**Paper title: A preliminary investigation on frequency dependant cues for human emotions**

**Authors:** Manish kumar, Thushara D. Abhayapala, and Prasanga Samarasinghe

**Manuscript ID:** acoustics-1696040

**Response to the reviewers**

This is the response to the reviewers' comments on the manuscript (acoustics-1696040) submitted to MDPI acoustics special issues on Human's Psychological and Physiological Responses to Sound Environment.

We would like to thank the reviewers for their time in giving constructive feedback on our manuscript. In this document, we have addressed each question/comment separately with a detailed response, along with references to the changes made in the revised manuscript. The major changes from the original manuscript are highlighted in red in the revised manuscript for convenience. Please note that due to the advised modifications, there have been a few changes in the figures and section numbers of the revised manuscript, and all the references we have made in our responses correspond to the revised manuscript unless explicitly mentioned otherwise. We sincerely hope that our attempts to improve the manuscript have done justice to the time and effort spent by the reviewers.

## 1 Response to the Reviewer 1

We would like to thank the reviewer for their valuable time.

***RC-1.1* Comment from the reviewer:**

*I suggest you can provide the listening duration for 1 frequency sound as such readers can imagine how the ratings were conducted.*

**Response:**

We thank the reviewer for the suggestion. We have revised Section 3.3 - Emotion rating Mechanism of the manuscript according to the reviewer's comment.

**Changes in the revised manuscript:**

> Each participant must listen to all 12 Solfeggio frequencies for the duration of 20 seconds each. For each primary emotion class (Happy, Sad, Calm, and Anger), the participants are required to rate these pure tones on a scale of 1 to 10 by using a comparative analysis. Here 1 represents "least significant emotion" and 10 represents "most significant emotion".

***RC-1.2* Comment from the reviewer:**

*The size of samples can be bigger instead of 30 samples per frequency. Referring to the emotion class - happy, the height of the median line of the box at 417 Hz is almost the same as the entire box at 440 Hz, which means there is no sharp or clear likelihood to show a difference between the two groups. So, the more data you can obtain, the more accurate the result would be.*

**Response:**

We thank the reviewer for the suggestion. We have now increased the number of samples per frequency to 37 and have updated all five plots and Section 3.1 (Participants) of the manuscript accordingly. Although the results did not improve significantly with addition of 7 new participants (samples), the trend indicates that more data we can obtain, the more accurate results would be.

**Changes in the revised manuscript:**

Thirty-seven adults in the age group of 20 – 55 yrs. from Canberra, Australia participated in the research voluntarily. The gender distribution among participants is 20:17 (i.e., 20 male and 17 female ), with median age of 27 years (SD 6.5987).

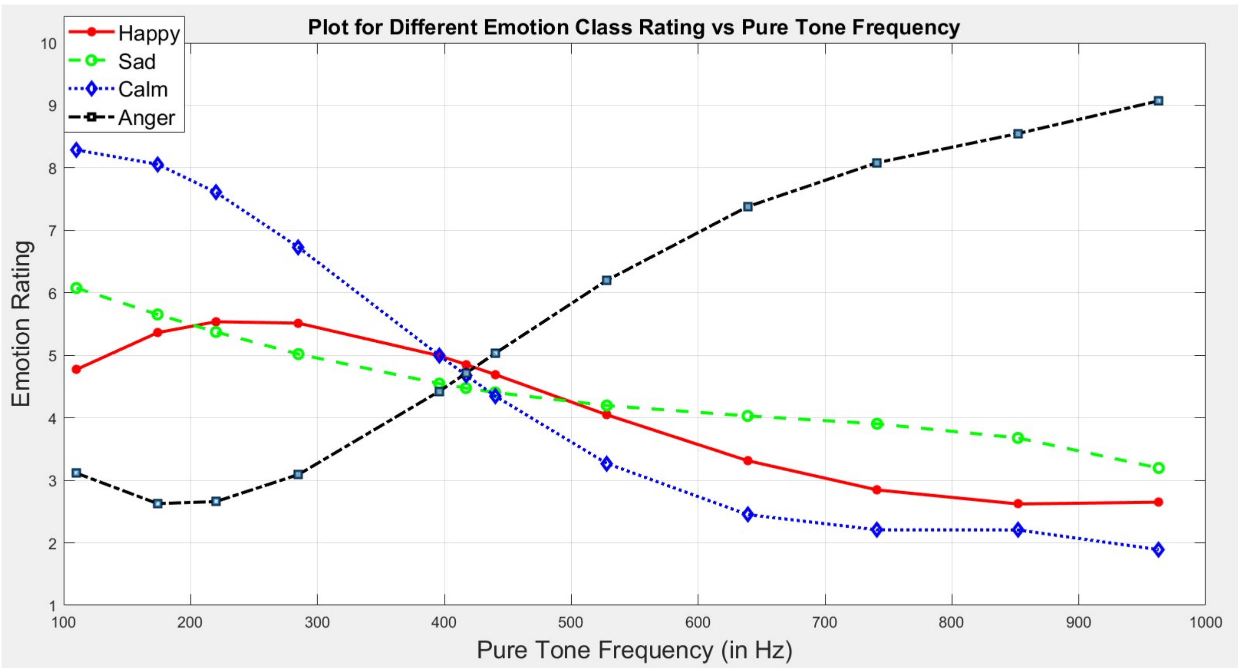


Fig. 1: Plot for **different emotion class rating vs Pure tone frequency.** Here, the line graph for all four primary emotions - Happy, Sad, Anger, and Calm are plotted w.r.t pure tone frequency (in Hz).

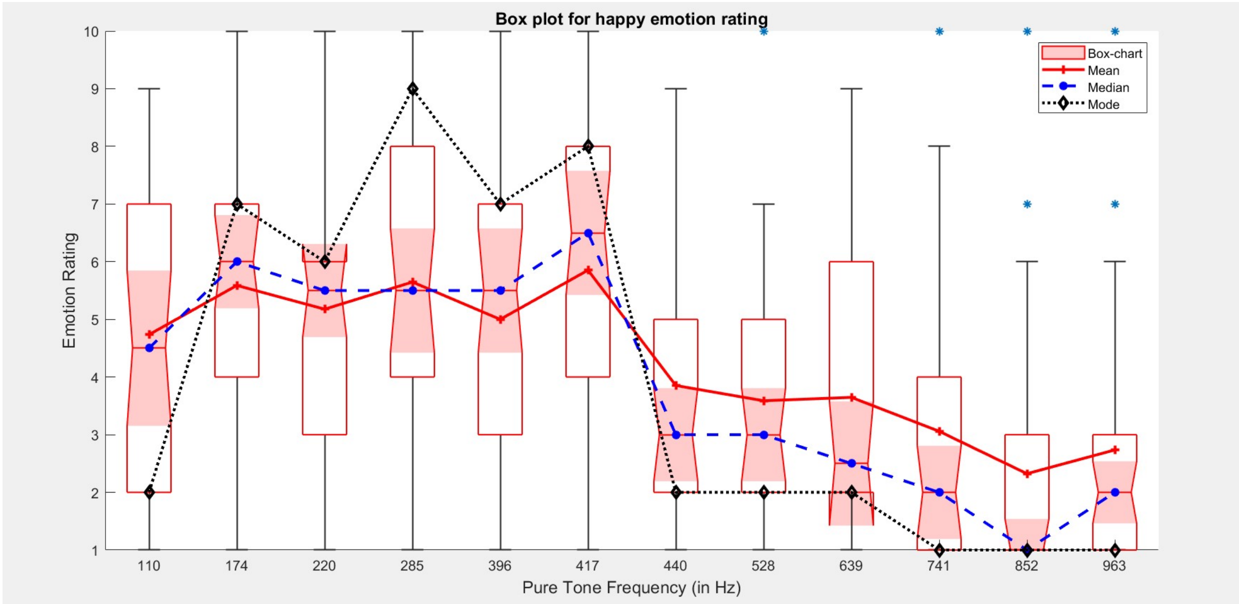


Fig. 2: Plot for **Happy emotion rating v/s Pure tone frequency.** Here, the shaded area (red in color) of box chart represents distribution for annotated emotion corresponding to each pure tone frequency. Also, blue dots in emotion rating plot represent outliers data.

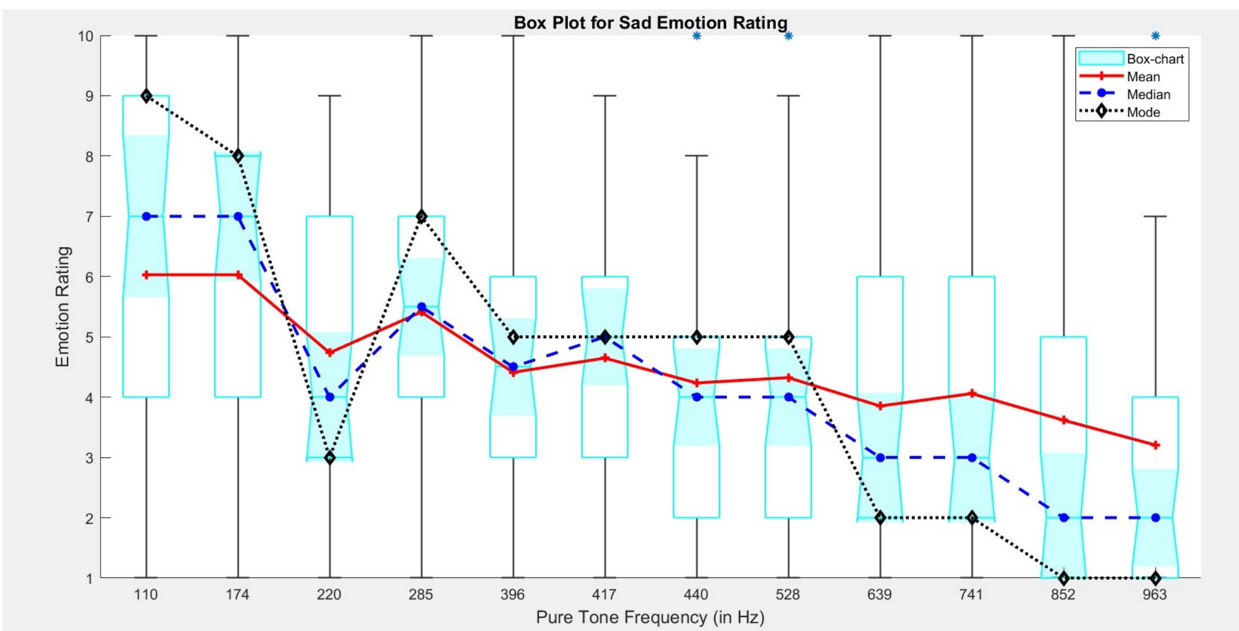


Fig. 3: Plot for **Sad emotion rating v/s Pure tone frequency.** Here, the shaded area (cyan in color) of box chart represents distribution for annotated emotion corresponding to each pure tone frequency. Also, blue dots in emotion rating plot represent outliers data.

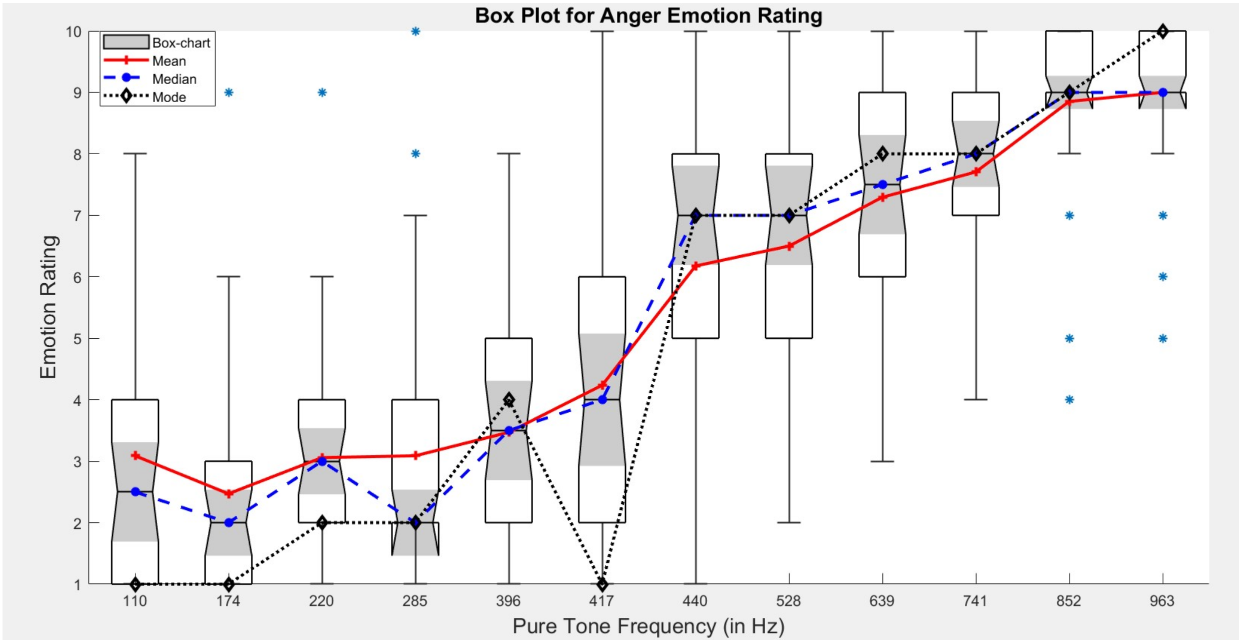


Fig. 4: Plot for **Anger emotion rating v/s Pure tone frequency.** Here, the shaded area (grey in color) of box chart represents distribution for annotated emotion corresponding to each pure tone frequency. Also, blue dots in emotion rating plot represent outliers data.

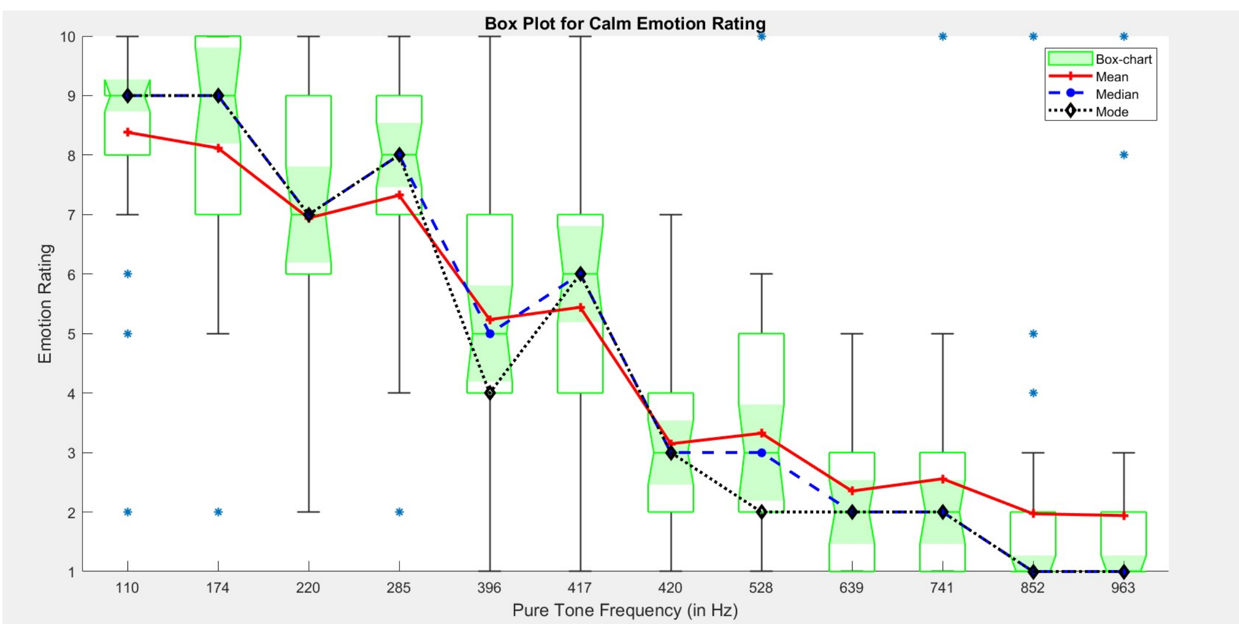


Fig. 5: Plot for **Calm emotion rating v/s Pure tone frequency.** Here, the shaded area (green in color) of box chart represents distribution for annotated emotion corresponding to each pure tone frequency. Also, blue dots in emotion rating plot.

***RC-1.3*** **Comment from the reviewer:**

*To clarify the specific preference on frequency choice, you can put some rules on picking the preferred tone frequency.*

**Response:**

We thank the reviewer for the suggestion. We designed the listening test according to a rule of thumb, which is now mentioned in Section 3.3.

**Changes in the revised manuscript:**

> Each participant must listen to all 12 Solfeggio frequencies for the duration of 20 seconds each. For each primary emotion class (Happy, Sad, Calm, and Anger), the participants are required to rate these pure tones on a scale of 1 to 10 by using a comparative analysis. Here 1 represents "least significant emotion" and 10 represents "most significant emotion". The detailed instructions for completing the emotion rating test are as follows:
>
> - For each emotion category, participants must listen to all 12 pure tones, perform a comparative analysis, and rate the significance of highlighted emotion on a scale of 1 to 10. Here, the rule of thumb for rating a particular tone is based on performing the comparative analysis of the previous and next tone, as they are in ascending order of frequency.
> - For an effective and consistent perception test, participants are asked to use headphones of good frequency responses.
> - Before starting the perception test, participants must adjust their speaker/ headphone sound level to a comfortable level (Recommended level: 15 to 20% ).

***RC-1.4*** **Comment from the reviewer:**

*You have cited the reference that the most preferred tones range from 400 to 750 Hz. However, it is observed that the emotion class - anger is triggered starting from 440 Hz to 528 Hz and your preferred tone is from 210 to 540 Hz. So, what is your underlying principle that includes the anger triggering frequency?*

**Response:**

We thank the reviewer for their valuable comment and observations regarding the emotion class - anger. Here, we would like to highlight two things -

- In the referenced paper [1], Paul C. Vitz identified preferred tones as an inverted function of frequency and intensity. However, with regards to our paper, we are investigating cues for four different primary emotions - Happy, Sad, Anger, and Calm over a frequency range of 110 Hz to 963 Hz. Hence, comparing the results of our paper directly with referenced paper will be difficult, as the motivation, and research methodologies of both papers are slightly different.
- Also, angry emotion being triggered in the 440 to 528 Hz frequency range strengthen our hypothesis that the frequency range 432 – 440 Hz is neutral from human emotion perspective.

We have revised Section 4 - Results and Discussions of the manuscript and have acknowledged the trend that emotion class - anger is triggered from 440 Hz to 528 Hz .

**Changes in the revised manuscript:**

> Apart from that, for higher frequencies ($\geq$ 440 Hz), the **emotion rating tends to dampen continuously for three primary emotions happy, sad, and calm**. However, for emotion class - anger, the rating gradually increases with an increase in pure tone frequency ($\geq$ 440 Hz). This clearly reflects the **dominance of anger emotion class in the higher frequency range**. Also, it is noteworthy that the emotion class anger is triggered in a frequency range starting from 440Hz to 528Hz.

## 2 Response to the Reviewer 2

We thank the reviewer for their recommendations and support.

***RC-2.1*** **Comment from the reviewer:**

*First, it is rather difficult to assess the validity of the main conclusions of the paper and their compliance with the world level, since the authors of the paper briefly described the main problems associated with the recognition of emotions, and the section with related work of the world scientific community is completely absent. Works related to the subject of emotion recognition, both in audio and video modalities, are constantly presented at such conferences as: ICASSP, INTERSPEECH, CVPR and others.*

**Response:**

We thank the reviewer for valuable feedback and the suggestion. We have revised the paper and included a separate section for related work of the world scientific community (Section 2 - Related Work).

**Changes in the revised manuscript:**

> This section provides a brief overview of related work on psychoacoustic parameters based auditory emotion recognition and discusses how psychoacoustic parameters can be vital in developing multi-modal emotion recognition algorithms.
>
> To date, studies on emotion recognition have been predominantly based on identifying cues from facial expression data [2] [3], voice and speech data [4], audio-visual data [3] [5], and physiological data [6] [7] [3]. Among these, the area of research involving sound without verbal information as the stimulus for emotion recognition is classified as psychoacoustics [8] [9]. It has the potential to be established as a distinct field of emotion recognition, as sound plays an important role in detecting cues for various primary emotions. Also, it is possible to accurately detect other's emotions from the tone of their voice without actually understanding/ listening to speech [10] [11]. Furthermore, the accuracy of emotion recognition algorithms improve with multi-modality (i.e., two or more sensory) inputs in comparison to uni-modality inputs [12] [13] [14].
>
> However, most of the emotion recognition studies are conducted either by extracting specific parameters that become apparent when emotion is evoked (e.g., facial, voice or physiological indicators) [15] [16], or by training the data to predict an emotion by applying mathematical algorithms [17] [18]. Also, it is important to understand that non-verbal sound is processed

differently than verbal sound/ speech in the brain and is relatively natural and universal means of human communication with higher cues for emotional responses [19] [20]. Hence, it is important to identify physical sound determinants of emotional responses through non-verbal sound, as the chances of emotional responses being convoluted with other factors will be minimal and will only be the result of present stimuli.

In the past, there have been few research studies to establish perceptual determinants for physical parameters in non-verbal sound such as mapping emotional reactions to auditory events onto a pleasantness-unpleasantness (Lust-Unlust) dimension [21], and identifying Valence and potency/control cues in the vocal expression of emotion [22]. Also, in a recent research Erkin Austy and Daniel mapped everyday emotional reactions to both sound characteristics, and the appraisal of sound/sound source [23]. In another research, Paul C. Vitz identified preferred tones as an inverted function of frequency and intensity [1]. Apart from that, other noteworthy researches in the field of psychoacoustics are the development of an auditory emotion recognition function using psychoacoustic parameters based on IADS [8], and emotional reactions to sounds without meaning [24].

However, to the best of our knowledge, there has been no research in the past aiming to establish a relationship between frequency (one of the most important physical parameters of sound) and perceived human emotions. Establishing such a relationship between physical parameters of sound and emotional responses can be important for various applications such as emotion induction with sound, development of new sound abatement approaches, sound and auditory interface design, and prediction and assessment of subjective noise experiences [25] [26] [27]. The findings of this research can also be helpful in establishing perceptual determinants for physical parameters in sound, along with setting directions for other research in emotion recognition such as detecting emotion cues from non-musical, and non-vocal sounds.

***RC-2.2*** **Comment from the reviewer:**

*Further, as stated in the paper, the average age of people who participated in online testing is 27 years old, if the maximum age is 55 years old, this may indicate that the sample is dominated by a young age group and this may lead to some bias in the perception of different types of pure tones in the specified range. However, the authors of the paper did not take this into account.*

**Response:**

We thank the reviewer for their suggestion. We have revised Section 4 - Results and Discussions of the manuscript, and have acknowledged that the sample is dominated by young age group (average age of participants is 27 years) which may leads to some bias in the perception of different pure tones in specified frequency range.

**Changes in the revised manuscript:**

> Since, the emotion rating test is conducted online via Qualtrics software as a survey tool, there is a possibility that differences in testing conditions (i.e., ambiance noise, head-phone response, and quality) may affect the true stimuli to the listener. To minimize the influence of these variations on the final observation, box-plots are used as a statistical tool that is capable of visually summarising and conveying the level, spread, outliers, and symmetry of data distribution [28]. Also, mean, median, and mode of emotion rating data is plotted along with box-plot to visualise the overall trend. Apart from that, the sample is dominated by a young age group (the average age of participants is 27 years) which may lead to some bias in the perception of different pure tones in the specified frequency range.

***RC-2.3*** **Comment from the reviewer:**

*In section 3, the authors of the paper write: "For that, we only consider primary emotions - Happy, Sad, Anger and Joy", where is Joy Calm?*

**Response:**

We thank the reviewer for the feedback. We have revised Section 4 (Results and Discussions) (which was Section 3 in earlier version) of the manuscript, and have corrected the typo of emotion class "calm" being referred as "joy".

**Changes in the revised manuscript:**

> One of the main intentions of this study is to establish a frequency-dependent inter-relationship amongst different emotion classes. For that, we only considered primary emotions - Happy, Sad, Anger, and Calm, as they almost cover the entire human emotion spectrum [29] [30]. Also, considering only the primary emotions will allow our results to be more generic and broadly applicable.

***RC-2.4* Comment from the reviewer:**

*The list of references is mainly represented by rather old works, which once again emphasizes the need for a section with a detailed analysis of previous studies.*

**Response:**

We thank the reviewer for their valuable feedback and the suggestions. We have revised the paper and included a separate section for related work of the world scientific community (Section 2 - Related Work). Also, we have included 21 new references (total 46) from recent works in the field of emotion recognition being published in journals and conferences such as ICASSP, CVPR, JASA, MDPI and others.

## REFERENCES

[1] P. C. Vitz, “Preference for tones as a function of frequency (hertz) and intensity (decibels),” *Perception & Psychophysics*, vol. 11, no. 1, pp. 84–88, 1972.

[2] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, “A survey of affect recognition methods: Audio, visual, and spontaneous expressions,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 1, pp. 39–58, 2008.

[3] S. Saganowski, “Bringing emotion recognition out of the lab into real life: Recent advances in sensors and machine learning,” *Electronics*, vol. 11, no. 3, p. 496, 2022.

[4] S. Steidl, M. Levit, A. Batliner, E. Noth, and H. Niemann, “" of all things the measure is man" automatic classification of emotions and inter-labeler consistency [speech-based emotion recognition],” in *Proceedings.(ICASSP’05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 1. IEEE, 2005, pp. I–317.

[5] F. Ma, W. Zhang, Y. Li, S.-L. Huang, and L. Zhang, “Learning better representations for audio-visual emotion recognition with common information,” *Applied Sciences*, vol. 10, no. 20, p. 7239, 2020.

[6] R. A. Calvo and S. D’Mello, “Affect detection: An interdisciplinary review of models, methods, and their applications,” *IEEE Transactions on affective computing*, vol. 1, no. 1, pp. 18–37, 2010.

[7] T. Wierciński, M. Rock, R. Zwierzycki, T. Zawadzka, and M. Zawadzki, “Emotion recognition from physiological channels using graph neural network,” *Sensors*, vol. 22, no. 8, p. 2980, 2022.

[8] Y. Choi, S. Lee, S. Jung, I.-M. Choi, Y.-K. Park, and C. Kim, “Development of an auditory emotion recognition function using psychoacoustic parameters based on the international affective digitized sounds,” *Behavior Research Methods*, vol. 47, no. 4, pp. 1076–1084, 2015.

[9] H. Brumm and S. A. Zollinger, “The evolution of the lombard effect: 100 years of psychoacoustic research,” *Behaviour*, vol. 148, no. 11-13, pp. 1173–1198, 2011.

[10] R. Adolphs, H. Damasio, and D. Tranel, “Neural systems for recognition of emotional prosody: a 3-d lesion study.” *Emotion*, vol. 2, no. 1, p. 23, 2002.

[11] P. N. Juslin and P. Laukka, “Communication of emotions in vocal expression and music performance: Different channels, same code?” *Psychological bulletin*, vol. 129, no. 5, p. 770, 2003.

[12] G. Castellano, L. Kessous, and G. Caridakis, “Emotion recognition through multiple modalities: face, body gesture, speech,” in *Affect and emotion in human-computer interaction*. Springer, 2008, pp. 92–103.

[13] K. R. Scherer and H. Ellgring, "Multimodal expression of emotion: Affect programs or componential appraisal patterns?" *Emotion*, vol. 7, no. 1, p. 158, 2007.

[14] T. Mittal, P. Guhan, U. Bhattacharya, R. Chandra, A. Bera, and D. Manocha, "Emoticon: Context-aware multimodal emotion recognition using frege's principle," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 234–14 243.

[15] C. S. Ooi, K. P. Seng, L.-M. Ang, and L. W. Chew, "A new approach of audio emotion recognition," *Expert systems with applications*, vol. 41, no. 13, pp. 5858–5869, 2014.

[16] H. Liu, Y. Zhang, Y. Li, and X. Kong, "Review on emotion recognition based on electroencephalography," *Frontiers in Computational Neuroscience*, p. 84, 2021.

[17] F. Ma, Y. Li, S. Ni, S. Huang, and L. Zhang, "Data augmentation for audio–visual emotion recognition with an efficient multimodal conditional gan," *Applied Sciences*, vol. 12, no. 1, p. 527, 2022.

[18] F. Zhu-Zhou, R. Gil-Pita, J. García-Gómez, and M. Rosa-Zurera, "Robust multi-scenario speech-based emotion recognition system," *Sensors*, vol. 22, no. 6, p. 2343, 2022.

[19] P. Belin, S. Fillion-Bilodeau, and F. Gosselin, "The montreal affective voices: a validated set of nonverbal affect bursts for research on auditory affective processing," *Behavior research methods*, vol. 40, no. 2, pp. 531–539, 2008.

[20] S. Paquette, I. Peretz, and P. Belin, "The "musical emotional bursts": a validated set of musical affect bursts to investigate auditory affective processing," *Frontiers in psychology*, vol. 4, p. 509, 2013.

[21] E. Glover, "The re-creating of the individual: By beatrice m. hinkle, md (george allen & unwin, ltd., london, 1924. pp. 465. price 18 s." *International Journal of Psycho-Analysis*, vol. 6, pp. 75–78, 1925.

[22] M. Goudbeek and K. Scherer, "Beyond arousal: Valence and potency/control cues in the vocal expression of emotion," *The Journal of the Acoustical Society of America*, vol. 128, no. 3, pp. 1322–1336, 2010.

[23] E. Asutay, D. Västfjäll, A. Tajadura-Jimenez, A. Genell, P. Bergman, and M. Kleiner, "Emoacoustics: A study of the psychoacoustical and psychological dimensions of emotional sound design," *Journal of the Audio Engineering Society*, vol. 60, no. 1/2, pp. 21–28, 2012.

[24] D. Västfjäll, "Emotional reactions to sounds without meaning," *Psychology*, vol. 3, no. 8, p. 606, 2012.

[25] P. Desmet, "Designing emotions. delft: Delft university of technology," Ph.D. dissertation, Dissertation, 2002.

[26] R. W. Picard, *Affective computing*. MIT press, 2000.

[27] D. Västfjäll, "Emotion induction through music: A review of the musical mood induction procedure," *Musicae Scientiae*, vol. 5, no. 1_suppl, pp. 173–211, 2001.

[28] D. F. Williamson, R. A. Parker, and J. S. Kendrick, "The box plot: a simple visual method to interpret data," *Annals of internal medicine*, vol. 110, no. 11, pp. 916–921, 1989.

[29] C. E. Izard, *Human emotions*. Springer Science & Business Media, 2013.

[30] J. A. Russell, M. Lewicka, and T. Niit, "A cross-cultural study of a circumplex model of affect." *Journal of personality and social psychology*, vol. 57, no. 5, p. 848, 1989.