Speech Emotion Perception in Depression

Guneesh Vats

Abstract

This study aims to investigate the perception of speech emotions in sub-clinical population compared to a healthy population. A considerable body of literature has reported on emotion perception deficits and the relevance of these impairments in people with clinical depression and its association. Whatever work done on speech impairments is mostly based on production and how it relates to depression severity levels and treatment and much less work on how emotion perception is affected in individuals with self reported depression. This research project aims to contribute to the under- standing of speech emotion perception in depression and its implications for emotional processing. By examining the relationships between depression, emotion perception, and other mental health measures along with the demo- graphic details provided such as Familial Mental history, personal history etc.

Keywords : Self Reported Depression, Anxiety, Stress, Speech Emotion Perception, Cross-Sectional Study, Demographics

1 Introduction

Speech is an integral part our communication and social interaction. It enables us to share complex thoughts and emotion. How we perceive speech depends not only on linguistic and acoustic features (e.g., words, and fundamental frequency), its content, or the paralinguistic cues (e.g., speaker's emotion) but it also influenced by the listeners' mood. Though mood and emotion are often used interchangeably, these two are different constructs [1] Emotion is more transient in nature and is triggered by the external stimulus, i.e., event, situation, or object [2, 3]. Whereas mood is considered more general and can be influenced by wide range of events than specific events or situations. It is more stable and long-lasting mental state that varies from hours to weeks [1, 2, 3]. Pleasant mood broadens the scope of attention and associated cognitive processing and builds resilience, whereas unpleasant or sad mood narrows the scope of attention and increase rumination [4, 5].

Studies investigating the role of mood in processing emotional stimuli have widely used visual information either static or dynamic [4], words [1] or music [6]. Whereas, only a few studies have examined the role of individual's mental state in speech affective perception [7-10] despite speech being critical component of our everyday behavior. The current study investigates the impact mood on affective speech perception by analyzing emotional reactions to the affective speeches.

The role of mood state becomes critical for the examination when it associates with clinical condition called depression. Depression or unipolar depression is characterized as a mood disorder or affective disorder [1], and affect our sleep, appetite, motivation to perform daily activities, and most importantly, depression leads to the loss of feeling happy and satisfied in performing those activities which were pleasurable once. The common features of all the depressive disorders are presence of sad, empty, or irritable mood, accompanied by somatic, cognitive, and psychomotor changes that significantly affect the individual's capacity to function [11]. Emotional reactivity, therefore, becomes a core evaluation tool for examining the role of depression in emotion processing. Positive and negative emotional reactivity has been associated with risk assessment of depression [1]. Emotional reactivity is defined as a change in individuals' affective perception or experience in correspondence with the emotionally evocative stimuli. Though, it is commonly assumed that depressive state is associated with reduced positive

1

emotional reactivity [12, 13], and increased negative emotional reactivity [12] the cognitive models of depression have contradictory views.

Beck's model of depression [12], commonly known as the Cognitive Triad, argues that an individual internal representation determines how individuals perceive themselves and view the world around them. The information processing model of depression [14] incorporating the Cognitive Triad model postulates that biased semantic representation of self and the event leads to biased attentional processing and biased memory formation. Further, the associative network theory [15] posits that negative interpretation of any events favors negative stimulus processing and inhibits or blocks attention towards positive events and memories, suggesting increased negative emotional reactivity, a mood-congruency effect. However, recently, a study with MDD patients reported contrasting results [13]. They observed attenuated emotional reaction to all the affective stimuli regardless of their nature, indicative of emotional context insensitivity (ECI). The ECI model argues that individual with depressive episode experience resistance to emotional alteration and fails to accommodate sudden environmental changes [13, 16-17]. Consistent to positive affective processing, the positive attenuation theory argues for reduced positive emotional reactivity. In a nutshell, the positive emotional reactivity to positive stimuli supports the positive model of depression. Whereas the negative emotional reactivity presents a conflict between mood congruency and ECI model. We use these two competing models to help us explain the affective speech processing by individuals reporting depressive and non-depressive symptoms.

2 The examination of speech affective perception by individuals with varying depressive symptoms will not only help us characterize depression better but it will also help us develop digital application to analyze the changes in affective speech perception associated with depression and comorbidities. Despite recognizing the importance of speech analysis, we lack objective assessment of speech related processing in clinical examination. One of the plausible explanations for the gap could be lack of technology driven study. The advancement in digital information technology allows unobtrusive examination of our everyday behaviour ranging from text to voice to walk and offers a fresh approach to mental health practices. Digital information technology not only brings objectivity to assessment and allow health monitoring more effectively. Currently, mental health assessment does not consist of any standardized technological/ digital support to evaluate the progress or deterioration caused by a given intervention. The consultants primarily rely on patients' self-discretion and report. Such services has a great impact on the low and middle-income countries (LIMICs) like India because of the obscured mental health practices and awareness, social stigma, and the paucity of mental health experts, i.e., 0.75:100 000:: psychiatrist: patient ratio [1, 18], that affect the urban and rural India differentially [1, 18].

The current project aims to use digital interface to examine the impact of depression and comorbid conditions on emotional reactivity to affective speech. Does depression or a depressive state influence speech perception? Does subclinical depression or self-reported depressive symptoms show a similar trend in speech production and speech perception?

2   Methodology

2.1   PARTICIPANTS

Participants of the study were the young adult population between the age of 18-25 years. Their participation was completely voluntary. The goal of this study is to investigate the emotion perception deficits in the young population with self-reported depression, anxiety or Stress and how differently they rate the emotions of audio stimuli than healthy adults.

2.2   STIMULI

Audio files were taken from an open source database - IEMOCAP. IEMOCAP database is an acted, multimodal and multispeaker database, collected at SAIL lab at USC. It consists of dyadic sessions where actors perform improvisations or scripted scenarios, specifically selected to elicit emotional expressions. IEMOCAP database is annotated by multiple annotators into categorical labels, such as anger, happiness, sadness, neutrality, as well as dimensional labels such as valence, activation and dominance. We took 25 files from each of the emotions - sad, happy, angry & neutral

2

randomly. Each audio stimulus was between 3-7 seconds of duration.

2.3 TASKS & MEASURE

This study consisted of 5 tasks:
It started with a Demographic information survey to ask age, gender, hearing abilities, personal and familial mental health conditions, followed by a Language fluency test to ensure each participant can understand the American accent audio files from the dataset and then current mood assessment through a self reported test called PANAS-10. Then comes the emotion perception task where participants were asked to perceive the speaker's emotions from a series of audio files presented to them on a SAM scale rated from 1 to 5 for valence and arousal. Clear Instructions about each task were provided before each task. Finally the study ended with a health survey to obtain general health measures(To assess self reported depressive symptoms - PHQ9, for Anxiety - GAD-7, For stress PSS - 4, STAI-T).

Participants completed the study via Labvanced a JavaScript web application developed for professional behavioral research. Participants were randomly assigned a set (out of 4). Each set had a different set of 25 audio files but almost equal number of files from each emotion (Sad, Happy, Angry and Neutral) sampled from IEMOCAP dataset. We sampled 100 files from the dataset, 25 from each emotion and distributed these into 4 sets(A, B, C, D) of 25 files, making sure that no 2 files of the similar rating come together in the order. In the end we repeated the first 5 files again to make the total of 30 files in each set so that we can ensure Intra Rater Reliability. This pseudo randomization helps us to ensure Inter rater reliability as well as help us validate the results we got are not fixed to the type of stimuli being presented to the participant
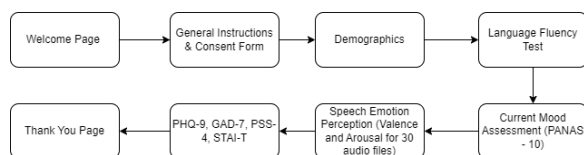
3 Findings & Results

3.1 Data Collection & Observations

We have completed the data collection for about 100 people and have started the analysis of the data to observe the differences in perception and conclude some of the possible reasons for it. Participants' gender distribution was approximately equal. Hearing abilities were assessed, and all participants demonstrated the ability to understand American-accented audio files. Participants reported a range of personal and familial mental health conditions.

Following are the mean observations of the data of approx 50 participants for valence and arousal of the audio file for set 1 out of 4 similar sets :
Following is the observation of valence ratings of the participants from

Valence (Data Collected)

Emotion	Depressed	Mild - Depressed	Non - Depressed
Happy	3.77	3.88	4.25
Sad	2.5	2.38	2.42
Angry	3.0	3.02	2.83
Neutral	2.6	2.73	2.71

Valence (Data Collected)

Emotion	Anxiety	Mild - Anxiety	Non - Anxiety
Happy	3.42	3.94	3.98
Sad	2.58	2.39	2.4
Angry	3.25	2.89	3.0
Neutral	2.64	2.57	2.8

Now the following table is for the Arousal Ratings of participants with their categorisation based on the levels of self reported Anxiety and Depression.

Arousal (Data Collected)

Emotion	Depressed	Mild - Depressed	Non - Depressed
Happy	3.33	3.58	4.58
Sad	2.17	2.1	2.17
Angry	2.97	2.98	3.0
Neutral	2.17	2.18	2.14

Arousal (Data Collected)

Emotion	Anxiety	Mild - Anxiety	Non - Anxiety
Happy	3.33	3.89	3.5
Sad	2.5	2.22	1.95
Angry	3.42	3.19	2.67
Neutral	2.21	2.26	2.08

I have generated such tables and mean/median values for all of the sets. And Based on the data obtained we can make the following observations:

Participants with self-reported depression tended to rate emotions with lower valence (more negative) compared to healthy participants. For

example, sad emotions received lower valence ratings. Participants with self-reported anxiety showed a tendency to rate emotions with higher arousal (more intense) compared to healthy participants. For instance, emotions like anger received higher arousal ratings. The healthy group demonstrated more accurate emotion perception across all categories, with a consistent trend of rating emotions closer to the true valence and arousal.

Participants with self-reported depression, anxiety, or stress tended to have higher scores on the negative affect items (e.g., sadness, fear, distress) compared to healthy participants. Healthy participants exhibited higher scores on positive affect items (e.g., joy, interest, excitement) compared to those with self-reported mental health conditions.

3.2 Statistical & Regression Analysis

Performing statistical analysis on the data, including Repeated Measure ANOVA, Mann-Whitney U test, and Kruskal-Wallis variance test, is important to gain insights and draw conclusions from the research findings. ANOVA helps determine if there are significant differences in ratings across emotion categories. The Mann-Whitney U test compares ratings between depressed and non-depressed participants, as well as anxious and non-anxious participants. The Kruskal-Wallis variance test examines differences in ratings among the four emotion categories. These tests provide statistical evidence to support interpretations and enhance our understanding of the data.

Regression analysis is essential in our study to examine the relationships between speech emotion perception and various mental health measures. By employing regression analysis, we can quantitatively assess the impact of factors such as depression, anxiety, and stress (and other demo- graphics as discussed earlier) on individuals' ability to perceive and interpret emotions in speech. This statistical technique allows us to identify significant predictors and determine the strength and direction of their influence on emotion perception. Through regression analysis, we can uncover valuable insights into the emotional experiences of individuals with depression, potentially shedding light on the underlying mechanisms and informing future assessment and treatment strategies for depressive disorders.

4   Conclusion and Discussion

Participants with self-reported depression, anxiety, or stress tended to have higher scores on the negative audio files (e.g., sadness or distress) compared to healthy participants. Healthy participants exhibited higher scores on positive affect audio files (e.g., joy, interest, excitement) compared to those with self-reported mental health conditions.

This study provided insights into the emotion perception deficits in young adults with self-reported depression, anxiety, or stress. Based on their valence and arousal ratings, the results imply that those who self-report mental health issues may have different emotional perception. Individuals in good health demonstrated a more constant and accurate ability to identify emotions from auditory cues or audio recordings.

The study's inter- and intra-rater reliability was enhanced by the pseudo-randomization of the audio recordings, giving confidence in the findings.

5   References

1. Sheoran, H., & Srivastava, P. (2022). Self-Reported Depression Is Associated With Aberration in Emotional Reactivity and Emotional Concept Coding. Frontiers in Psychology, 13, 814234.

2. Kaplan, R. L., Levine, L. J., Lench, H. C., & Safer, M. A. (2016). Forgetting feelings: Opposite biases in reports of the intensity of past emotion and mood. Emotion, 16(3), 309.

3. Beedie, C., Terry, P., & Lane, A. (2005). Distinctions between emotion and mood. Cognition & Emotion, 19(6), 847-878.

4. Goel, M., Srivastava, P., Agrawal, M., Singhal, R., Chand, R., & Baijesh, A. R. (2021, May). Head-movement analysis of 360° affective experience. In 2021 IEEE 7th International Conference on Virtual Reality (ICVR) (pp. 231-239). IEEE.

5. Fredrickson, B. L. (2004). The broaden–and–build theory of positive emotions. Philosophical transactions of the royal society of London. Series B: Biological Sciences, 359(1449), 1367-1377.

6.Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., & Tervaniemi, M. (2011). A functional MRI study of happy and sad emotions in music with and without lyrics. Frontiers in psychology, 2, 308.

7.Chandrasekaran, B., Van Engen, K., Xie, Z., Beevers, C. G., & Maddox, W. T. (2015). Influence of depressive symptoms on speech perception in adverse listening conditions. Cognition and Emotion, 29(5), 900-909.

8.Xie, Z., Zinszer, B. D., Riggs, M., Beevers, C. G., & Chandrasekaran, B. (2019). Impact of depression on speech perception in noise. Plos one, 14(8), e0220928.

9.Liu, G., Cai, S., & Wang, C. (2023). Speech emotion recognition based on emotion perception. EURASIP Journal on Audio, Speech, and Music Processing, 2023(1), 22. 10.Yi, H., Smiljanic, R., & Chandrasekaran, B. (2019). The effect of talker and listener depressive symptoms on speech intelligibility. Journal of Speech, Language, and Hearing Research, 62(12), 4269-4281.

11.First, M. B. (2013). DSM-5 handbook of differential diagnosis. American Psychiatric Pub.

12.Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. American Journal of Psychiatry, 165(8), 969-977.

13.Bylsma, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional reactivity in major depressive disorder. Clinical psychology review, 28(4), 676-691.

14.Disner, S. G., Beevers, C. G., Haigh, E. A., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. Nature Reviews Neuroscience, 12(8), 467-477.

15. Bower, G. H. (1981). Mood and memory. American psychologist, 36(2), 129.

16. Kuppens, P., Allen, N. B., & Sheeber, L. B.(2010).Emotional inertia and psychological maladjustment. Psychological science, 21(7), 984-991.

17. Rottenberg, J. (2017). Emotions in depression: What do we really know?. Annual Review of Clinical Psychology, 13, 241-263.

18. Garg, K., Kumar, C. N., & Chandra, P. S. (2019). Number of psychiatrists in India: Baby steps forward, but a long way to go. Indian journal of psychiatry, 61(1), 104.

19. Xie, Z., Zinszer, B. D., Riggs, M., Beevers, C. G., Chandrasekaran, B. (2019). Impact of depres- sion on speech perception in noise. PLOS ONE, 14(8). https://doi.org/10.1371/journal.pone.0220928

20. Bradley, M. M., Lang, P. J. (1994). Measuring emo- tion: The self-assessment manikin and the semantic dif- ferential. Journal of Behavior Therapy and Experimental Psychiatry, 25(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

21. Livingstone SR, Russo FA (2018) The Ryerson Audio- Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multi-modal set of facial and vocal expressions in North American English. PLoS ONE 13(5): e0196391. https://doi.org/10.1371/journal.pone.0196391.

22. Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. Speech Communication, 71, 10–49. https://doi.org/10.1016/j.specom.2015.03.004

23. Robin, J., Harrison, J. E., Kaufman, L. D., Rudzicz, F., Simpson, W., Yancheva, M. (2020). Evaluation of speech-based Digital Biomarkers: Review and recommendations. Digital Biomarkers, 4(3), 99–108. https://doi.org/10.1159/000510820

24. Watson, D., Clark, L. A., Tellegen, A. (1988). Development and validation of brief measures of pos- itive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063