IMPACT OF MUSIC STIMULI ON DECISION MAKING

Utkarsh
Center for Computational Natural Sciences and Bioinformatics,
International Institute of Information Technology Hyderabad
utkarsh.azad@research.iiit.ac.in

Nitesh Gupta
Robotics Research Center,
International Institute of Information Technology Hyderabad
nitesh.gupta@students.iiit.ac.in

ABSTRACT

Music affects us in many ways in our everyday life. In this article, we study our experiment in which participants scored words depending upon how they perceived them emotionally between positive or negative while listening to music that was chosen to induce positive, neutral or negative mood. Through its outcomes, we discuss how music affects our decision making in various scenarios. We use analysis of variance (ANOVA) to test whether music individually affects the decision making or it interacts with other factors to make an impact on the result. Post-Hoc tests are then used to analyze how this interaction between music and other factors changes in different conditions. Finally, using a stochastic sequential model of simple decisions, the drift-diffusion model (DDM), we study which parts of decision making like stimulus evaluation, response caution, and response bias, are actually affected by music. Here, we note that the music manipulation was indeed effective and our results have implications for future studies of the connection between music and mood.

1. AUXILIARY GRADING INFORMATION

We think both of us deserve to be graded between the range: **8-10** for the class participation. When we say this, we are not randomly shooting for the stars but because unlike our other courses at this college, we both attended almost all the classes despite lenient attendance policy, did a lot of class notes, and got engaged with discussion whenever possible. Despite having a little/no prior musical training, we really enjoyed participating in class because we think the class environment and other participants were always open for learning, and also to new ideas. In fact, the motivation for the proposals we had kept forward as our possible research projects, like *Plagiarism Detector*, *Quantum Music*, or even the current one itself comes from some such active discussions during the lectures and mess tables.

2. INTRODUCTION

In the recent past, a considerable amount of studies have shown that mood can affect emotional processing. This phenomenon is usually spoken as mood-congruent processing or bias, reflecting the finding that negative mood induces a relative preference for negative emotional content and similarly for a positive mood. In fact, studies that induce mood, either through being attentive to happy/sad music or having participants write passages or see pictures supported a specific emotion, have shown mood-congruent bias across a spread of tasks [1–3]. Similarly, this has also being studied from a neurophysical perspective where music was found to be strongly linked to brain regions linked with emotion and reward, and different musical patterns are shown to own meaningful associations to emotional affectations [4–6]. All this evidence indicates a deep and profound two-way connection between music and emotional perception. Therefore, working with these pieces of evidence, we designed a study to:

1. Provide a useful evidence that music stimuli affect decision making.
2. Analyze the observed affects of music stimuli and its interaction with other stimuli.
3. Study which parts of decision making like stimulus evaluation, response caution, and response bias, are actually affected by music stimuli or its interaction with other stimuli.

The structure of the paper is as follows. In Section 3 we discuss our data collection procedure along with details of the experiment. In Section 4 we present our initial results on the basis of which we gain the confidence to investigate further. In Section 5 we introduce the analysis of variance (ANOVA) and its results. In Section 6, we give the characteristics of the drift-diffusion model and its results. In Section 7 we conclude our findings, and finally in Section 8 we discuss them in a broader context with our current limitations and further possibilities.

3. DATA COLLECTION

For collecting data, we made a form from scratch and hosted it online on Google App Script for accepting submissions [1] from our participants. The form had 4 sections with 10 questions each, i.e. 40 questions in total. At the start of each section, a song (without video) is auto-played in the background. Once, the user is sufficiently comfortable, he was allowed to begin answering the question using a toggle button. Each question consists of a word and

[1] Form Link

Note: Both Utkarsh, Nitesh Gupta have contributed equally to this work.
Attribution: Submitted as a final course project for the course [CSE588] Music, Mind and Technology.

WS\MIM	*Negative*	*Neutral*	*Positive*
Negative	**0.966833**	0.765625	0.6625
Neutral	0.14583333	**0.34375**	0.20833333
Positive	0.65833333	0.86979167	**0.95979167**

Table 1. % Accuracy - WS v/s MIM

WS\MIM	*Negative*	*Neutral*	*Positive*
Negative	**0.03035**	0.03123	0.21875
Neutral	0.04117	**0.30208333**	0.79166667
Positive	0.65833333	0.86979167	**0.95979167**

Table 2. % Positive Response for - WS v/s MIM

WS\MIM	*Negative*	*Neutral*	*Positive*
Negative	**2.135417**	3.083333	4.343750
Neutral	3.416667	**5.375000**	4.937500
Positive	5.291667	7.165625	**7.572917**

Table 3. Mean Response Score - WS v/s MIM

WS\MIM	*Negative*	*Neutral*	*Positive*
Negative	**1.110901**	1.422826	1.375179
Neutral	0.918679	**1.093642**	1.002656
Positive	1.457588	1.394554	**1.102579**

Table 4. STDEV of Response Score - WS v/s MIM

participants were asked to mark it with a number in the range 1-9, where 1 indicates a strong negative feeling for that word, and 9 indicates a strong positive feeling for that word. Each person was given 10 seconds to respond to the answer, otherwise, 5 was chosen as the default response.

3.1 Selection Procedure Of Music

The selection of music is an important step to take into account the effect of our first stimuli: **Music Induced Mood (MIM)**. To do so, we selected 4 songs from the Image-Music Affective Correspondence (IMAC) Dataset [7]. This dataset collected songs along with their unique tags from LAST.FM (or Million Song Dataset). Tags that depicted emotions like 'happy', 'energetic', 'soothing', 'sad' were shortlisted and were used to assign the emotional score to the songs. Based on this emotional score, i.e. the capability of a song to induce a certain emotion, and we selected the following songs:

1. *No Such Thing* by *John Mayer*: High positive emotional score.
2. *Everything Reminds Me of Her* by *Elliot Smith*: High negative emotional score.
3. *Dreamland* by *Robert Miles*: Neutral emotional score.
4. *Lower Your Eyelids to Die With The Sun* by *M83*: Neutral emotional score.

3.2 Selection Procedure Of Words

Another important stimulus is the **Word Stimuli (WS)** [8] which is accounted for by the selection of appropriate words. It comprises an equal set of Positive (like a laugh, love, and reward), Negative (like Terrorist, Quiver and Timid), and Neutral (like Planet and alien) words for each section. In each section, the order of the words is chosen such that the immediate predecessor is not directly related to it. These words were chosen from the following datasets:

1. *Bing Liu's Opinion Lexicon*
2. *Saif M. Mohammad's NRC Emotion Lexicon*
3. *SentiWordNet sentiment lexicon*

3.3 Alternate Proposals

As we will learn in the next section, one major requirement for our We didn't use Google or Microsoft because they are static forms and therefore we can not get response times for each question. Also, they do not provide the facility to autoplay music in the background for each section without revealing questions. Though both these features are available in Inquisit software by millisecond which is generally used in collecting data in Music research, we couldn't use it because it is not available for Linux and most of the participants use the Linux operating system.

4. INITIAL RESULTS

We have accumulated our curated response data in the Tables 1-6. Each table We have analyzed the response data on the basis of three criteria (1) Accuracy, (2) Response Score, and (3) Response Time, as given below:

4.1 Accuracy

In the Tables 1 and 2, we have reported the data regarding accurate response % and positive response % for each type of WS against different MIM. From Table 1, it can be clearly seen that the accuracy in diagonal elements, i.e. when MIM was equal to the given WS, was much more than other elements in their respective rows. Moreover, we saw a decrease in accuracy as MIM became more contrasting w.r.t the given WS. Similarly in Table 2, we see that the % positive response increased in the case of every given WS (each row) with a positive MIM. This can also be understood as the % positive responses decreased as MIM became non-positive.

4.2 Response Score

In the Tables 3 and 4, we have reported the mean response score and the standard deviation for each WS against the MIM. Similar to accuracy, the diagonal elements give the extremity (or one can say more idealistic) score in their respective rows. Therefore, participants were biased to score words positive in the positive MIM and vice versa. Moreover, the deviation values show that this is more significant in the case of contrasting WS and MIM type.

Response\MIM	*Negative*	*Neutral*	*Positive*
Negative	**4.899853**	6.683169	6.207500
Neutral	6.329645	**6.349489**	6.253775
Positive	6.689684	6.591271	**5.114316**

Table 5. Mean Response Times (seconds) Response v/s MIM

Response\MIM	*Negative*	*Neutral*	*Positive*
Negative	**2.151572**	3.164059	3.750442
Neutral	2.518355	**2.160259**	2.442990
Positive	2.797048	2.804592	**2.241320**

Table 6. STDEV of Response Times (seconds) Response v/s MIM

Dep. Variable:	Score	**R-squared:**	0.682
Model:	OLS	**Adj. R-squared:**	0.679
Method:	Least Squares	**F-statistic:**	255.1
Date:	Fri, 17 Apr 2020	**Prob (F-statistic):**	1.25e-230
Time:	19:56:34	**Log-Likelihood:**	-1599.2
No. Observations:	960	**AIC:**	3216.
Df Residuals:	951	**BIC:**	3260.
Df Model:	8	**Covariance Type:**	nonrobust
Omnibus:	227.175	**Durbin-Watson:**	1.925
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	41.768
Skew:	0.030	**Prob(JB):**	8.52e-10
Kurtosis:	1.980	**Cond. No.**	4.88

Table 7. Complete Summary for the interaction model being fit and its test analysis for the presence of autocorrelation and normality, homogeneity of variance and assessment of multicollinearity.

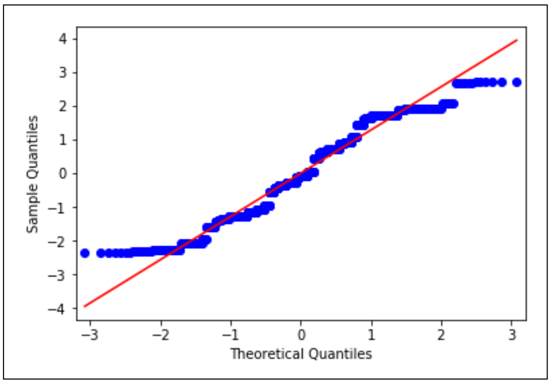

Figure 1. Q-Q plot for Interaction model linear fit.

4.3 Response Time

In the Tables 5 and 6, we have reported the mean response time and the standard deviation for each response against the MIM. Similar to our discussion about the response score, the diagonal elements give the extremity (or one can say more lowest) time values in their respective rows. Therefore, participants were biased to score words positive quicker in the case of positive MIM and vice versa. Moreover, the deviation values show that this is more significant in the case of contrasting responses and MIM type.

5. ANALYSIS OF VARIANCE (ANOVA)

An ANOVA test is a way to find out if the survey or experiment results are significant. In other words, they help us to figure out if we need to reject the null hypothesis or accept the alternate hypothesis. Following are the hypothesis we want to study:

1. **Main effect of Music Induced Mood (MIM)**
   - H_0: MIM does not affect the response
   - H_a: MIM affects the response
2. **Main effect of Word Stimuli (WS)**
   - H_0: WS affects the response
   - H_a: WS does not affects the response
3. **Interaction of MIM and WS**
   - H_0: There is no interaction
   - H_a: There is an interaction

In order to study the following hypothesis, we require two models: the Interaction Model and the Addition Model. We initially begin with the analysis of the interaction model and generate a prediction model for it through a linear fit [9]. Table 7 contains a summary of the results for the fit produced for this model. In fact, we can see using Figure 1 that our model fits almost perfect for most of the data points. Therefore, we can summarize our overall model as:

$$F(8, 951) = 255.073; \quad p = 0.0000 \tag{1}$$

5.1 Results for Two-Way ANOVA: Interaction Model

We have used Two Way ANOVA with Word Stimulus (WS) and Music Induced Mood (MIM) as independent variables and Response as dependent variables. Both the independent variables have 3 levels each, i.e., Positive, Negative, and Neutral.

We show the results of the Two-Way ANOVA table for the interaction model in Table 8. From this table, we can clearly reject the null hypothesis that "*MIM does not affect the response*" because the f-value and p-value of MIM are significant. Similarly, we reject the null hypothesis that "*There is no interaction between MIM and WS*" as their f-value and p-value is also significant. Finally, we accept the null hypothesis that "WS affects the response" due to the same reason. Therefore, the adequacy of the additive model can be rejected as interaction had a significant value. Our next task is to infer the strength of the interaction effect for different pairs of WS and MIM.

5.1.1 *Profile Plot: Interaction Model*

1. An interaction effect means that the effect of one factor depends on the other factor and it's shown by the lines in our profile plot not running parallel.
2. In this case, the effect for Mood-Induced Mood (MIM) interacts with Word Stimulus (WS). This means the effect of MIM on each WS is different.

	Sum Sq	*Mean Sq*	*Df*	*F Value*	*P Value*	η^2	ω^2
C(WS)	2611.954167	1305.977083	2.0	789.624156	**8.332225e-203**	0.527897	0.527052
C(MIM)	702.722917	351.361458	2.0	362.441320	**5.362712e-77**	0.142026	0.141310
C(WS):C(MIM)	60.292708	15.073177	4.0	9.113594	**3.165397e-07**	0.012186	0.010845
Residual	1572.880208	1.653922	951.0	NaN	NaN	NaN	NaN

Table 8. Two-Way ANOVA table for the interaction model is shown above. Since the p-value of interaction between the variables is much smaller than 0.005, the adequacy of the additive model can be rejected. This means that interaction is significant.

	sum_sq	**mean_sq**	**df**	**F**	**PR(>F)**	η^2	ω^2
C(Mood)	236.42708	118.21354	2.0	65.88916	**2.658327e-25**	**0.256989**	0.252596
Residual	683.56250	1.794127	381	NaN	NaN	NaN	NaN

Table 9. Between Subjects Effects: One-Way ANOVA table with Negative WS

	sum_sq	**mean_sq**	**df**	**F**	**PR(>F)**	η^2	ω^2
C(Mood)	316.46354	158.23177	2.0	87.52671	**5.266352e-32**	**0.314814**	0.310659
Residual	688.77604	1.807811	381	NaN	NaN	NaN	NaN

Table 10. Between Subjects Effects: One-Way ANOVA table with Positive WS

	sum_sq	**mean_sq**	**df**	**F**	**PR(>F)**	η^2	ω^2
C(Mood)	210.12500	105.06250	2.0	99.01589	**3.833392e-30**	**0.511668**	0.505195
Residual	200.54166	1.061067	189	NaN	NaN	NaN	NaN

Table 11. Between Subjects Effects: One-Way ANOVA table with Neutral WS

3. In Figure 2, we plot the profile for the interaction. From these results, we can see that the orange line (MIM-Positive) ascent quite steeply from Negative to Neutral WS.

4. However, from Neutral to Positive WS, the ascent is not that steep. This variation, but in the opposite order, can be seen for the blue line (MIM-Negative) too. In the case of the green line (MIM-Neutral), however, there's a uniform accent in both the steps.

5. Therefore, we might want to ignore the individual effect of MIM or WS. Instead, this main effect "lumps together" their different effects and shows how actually Response Scores are affected

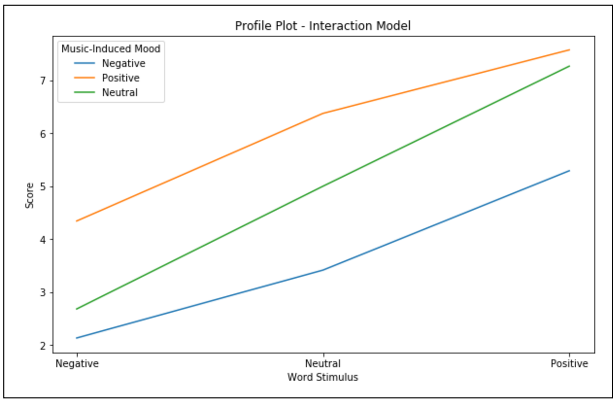

Figure 2. Profile Plot for Interaction Model

5.1.2 *Between Subjects Effects: One-Way ANOVA*

Now, we will analyze the effect of MIM on each WS separately to see how differently it affects them. The results in Tables: 9-11 show us that the effect of MIM is statistically significant for all of WS. However, this just means it's probably not zero. To have more clarity about its strength we check the η^2 values. We see that for both Negative and Positive WS it is not that strong. However, for Neutral WS it is considerably stronger than the latter two

5.1.3 *ANOVA Output - Post Hoc Tests*

Our above analysis of the response data suggests that our MIM doesn't perform similarly on all the WS. But the amount by which they differ from each other (in a pairwise fashion) can be seen via a post hoc test called *Tukey's Honesty Significant Differences (HSD)* comparisons. The results given in the Tables: 12-13, suggest all the pair differ considerably to express an effect that is statistically significant.

5.1.4 *ANOVA Final Conclusion*

All the analysis performed [9] above on our response data shows there's a substantial interaction effect between MIM and WS on Response Score. Moreover, this effect was much larger for the Neutral WS than the Positive/Negative WS.

6. DRIFT DIFFUSION MODEL (DDM)

By ANOVA, we understood that music-induced mood interacts with word stimuli to impact the decision making.

WS-1	WS-2	meandiff	p-adj	*reject*
N	P	3.6875	0.001	True
N	U	1.7552	0.001	True
P	U	-1.9323	0.001	True

Table 12. Tukey's Honesty Significant Differences (HSD) comparisons for WS - (Pairwise)

MIM-1	MIM-2	meandiff	p-adj	*reject*
N	P	2.3875	0.001	True
N	U	1.4729	0.001	True
P	U	-0.9146	0.001	True

Table 13. Tukey's Honesty Significant Differences (HSD) comparisons for MIM - (Pairwise)

But we didn't get to know which part of decision making is affected. The drift-diffusion model allows detailed explanations of behavior in two-choice discrimination tasks. It extracts theoretically relevant components of decision making from our Response and Response Time (RT) data. The model provides a decomposition of data that isolates components so that they can be individually studied.

In this model, the decision process starts between the two boundaries that correspond to the positive and negative responses. Evidence is accumulated over time to drive the process towards one of the boundaries. Once a boundary is reached, it signals a commitment to that response. The time taken to reach the boundary denotes the decision time, and the overall response time is given by the decision time plus the time required for processes outside the decision process like encoding and motor execution. The model includes a parameter for this non-decision time (Ter), to account for the duration of these processes. The primary components [10–12] of the decision process in the DDM which are also shown in Figure 3 are:

1. **Boundary separation (a)**: It provides an index of responses caution or speed/accuracy settings. Wide boundaries indicate a cautious response style where more evidence needs to be accumulated before the choice is made. The need for more evidence makes the decision process slower and more accurate.

2. **Drift rate (v)** It provides an index of the evidence from the stimulus that drives the accumulation process. Positive values indicate evidence for the top boundary and negative values for the bottom boundary. Further, the absolute value of the drift rate indexes the strength of the stimulus evidence, with larger values indicating strong evidence and leading to fast responses.

3. **Starting point (z)**: It indicates whether there is response bias. If z is closer to the top boundary, it means less evidence is required to reach that boundary, so *positive* responses will be faster and more probable than *negative* responses.

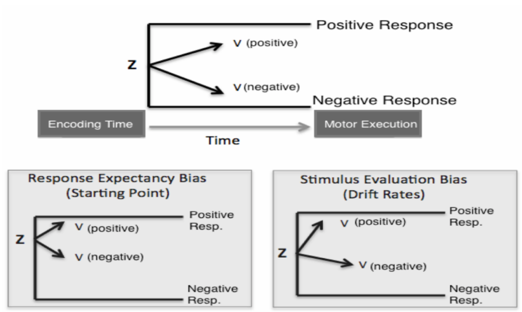

Figure 3. Drift-Diffusion Model

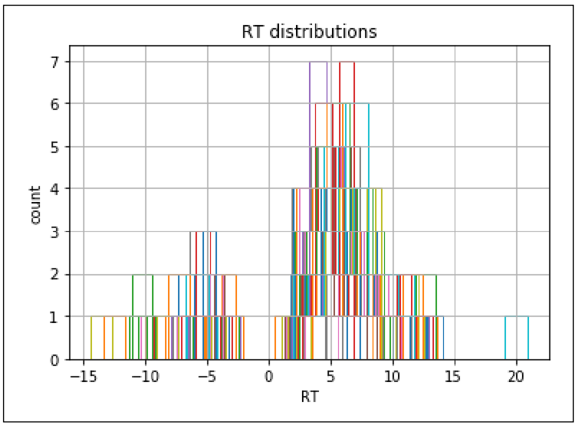

Figure 4. RT distributions of each individual during positive musical stimuli

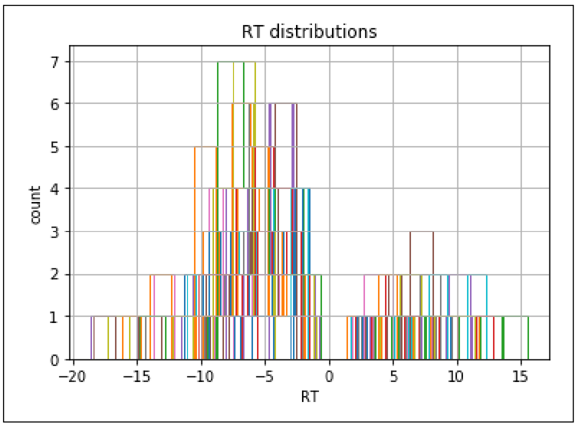

Figure 5. RT distributions of each individual during negative musical stimuli

Figure 4 and Figure 5 shows the Response Time distributions for each individual in different conditions. These are made using accuracy coding. In this type of coding the positive answers irrespective of the stimuli, bias has more reward (1) than the negative answer (0). Response times for the negative responses were flipped to the negative domain. The formal model for the DDM is the following, where dx represents the accumulated evidence at each time step with A being some evidence for one of the two choices, and cdW being the noise term:

$$dx = A(a, v)dt + c(a, v)dW \qquad x(0) = z \tag{2}$$

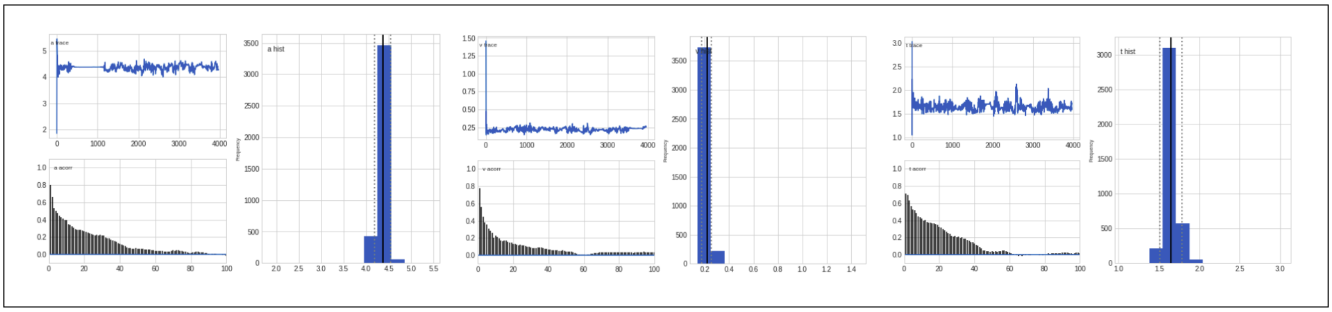

Figure 6. Convergence of MCMC algorithm was tested by plotting the posteriors for the estimated all the parameters. (a) Problematic patterns in the trace like drifts are absent here. (b) Also, the autocorrelation drops to around zero rather quickly (i.e., before 50). (c) Spread of marginal posterior histogram is contained between 2.5 and 97.5% percentiles. This shows that convergence was met, and the model used for estimation is giving the right results.

P	*mean*	*std*	*err*
a	4.0769	0.580325	0.0583936
v	0.226518	0.389558	0.0392595
t	1.60662	0.290486	0.0286334
z	***0.569457***	***0.0217889***	***0.00210441***

Table 14. Estimated mean DDM parameters for the participants (N=24) under positive stimulus.

N	*mean*	*std*	*err*
a	4.01185	0.442589	0.0441975
v	0.245041	0.28314	0.0282704
t	1.74586	0.243499	0.0197383
z	***0.473327***	***0.0246451***	***0.00241933***

Table 15. Estimated DDM parameters for the participants (N=24) under negative stimulus.

6.1 DDM Results

Based on Equation 2, an analytic solution to the resulting probability distribution of the response times is provided by the following:

$$f(t|v,a,z) = \frac{\pi}{a^2}\exp\left(-vaz - \frac{v^2t}{2}\right) \times \sum_{k=1}^{\infty} k\exp\left(-\frac{k^2\pi^2t}{2a^2}\right)\sin k\pi z \tag{3}$$

We performed a hierarchical Bayesian estimation [13] for an approximated version of Equation 3 to model a fit on our given response time data for both positive and negative MIM using the Markov Chain Monte Carlo (MCMC) algorithm [13, 14]. We show the convergence of our algorithm in Figure 6 and the approximated parameters: a, v, z, are shown in the Tables: 14-15. These results show that *boundary separation (a)* and *drift velocity (v)* remains almost same under both positive and negative mood. Response time is a little less during the positive music-induced mood. But the main thing that the music affected is the *starting point (z)*, i.e, there is response expectancy bias. So, music does not significantly affect how participants evaluate the emotional content of the stimuli, but rather it affects how they favor one response option over others, i.e., in other words, the expectancy of one response is much higher.

7. CONCLUSION

Response data showed a mood-congruent emotional bias [10] based on the music conditions. Two-Way ANOVA results confirmed that the interaction between music-induced mood and word stimuli does have a significant contribution to the selection of response. However, it cannot then explain how the components of the interactive model drive the mood-congruent bias. To identify that, we performed the DDM analysis. Its results suggest that music-induced mood does not significantly affect how participants evaluate the emotional content of the stimuli, but rather it affects how they favor one response option. Thus the mood-congruent bias appears to be driven more by the preference/expectancy of one response option (over the other), rather than the emotional processing of the word stimulus.

8. DISCUSSION

We note that the smaller number of participants limits the generalizability of the results produced by the study. So, we suggest to substantiate our results further, we would require a much larger group of participants along with a significant increase in the number of trials. Also, during music selection, we had assumed that the emotional scores represented mood-induced by that piece of music. To verify this assumption, we could have used a separate group of participants to see the effectiveness of these pieces for inducing the appropriate mood. However, due to limited resources, we were not able to do so. Moreover, we would like to point out that our results show that there's an opportunity of using more rigorous computational models of decision making in a *two-interval forced-choice (2IFC)* paradigm such as Race model. Such models can be studied with a combination of complex tasks like as cooperative, quantitative, etc and once trained for the individual participants, they can use them directly to produce data for a future experiment.

9. REFERENCES

[1] Shannon K de lEtoile. *The effectiveness of music therapy in group psychotherapy for adults with mental illness*. The Arts in Psychotherapy, 29(2):69–78, 2002.

[2] Jeong-Won Jeong, Vaibhav A Diwadkar, Carla D Chugani, Piti Sinsoongsud, Otto Muzik, Michael E Behen, Harry T Chugani, and Diane C Chugani. *Congruence of happy and sad emotion in music and faces modifies cortical audiovisual activation*. NeuroImage, 54(4):2973–2982, 2011.

[3] Christof Kuhbandner and Reinhard Pekrun. *Joint effects of emotion and color on memory*. Emotion, 13(3):375, 2013.

[4] Anne J Blood and Robert J Zatorre. *Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion.* Proceedings of the National Academy of Sciences, 98(20):11818–11823, 2001.

[5] Sebastien Paquette, Isabelle Peretz, and Pascal Belin. *The musical emotional bursts: a validated set of musical affect bursts to investigate auditory affective processing.* Frontiers in psychology, 4, 2013.

[6] Carol L Krumhansl. *Music: A link between cognition and emotion. Current Directions in Psychological Science*, 11(2):45–50, 2002.

[7] G. Verma, E. G. Dhekane and T. Guha, *Learning Affective Correspondence between Music and Image,* ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019, pp. 3975-3979.

[8] *Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon*, Saif Mohammad and Peter Turney, In Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, June 2010, LA, California.

[9] Seabold, Skipper, and Josef Perktold. *"statsmodels: Econometric and statistical modeling with python."* Proceedings of the 9th Python in Science Conference. 2010.

[10] Liebman, Elad, Corey N. White and Peter Stone. "*On the Impact of Music on Decision Making in Cooperative Tasks."* ISMIR (2018).

[11] Liebman, Elad, Peter Stone and Corey N. White. *"Impact of Music on Decision Making in Quantitative Tasks."* ISMIR (2016).

[12] Liebman, Elad, Peter Stone and Corey N. White. *"How Music Alters Decision Making - Impact of Music Stimuli on Classification Tasks"* ISMIR (2015).

[13] Wiecki TV, Sofer I and Frank MJ (2013) *HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python.* Front. Neuroinform. 7:14. doi: 10.3389/fninf.2013.00014

[14] Navarro, D. D. J., and Fuss, I. I. G. (2009). *Fast and accurate calculations for first-passage times in Wiener diffusion models.* J. Math. Psychol. 53, 222–230. doi: 10.1016/j.jmp.2009.02.003