Investigating Prosodic Accommodation in Clinical Interviews with Depressed Patients

Brian Vaughan[1](✉), Carolina De Pasquale[1], Lorna Wilson[2], Charlie Cullen[3], and Brian Lawlor[4]

[1] Dublin Institute of Technology, Dublin, Ireland
brian.vaughan@dit.ie
[2] St. James's University Hospital Dublin, Dublin, Ireland
[3] University of the West of Scotland, Hamilton, Scotland
[4] Trinity College Dublin, Dublin, Ireland

Abstract. Six in-depth clinical interviews, involving six elderly female patients (aged 60+) and one female psychiatrist, were recorded and analysed for a number of prosodic accommodation variables. Our analysis focused on pitch, speaking time, and vowel-space ratio. Findings indicate that there is a dynamic manifestation of prosodic accommodation over the course of the interactions. There is clear adaptation on the part of the psychiatrist, even going so far as to have a reduced vowel-space ratio, mirroring a reduced vowel-space ratio in the depressed patients. Previous research has found a reduced vowel-space ratio to be associated with psychological distress; however, we suggest that it indicates a high level of adaptation on the part of the psychiatrist and needs to be considered when analysing psychiatric clinical interactions.

Keywords: Speech analysis · Clinical interviews · Depression Prosody · Accommodation · Interaction · Vowel-space

1 Introduction

Clinical depression is a common illness with a negative impact on several aspects of a patient's life, accounting for 4.3% of the global burden of disease, and is among the single largest contributor to global disability [1]. In Europe, depression has a 9% prevalence among men and a 17% prevalence among women; the economic costs of depression amounts to 136.3 billion Euro in the European Economic Area [2]. There are some objectively measurable markers for depression [3], but none that can be unobtrusively and readily measured during diagnosis, which at this time depends on subjective assessments by expert practitioners. These practitioners have to rely on patients' self-reported symptoms and have to perform clinical interviews to assess the impact and prevalence of symptoms on the individual [4]. In an effort to find an easily measured objective marker for depression, recent studies have explored the impact that it has on a number

P. Cipresso et al. (Eds.): MindCare 2018, LNICST 253, pp. 150–159, 2018.
https://doi.org/10.1007/978-3-030-01093-5_19

of prosodic variables, which are affected by a variety of symptoms that occur in depressed individuals, such as psychomotor retardation, muscle tension and cognitive impairment [5]; indeed, changes to the prosodic behaviour of individuals with mental health issues are well known (for a review see [5]). These studies have found that depression induces measurable, manifest changes in a person's speech that can be indicative of depressive severity or even suicidal ideation. Some of the symptoms required for a diagnosis of depression can lead to a variety of behaviours that impact interactions. Individuals affected by a loss of motivation can experience social withdrawal, excessive negative thoughts, and feelings of guilt, which can lead to socio-communicative disruptive behaviours. During an interaction, participants will subconsciously adapt their communicative behaviour to that of their interlocutor, a phenomenon that is described using many terms, particularly prosodic adaptation, accommodation, synchrony, mimicry, convergence, alignment and entrainment [6]. Prosodic accommodation occurs between speakers when changes in their prosodic parameters move in synchronous alignment or when they converge towards a common point [7], and is an important factor in social interactions, as it aids comprehension and understanding between participants [8].

2 Speech and Depression

Various researchers have found that depression results in measurable changes in prosody (tone, intensity and rhythm of the voice); these changes have been found to be somewhat indicative of the severity of depression and have been used to differentiate between depressed and non-depressed patients. Moore et al. found that pitch, energy and speech rate feature statistics could be used to differentiate between patients with depression from those without [9]. Likewise, Cannizzaro et al. found speaking rate and pitch variability to be strongly negatively correlated with the severity of major depression [10]. Ozdas et al. observed that jitter (short-term perturbations of pitch) are different in patients at imminent risk of suicide[1] as compared to those in a control group of non-depressed patients. Moreover, they developed a Machine Learning (ML) classifier, using these findings, which performed well in discriminating between suicidal speech and speech from the control group [11]. Similarly, France et al. were able to discriminate between the speech of depressed patients and the speech of suicidal patients using prosodic analysis [12].

While interpersonal effects have not been investigated to the same extent, promising results have been obtained in some studies. Yang et al. observed that, as the severity of a patients depression changed, so too did the prosodic parameters of the trained clinical interviewers interviewing the patient. They suggest that the observed effect is not simply caused by behavioural mimicry, but rather a reflection of how the "pitch", measured by the fundamental frequency (f_0), is influenced by the intentions and goals of the speaker (in this case, to express

[1] These patients comprised of people who had recorded suicide notes and people who had subsequently made potentially lethal suicide attempts.

sympathy). Scherer et al. found that vowel-space was reduced in depression, post-traumatic stress disorder (PTSD), and suicidality. Vowel-space is defined as the frequency space covered by the first and second formants of the vowels /i/, /a/, and /u/ as these are at the extremes of a triangular shaped frequency space [13]. Scherer et al. [14] also found that the interviewers' acoustic features were strongly correlated with the depression severity of the interaction partner: interviewers' f_0 mean and variability were correlated to the partner's depression severity, and interviewers displayed a "breathy" voice quality (associated with a more empathetic demeanour, and with sad behaviours [14]). The findings of both Yang and Scherer, as discussed above, suggest that an interviewer's behaviour might also be an indicator of the depression severity of the interviewee, and thus merits further investigation.

2.1 Prosodic Accommodation

While the majority of research on prosodic accommodation describes it as a mono-tonic communicative property, recent work has demonstrated that prosodic accommodation is a dynamic phenomenon that increases and decreases over the course of an interaction [8,15]. Moreover, the dynamic manifestation of prosodic accommodation is directly related to the perceived naturalness of conversational flow, mutual understanding and affinity, and the overall level of engagement between speakers. Studies on the effects of depression on speech and prosodic characteristics tend to focus on the individual speech parameters of the depressed patient and their interlocutor, and neglect the socio-communicative aspect of prosodic accommodation. Previous work has shown this to be a dynamic, socio-communicative process that is indicative of the level of engagement and communication between interlocutors, and demonstrated that higher levels of prosodic accommodation were associated with a higher level of engagement and greater affinity between interlocutors [8]. Prosodic accommodation was found to be positively correlated with the level of communication and global coordination between interlocutors. Previous research also found that interlocutors accommodate across a range of prosodic parameters: intensity [16], speech rate [17,18], and pitch [19]. Collins [20] and Gregory et al. [21] observed global pitch to be indicative of a level of accommodation (in terms of mean f_0), using unconstrained conversations and interviews of English. Accommodation in average vocal intensity and intensity range has been observed both at the global and local levels of task-based and unconstrained dialogues of English and Swedish [15,22]. Accommodation is important in decreasing communicative misunderstandings, and facilitating faster goal attainment [23–25], while also increasing the level of rapport and overall social success of an interaction [26–28].

3 Methodology

In order to investigate prosodic accommodation in psychiatrist/patient interactions, noise-free, channel separated audio recordings were needed. Six psychiatrist/patient interviews were recorded over the course of a few weeks, at an outpatient clinic for elderly sufferers of depression at St. James's Hospital Dublin; patients were six elderly female individuals (aged 60+). Ethics approval was sought and granted from the hospital. All participant data was anonymised: participants were assigned coded names (participant 1, 2, 3 etc), and all medical records, including details on the depressive disorders, remained with the psychiatrist and hospital. Only basic demographic data was collected alongside the speech recordings (duration of the disorder, number of major episodes, age, and gender). All participants were broadly classed by the psychiatrists as mild to severely depressed, with no comorbid cognitive impairment (scores >24 on the Mini-Mental State Examination); HAM-D scores were not obtained, this is an aspect that future work will address, by ensuring a more granular measure of depression is captured and used as part of the analysis. The average length of the recordings was 20 min. In order to preserve the normal psychiatrist/patient interview environment, it was important to ensure that any recording set-up was as non-intrusive as possible. To this end, a small, portable audio recorder, the ZOOM H4N, was used, in conjunction with discrete clip-on lapel mics. No constraints were placed on the topics of conversation; the clinical interviews were part of the normal course of treatment for all participants and were not in addition to ongoing treatment.

Due to the nature of the recordings in the clinical setting, a lot of cross talk was present between channels: the setting and the physical space in which the recording took place did not allow for the speakers to be sufficiently acoustically isolated from each other; therefore, each microphone captured both speakers, albeit at different amplitudes. This made it necessary to prepare the audio for analysis before feature extraction, so that each speaker was clearly labelled and could be separated into different audio channels, and overlapping speech and extraneous noise removed. Each recording was annotated using Textgrids in PRAAT, and where then exported as separated audio files. This ensured that we had noise-free, channel separated audio, ready to be analysed.

4 Analysis

Once the files were separated, acoustic features and prosodic accommodation measures were extracted and computed using Vocavio Matlab software [29][2], and the COVAREP Matlab toolbox [30]. This is a set of open source Matlab scripts for extracting and computing a number of acoustic features. f_0 (pitch) was the main focus of our analysis; as per [7,8], it has long been observed that conversation partners exhibit similar pitch and intonation contours, and f_0 contours are the basis of the AUC calculations (discussed below); moreover, other frequency based components, f_1 and f_2 measurements, are the basis of the vowel-space calculations, which are also discussed below. Intensity based measures and their relation to frequency based measures are currently being explored.

[2] http://vocavio.com/.

A Time Aligned Moving Average (TAMA) method [31] was used to obtain a smoother prosodic contour for analysis. In the TAMA method, moving overlapping windows, large enough to extract useful chunks of speech, are used to extract prosodic parameters. Average values for a number of prosodic parameters are calculated for each window, then the window is moved so that the new time window overlaps by a certain degree with the old window (see Fig. 1): the result is a smoother contour, with no significant loss of accuracy in the capture of prosodic dynamics. For our analysis, we used a window length of ten seconds, with a five second overlap. The TAMA method solves the problem of measuring accommodation in conversations, as speakers do not accommodate immediately due the reactive temporal nature of conversation, and their inherent turn-based structure.

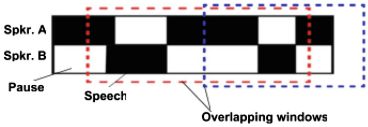

Fig. 1. Time Aligned Moving Window. The diagram illustrates the TAMA method, showing how the moving windows operate.

From each average obtained with the TAMA method, a correlation window of thirty seconds was used to calculate the correlation of prosodic measures in the dyad. The level of accommodation in dyads are measured with the Spearman's correlation coefficient $\rho\epsilon[-1,1]$. Large $\rho_{xy} \gg 0$ can be considered indicators of high levels of accommodation, while small $\rho_{xy} \ll 0$ indicate a low level of accommodation [32]. Our TAMA settings, of a window length of ten seconds, with a five second overlap, resulted in six values per 30 second correlation window. As per [32] a normalised area-under-curve (AUC) calculation is used to make an approximate comparison between the two speakers regarding the similarity of their pitch contours over the course of an interaction; this is taken to be an indication of the balance of effort of each speaker in adapting their prosodic parameters such that a larger AUC indicates stronger changes in pitch by the speaker. The normalised values were obtained by dividing the individual AUC calculations with the summed AUC of both speakers, and then multiplied by 100; this gives individual percentage values, enabling differences to be expressed as a percentage of overall effort:

$$\mathrm{NORM}^{A}_{auc} = \frac{auc^{A}}{auc^{A} + auc^{B}} * 100 \tag{1}$$

where aucA is the AUC for speaker A and aucB is the AUC for speaker B. The combined $\mathrm{NORM}^{A}_{auc} + \mathrm{NORM}^{B}_{auc}$ will always be equal to 100 percent. Formant analysis and vowel-space calculations were carried out using the COVAREP Matlab toolbox [30], as per [13]. Speaking time was calculated for each speaker using an intensity based Voice Activity Detector (VAD), with the intensity threshold set to a level that ensured all speech was accurately extracted. The overall speaking time for each speaker was calculated as a percentage of total speaking time for each conversation.

5 Results

5.1 Pitch

All the conversations are positively correlated. There are no particularly strong values (close to +1), and the overall accommodation level across all conversations is weak (see Table 1 for overall accommodation values). More pertinent, however, is the fact that the dynamic nature of prosodic accommodation means that there will be moments of high and low accommodation throughout, so a single accommodation score will not capture this dynamic aspect, as discussed in Sect. 2.1). This dynamic aspect is evident from the graphical representations of the accommodation of each interaction, which show a huge variation in pitch accommodation level. Section 2.1 shows a graphical representation from one conversation that is typical of all the conversations: moments of high and low pitch accommodation throughout and an overall high degree of variability (Fig. 2).

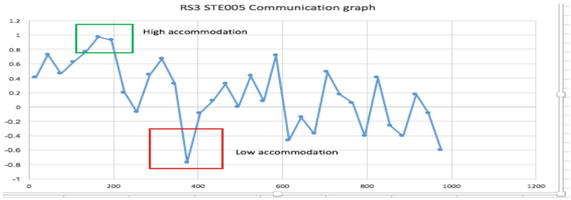

Fig. 2. Accommodation Graph. The graph shows the accommodation dynamics of the dyad: clear moments of high and low accommodation can be observed (highlighted), showing that accommodation during the clinical interactions was dynamic.

5.2 AUC and Speaking Time

The area-under-the-curve (AUC) calculations (see Sect. 4), show that in two of the six conversations, the psychiatrist made most of the effort to adapt to the patient. In two of the six conversations, the patient was making the most effort to accommodate. In two conversations, there is a fairly even balance of effort between the psychiatrist and patient. In all conversations the patient was doing the majority of the talking. Table 1 shows the overall accommodation scores and the speaking time and AUC value for each of the six conversations. The reduced speaking time across all interactions of the psychiatrist can potentially be explained by the nature of the clinical interaction, where the psychiatrist asks questions that are then answered at length by the patient (Table 2).

Table 1. Table showing the overall accommodation for each interaction, as well as speaking time, and AUC values for the patient and psychiatrist in each interaction.

Team accom.		Speaking time			AUC Values	
		Overall	Patient	Psychiatrist	Patient	Psychiatrist
1	0.106259	67%	49%	18%	35%	65%
2	0.199743	66%	42%	25%	70%	30%
3	0.184069	70%	53%	16%	47%	53%
4	0.305988	71%	56%	15%	52%	48%
5	0.108911	61%	35%	26%	44%	56%
6	0.0716094	69%	53%	16%	58%	42%

Table 2. Table of the vowel-space ratios for patients and psychiatrist. No vowel-space reduction would be a 1:1 ratio. Values less that this indicate a reduction in vowel-space.

Vowel space ratio values		
	Patient	Psychiatrist
1	0.607	0.612
2	0.496	0.673
3	0.451	0.409
4	0.601	0.549
5	0.637	0.662
6	0.733	0.618

5.3 Vowel Space Ratio

In agreement with [13], observed vowel space ratio was reduced in the patients' speech: this preliminary result supports their findings of reduced vowel-space in depressed patients. However, analysis also showed significant reduction in the vowel space ratio of the psychiatrist involved in the interactions. While Scherer et al. did find some vowel space ratio reduction in non-depressed participants [13], what is potentially significant about our results is the consistency across all interactions. Vowel-space ratio reduction on the part of the psychiatrist has not been previously investigated, and it represents a potential indication of adaptive behaviour in response to that of the patient. Due to ethical requirements, it is extremely unlikely that a depressed psychiatrist would practice and care for depressed patients, therefore it is more likely that the vowel-space reduction in the psychiatrist's speech indicates a global measure of prosodic accommodation.

6 Discussion and Future Work

Our investigation shows that prosodic accommodation occurs in clinical interactions, and is a dynamic phenomenon, as per previous findings (see Sect. 2.1).

Using a combination of speaking time and normalised AUC values, we show that, in two of the six conversations, the psychiatrist is making the most effort to accommodate; in two of the conversations the patient makes the most effort, and in two conversations there is a balance of effort. There is no clear reason for the variance in effort across conversations; it is possible that a more granular measure of depression could account for these variances, with increased depression severity being linked to a reduced AUC value. This is an aspect of the results that will be further investigated.

Vowel-space ratio measurements are more consistent, and the results suggest that vowel-space ratio may be used as a measure of global interpersonal adaptation in the context of clinical interaction; vowel-space ratio has not been investigated as a measure of prosodic adaptation, and this result will be explored further. We must also consider possible age related effects across all parameters, but especially vowel space ratio measures, as there may be age related aspects that can cause a reduced vowel-space ratio. However, given the age difference between the patients and the psychiatrist (25+years), age related effects were not likely a factor in her reduced vowel space ratio, even if they were for the patient group. Moreover, even if patient vowel-space ratio was reduced due to age related effects, the reduced vowel-space ratio of the psychiatrist can still be indicative of strong adaptation on her part. Further work will be conducted with a more granular measurement of depression severity (using the Hamilton depression rating scale (HAM-D) [33]), as well as an age and gender matched control group. Future work will also focus on further examining vowel-space ratio reduction across a larger data-set of psychiatrist/patient interactions, as well as exploring the relationship between speaking time and AUC calculations, and their relation to aspects of the conversation, such as topic, and depression severity.

References

1. World Health Organization: Depression and other common mental disorders: global health estimates. Technical report, World Health Organization, Geneva (2017)
2. Smit, F., Shields, L., Petrea, I.: Preventing depression in the WHO European region. Technical report, World Health Organization (2016)
3. Strawbridge, R., Young, A.H., Cleare, A.J.: Biomarkers for depression: recent insights, current challenges and future prospects. Neuropsychiatr. Dis. Treat. **13**, 1245–1262 (2017)
4. Asgari, M., Shafran, I., Sheeber, L.B.: Inferring clinical depression from speech and spoken utterances. In: 2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–5. IEEE, September 2014
5. Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T.F.: A review of depression and suicide risk assessment using speech analysis. Speech Commun. **71**, 10–49 (2015)
6. De Looze, C., Oertel, C., Rauzy, S., Campbell, N.: Measuring dynamics of mimicry. In: ICPhS, vol. 1, pp. 1294–1297, August 2011

7. De Looze, C., Rauzy, S.: Measuring speakers' similarity in speech by means of prosodic cues: methods and potential. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, pp. 1393–1396 (2011)
8. De Looze, C., Scherer, S., Vaughan, B., Campbell, N.: Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. Speech Commun. **58**, 11–34 (2014)
9. Moore II, E., Clements, M., Peifer, J., Weisser, L.: Analysis of prosodic variation in speech for clinical depression. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439), vol. 3, pp. 2925–2928 (2003)
10. Cannizzaro, M.S., Harel, B., Reilly, N., Chappell, P., Snyder, P.J.: Voice acoustical measurement of the severity of major depression. Brain Cogn. **56**, 30–35 (2004)
11. Ozdas, A., Shiavi, R., Silverman, S., Silverman, M., Wilkes, D.: Analysis of fundamental frequency for near term suicidal risk assessment. In: SMC 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics. 'Cybernetics Evolving to Systems, Humans, Organizations, and Their Complex Interactions' (Cat. No. 00CH37166), vol. 3, pp. 1853–1858. IEEE (2000)
12. France, D.J., Shiavi, R.G., Silverman, S., Silverman, M., Wilkes, D.M.: Acoustical properties of speech as indicators of depression and suicidal risk. IEEE Trans. Biomed. Eng. **47**, 829–837 (2000)
13. Scherer, S., Morency, L.-P., Gratch, J., Pestian, J.: Reduced vowel space is a robust indicator of psychological distress: a cross-corpus analysis. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4789–4793. IEEE, April 2015
14. Scherer, S., Hammal, Z., Yang, Y., Morency, L.-P., Cohn, J.F.: Dyadic behavior analysis in depression severity assessment interviews. In: Proceedings of the 16th International Conference on Multimodal Interaction - ICMI 2014, pp. 112–119 (2014)
15. Levitan, R., Hirschberg, J.: Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In: Proceedings of Interspeech 2011, pp. 3081–3084. ISCA (2011)
16. Coulston, R., Oviatt, S., Darves, C.: Amplitude convergence in children's conversational speech with animated personas. In: 7th International Conference on Spoken Language Processing, ICSLP 2002 - INTERSPEECH 2002 (2002)
17. Kousidis, S., Dorran, D., McDonnell, C., Coyle, E.: Times series analysis of acoustic feature convergence in human dialogues. In: Proceedings of Interspeech (2008)
18. Edlund, J., Heldner, M., Hirschberg, J.: Pause and gap length in face-to-face interaction. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, pp. 2779–2782 (2009)
19. Babel, M., Bulatov, D.: The role of fundamental frequency in phonetic accommodation. Lang. Speech **55**, 231–248 (2011)
20. Collins, B.: Convergence of fundamental frequencies in conversation: if it happens, does it matter? In: Proceedings of ICSLP, vol. 98, (1998)
21. Gregory, S.W., Webster, S.: A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. J. Pers. Soc. Psychol. **70**(6), 1231–1240 (1996)
22. Heldner, M., Edlund, J.: Pauses, gaps and overlaps in conversations. J. Phon. **38**, 555–568 (2010)
23. Parrill, F., Kimbara, I.: Seeing and hearing double: the influence of mimicry in speech and gesture on observers. J. Nonverbal Behav. **30**, 157–166 (2006)

24. Pickering, M.J., Garrod, S.: Alignment as the basis for successful communication. Res. Lang. Comput. **4**, 203–228 (2006)
25. Boylan, P.: Accommodation Theory Revisited Again. Lingua e società, pp. 287–305 (2009)
26. Tickle-Degnen, L., Rosenthal, R.: The nature of rapport and its nonverbal correlates. Psychol. Inq. **1**(4), 285–293 (1990)
27. Shepard, C., Giles, H., Le Poire, B.: Communication accommodation theory. In: Robinson, W., Giles, H. (eds.) The New Handbook of Language and Social Psychology, pp. 33–56. Wiley, New York (2001)
28. Miles, L.K., Nind, L.K., Macrae, C.N.: The rhythm of rapport: interpersonal synchrony and social perception. J. Exp. Soc. Psychol. **45**(3), 585–589 (2009)
29. Vocavio (2017)
30. Degottex, G., Kane, J., Drugman, T., Raitio, T., Scherer, S.: COVAREP - a collaborative voice analysis repository for speech technologies. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 960–964. IEEE, May 2014
31. Kousidis, S., Dorran, D., McDonell, C., Coyle, E.: Time series analysis of acoustic feature convergence in human dialogues. In: Specom 2009, St. Petersburg, Russian Federation, pp. 1–6 (2009)
32. De Looze, C., Vaughan, B., Kelly, F., Kay, A.: Providing objective metrics of team communication skills via interpersonal coordination mechanisms. In: INTERSPEECH, Dresden, Germany, pp. 1–5 (2015)
33. Hamilton, M.: Development of a rating scale for primary depressive illness. Br. J. Soc. Clin. Psychol. **6**, 278–96 (1967)