Tracking depressed mood using speech pause patterns

TRACKING DEPRESSED MOOD USING SPEECH PAUSE PATTERNS

Maria K Wolters[1], Luis Ferrini[2], Elaine Farrow[1], Aurora Szentagotai Tatar[3], Christopher D Burton[4]

[1]University of Edinburgh, UK; [1]FVA, Italy; [1] Babes-Bolyai University, Romania; [1] University of Aberdeen, UK
maria.wolters@ed.ac.uk

ABSTRACT

The speech of people with depression often shows clear signs of their condition (e.g., flat intonation, slow speech, long pauses), but it is not clear to what extent these signs covary with diurnal fluctuations in mood. In this paper, we report results from a pilot longitudinal study where 11 people with depression tracked various aspects of their mental health for a month. This included a daily mood tracker and regular completion of speech tasks. Speech tasks were designed to be emotionally neutral and require different levels of automaticity. We found that participants differed in their willingness to complete the speech tasks, and that preliminary analyses show no clear link between mood and prosody. We discuss implications of this study for tracking depressed mood using speech in real-life applications.

Keywords: emotion; depression; prosody; pauses

1. INTRODUCTION

At any one time, 4.4% of the population world wide experiences depression [11], which is characterised by two main symptoms, depressed mood and loss of interest or pleasure in activities [2]. Depression can affect the whole body. Many people with depression move less and more slowly, and speech is often slow and flat in intonation and affect [3, 20, 21].

Mood disorders are generally diagnosed and tracked using measures that rely on self-report [2]. To supplement and potentially replace these measures, researchers have been investigating objective biomarkers for mood. Metrics that can be extracted from the speech waveform are one potential candidate [1, 9, 17, 22].

While most studies (e.g. [7, 17, 25]) focus on detecting presence and severity of depression from longer speech samples collected at 1-3 time points, others (e.g. MONARCA, [10]) use shorter, longitudinal speech data as part of ongoing symptom monitoring.

Help4Mood [24] also follows a longitudinal approach. Whereas MONARCA is intended for people with bipolar depression, who experience highs as well as lows, Help4Mood focuses on people with unipolar depression, where lows dominate.

In this paper, we investigate the relation between speech acoustics, in particular patterns of pausing, and daily mood changes, in a data set generated during a small pilot trial of the Help4Mood system. Since Help4Mood covers three languages, British English, Romanian, and Castilian Spanish, only suprasegmental features are considered. We discuss the challenges to data collection, analysis, and interpretation that we encountered and conclude with suggestions for incorporating speech data into a rich longitudinal picture of a person's mental health.

2. BACKGROUND

2.1. Voice Markers of Depression

Most of what is known about the acoustic markers of depression comes from two types of studies, cross-sectional studies that compares the speech of depressed versus non-depressed people (e.g., [5, 7]), and longitudinal studies, often in conjunction with intervention trials, that collect samples of a person's speech before, during, and after the intervention (e.g. [1, 17, 18]).

Across studies, measures related to pitch range, pitch variation, pause patterns, and speaking rate have been shown to be sensitive to changes in mood. As depression lifts, speakers pause less, speak faster, with more varied pitch and using a larger pitch range. Pauses, in particular total pause duration, have been found to be a potential indicator of psychomotor retardation due to depression [12, 14, 23], which also affects overall speed of movement and reaction times.

The speech samples from which these findings are derived are usually substantial and comprise read speech (e.g., Grandfather passage), spontaneous speech (e.g., talking about one's day), and semiautomatic speech (e.g., counting). For daily tracking of speech patterns linked to mood, such a complex set up is not feasible.

2.2. Help4Mood

Help4Mood is a system for supporting the treatment of people with depression in the community. It collects three types of data: (1) self-reports of mood, thinking patterns, and behaviour; (2) psychomotor symptom data (speech); and (3) activity data. Activity data is collected through an accelerometer that can be worn on the wrist or waist, while psychomotor and self-report data is collected through interaction with a Virtual Agent.

Sessions with Help4Mood are planned by a patient-side Decision Support System which is designed to create varied, interesting sessions while ensuring that sufficient data is collected for planning and interpretation. Since patients are supposed to interact with the system every day, total session duration is designed to be around 15 minutes, unless the user requests a long session.

2.3. Speech in Help4Mood

Three brief speech tasks were created for Help4Mood.

checking in: saying one's name and the current date, probes semi-automatic speech

counting: Counting from one to ten and ten to one, another probe of semi-automatic speech often used in the literature

CFT: naming as many animals as possible in 60 seconds. This task is the semantic category fluency task (CFT), also known as verbal fluency, which has been used extensively to characterize executive function [13].

All speech tasks could be completed in 60 seconds or less. Speech was recorded using a simple graphical interface that allowed patients to start and stop recordings themselves.

Patients were asked to complete a speech task at least three times per week, with a week consisting of seven Help4Mood sessions. The task that was presented to patients was selected randomly among the three candidate tasks. This strategy is a compromise between the need for rich speech data and the time constraints of Help4Mood sessions. We deliberately excluded prompted spontaneous speech tasks that could have yielded emotional speech, because we wanted to avoid rumination and induction of negative mood.

As Help4Mood was potentially used by speakers of four languages (English, Romanian, Spanish, Catalan), we focused on language-independent measures of prosody. Since speech measures were only a small part of the overall Help4Mood package, and users of previous prototype iterations had asked for a system with as few components as possible, speech was recorded by the in-built microphone of the Help4Mood laptop. This often resulted in a relatively noisy signal.

3. METHOD

3.1. Mood Data

The data used in this study comes from a pilot randomised controlled trial of Help4Mood that was conducted in the UK, Romania, and Spain. Within the trial, 14 people (11 female, 80%) were allocated to the Help4Mood condition and used the system for a month. All participants had been diagnosed with depression using a structured clinical interview [2] and had mild to moderate depression. Patients with severe depression were excluded because Help4Mood was not designed for this target group.

Although participants were asked to interact with the system every day, most used it as they saw fit, ranging from once a week to several times a day. At the end of their time with Help4Mood, participants were debriefed in a semi-structured interview.

Participants were provided with a laptop and an actigraph. Extensive system logs, including the speech files, were recorded on the laptop and downloaded on its return to the research team. One laptop was stolen, files from one were lost, and one Help4Mood system was not initialised properly, which leaves us with 11 data sets for analysis.

The mood data used for analysis comes from the Daily Mood Check, a set of 4 validated items that are assessed on a visual analogue scale [16] together with an indication of overall mood. For our analysis, we focused on the overall mood rating, which ranged from 0 to 100, with higher values indicating better mood. If users had provided several values for a day when a speech sample was available, we used the mood rating from the session that had also yielded the speech sample.

3.2. Analysis of Speech Data

Pauses were annotated semi-automatically using Praat's silence detection function with a minimum pause duration of 200ms. Those segments were then adjusted manually. Hesitations and filled pauses were treated as speech, audible breathing was treated as part of pauses.

We computed total speaking time (sum of the du-

ration of all speech segments), total pause time (sum of the duration of all pauses), number of pauses, mean pause duration, and % speaking time (total speaking time / file length in seconds). Pauses that occurred at the start or end of a speech file were not used for analysis.

Pitch and intensity were automatically calculated using Praat version 5.3.01 [4]. Due to low signal quality and substantial background noise, the pitch measures were badly affected by outliers. Our analysis will therefore focus on pausing patterns.

4. RESULTS

All in all, 59 speech files were recorded from 11 participants. Of these, only 46 could be used for analysis. One file contained conversation with the Virtual Agent, one file was too noisy, one file had undefined content, and ten files were empty.

The median number of files per participant was 4 (range: 1-9). Four participants produced less than four speech files, while the remaining 7 produced between 4 and 10 samples, repeating at least one of the three speech tasks (c.f. Table 1).

Despite instructions to use Help4Mood in a quiet, private area, many recordings featured background noise ranging from dogs barking to television. Some patients were not alone while they used the system. In one recording, a female voice can be heard in the background, suggesting animals for the CFT task, and in another, a child's voice can be heard. Most samples were recorded in the evening; 18 were recorded between 22 o'clock at night and 3 o'clock in the morning.

Table 1: Speech Files Produced Per Participant.

Language	# Part.	Samples Produced			
		CFT	Check	Count	All
English	2	2	4	4	10
Romanian	8	16	10	9	35
Spanish	1	0	1	0	1
Total	11	18	15	13	46

4.1. Pauses and Mood

Table 2 gives an overview of speaking and pausing patterns. Participants only paused once when stating their name and the current date (checking in), but paused frequently between numbers when counting, even though the task should be fast and automatic, and between groups of animals when generating the list of animals. While pauses between numbers were brief, reflecting the automaticity of the task, pauses between groups of animals were far longer.

We tested whether any of our three core measures correlated with overall mood using the Spearman test of independence as implemented in the R package coin [15]. The test was stratified by Speech task. None of the associations are significant (mean pause duration: $Z = 0.2477, p < 0.85$; percentage of speaking time: $Z = -0.2603, p < 0.66$, number of pauses: $Z = 1.5441, p < 0.15$).

Table 2: Key Measures of Pause and Speech Duration

Measure	Statistic	Speech Task		
		CFT	Check In	Count
Number of Pauses	Median	15.5	1	14
	1st Quart.	13	0.5	3.5
	3rd Quart.	18	2	19
Mean Pause Duration	Median	2.270	0.470	0.543
	1st Quart.	1.817	0.128	0.417
	3rd Quart.	3.019	1.235	0.746
% Speech	Median	32.3%	87.3%	60.2%
	1st Quart.	24.3%	65.9%	53.6%
	3rd Quart.	53.6%	60.2%	73.0%

4.2. Participant Comments

In the debriefing interviews, participants' opinion of the speech tasks differed. Some were completely averse to speaking to a machine, even though the interface showed the virtual agent listening and looking up to the recording interface.

"I didn't like it. I don't like to speak. I don't like to speak at the laptop, even when I have to Skype or use other similar programs where I have to speak and I don't see a real person and I can't express myself both verbally and non-verbally, like in a classical conversation. I feel embarrassed in situations like that." (RO09)

Some participants also found the speech tasks disruptive, and were not sure why they were being asked to do them.

"That part really annoyed me. The part with the animals ... it was dreadful. It was so hard to concentrate. I was trying to concentrate on my own thoughts and feelings and answer all these questions, and then I had to switch and think animals." (RO03)

Others appreciated the scope for reflection that the tasks afforded them.

"I didn't like [the speech tasks] at first because it showed how really depressed I felt but it was very

insightful when it showed me the progress I made ... I mean the volume, the pace, the energy." (RO17)

There were also problems with the user interface. Several participants asked for more feedback or at least a message indicating that the speech sample had been recorded successfully.

"Sometimes I was not sure if I what I had entered, or what I had recorded was actually there. And I think it would maybe be nice to be able to check to see your answer, or to hear your recording. Even, or just to see something like a tick, like your recording is done, or something." (SCO06)

5. DISCUSSION

Despite the substantial amount of studies that have investigated the prosody of depression, it is not clear how best to design a protocol that allows patients and their clinicians to track the traces of depression in their day-to-day speech.

In Help4Mood, we used a paradigm that relied on short, neutral, easy to analyse passages of speech which preserved patient privacy. Unfortunately, we were unable to see statistically significant trends and collected far less data than expected. This is only partially due to usability issues; it is also a reflection of the way in which patients appropriated Help4Mood. In our preliminary analysis of the full trial data, which is yet to be published, we found that those who engaged with Help4Mood did not use it for daily monitoring, but repurposed it as a way to facilitate reflection and coping. This means that they use it far less frequently than originally intended.

Our results also show that speech tracking is not for everyone, with some participants unwilling to speak to a machine. We also need to consider that most patients checked in with Help4Mood at night, towards the end of their day, when hearing them speak might be disruptive for the people they live with, especially when the Help4Mood laptop is kept in a shared bedroom.

Some of these concerns might be addressed through a smartphone implementation, or an Interactive Voice Response System like the one used by Mundt et al. [17]. A telephone-based app is more portable, and the communication situation, speaking into a telephone, is far more natural than speaking to a computer. The microphones of modern smartphones are also bound to yield better sound quality, especially if the speech is recorded and analysed in situ and not transferred over a mobile phone connection with potentially low sound quality.

Better sound quality would also open up the possitiblity of more in-depth analyses of pitch and voice quality, for which one needs reliable estimates of fundamental frequency, jitter and shimmer.

In addition, one might want to consider detecting breathing, crying, and laughter. In our sample, one participant was close to tears or cried while performing the CFT task, which is a much clearer indicator of mood than any prosodic pattern.

Speech tasks could be designed to evoke specific positive or negative emotions (e.g. [6]), which would further highlight any flat affect. Such tasks need to be designed carefully, though. If negative emotions are evoked or explored, the mood of the person of depression might take a turn for the worse.

Finally, the psychomotor symptoms of depression are complex, and while many people show signs of psychomotor retardation, others become more agitated and angry. Therefore, instead of attempting to detect one particular signature of low mood, it might be more fruitful to use an emotion detection system that includes anger (e.g. [8, 19]).

6. CONCLUSION

Tracking mood through speech acoustics for clinical use is difficult. We need to recognise that not everybody will be amenable to having their speech analysed for signs of mental illness, no matter how the data collection is framed. For those who are willing to explore the effect of mental illness on their speech, we need to design appropriate, unobtrusive tasks that can be analysed efficiently without compromising a person's privacy or inducing a negative mood that then proves difficult to shake off.

We also need a database of speech that reflects typical variations in depressed mood over time, captured in a way that is amenable to unobtrusive tracking and that does not require interviews or samples of speech that are longer than five minutes in total per session.

In future work, we plan to explore variants of the CFT task. Even though some patients might find it hard to complete, it does not require them to verbalise private information, is emotionally neutral, and is amenable to gamification. Additional analysis of the speech content, which covers a limited vocabulary (animals, food items, etc.) yields a clinically meaningful measure of verbal fluency, which can be used to track the cognitive effects of depression.

ACKNOWLEDGEMENTS

This work was funded by EU FP7 grant Help4Mood, Grant Agreement No. 248765. We thank our participants, the Help4Mood team, and the data collection team for the pilot RCT.

7. REFERENCES

[1] Alpert, M., Pouget, E. R., Silva, R. R. 2001. Reflections of depression in acoustic measures of the patient's speech. *J Affect Disord* 66(1), 59–69.

[2] American Psychiatric Association, 2000. *DSM-IV*.

[3] Bennabi, D., Vandel, P., Papaxanthis, C., Pozzo, T., Haffen, E. 2013. Psychomotor retardation in depression: a systematic review of diagnostic, pathophysiologic, and therapeutic implications. *BioMed research international* 2013, 158746.

[4] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10), 341–345.

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

[6] Cohen, A. S., Docherty, N. M. July 2004. Affective reactivity of speech and emotional experience in patients with schizophrenia. *Schizophrenia research* 69(1), 7–14.

[7] Cummins, N., Epps, J., Breakspear, M., Goecke, R. 2011. An Investigation of Depressed Speech Detection : Features and Normalization. *Interspeech* 6–9.

[8] El Ayadi, M., Kamel, M. S., Karray, F. Mar. 2011. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition* 44(3), 572–587.

[9] Ellgring, H., Scherer, K. R. June 1996. Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior* 20(2), 83–110.

[10] Faurholt-Jepsen, M., Vinberg, M., Christensen, E. M., Frost, M., Bardram, J., Kessing, L. V. Jan. 2013. Daily electronic self-monitoring of subjective and objective symptoms in bipolar disorder–the MONARCA trial protocol (MONitoring, treAtment and pRediCtion of bipolAr disorder episodes): a randomised controlled single-blind trial. *BMJ open* 3(7).

[11] Ferrari, A. J., Charlson, F. J., Norman, R. E., Flaxman, A. D., Patten, S. B., Vos, T., Whiteford, H. A. Jan. 2013. The epidemiological modelling of major depressive disorder: application for the global burden of disease study 2010. *PloS one* 8(7), e69637–e69637.

[12] Hardy, P. Feb. 1984. Speech pause time and the retardation rating scale for depression (ERD) Towards a reciprocal validation. *Journal of Affective Disorders* 6(1), 123–127.

[13] Henry, J., Crawford, J. R. 2005. A meta-analytic review of verbal fluency deficits in depression. *Journal of Clinical and Experimental Neuropsychology* 27(1), 78–101.

[14] Hoffmann, G. M., Gonze, J. C., Mendlewicz, J. May 1985. Speech pause time as a method for the evaluation of psychomotor retardation in depressive illness. *The British Journal of Psychiatry* 146(5), 535–538.

[15] Hothorn, T., Hornik, K., van de Wiel, M., Zeileis, A. 2008. Implementing a class of permutation tests: The coin package. *Journal of Statistical Software* 28(8), 1–23.

[16] Moullec, G., Maïano, C., Morin, A. J. S., Monthuy-Blanc, J., Rosello, L., Ninot, G. 2010. A very short visual analog form of the Center for Epidemiologic Studies Depression Scale (CES-D) for the idiographic measurement of depression. *Journal of Affective Disorders* 128(3), 220–234.

[17] Mundt, J. C., Vogel, A. P., Feltner, D. E., Lenderking, W. R. Apr. 2012. Vocal Acoustic Biomarkers of Depression Severity and Treatment Response. *Biological psychiatry* 72(7), 587–580.

[18] Nilsonne, A., Sundberg, J., Ternström, S., Askenfelt, A. Feb. 1988. Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression. *The Journal of the Acoustical Society of America* 83(2), 716–28.

[19] Polzehl, T., Schmitt, A., Metze, F., Wagner, M. Nov. 2011. Anger recognition in speech using acoustic and linguistic cues. *Speech Communication* 53(9-10), 1198–1209.

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

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

[22] Stassen, H. H., Bomben, G., Gunther, E. 1991. Speech characteristics in depression. *Psychopathology* 24(2), 88–105.

[23] Szabadi, E., Bradshaw, C. M., Besson, J. A. Dec. 1976. Elongation of pause-time in speech: a simple, objective measure of motor retardation in depression. *The British Journal of Psychiatry* 129(6), 592–597.

[24] Wolters, M. K., Martînez-Miranda, J., Estevez, S., Hastie, H. F., Matheson, C. 3 2013. Managing data in help4mood. *EAI Endorsed Transactions on Ambient Systems* 13(01-06).

[25] Yang, Y., Fairbairn, C., Cohn, J. F. 2013. Detecting Depression Severity from Vocal Prosody. *IEEE Transactions on Affective Computing* 4(2), 142–150.