

# Articulation rate in psychotherapeutic dialogues for depression: patients and therapists

Laurence White<sup>1</sup>, Hannah Grimes<sup>1</sup>

<sup>1</sup>School of Education, Communication and Language Sciences, Newcastle University, UK

laurence.white@newcastle.ac.uk

## Abstract

Prosodic features anecdotally associated with the speech of people with clinical depression include slower rate, lower pitch range and reduced loudness, but there is a significant degree of contradiction in the literature regarding depressed prosody. This complex picture reflects the heterogeneity of depression aetiology, symptomatology and prognosis. It is also likely to be influenced by elicitation methods, in particular, whether natural dialogue contexts are employed and whether the interlocutor's prosody is also considered.

We analysed 40 patient-therapist dialogues from the first and last of 29 weekly sessions of a behavioural therapy for refractory depression, sampling early and late in both sessions. Across all dialogues, we found that therapists spoke faster than patients, as expected, but in female-female therapist-patient dialogues (the majority of our sample), patients' articulation rate increased substantially over the first session. Moreover, and contrary to expectations, there was a positive correlation between articulation rate and assessed depression severity (PHQ-9 scale) in the final therapy session, also evident in therapists' speech for female-female dialogues. We suggest that this may reflect features of anxiety in speakers with ongoing depression and possibly also personality characteristics. We also consider evidence for prosodic convergence between patients and therapists.

**Index Terms:** clinical depression, articulation rate, dialogue, prosodic convergence

## 1. Introduction

Listeners use prosody to interpret speakers' emotional states [1]. For example, happiness is vocally expressed and recognised through a combination of higher pitch or pitch range, faster articulation rate and increased loudness [2]. By contrast, relatively low pitch and pitch range, slower speech and reduced loudness may be interpreted as indicating sadness [3].

Transient emotional expression is mediated by longer-term psychological states and personality traits [4]. In clinical mood disorders, neurophysiological disturbances, affecting somatic and autonomic nervous systems, can affect vocal production, e.g., through their effects on respiratory function and muscular tension [5]. Clinical depression is associated with prosodic markers that, to a degree, mirror more transient states of sadness, i.e., slower rate, lower pitch/pitch range, and reduced loudness [5], [6]. Moreover, listeners can use these features to identify speakers with clinical depression [7], although the severity of depression is not consistently associated with the strength of such features. Thus, for example, some studies find lower pitch and/or pitch range to be indicative of more severe depression [8], [9], but others report a positive correlation between pitch features and depression severity [10].

Multiple interacting factors contribute to heterogeneity in the vocal expression of depression. Firstly, the statistical relationship between depression severity and prosodic features is likely to be non-linear, at least for relatively extreme values. Secondly, depression itself is heterogeneous, and interacts with personality factors and use of antidepressant medication. Comorbid anxiety states, for example, tend to increase pitch and pitch range, contrary to the typical depressive pattern [11]. Thirdly, anatomical, physiological, cultural and linguistic factors all impact on the vocal expression of emotional states. Some studies find female/male differences in associations between depression and pitch features [12], although other find more consistency between sexes in such relationships [9].

Variable speech elicitation are another complicating factor. Some studies find that spontaneous samples are more likely than read speech to be correctly identified as depressed [13], whilst others find greater reliability of depressive features in read speech and well-practiced recitation [14].

Going beyond the elicitation issue, the nature of speakers' relationships and the prosodic dynamics of dialogue are critical factors to consider. Here we use "convergence" specifically for the phenomenon wherein dialogue partners become more similar in linguistic forms/structures and in phonetic features, including prosody. Conversational dynamics (also embracing *accommodation*, *entrainment*, *coordination*, etc. – [15] for discussion of terminology) have long been investigated in the literature on social aspects of language. Thus, Communication Accommodation Theory [16] proposes that interlocutors use convergence to reduce social distance (or create distance via divergence). Others argue that convergence phenomena are automatic, unconscious consequences of our cognitive mechanisms for comprehension and production of linguistic interactions [17].

Regardless of the level of conscious speaker control over dialogue convergence, it is associated with perceptions of positive interactional outcomes [18]–[20]. Thus, the impact of personality factors on dialogue dynamics is clearly highly relevant to the psychotherapeutic context. In one of the few relevant studies, speakers with higher behavioural avoidance were found to show lower levels of convergence, whilst both higher openness and higher neuroticism are associated with greater negatively correlated with convergence [21].

Recordings of depressed speakers in clinical interactions have sometimes been used, for example, to demonstrate the association between slower articulation rate and higher levels of depression [22]. Lacking thus far, however, is consideration of the full dynamic context of interactions, considering both the depressed speaker and their interlocutor. More work is required on how prosodic features may change over time, both in individual dialogues with depressed speakers and over the longer-term development of a therapeutic relationship.

### 1.1. Articulation rate in psychotherapeutic dialogues

The present exploratory study focuses on articulation rate, as part of a broader examination of prosody in dialogues between depressed speakers and psychotherapists [Grimes et al., in prep]. Lower speech rate in people with depression is partly attributable to their production of longer and more variable pauses [14], [23]. However, even when considering articulation rate (syllables per second excluding pauses) rather than raw speech rate, depressed speakers are consistently found to talk more slowly [9]. Indeed, rate measures are proposed to be the “most promising prosodic features for recognising depression” [5, p. 23].

The purpose of this study is to examine the evidence for reduced articulation rate in speakers with depression in a dialogue context, specifically, during psychotherapy sessions to treat their depression. We measured articulation rate of patients being treated for refractory depression, i.e., clinical depression maintained over an extended period and resistant to earlier treatment. Dialogue samples were taken from early and late in the first and last of 29 therapy sessions over six months. We also measured the articulation rate of the psychotherapists in dialogue turns adjacent to those of the patients. Thus we looked for evidence of articulation rate changes, by patients and therapists, over the short-term course of therapeutic dialogues and over the longer-term course of treatment. We also used a clinical measure, PHQ-9 [24], to test whether articulation rate changed with levels of depression.

## 2. Method

### 2.1. Participants

Patients were 40 volunteers (26 female) in the treatment group of a randomized control trial testing the effectiveness of Radically Open Dialectical Behaviour Therapy (RO DBT) [25]. All patients were 18 years of age or older, had been diagnosed with refractory depression, had a Hamilton Rating Scale for Depression (HRSD) [26] score of at least 15, and did not have diagnoses of bipolar disorder, psychosis or substance dependence. To take part in the trial, patients must have been taking antidepressant medication for at least six weeks prior to the start of therapy. Most patients had been diagnosed with comorbid personality disorder, typically characterised by overcontrol, with features including behavioural rigidity, restrained emotional expression and limited social interaction. Participants were recruited from UK National Health Service (NHS) secondary care centres in Dorset, Hampshire, and North Wales. Their English language proficiency had to be sufficient to take part in dialogue-based treatment in English.

There were nine therapists (eight female), all dialectical behaviour therapists with a minimum of three years’ clinical experience (minimum age 32 years). Dialectical behaviour therapy (DBT) is a modified form of cognitive behavioural therapy (CBT). Each therapist worked with between two and five patients, and all patient-therapist pairings remained consistent between the first and last therapy sessions.

### 2.2. Design and Procedure

#### 2.2.1. Selection of speech samples

Psychotherapy sessions were originally recorded by video camera for the analysis of the therapeutic benefit of RO DBT treatment [25]. Ethical approval to analyse the speech samples for the

current project was obtained as an extension to the original randomized control trial NHS ethics plan (National Research Ethics Service Committee South Central – Southampton A; 11/SC/0146), and also subject to scrutiny by university school/faculty ethics committees (Plymouth and Newcastle).

Patient-therapist pairs were randomly extracted from the larger study, with the constraints that patients must have completed the six-month course of treatment and that the same therapist was present in the first and last sessions for a specific patient. The audio from these two therapy sessions for each patient-therapist pair was extracted from the video and automatically cut into four-minute sections. The first patient-therapist *exchange* (see below) from the first and last four-minute clips of the two sessions for each patient-therapist pair were then identified by manual inspection of the audio.

An *exchange* was defined as sequence of patient and therapist turns, in either order, where a *turn* was defined as one person speaking for at least three seconds, following a silent interval of at least 500ms (to avoid conversational overlap). If a turn extended for longer than five seconds, it was cut at the first natural break after five seconds. Therapists spoke first in 112 of the 160 exchanges identified using the above protocol.

#### 2.2.2. Articulation rate measurements

Each identified exchange was transcribed by the second author, who then labelled the speech into vocalic and consonantal intervals using Praat [27] and following standard criteria [28]. Articulation rate (syllables per second) was calculated as the number of vocalic intervals (as a proxy for syllables), divided by the duration of the turn excluding pauses. Pauses were defined as periods of silence longer than 200ms that were not associated with articulatory gestures

#### 2.2.3. Depression measure

The PHQ-9 (Patient Health Questionnaire) [24] is a nine-item self-report questionnaire for assessment of depression, with scores range from 0 to 28, and higher scores indicating stronger depressive symptoms. Questionnaires were administered in both of the treatment sessions for which speech was analysed.

### 2.3. Statistical analysis

The effects of Role (patient/therapist), Session (first/last) and Sample (early/late within session) were analysed with mixed-effects regression models using the lme4 package in R [29]. Pairing (patient-therapist) was a random intercept. Fixed effects were assessed in model comparisons via likelihood ratio tests.

## 3. Results

### 3.1. Patient and therapist articulation rate

#### 3.1.1. All patient-therapist pairings

Mean articulation rates by Role, Session and Sample are shown in Figure 1. As expected, there was an effect of Role, with therapist articulation rate faster than patient rate,  $X^2(1) = 9.079$ ,  $p = .0026$ . There was no effect of Session,  $X^2(1) = 0.012$ ,  $p = .91$ , or Sample (early/late within session)  $X^2(1) = 0.39$ ,  $p = .53$ . There were no interactions between Role and Session,  $X^2(1) = 0.18$ ,  $p = .67$ , or Role and Sample,  $X^2(1) = 0.25$ ,  $p = .62$ . There was a marginal interaction between Session and Sample,  $X^2(1) = 0.18$ ,  $p = .089$ . Figure 1 indicates that patients tended to speak

more quickly towards the end of the first session than at the start, which is the most likely reason for this trend, explored in more depth in the analysis of the female-female dialogues. There was no three-way interaction between Role, Session and Sample,  $\chi^2(1) = 1.54, p = .21$ .

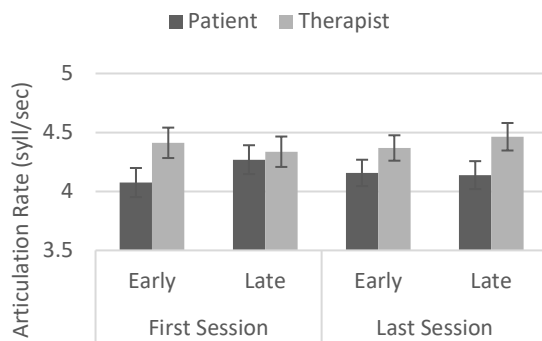


Figure 1: All dialogue pairs – patient and therapist articulation rate by session (first/last) and sample within session (early/late).

### 3.1.2. Female-female patient-therapist pairings

As part of the broader prosodic analysis [Grimes et al., in prep], we separately analysed the subset of dialogues that featured both a female patient and a female therapist (23 of the 40 dialogue pairs). This was in part to achieve greater comparability on pitch measures for the extended analysis, but also because sex differences have been reported in phonetic convergence and/or accommodation [30], [31].

Therapists in female-female dialogues (Figure 2) had faster articulation rate than patients, shown by a main effect of Role,  $\chi^2(1) = 19.34, p < .0001$ . There was no effect of Session,  $\chi^2(1) = 0.23, p = .63$ , or Sample,  $\chi^2(1) = 0.60, p = .44$ . There were no interactions between Role and Session,  $\chi^2(1) = 0.033, p = 0.86$ , Role and Sample,  $\chi^2(1) = 0.78, p = .38$ , or Session and Sample,  $\chi^2(1) = 1.33, p = .25$ . There was a three-way interaction between Role, Session, and Sample within session,  $\chi^2(1) = 4.90, p = .027$ . Figure 2 indicates that this is due to patient rate increasing substantially between early and late samples in the first therapy session,  $t(22) = -2.40, p = .025$  (two-tailed), a trend that was also evident, to a lesser degree, in the full sample (Figure 1).

### 3.2. Depression scores and articulation rate

Patients' mean PHQ-9 scores were 19.8 in the first session and 13.5 in the final session, indicating that many patients (although not all) had an alleviation of depressive symptoms over the six months of therapy,  $t(39) = 5.349, p < .001$ .

To test for the expected association between higher levels of depression and lower articulation rate, we looked separately at the first and last sessions (taking the mean rate within sessions). For the first therapy session, considering all 40 patient-therapist pairs, PHQ-9 scores were not correlated with articulation rate,  $r(38) = -0.13, p = .42$ . There was an unexpected *positive* correlation in the last session, however,  $r(38) = 0.36, p = .021$ . Similar patterns were found for female-female patient-therapist pairs, although the last session correlation was only marginally significant: first session,  $r(21) = -0.19, p = .40$ ; last session,  $r(21) = 0.38, p = .076$ .

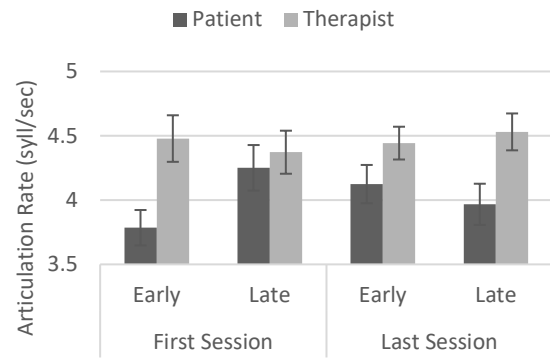


Figure 2: Female-female dialogue pairs – patient and therapist articulation rate by session (first/last) and sample within session (early/late).

Finally, we examined whether therapists' articulation rate reflected patients' level of depression, again taking mean rate separately for the first and last therapy sessions. Considering all patient-therapist pairs, there were no correlations between therapists' articulation rate and patients' PHQ-9 scores for either session: first,  $r(38) = -0.03, p = .87$ ; last session,  $r(38) = 0.21, p = .19$ . For female-female patient-therapist pairs, however, whilst there was no association between therapists' articulation rate and patients' PHQ-9 scores for the first session,  $r(21) = 0.15, p = .51$ , there was a *positive* correlation for the last session,  $r(21) = 0.50, p = .015$ .

Thus, patients with higher levels of depression spoke more quickly in the final session. Similarly, therapists in female-only dialogue pairs spoke more quickly in the final session to patients with higher levels of depression. Possible explanations for this unexpectedly positive association between articulation rate and depression level are considered below.

## 4. Discussion

We measured the articulation rates of people with refractory depression and their psychotherapists in the first and last psychotherapy sessions of a six-month course of experimental RO DBT treatment [25]. As expected, patients had lower articulation rate than psychotherapists, a difference that did not change over the course of therapy (as indicated by the lack of interaction between Role and Session). The only evidence of a significant change in articulation rate was in the first session, where a three-way interaction indicated that patients spoke more quickly towards the end of the session than at the beginning, a difference that was statistically reliable in female-female dialogues (Figure 2).

The finding of lower articulation rate in speakers with depression is expected and concurs with most of the foregoing literature regarding the prosodic characteristics of depressed speech [5]. The lack of change in articulation rate over the six-month course of treatment appears surprising, however, bearing in mind that, overall, patients do show evidence of relief in depressive symptoms between the start and end of therapy, as indicated by PHQ-9 scores. Moreover, there were no correlations between individual patients' PHQ-9 scores and their articulation rate in the first session of therapy.

The final, and unexpected, finding with regard to patients' speech was a *positive* correlation, in the last therapy session, between PHQ-9 depression scores and articulation rate: i.e.,

more depressed patients spoke more quickly. This trend was also mirrored in the articulation rate of therapists in female-female dialogues, who spoke more quickly to patients with higher PHQ-9 depression scores in the last therapy session.

A number of factors are likely to contribute to the general lack of change in articulation rate over the six months of therapy. Firstly, while the majority patients did experience an alleviation in symptoms of depression, that was not true of all of those involved therapy. Secondly, the depressed speakers in this study were specifically selected because of their longstanding depressive symptoms, which had previously proved resistant to treatment. Whilst PHQ-9 scores indicate some relief, that does not mean that most speakers were free of depressive symptoms at the end of treatment.

Thirdly, and perhaps most importantly, most of the depressed speakers had diagnoses of comorbid personality disorders, typically characterised by overcontrol, including features of social isolation and restrained emotional expression. This is a critical point for future studies of depressive prosody and its development over time: it may be that certain speakers tend to exhibit prosodic patterns – such as reduced articulation rate, increased pause length and greater pause frequency – that are characteristic of depressed speech, even in periods when they are not diagnosed as clinically depressed.

As the literature on conversational dynamics demonstrates, features of spoken interaction such as prosodic convergence can contribute to positive feelings about the outcome of interactions, and – conversely – positive attitudes towards conversational partners are likely to promote the emergence of convergence features [32], [33]. Whether convergence facilitates positive interpersonal feelings or is a consequence of them is beyond the scope of the present discussion, although the evidence suggests that the relationship is reciprocal. A hypothesis worthy of further investigation, however, is that some speakers who consistently fail to manifest convergence in prosody and other dialogue features – along with accommodation and other elements of interactional dynamics hypothesised to be prosocial – may tend to experience less rewarding interactions that over time, could in themselves be depressogenic. Indeed, RO DBT therapy – investigated in the original study for which the recordings used here were made – had the social nature of interaction as a key aspect of its therapeutic purpose: “RO DBT is the first treatment known to us to target deficits in social signalling as the main problem underlying overcontrolled emotional loneliness” [25, p. 210].

The above discussion addresses the overall lack of change in patients’ articulation rate over time, despite the lowering of depression scores. This might be taken as evidence of atypical prosodic features that are relatively resistant to general-purpose treatment for depression. This exploratory study was not specifically designed, however, to examine prosodic convergence features of therapeutic dialogues. More samples over the course of dialogues would be required, allied to specific statistical techniques to map point-by-point associations between patients’ and therapists’ speech. Moreover, the first and last therapy sessions may not be the most typical for a study of natural prosodic dynamics, and future work should sample from a wider range of sessions.

It is worthy of note, however, that patients in female-female dialogues do show an increase in rate over the course of the first therapy session. Although this is obviously in the direction of convergence with the therapists’ articulation rate, we would not argue that this is evidence of convergence *per se*. The most

parsimonious explanation is that articulation rate increases as speakers become more engaged in the interaction, in line with the observation that the manifestation of depressive prosody may decrease with greater task engagement [13].

Unexpectedly, we found a positive correlation, in the last session, between PHQ-9 depression scores and articulation rate. We offer only tentative interpretations at this stage. Firstly, patients with ongoing depression are perhaps more likely to feel high levels of anxiety concerning the end of treatment, and anxiety states can be associated with prosodic patterns that are rather different from those typical of depression [11]. Secondly, speeding up speech is an obvious strategy for increasing the amount of information transmitted when time is short, which could be more of an imperative for patients with higher ongoing levels of depression at the end of a course of treatment.

Whilst there was little evidence of change in therapists’ articulation rate either within sessions or over the course of treatment, we did find that therapists spoke more quickly to patients with higher levels of depression in the last session. As with the similar finding for patients, this may reflect a need to maximise the utility of the remaining time for those patients for whom further treatment might be desired.

## 5. Conclusion

There are many aspects of the prosody of therapeutic interactions that remain to be thoroughly investigated, but this exploratory study points to the importance of considering the speech patterns of both patients and therapists, and examining the dynamics of their development over shorter and longer timescales. Strong patient-therapist relationships are key predictor of positive outcomes in treatment for depression [34]. Given the established role of prosodic convergence in promoting and reflecting positive dialogue outcomes, we recommend closer investigation of therapists’ prosody and its impact on the therapeutic relationship.

## 6. Acknowledgements

We gratefully acknowledge the essential contributions of Ben Whalley of the University of Plymouth, Thomas Lynch of the University of Southampton, Roelie Hempel of Radically Open Ltd. and the additional co-authors of Lynch et al. (2020) [25]. We also thank Tess Saunders, James Manthorp and Rosie Lucas for assistance in speech analysis and data processing.

## 7. References

- [1] R. W. Frick, ‘Communicating emotion: The role of prosodic features.’, *Psychological Bulletin*, vol. 97, no. 3, pp. 412–419, 1985.
- [2] M. D. Pell, L. Monetta, S. Paulmann, and S. A. Kotz, ‘Recognizing emotions in a foreign language’, *Journal of Nonverbal Behavior*, vol. 33, no. 2, pp. 107–120, Jun. 2009, doi: 10.1007/s10919-008-0065-7.
- [3] C. Sobin and M. Alpert, ‘Emotion in speech: the acoustic attributes of fear, anger, sadness, and joy’, *J Psycholinguist Res*, vol. 28, no. 4, pp. 347–365, Jul. 1999, doi: 10.1023/A:1023237014909.
- [4] R. Plutchik, ‘The nature of emotions’, *American Scientist*, vol. 89, no. 4, pp. 344–350, 2001.
- [5] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, ‘A review of depression and suicide risk assessment using speech analysis’, *Speech*

- Communication*, vol. 71, pp. 10–49, Jul. 2015, doi: 10.1016/j.specom.2015.03.004.
- [6] B. Schuller and A. Batliner, *Computational paralinguistics: emotion, affect and personality in speech and language processing*. John Wiley & Sons, 2013.
- [7] Å. Nilsson and J. Sundberg, ‘Differences in Ability of Musicians and Nonmusicians to Judge Emotional State from the Fundamental Frequency of Voice Samples’, *Music Perception: An Interdisciplinary Journal*, vol. 2, no. 4, pp. 507–516, 1985, doi: 10.2307/40285316.
- [8] Z. Breznitz, ‘Verbal indicators of depression’, *The Journal of general psychology*, vol. 119, no. 4, pp. 351–363, 1992.
- [9] F. Hoenig, A. Batliner, E. Noth, S. Schnieder, and J. Krajewski, ‘Automatic modelling of depressed speech: relevant features and relevance of gender’, *Interspeech 2014*, pp. 1248–1252, 2014.
- [10] R. Horwitz, T. F. Quatieri, B. S. Helfer, B. Yu, J. R. Williamson, and J. Mundt, ‘On the relative importance of vocal source, system, and prosody in human depression’, in *2013 IEEE International Conference on Body Sensor Networks*, 2013, pp. 1–6.
- [11] P. N. Juslin and K. R. Scherer, ‘Vocal expression of affect’, in *The New Handbook of Methods in Nonverbal Behavior Research*, J. Harrigan, R. Rosenthal, and K. Scherer, Eds. Oxford: Oxford University Press, 2008. doi: 10.1093/acprof:oso/9780198529620.001.0001.
- [12] H. Ellgring and K. R. Scherer, ‘Vocal indicators of mood change in depression’, *J Nonverbal Behav*, vol. 20, no. 2, pp. 83–110, Jun. 1996, doi: 10.1007/BF02253071.
- [13] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, M. Breakspear, and G. Parker, ‘Detecting depression: A comparison between spontaneous and read speech’, in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, May 2013, pp. 7547–7551. doi: 10.1109/ICASSP.2013.6639130.
- [14] J. C. Mundt, A. P. Vogel, D. E. Feltner, and W. R. Lenderking, ‘Vocal acoustic biomarkers of depression severity and treatment response’, *Biol Psychiatry*, vol. 72, no. 7, pp. 580–587, Oct. 2012, doi: 10.1016/j.biopsych.2012.03.015.
- [15] M. M. Louwerse, R. Dale, E. G. Bard, and P. Jeuniaux, ‘Behavior matching in multimodal communication is synchronized’, *Cognitive Science*, vol. 36, no. 8, pp. 1404–1426, Nov. 2012, doi: 10.1111/j.1551-6709.2012.01269.x.
- [16] H. Giles and P. F. Powesland, *Speech Style and Social Evaluation*. Oxford, England: Academic Press, 1975.
- [17] S. Garrod and M. J. Pickering, ‘Why is conversation so easy?’, *Trends in Cognitive Sciences*, vol. 8, no. 1, pp. 8–11, Jan. 2004, doi: 10.1016/j.tics.2003.10.016.
- [18] C. M. Mitchell, K. E. Boyer, and J. C. Lester, ‘From strangers to partners: examining convergence within a longitudinal study of task-oriented dialogue’, in *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, 2012, pp. 94–98.
- [19] R. Levitan, A. Gravano, L. Willson, S. Benus, J. Hirschberg, and A. Nenkova, ‘Acoustic-prosodic entrainment and social behavior’, in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Stroudsburg, PA, USA, 2012, pp. 11–19.
- [20] T. L. Chartrand and J. A. Bargh, ‘The chameleon effect: the perception–behavior link and social interaction.’, *Journal of personality and social psychology*, vol. 76, no. 6, p. 893, 1999.
- [21] N. Lewandowski and M. Jilka, ‘Phonetic convergence, language talent, personality and attention’, *Front. Commun.*, vol. 4, p. 18, May 2019, doi: 10.3389/fcomm.2019.00018.
- [22] M. Cannizzaro, B. Harel, N. Reilly, P. Chappell, and P. J. Snyder, ‘Voice acoustical measurement of the severity of major depression’, *Brain and Cognition*, vol. 56, no. 1, pp. 30–35, Oct. 2004, doi: 10.1016/j.bandc.2004.05.003.
- [23] Y. Yang, C. Fairbairn, and J. F. Cohn, ‘Detecting depression severity from vocal prosody’, *IEEE Trans. Affective Comput.*, vol. 4, no. 2, pp. 142–150, Apr. 2013, doi: 10.1109/T-AFFC.2012.38.
- [24] K. Kroenke, R. L. Spitzer, and J. B. Williams, ‘The PHQ-9: validity of a brief depression severity measure’, *Journal of general internal medicine*, vol. 16, no. 9, pp. 606–613, 2001.
- [25] T. R. Lynch *et al.*, ‘Refractory depression – mechanisms and efficacy of radically open dialectical behaviour therapy (RefraMED): findings of a randomised trial on benefits and harms’, *British Journal of Psychiatry*, vol. 216, pp. 204–212, 2020.
- [26] J. B. Williams *et al.*, ‘The GRID-HAMD: standardization of the Hamilton depression rating scale’, *International Clinical Psychopharmacology*, vol. 23, no. 3, pp. 120–129, 2008.
- [27] P. Boersma and D. Weenink, *Praat: doing phonetics by computer [Computer program]*. 2021.
- [28] L. White and S. L. Mattys, ‘Calibrating rhythm: First language and second language studies’, *Journal of Phonetics*, vol. 35, no. 4, pp. 501–522, Oct. 2007, doi: 10.1016/j.wocn.2007.02.003.
- [29] R Core Team, *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2021. [Online]. Available: <https://www.R-project.org/>
- [30] J. S. Pardo, ‘On phonetic convergence during conversational interaction’, *The Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2382–2393, 2006.
- [31] L. L. Namy, L. C. Nygaard, and D. Sauerterg, ‘Gender Differences in Vocal Accommodation: The Role of Perception’, *Journal of Language and Social Psychology*, vol. 21, no. 4, pp. 422–432, Dec. 2002, doi: 10.1177/026192702237958.
- [32] A. Schweitzer and N. Lewandowski, ‘Convergence of articulation rate in spontaneous speech.’, in *INTERSPEECH*, 2013, pp. 525–529. Accessed: Dec. 05, 2016.
- [33] K. Schweitzer, M. Walsh, and A. Schweitzer, ‘To see or not to see: interlocutor visibility and likeability influence convergence in intonation.’, in *Interspeech*, 2017, pp. 919–923.
- [34] M. J. Lambert and D. E. Barley, ‘Research summary on the therapeutic relationship and psychotherapy outcome.’, *Psychotherapy: Theory, research, practice, training*, vol. 38, no. 4, p. 357, 2001.