Modeling the Emotional Judgement Pathway of Auditory Stimuli

Joseph Kim[1,3], Emily Choe[2,3], Marianne Bezaire, Ph.D[3]

Portola High School, Irvine, CA 92618[1], Brooks School, North Andover, MA 01845[2], Boston University, Boston, MA 02215[3]

Introduction

A variety of sensory stimuli engage emotional circuits in the brain. Sensory representations are integrated with socio-emotional relevance (Figure 1) to guide our cognitive behavior.
Exactly how different neuroanatomical regions interact during this process remains unknown.

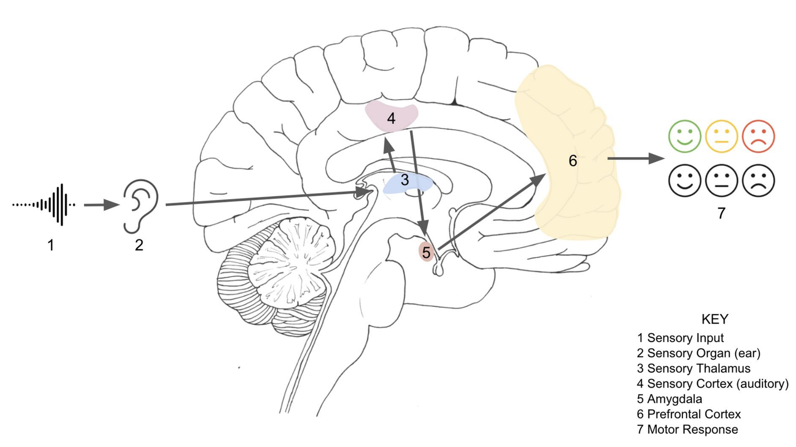

Figure 1. Emotional processing of auditory stimuli

Two theoretical frameworks try to explain the nature of sensory-emotion interactions:

1) discrete emotion theory: specific core emotions have biologically determined emotional responses
2) constructed/continuous emotion theory: scalar valuations of stimuli guide emotional responses

Arousal and valence are useful features to describe sensation-induced emotional responses that measure the level of intensity (less vs. more important) and attractiveness (positive vs. negative), respectively.

Neural networks are useful to model and understand networks of neurons involved in auditory-emotional processing.

We utilized the hypothesis of constructed emotion theory with the dimensions valence and arousal.

Methods

In this investigation, we create a functional abstract model of the neuroanatomical pathways involved in evaluating auditory stimuli to emotion effects. We model the cochlea and initial auditory processing using a Short-Term Fourier Transform (STFT). The STFT emulates the anatomical network of the cochlea by transforming the audio by the frequencies and amplitudes present in the stimuli. This initial processing of the audio yields a spectrogram of the STFT.

The auditory cortex, amygdala, OFC, VLPFC, and rVLPFC are then modeled with the use of a deep learning neural network, specifically the libraries tensorflow and keras. To emulate aspects of these neuroanatomical networks and their motifs, we utilize the Convolutional Neural Network (CNN) Xception, and various types of activation, dense, comparison layers and the like. The network is then trained on a Quadro RTX 4000 equivalent through indirect supervised learning with the rated audio database Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) using the 8 emotional categories Angry, Calm, Disgust, Fearful, Happy, Neutral, Sad, and Surprise.

Results

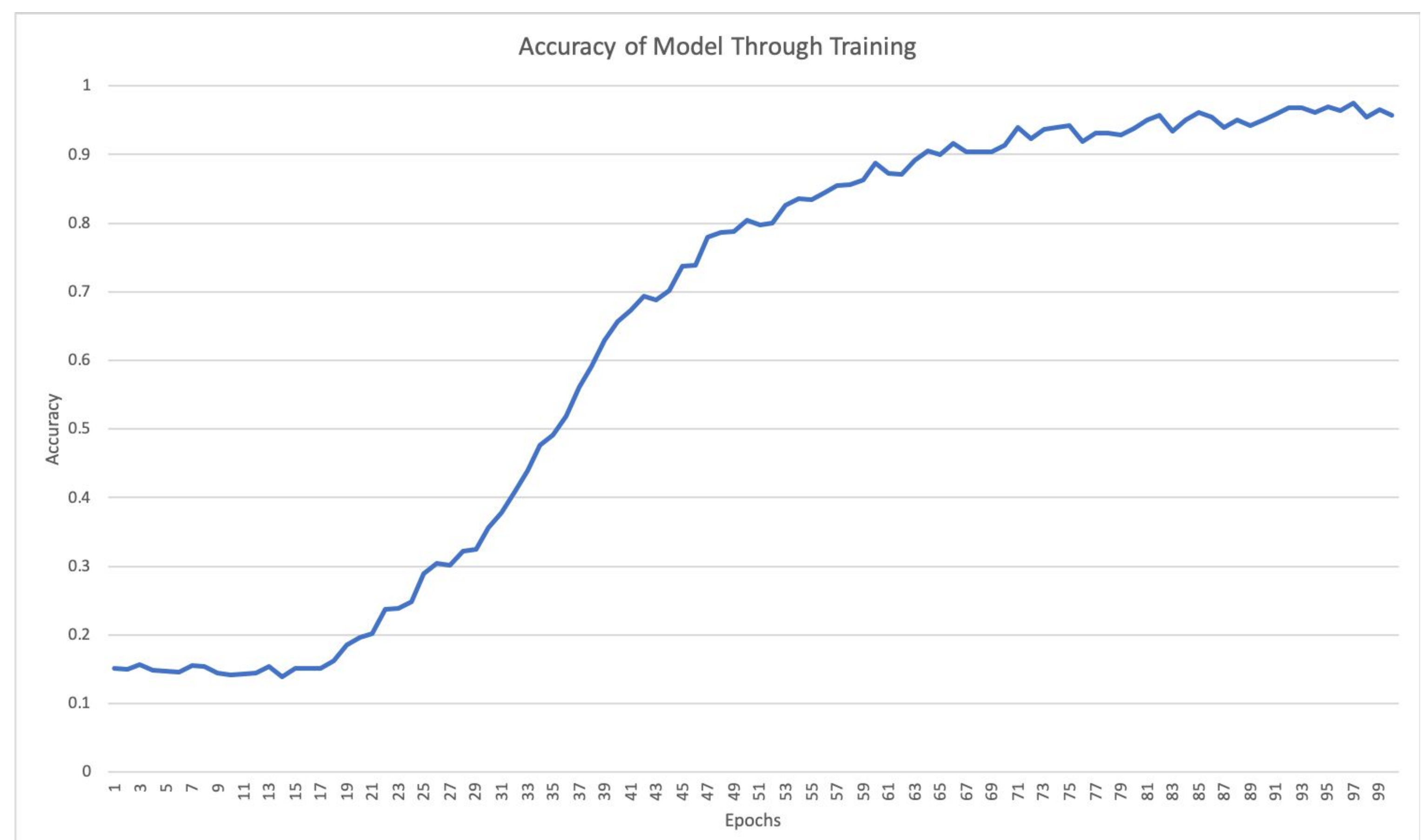

Figure 2. Learning curve of model in categorical accuracy across 100 epochs

The model successfully learns through the indirect supervised learning. This is evidenced by the plateauing of the accuracy shown in Figure 2 and the high final categorical accuracy of 0.9655485.

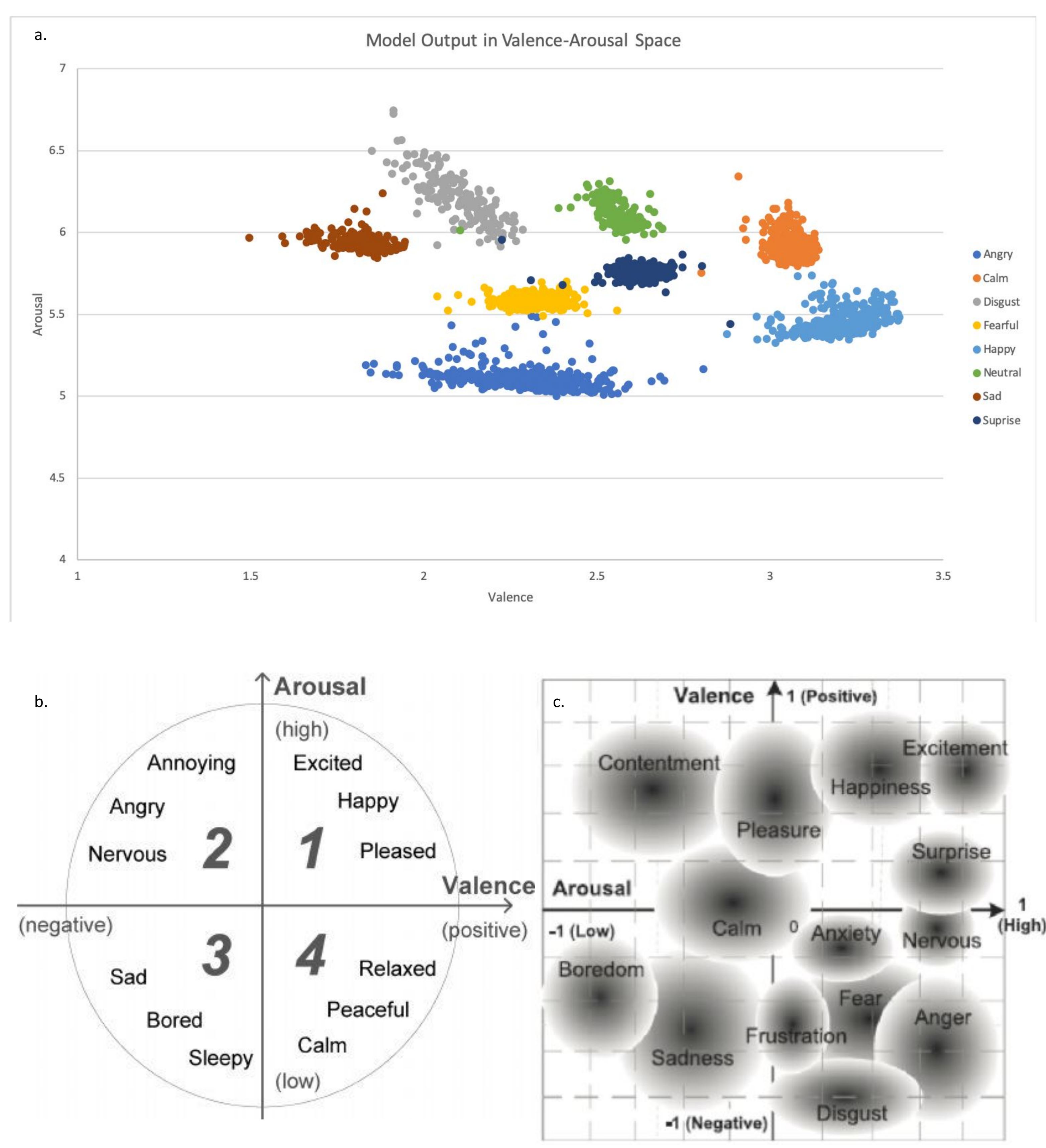

Figure 3. (a) Model output in Valence–Arousal dimensions; scalar values translate to low to high valence and high to low arousal (b) Approximation of emotions in 2D valence-arousal emotion space proposed by Russell 1980 (c) Emotions in 2D valence-arousal emotion space from an empirical study

The model outputs the 8 categories of emotions considered in this investigation (Angry, Calm, Disgust, Fearful, Happy, Neutral, Sad, and Surprise) in clear areas within the valence-arousal space. To objectify the uniqueness of the emotions in the valence-arousal space, we run a one-way analysis of variance test (ANOVA) for both valence and arousal. For the ANOVA test, we get a F-critical value of 2.013742036 and 2.013742036 for valence and arousal respectively, yielding a p-value of effectively 0, making the difference statistically significant.

We compare the output of our model in the valence-arousal space to other studies. Russell's 1980 investigation of emotional theory provide an approximation of discrete emotional categories in the valence-arousal space presented in Figure 3b. Note that positive emotions, neutral emotions, and negative emotions share a similar distribution through the valence scale. Figure 3c presents the area discrete emotional categories occupy in the valence-arousal space.from empirical data. Note the relationship of emotional categories considered in this investigation relative to each other in comparison to the empirical data (e.g. Anger, Fear, and Disgust, Happiness and Calm, and other relative similarities). This similarity underlines the significance of the axis in the model.

Discussion/ Conclusions

- Results are comparable to empirical data and previous studies
  - Figure 3a shows valence and arousal levels
  - Compare experimental results of 3a with published data in 3b and 3c, which provide a discrete area of each emotion in the valence-arousal space that follow those of empirical data
- The known anatomical motifs and connections of the theory of constructed emotion were utilized in the model
  - The results therefore provide further credibility to constructed emotion theory
- The model is generalizable to other auditory stimuli and usable in big-data investigations
- Future implications:
  - Provide emotional auditory processing for people with auditory disabilities
  - Studies involving auditory processing across different populations
  - Artificial Intelligence with emotional intelligence

References

Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)***2017**. https://arxiv.org/abs/1610.02357

Ledoux, J. E. Brain Mechanisms of Emotion and Emotional Learning. *Current Biology***1992**, *2*(4), 199. https://doi.org/10.1016/0959-4388(92)90011-9

Livingstone, S. R.; Russo, F. A. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A Dynamic, Multimodal Set of Facial and Vocal Expressions in North American English. *Plos One***2018**, *13*(5). https://doi.org/10.1371/journal.pone.0196391

Markov, K.; Matsui, T. Music Genre and Emotion Recognition Using Gaussian Processes. *IEEE Access***2014**, *2*, 688–697. https://doi.org/10.1109/ACCESS.2014.2333095

Picou, E. M.; Singh, G.; Goy, H.; Russo, F.; Hickson, L.; Oxenham, A. J.; Buono, G. H.; Ricketts, T. A.; Launer, S. Hearing, Emotion, Amplification, Research, and Training Workshop: Current Understanding of Hearing Loss and Emotion Perception and Priorities for Future Research. *Trends in Hearing***2018**, *22*, 233121651880321. https://doi.org/10.1177/2331216518803215

Russell, J. A. A Circumplex Model of Affect. *Journal of Personality and Social Psychology***1980**, *39*(6), 1161–1178. http://dx.doi.org/10.1037/h0077714

Schirmer, A.; Gunter, T. C. Temporal Signatures of Processing Voiceness and Emotion in Sound. *Social Cognitive and Affective Neuroscience***2017**, *12*(6), 902–909. www.doi.org/10.1093/scan/nsx020

Tye, K. M. Neural Circuit Motifs in Valence Processing. *Neuron***2018**, *100*(2), 436–452. https://doi.org/10.1016/j.neuron.2018.10.001

Yang, Y.-H.; Chen, H. H. Machine Recognition of Music Emotion. *ACM Transactions on Intelligent Systems and Technology***2012**, *3*(3), 1–30. http://doi.acm.org/10.1145/2168752.2168754

Acknowledgements

We would like to sincerely thank Johan Sebastian Martinez-Fuentes and Justin Lee for their guidance and help throughout the program and the presented project.