each pair of recordings was transcribed manually using Transcriber software [7] and then force-aligned using CMU Sphinx III [2] post-processed using Praat [8]. Because session recordings exceeded the memory limits of Sphinx, it was necessary to segment recordings prior to forced alignment. While several approaches to segmentation were possible, we segmented recordings at transcription boundaries; that is, whenever a change in speaker occurred. Except for occasional overlapping speech, this approach resulted in speaker-specific segments. Forced alignment produced a matrix of four columns: speaker (which encoded both individual and simultaneous speech), start time, stop time, and utterance. To assess the reliability of the forced alignment, audio files from 30 sessions were manually aligned and compared with the segmentation yielded by Sphinx. Mean error (s) for onset and offset, respectively, were .097 and .010 for participants and .053 and .011 for interviewers.

The forced alignment timings were used to identify speaker-turns and speaker diarization for the subsequent automatic feature extraction (cf. Section 2.4).

2.4 Acoustic Features

For the processing of the speech signals, we use the freely available COVAREP toolbox, a collaborative speech analysis repository available for Matlab and Octave [16][1]. COVAREP provides an extensive selection of open-source ro-

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

bust and tested speech processing algorithms enabling comparative and cooperative research within the speech community.

The extracted acoustic features were chosen based on previous encouraging results in classifying voice patterns of suicidal adolescents and distressed adults [58, 59] as well as the features' relevance for characterizing voice qualities on a breathy to tense dimension [57, 42]. Below we introduce the utilized speech features in more detail.

2.4.1 Fundamental Frequency (f_0)

In [18], a method for fundamental frequency f_0 tracking and simultaneous voicing detection based on residual harmonics is introduced. The method is especially suitable in noisy and unconstrained conditions. The residual signal $r(t)$ is calculated from the speech signal $s(t)$ for each frame using inverse filtering, for all times t. In particular, we utilize a linear predictive coding (LPC) filter of order $p = 12$ estimated for all Hanning windowed speech segments. Each speech segment has the length of 25 ms and is continuously shifted by 5 ms. This process removes strong influences of noise and vocal tract resonances. For each $r(t)$ the amplitude spectrum $E(f)$ is computed, revealing peaks for the harmonics of f_0, the fundamental frequency. Then, the summation of residual harmonics (SRH) is computed as follows [18]:

$$SRH(f) = E(f) + \sum_{k=2}^{N_{harm}} [E(k \cdot f) - E((k - \frac{1}{2}) \cdot f)], \quad (1)$$

for $f \in [f_{0,\min}, f_{0,\max}]$, with $f_{0,min} = 50$, $f_{0,max} = 500$, and $N_{harm} = 5$. The frequency f for which $SRH(f)$ is maximal $f_0 = \arg\max_f (SRH(f))$ is considered the fundamental frequency of the investigated speech frame. By using a simple threshold $\theta = 0.07$, the unvoiced frames can be discarded as in [18]. Unvoiced samples, i.e. times when no vocal fold vibration is present, are not analyzed for any of the extracted features.

2.4.2 Maxima Dispersion Quotient (MDQ) and Peak Slope (PS)

The Maxima Dispersion Quotient (**MDQ**) and the Peak Slope (**PS**) measure involve a dyadic wavelet transform using $g(t)$, a cosine-modulated Gaussian pulse similar to that used in [12] as the mother wavelet:

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

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:

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

where $*$ denotes the convolution operator and $s_i = 2^i$. This functions essentially as an octave band zero-phase filter bank. For the PS feature [40], 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 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. Finally, the measurement of the maxima dispersion quotient (MDQ, [41]) uses the Linear Prediction (LP) residual as $x(t)$ in Eq. (3). Then using the GCIs, located using SE-VQ, the dispersion of peaks in relation to the GCI position is averaged across the different frequency bands and then normalized to the local glottal period. For tense voice, where the sharp closing of the glottis is analogous to an impulse excitation the maxima are tightly aligned to the GCI, whereas for laxer phonation the maxima become highly dispersed.

2.4.3 Normalized Amplitude Quotient (NAQ)

The normalized amplitude quotient (**NAQ**) feature is derived from the glottal source signal estimated by iterative adaptive inverse filtering (IAIF, [4]). The output is the differentiated glottal flow. The normalized amplitude quotient (NAQ, [5]) is calculated using:

$$\mathrm{NAQ} = \frac{f_{ac}}{d_{peak} \cdot T_0}, \quad (4)$$

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.

NAQ is a direct measure of the glottal flow and glottal flow derivative and as an amplitude based parameter, was shown to be more robust to noise disturbances than parameters based on time instant measurements and has, as a result, been used in the analysis of conversational speech [10], which is frequently noisy. The parameter, however, may not be as effective as a voice quality indicator when a speaker is using a wide f_0 range [29].

2.4.4 Quasi Open Quotient (QOQ)

The quasi-open quotient (**QOQ**, [32]) is also derived from amplitude measurements of the glottal flow pulse and is a frequently used correlate of the open quotient OQ, i.e. the period the vocal folds are open. OQ is a salient measurement of the glottal pulse, thought to be useful for discriminating breathy to tense voice [35, 36, 64]. OQ can be defined as the duration of the glottal open phase normalized to the local glottal period. 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.

Below the investigated speech parameters are specified with subscripts $_P$ and $_I$ to specify if they relate to participants or interviewers speech respectively.

2.5 Accommodation Analysis

Accommodation, also described under the terms convergence [27] [53], entrainment [9], or mimicry [54], refers to the tendency of interactants to adapt their communicative behavior to each other.

To measure the synchrony in the development of the extracted speech parameters for each interactant we utilized the standard Pearson correlation coefficient $\rho \in [-1, 1]$, that measures linear dependencies between two sets of observations x and y:

$$\rho_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)\sum_{i=1}^{N}(y_i - \mu_y)}{(N-1)\sigma_x\sigma_y}, \quad (5)$$

where $|x| = |y| = N$ the length of the observation set, μ_x the mean value of x (respectively μ_y) , σ_x the standard deviation of x (respectively σ_y), and $x_i \in x \quad \forall i = 1, ..., N$ (respectively y_i). For large $\rho_{xy} > 0$ we have strong linear dependencies, which indicates a synchronous behavior of the prosodic parameters over the analyzed fragment. Small values $\rho_{xy} < 0$ indicate strong asynchronous developments of the observed parameters. For values close to zero no linear correlation is observed, i.e. a state of maintenance is present.

Within this study, we investigate correlation between participants' and interviewers' speech characteristics using a time-aligned moving average approach [15]. Interviews were subdivided into successive segments of 50 seconds with 10 second shifts (i.e. 40 seconds overlap). For each segment, characteristics for both the participant and the respective interviewer are computed. This results in a time series of observations for both the interviewer and the participant. A Pearson correlation ρ was computed for each observed feature for each interview.

The observed accommodation between the participants' and interviewers' acoustic parameters are denoted as ρ with the respective acoustic parameter as subscripts.

3. DATA ANALYSES AND RESULTS

We investigated the acoustic characteristics of participants, interviewers, and the accommodation between participants and interviewers during high depression severity and low severity, respectively.

3.1 Condition Effects on Participant Behavior

Do acoustic characteristics of participants vary with depression severity? In particular, we investigate acoustic parameters characterizing the participants' voice quality on a breathy to tense dimension.

Table 1: Participant acoustic parameter correlation. P for participants.

	$\mathbf{f}_{0,P}$	$\mathbf{MDQ}_P$	$\mathbf{PS}_P$	$\mathbf{NAQ}_P$	$\mathbf{QOQ}_P$
$\mathbf{f}_{0,P}$	1	$.711^{**}$	$.614^{**}$	$-.402^{**}$	$-.520^{**}$
$\mathbf{MDQ}_P$		1	$.400^{**}$	-.020	$-.315^{**}$
$\mathbf{PS}_P$			1	$-.507^{**}$	$-.540^{**}$
$\mathbf{NAQ}_P$				1	$.847^{**}$
$\mathbf{QOQ}_P$					1

Because speech parameters were highly correlated (Table 1[2]), they were analyzed using MANOVA. Depression severity, sex, and their interaction were included as covariates. There was no effect for depression severity (F(5, 55) = 1.759, p = .137) or sex-by-depression severity interaction (F(5, 55) = 0.964, p = .448). There was a significant effect for sex (F(5, 55) = 5.339, p < .01).

Follow-up F tests revealed significant sex effects for all speech parameters with the exception of NAQ_P. As expected, fundamental frequency was higher for women than for men. Further, significantly more breathy voice characteristics were observed as measured with MDQ_P, and more tense voice characteristics as measured using PS_P as well as QOQ_P, for details see Table 2

Table 2: Follow-up F test results for participant behavior depending on sex.

	Women		Men	
	μ	σ	μ	σ
$\mathbf{f}_{0,P}$	191.77	26.24	145.05	38.26^{**}
$\mathbf{MDQ}_P$	0.130	0.007	0.121	0.010^{**}
$\mathbf{PS}_P$	-0.422	0.034	-0.465	0.028^{**}
$\mathbf{QOQ}_P$	0.380	0.072	0.444	0.082^{*}

3.2 Condition Effects on Interviewer Behavior

Do acoustic characteristics of interviewers vary with depression severity?

Table 3: Interviewer acoustic parameter correlation. I for Interviewer.

	$\mathbf{f}_{0,I}$	$\mathbf{MDQ}_I$	$\mathbf{PS}_I$	$\mathbf{NAQ}_I$	$\mathbf{QOQ}_I$
$\mathbf{f}_{0,I}$	1	$.579^{**}$	$.546^{**}$	-.134	$-.239^{*}$
$\mathbf{MDQ}_I$		1	.202	$.442^{**}$	$.281^{**}$
$\mathbf{PS}_I$			1	$-.224^{*}$	-.208
$\mathbf{NAQ}_I$				1	$.934^{**}$
$\mathbf{QOQ}_I$					1

Speech parameters for interviewers were highly correlated (Table 3) and were analyzed using MANOVA. Depression severity, sex, and their interaction were included as covariates. There was a significant effect of depression severity on interviewers' acoustic characteristics (F(5, 55) = 2.609, p = .035). There was a marginally significant effect for sex (F(5, 55) = 2.121, p = .077) and no sex-by-depression severity interaction (F(5, 55) = 0.189, p = .966).

Table 4: Follow-up F test results for interviewer behavior depending on depression severity and sex.

	High Severity		Low Severity	
	μ	σ	μ	σ
$\mathbf{PS}_I$	-0.407	0.022	-0.383	0.027^{**}
$\mathbf{NAQ}_I$	0.099	0.022	0.086	0.015
	Women		**Men**	
	μ	σ	μ	σ
$\mathbf{f}_{0,I}$	195.10	26.24	170.41	38.26^{*}
$\mathbf{MDQ}_I$	0.130	0.007	0.125	0.010^{*}

Follow-up F tests for the individual speech parameters revealed significant effects for breathier voice quality as measured by PS_I and a trend to more breathy voice as measured with NAQ_I. Effects for other parameters failed to reach significance. PS_I and NAQ_I per condition and their standard errors are visualized in Figure 1; for details see Table 4.

Several follow-up F tests for effects of participant sex on interviewers were significant. Fundamental frequency of the interviewer and breathiness were higher when the participant was a woman (cf. Table 4)

3.3 Accommodation Between Interactants

Does accommodation vary with depression severity?

As discussed in Section 2.5 we compute Pearson's ρ for all speech parameters. The extracted accommodation parameters were moderately correlated (cf. Table 5). As above,

[2] All p-values of statistical tests < .05 and < .01 are denoted with * and ** respectively.

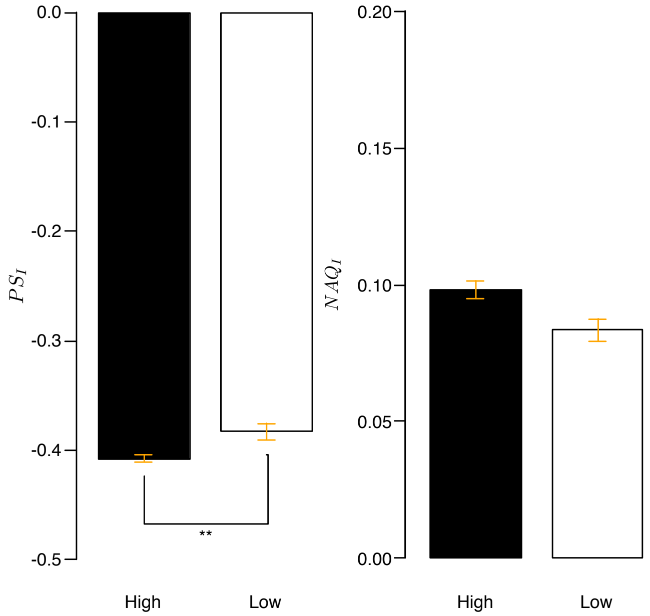

Figure 1: Acoustic Features of Interviewer Speech Characteristics. Observed acoustic features of interviewers across conditions high depression severity (High) vs. low severity (Low). The displayed whiskers signify standard errors and the bracket show significant results with ** ... $p < .01$.

Table 5: Correlation of acoustic accommodation measurements between participants and interviewers. ρ for the observed accommodation.

	ρ_{f_0}	ρ_{MDQ}	ρ_{PS}	ρ_{NAQ}	ρ_{QOQ}
ρ_{f_0}	1	.362**	.194	.279*	.269*
ρ_{MDQ}		1	-.024	.236*	.033
ρ_{PS}			1	.318**	.532**
ρ_{NAQ}				1	.479**
ρ_{QOQ}					1

they were analyzed using MANOVA with depression severity, sex, and sex-by-depression severity interaction entered as covariates. The effect of depression severity was significant (F(5, 55) = 2.508, p = .037). There was no main effect or interaction for sex (F(5, 55) = 1.648, p = .16; and F(5, 55) = 1.195, p = .320, respectively).

Table 6: Follow-up F test results for interactant accommodation depending on depression severity.

	High Severity		Low Severity	
	μ	σ	μ	σ
ρ_{PS}	0.160	0.300	0.350	0.281*
ρ_{QOQ}	0.093	0.274	0.269	0.331**

Follow-up F tests for the individual speech variables and the observed accommodation measured as Pearson's ρ revealed significant effects for higher accommodation for low depression severity for ρ_{PS} and ρ_{QOQ}. Effects for other parameters failed to reach significance, for details see Table 6.

The observed accommodation for ρ_{PS} and ρ_{QOQ} per condition and the standard errors are visualized in Figure 2.

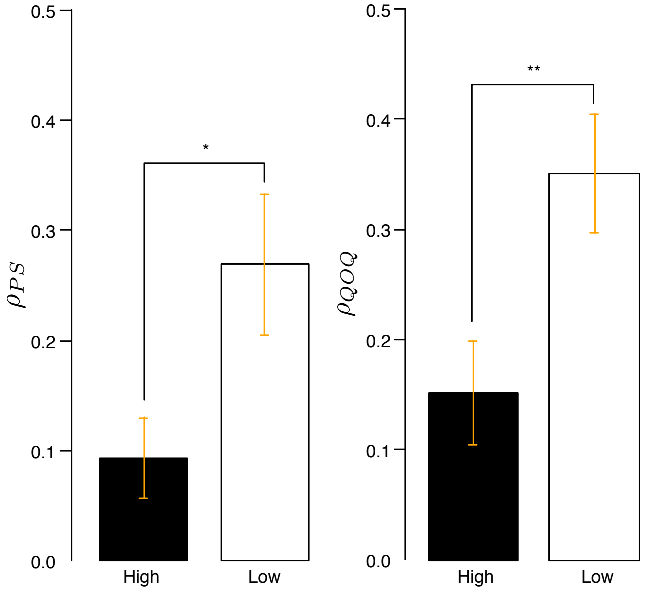

Figure 2: Acoustic Accommodation between Participants and Interviewers. Observed acoustic accommodation as measured as Pearson's ρ between participants and interviewers across conditions high depression severity (High) vs. low severity (Low). The displayed whiskers signify standard errors and the brackets show significant results with * ... $p < .05$ and ** ... $p < .01$.

4. DISCUSSION

We investigated acoustic characteristics in depressed participants and their interviewers. The participants varied in their level of depression severity (that is, high versus low). For participants, of the five acoustic parameters evaluated, none varied with depression severity. The only obtained effects were expected ones for sex. Women for example had higher fundamental frequency. These findings for sex effects are consistent with [45]. The lack of findings for depression effects could be due to several factors. A methodological reason may be that we used only a subset of the data used by [66], as well as the use of SSRI might have influenced speech characteristics. Further, most previous findings for f_0 effects in depression were from comparisons with non-depressed participants [3]. Because participants with depression differ in many ways from those without depression [46], such results lack specificity for depression. Those studies that have investigated change in depression severity over time have reported very small effect sizes that depended on large numbers of participants [49]. Small effect sizes are unlikely to have much utility for machine learning.

Another factor may be differences in participant characteristics across studies. With the exception of [66], previous work has compared depressed with non-depressed participants [13, 59, 65]; whereas we made comparisons within a depressed sample. All participants met criteria for Major Depression Disorder, all had histories of multiple, protracted episodes, and they varied only in their current level of severity. People with chronic history of depression may be very different than those that have never been depressed or only depressed in a limited way. High trait neuroticism and low trait extraversion, for instance, are much more likely in people with histories of chronic depression [46] These differences may contribute to relative stability in acoustic parameters robust to variation in depression severity.

Other factors may have been context related. In [59], participants were interviewed by virtual humans rather than humans and the interviews included broader range of questions [31, 17]. The latter may have occasioned more diversity in vocal samples, and vocal timing varied markedly from that with human interviewers. Accommodation was not possible with the virtual humans. [48] studied depressed adolescents in extended conversations with their two parents. In [65], participants were involved in human-computer interaction tasks. These factors, too, may explain the differences across studies. Greater attention is needed to the observational contexts and individual differences in studies.

Consistent with previous findings [66], interviewer acoustic characteristics showed a strong relationship with severity of depression. In particular, interviewers exhibited significantly more breathy voice characteristics when interacting with highly depressed participants. Our findings extend the findings of Yang et al. [66] who found that interviewer f_0 mean and variability were strongly related to depression severity. The obtained results can be explained by the increased felt empathy towards the interviewee. As previously investigated, breathy voice qualities were associated with sad emotion, friendly attitude, and intimate interpersonal relation in perceptual experiments [30]. Further, Scherer et al. have previously reported more breathy voice qualities in interviewers' voice and back-channels when interviewing suicidal adolescents [58]. Findings such as these suggest that depression effects are bidirectional and that nonverbal behavior in the non-depressed interactant could provide a sensitive barometer of depression. Behavioral differences of interviewers may be more indicative of a participant's condition than the participant's own behavior [67]. This further underlines the importance to investigate the behaviors jointly within dyadic HRSD interviews as argued in [23].

We utilized a time-aligned moving average approach to assess acoustic accommodation within dyadic HRSD screening interviews between the interviewer and the participant. This approach overcomes the issue of asynchronously observed speech in dyadic interactions and enables the investigation of accommodation between the interactants [15]. With respect to the accommodation between interactants, we identified a main effect of depression severity on PS and QOQ (cf. Section 2.4). They both were inversely correlated with depression severity. Hence, accommodation is significantly increased for interviews where the participants have low depression severity. These findings are consistent with the hypothesis that a function of depression is to attenuate interpersonal coordination in the service of social isolation [28]. Further, prosodic accommodation has been found to be perceptually correlated with increased perceived flow of conversation, speaker engagement, and liking [15].

We found accommodation effects, a remaining research question is how these effects are achieved. We might assume that the direction of effects is from participant to interviewer. That is, interviewers are responding to the participants' restricted acoustic characteristics. Alternatively, there may be dynamic adjustments made over the course of the interviews. Participants and interviewers may each be both cause and effect of the other person's behavior [43]

5. CONCLUSIONS

We investigated acoustic characteristics of depressed participants and their interviewers in a clinical interview to assess depression severity. All participants met stringent criteria for Major Depressive Disorder and had histories of multiple and often lengthy episodes. They differed in whether their symptoms were high or low at the time of the interview. We found no variation in acoustic characteristics in relation to severity for the participants. We found strong effects in interviewers of participants' depression severity, and we found strong accommodation effects. To the best of our knowledge, this study is the first to reveal accommodation for acoustic parameters related to depression severity.

Acknowledgements

This work was supported in part by US National Institutes of Health grants MH096951 and MH051435 to the University of Pittsburgh. This work is supported by DARPA under contract (W911NF-04-D-0005) and U.S. Army RDECOM. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

6. REFERENCES

[1] *Diagnostic and statistical manual of mental disorders.* American Psychiatric Association, Washington, DC, 1994.

[2] Cmu sphinx: Open source toolkit for speech recognition. Technical report, Carnegie Mellon University, Undated.

[3] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Breakspear, and G. Parker. From joyous to clinically depressed: Mood detection using spontaneous speech. In *Proceedings of International Florida Artificial Intelligence Research Society Conference (FLAIRS 2012)*, pages 141–146, 2012.

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

[5] P. Alku, T. Bäckström, and E. Vilkman. Normalized amplitude quotient for parameterization of the glottal flow. *Journal of the Acoustical Society of America*, 112(2):701–710, 2002.

[6] M. Alpert, E. R. Pouget, and R. R. Silva. Reflections of depression in acoustic measures of the patient's speech. *Journal of Affective Disorders*, 66(1):59–69, 2001.

[7] C. Barras, E. Geoffrois, Z. Wu, and M. Liberman. Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication special issue on Speech Annotation and Corpus Tools*, 33:5–22, 2001.

[8] P. Boersma. Praat, a system for doing phonetics by computer. *Glot International*, 5(9):341–345, 2001.

[9] S. Brennan. Lexical entrainment in spontaneous dialog. *Proceedings of ISSD*, pages 41–44, 1996.

[10] N. Campbell and P. Mokhtari. Voice quality: The 4th prosodic dimension. In *Proceedings of the 15th international congress of phonetic sciences (ICPhS'03)*, pages 2417–2420. ICPhS, 2003.

[11] M. Cannizzaro, B. Harel, N. Reilly, P. Chappell, and P. J. Snyder. Voice acoustical measurement of the severity of major depression. *Brain and Cognition*, 56(1):30–35, 2004.

[12] C. d'Alessandro and N. Sturmel. Glottal closure instant and voice source analysis using time-scale lines of maximum amplitude. *Sadhana*, 36(5):601–622, 2011.

[13] J. K. Darby, N. Simmons, and P. A. Berger. Speech and voice parameters of depression: a pilot study. *Journal of Communication Disorders*, 17(2):75–85, 1984.

[14] R. J. Davidson, editor. *Anxiety, Depression, and Emotion*. Oxford University Press, 2000.

[15] C. De Looze, S. Scherer, B. Vaughan, and N. Campbell. Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*, 58:11–34, 2014.

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

[17] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgilia, J. Gratch, A. Hartholt, M. Lhommet, G. Lucas, S. Marsella, F. Morbini, A. Nazarian, S. Scherer, G. Stratou, A. Suri, D. Traum, R. Wood, Y. Xu, A. Rizzo, and L.-P. Morency. Simsensei: A virtual human interviewer for healthcare decision support. In *Proceedings of Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1061–1068, 2014.

[18] T. Drugman and A. Abeer. Joint robust voicing detection and pitch estimation based on residual harmonics. In *Proceedings of Interspeech 2011*, pages 1973–1976. ISCA, 2011.

[19] S. Duncan. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 23(2):283–292, 1972.

[20] R. Elliott, R. Zahn, J. F. W. Deakin, and I. M. Anderson. Affective cognition and its disruption in mood disorders. *Neuropsychopharmacology*, 36:153–182, 2011.

[21] M. B. First, R. L. Spitzer, M. Gibbon, and J. B. W. Williams. *Structured clinical interview for DSM-IV axis I disorders*. Biometrics Research Department, New York State Psychiatric Institute, New York, patient edition edition, 1995.

[22] A. J. Flint, S. E. Black, I. Campbell-Taylor, G. F. G. Gailey, and C. Levinton. Abnormal speech articulation, psychomotor retardation, and subcortical dysfunction in major depression. *Journal of Psychiatric Research*, 27(3):309–319, 1993.

[23] G. N. Foley and J. P. Gentile. Nonverbal communication in psychotherapy. *Psychiatry*, 7(6):38–44, 2010.

[24] J. C. Fournier, R. J. DeRubeis, S. D. Hollon, S. Dimidjian, J. D. Amsterdam, R. C. Shelton, and F. J. Antidepressant drug effects and depression severity: A patient-level meta-analysis. *Journal of the American Medial Association*, 303(1):47–53, 2010.

[25] 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 Transactions on Biomedical Engineering*, 47(7):829–837, 2000.

[26] R. W. Frick. Communicating emotion: The role of prosodic features. *Psychological Bulletin*, 97(3):412–429, 1985.

[27] H. Giles, N. Coupland, and J. Coupland. Accommodation theory: Communication, context, and consequence. *Contexts of accommodation: Developments in applied sociolinguistics*, pages 1–68, 1991.

[28] J. M. Girard, J. F. Cohn, M. H. Mahoor, S. M. Mavadati, Z. Hammal, and D. P. Rosenwald. Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses. *Image and Vision Computing Journal*, 2014.

[29] C. Gobl and A. N. Chasaide. Amplitude-based source parameters for measuring voice quality. In *Proceedings of ISCA Tutorial and Research Workshop on Voice Quality: Functions, Analysis and Synthesis (VOQUAL'03)*, pages 151–156. ISCA, 2003.

[30] C. Gobl and A. Ní Chasaide. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, 40:189–212, 2003.

[31] J. Gratch, R. Artstein, G. Lucas, S. Scherer, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella, D. Traum, A. Rizzo, and L.-P. Morency. The distress analysis interview corpus of human and computer interviews. In *Proceedings of Language Resources and Evaluation Conference (LREC)*, 2014.

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

[33] J. A. Hall, J. A. Harrigan, and R. Rosenthal. Nonverbal behavior in clinician-patient interaction. *Applied and Preventive Psychology*, 4(1):21–37, 1995.

[34] M. Hamilton. A rating scale for depression. *Journal of Neurology and Neurosurgery*, 23:56–61, 1960.

[35] H. Hanson, K. Stevens, H. Kuo, M. Chen, and J. Slifka. Towards models of phonation. *Journal of Phonetics*, (29):451–480, 2001.

[36] N. Henrich, C. d'Alessandro, and B. Doval. Spectral correlates of voice open quotient and glottal flow asymmetry: theory, limits and experimental data. *Proceedings of EUROSPEECH, Scandanavia*, pages 47–50, 2001.

[37] S. D. Hollon, M. E. Thase, and J. C. Markowitz. Treatment and prevention of depression. *Psychological Science in the Public Interest*, 3(2):38–77, 2002.

[38] J. Jaffe, B. Beebe, S. Feldstein, C. L. Crown, and M. Jasnow. Rhythms of dialogue in early infancy. *Monographs of the Society for Research in Child Development*, 66:1–8, 2001.

[39] P. N. Juslin and P. Laukka. Communication of emotions in vocal expression and music performance:

Different channels, same code? *Psychological Bulletin*, 129:770–814, 2003.

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

[41] J. Kane and C. Gobl. Wavelet maxima dispersion for breathy to tense voice discrimination. *IEEE Transactions on Audio Speech and Language processing*, Under Review.

[42] 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 *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 7982–7986. IEEE, 2013.

[43] D. A. Kenny, L. Mannetti, A. Pierro, S. Livi, and D. A. Kashy. The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, 83(1):126–137, 2002.

[44] R. Kessler, W. Chiu, O. Demler, and E. E. Walters. Prevalence, severity, and comorbidity of 12-month dsm-iv disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62:617–627, 2005.

[45] D. H. Klatt and L. C. Klatt. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87(2):820–857, 1990.

[46] R. Kotov, W. Gamez, F. Schmidt, and D. Watson. Linking big personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, 136(5):768–821, 2010.

[47] J. Leff and E. Abberton. Voice pitch measurements in schizophrenia and depression. *Psychological Medicine*, 11(4):849–852, 1981.

[48] L.-S. Low, N. C. Maddage, M. Lech, L. B. Sheeber, and N. B. Allen. Detection of clinical depression in adolescents? speech during family interactions. *IEEE Transactions on Biomedical Engineering*, 58(3):574–586, 2010.

[49] J. C. Mundt, A. P. Vogel, D. E. Feltner, and W. R. Lenderking. Vocal acoustic biomarkers of depression severity and treatment response. *Biological Psychiatry*, 72(7):580–587, 2012.

[50] A. Nilsonne. Speech characteristics as indicators of depressive illness. *Acta Psychiatrica Scandinavica*, 77(3):253–263, 1988.

[51] J. P. O'Reardon, P. Cristancho, and A. D. Peshek. Vagus nerve stimulation (vns) and treatment of depression: To the brainstem and beyond. *Psychiatry*, 3(5):54–63, 2006.

[52] W. H. Organization. The global burden of disease: 2004 update. World Health Organization, 2008.

[53] J. Pardo. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119:2382, 2006.

[54] A. Pentland. *Honest signals: How they shape our world*. MIT Press, Cambridge, 2008.

[55] B. S. Reed. Speech rythm across turn transitions in cross-cultural talk-in-interaction. *Journal of Pragmatics*, 42(4):1037–1059, 2010.

[56] A. D. Rubin and R. T. Sataloff. Vocal fold paresis and paralysis. *Otolaryngologic Clinics of North America*, 40(5):1109–1131, 2007.

[57] 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*, 27(1):263–287, 2013.

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

[59] S. Scherer, G. Stratou, J. Gratch, and L.-P. Morency. Investigating voice quality as a speaker-independent indicator of depression and ptsd. In *Proceedings of Interspeech 2013*, pages 847–851. ISCA, 2013.

[60] D. Schrijvers, W. Hulstijn, and B. G. Sabbe. Psychomotor symptoms in depression: a diagnostic, pathophysiological and therapeutic tool. *Journal of Affective Disorders*, 109(1-2):1–20, 2008.

[61] B. Schuller, A. Batliner, S. Steidl, and D. Seppi. Recognising realistic emotions and affect in speech: State of the art and lessons learned from the first challenge. *Speech and Communication*, 53(9/10):1062–1087, 2010.

[62] T. Shimizu, N. Furuse, T. Yamazaki, Y. Ueta, T. Sato, and S. Nagata. Chaos of vowel /a/ in japanese patients with depression: a preliminary study. *Journal of Occupational Health*, 47(3):267–269, 2005.

[63] C. Sobin and H. A. Sackheim. Psychomotor symptoms of depression. *American Journal of Psychiatry*, 154(1):4–17, 1997.

[64] R. Timcke, H. von Leden, and P. Moore. Laryngeal vibrations: measurements of the glottic wave. Part 1: The normal vibratory cycle. *Archives of Otolaryngology - Head and Neck surgery*, 68(1):1–19, 1958.

[65] J. R. Williamson, T. F. Quatieri, B. S. Helfer, R. Horwitz, B. Yu, and D. D. Mehta. Vocal biomarkers of depression based on motor incoordination. In *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, AVEC '13, pages 41–48. ACM, 2013.

[66] Y. Yang, C. Fairbairn, and J. F. Cohn. Detecting depression severity from vocal prosody. *IEEE Transactions on Affective Computing*, 4(2):142–150, 2013.

[67] M. Yarczower, J. E. Kilbride, and A. T. Beck. Changes in nonverbal behavior of therapists and depressed patients during cognitive therapy. *Psychological Reports*, 69(3):915–919, 1991.

[68] Z. Zeng, M. Pantic, G. Roisman, and T. S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 31(1):31–58, 2009.

[69] A. J. Zlochower and J. F. Cohn. Vocal timing in face-to-face interaction of clinically depressed and nondepressed mothers and their 4-month-old infants. *Infant Behavior and Development*, 19:373–376, 1996.