Modeling of Perceived Musical Rhythms Using Electrocorticography

Modeling of Perceived Musical Rhythms using Electrocorticography

Michael Dexheimer[1], Garett D. Johnson[2], Jerry J. Shih[3], Christian Herff[4], Dean J. Krusienski[1]

***Abstract*— Numerous studies have explored the neural correlates of musical rhythms using various neuroimaging modalities. Non-invasive neuroimaging modalities lack either the spatial or temporal resolution to reveal the nuances of neural processes involved in perception of musical rhythms. Intracranial recordings of electrophysiological activity such as electrocorticography (ECoG) can jointly provide spatial and temporal resolution for improved characterization and modeling of the underlying processes. The present study examines anticipatory and perceptual models that use ECoG recordings to estimate simple perceived and imagined musical rhythms in human participants. The resulting models are characterized and compared across participants. The results show that the anticipatory and perceptual models can reconstruct the auditory stimulus envelope with statistically-significant correlations when trained and tested on independent listening data. However, these models are unable to reliably reconstruct the expected rhythm pattern when trained on listening data and applied to imagining data. This suggests, similar to recent findings in overt and imagined speech decoding using intracranial signals, that there are likely distinct neural substrates activated during listening and imagining of musical rhythms.**

I. Introduction

Recognition of rhythmic patterns is instinctive in humans and can evoke many neurological and psychological reactions ranging from anticipatory to emotional. The perception of syntax, pitch, beat, and timbre all contribute to the appreciation of music and the processing of language [1], [2]. Music and language share networks and have their own distinct mechanisms of neural encoding [2], [3]. Simple musical rhythms have a consistent, repetitive structure and are significantly less complex than other commonly studied auditory stimuli such speech. Through the examination of simple rhythms, it may be possible to gain insights about the intricate networks involved in both language and music processing.

Because a variety of complex speech and auditory signals (both perceived and imagined) have been successfully decoded using intracranial recordings [4]–[7], it is conceivable to decode perceived and imagined musical rhythms from intracranial signals for application to a brain-computer interface [8]. Moreover, since the perception of rhythm is generally ubiquitous in humans across race and culture, such stimuli can facilitate the identification of a common neural substrate of rhythm processing.

Numerous previous neuroscientific studies have examined various aspects of music using scalp EEG [9]–[18] or functional magnetic resonance imaging (fMRI) [19]–[22]. Electrocorticographic (ECoG) recordings from electrodes placed directly on the surface of the cortex, provide access to broadband gamma activity ($\sim$70-250 Hz), which has been shown to be highly correlated with a wide variety of behavioral, sensory, and cognitive functions [8]. Recent studies have demonstrated that ECoG is capable of uniquely resolving and decoding complex spatiotemporal speech [4], [5], [23]–[25] and music signals [6], [7], [26]–[31] to a degree that was previously unattainable by fMRI or EEG/MEG.

The present study develops participant-specific models to reconstruct perceived imagined rhythms from ECoG recordings using both anticipatory (causal) and perceptual (noncausal) models trained on ECoG data acquired during listening to the rhythm stimulus. The motivation for these distinct models is to elucidate different neural processes underlying pure auditory rhythm perception versus the anticipation of upcoming beats, respectively. In contrast to prior work, the presented stimuli represent very basic rhythmic drum patterns to better facilitate identification of consistent neural correlates, while maintaining musicality to promote user engagement in the task compared to alternate stimuli such as clicks [12]. The performance of the two models are evaluated using independent ECoG test data from listening and imagining conditions, respectively. The resulting models are characterized and compared across participants.

II. Methodology

A. Participants

ECoG data were recorded from six patients (ages 22-27, one female) with intractable epilepsy undergoing the localization of epileptogenic zones and eloquent cortex prior to surgical resection. No patients reported having hearing deficits or any form of formal musical training. No brain tumors or lesions were indicated in the clinical evaluations and therefore did not impact the ECoG recordings. All participants in this study gave written informed consent and the study protocol was approved by the institutional review boards of Mayo Clinic Florida and Old Dominion University.

Each patient was implanted with subdural electrode grids based exclusively on their clinical need. All electrode locations were verified by co-registering preoperative MRI

This work was supported by the National Science Foundation (NSF) [grant numbers 1451028, 2011595]

[1]M. Dexheimer and D.J. Krusienski are with the Department of Biomedical Engineering, Virginia Commonwealth University, Richmond, VA, USA dexheimermg@vcu.edu, djkrusienski@vcu.edu

[2]G.D. Johnson is with the Department of Biomedical Engineering, Old Dominion University, Norfolk, VA, USA gjohn037@odu.edu

[3]J.J. Shih is with the Department of Neurology, UCSD Health, San Diego, CA, USA jerryshih@health.ucsd.edu

[4]C. Herff is with the Department of Mental Health and Neuroscience, Maastricht University, Maastricht, Netherlands c.herff@maastrichtuniversity.nl

and postoperative computerized tomography scans. The electrode locations for each participant are shown in Figure 1. Electrode locations and activations were generated using the Neural Act package [32].

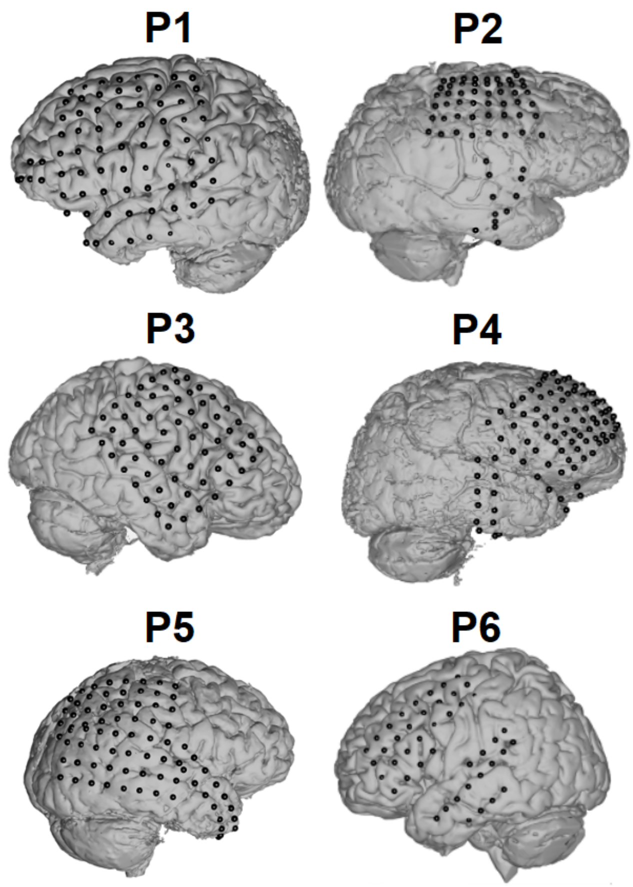

Fig. 1. Electrode grid locations for the six participants.

B. Stimuli and Task

In order to maintain simplicity while promoting ecological validity as musical stimuli, only fixed-pitch kick and snare drum sound samples were used to create the musical rhythms, each being of simple meter [26]. An experimental trial lasted 120 seconds, consisting of six continuous 20-second blocks. The first block represented the fundamental rhythm pattern, the second block added a single beat to the fundamental rhythm, and the third block added an additional beat to the rhythm from the second block. Blocks 4-6 repeated the patterns from blocks 1-3 in reverse order.

Participants were instructed to passively listen to and follow the rhythm being presented through research-grade earbuds while ECoG was simultaneously recorded. During each block, after at least 8 measures, the audio randomly ceased (i.e., became silent) for 4-8 measures during which the participant was instructed to continue imagining the beat. The intervals containing the audible rhythm are herein refereed to as *listen* and the silent intervals are referred to as *imagine*.

Four trials were performed, each consisting of a rhythm with a distinct tempo (120 or 140 bpm) and fundamental rhythmic pattern. Additional trials were performed representing silent, resting baseline; white noise; active tapping with the beat; and random sequences of beats using the same drum samples, respectively. For this preliminary investigation, only the data corresponding to blocks 1 and 6 (the fundamental rhythm) for the 120 bpm condition were analyzed. For this condition, a measure is 1 second in duration.

C. Data Collection

Data from the electrode grids or strips (Ad-Tech Medical Instrument Corporation, 1-cm spacing) were digitized at 1200 Hz using g.USB amplifiers (g.tec Medical Engineering, Austria). Data recording and stimulus presentation were controlled by BCI2000 [33].

D. Data Analysis

A high-pass filter with a cutoff frequency of 0.01 Hz was applied to the ECoG data to remove any low-frequency trends. For each participant, the data was re-referenced using a spatial common average reference to suppress activity that is common across the channels. To extract the broadband gamma power, each ECoG channel was bandpass filtered between 70 and 170 Hz using a 360-tap FIR bandpass filter with zero-phase filtering. An IIR elliptic notch filter was applied to attenuate the second harmonic of the 60-Hz line noise at 120 Hz. The Hilbert transform was then used to extract the instantaneous amplitude envelope. The resulting broadband power envelope was smoothed using a low-pass filter with a cutoff frequency of 8 Hz and zero-phase filtering.

The identical procedure was performed to compute the envelope of the acoustic waveform of the musical rhythm stimulus, except the Hilbert envelope was smoothed using a low-pass filter with a cutoff frequency of 6 Hz and zero-phase filtering based on the characteristics of the signal.

1) Model Development and Training: For each participant, two linear multiple regression models were developed and trained using Lasso regression [34]. The first is a causal, anticipatory model that uses the past 250 ms of ECoG data to reconstruct the current stimulus envelop (listened or imagined). The second is a non-causal, perceptual model that uses the future 250 ms of ECoG data to reconstruct the current stimulus envelop. The anticipatory model examines the potential to predict the rhythms in real-time for application to brain-computer interfaces, while the perceptual model highlights the effects of actual or imagined perceptual feedback on the ongoing brain activity.

For each participant, ECoG segments corresponding to the *listen* intervals of stimulus blocks 1 and 6 were used for training the model. Continuous ECoG data corresponding to the listen intervals were further segmented into three non-overlapping sets: (1) $\sim$18 seconds for training the model, (2) $\sim$4 seconds for optimizing the regularization parameter of the Lasso regression, and (3) $\sim$4 seconds for testing the model on listen data. Additionally, the model was also

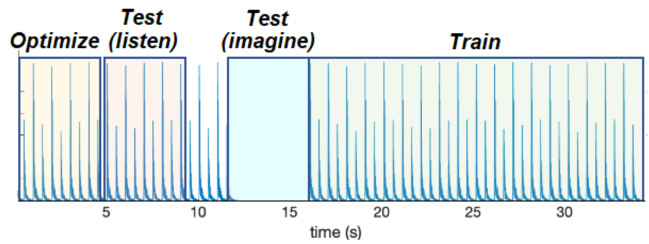

Fig. 2. Example of trial data segmentation with respect to the stimulus envelope for training, model optimization, and testing (listen and imagine).

tested on an equivalent duration of imagine data. An example segmentation of the stimulus data is shown in Figure 2.

For the anticipatory model, the instantaneous broadband gamma envelope for each ECoG channel was lagged at 0, 50, 100, 150, 200, and 250 ms prior to the current time point and used as inputs to the model to predict the current value of the acoustic stimulus envelope. This yielded 5 x # channels features for training the respective models. The perceptual model setup was identical except that the ECoG data at 0, 50, 100, 150, 200, 250 ms *after* to the current time point and used as inputs to the model to predict the current value of the acoustic stimulus envelope.

The regularization parameter of Lasso regression [34] was optimized for each model by iterating the parameter on the independent validation data, labeled as *'Optimize'* in Figure 2, and selecting the value that achieved the maximum Spearman correlation of the model output and rhythm envelope.

2) Model Evaluation: The models were applied to both the listen and imagine test segments corresponding to the same trial data used to train the respective model (see Figure 2). The resulting reconstructions were compared to the expected rhythm envelope using Spearman's rank correlation. The expected rhythm envelope for the imagine test data simply represents the original stimulus pattern before the silence intervals were inserted.

To evaluate the statistical significance of the resulting Spearman correlation coefficients, a randomization test was performed where the ECoG data from the silent, resting baseline trial for each participant were used as the input to the respective models and the Spearman's correlation coefficient was computed between the model output and the stimulus envelope, which does not have a natural temporal alignment with the baseline ECoG data. The baseline ECoG data was circularly shifted, temporally, by a random factor and the process was repeated 1000 times. The empirical distribution of the resulting correlation coefficients was used to compute the p-value of the correlation coefficients obtained during the actual listing and imagine test conditions.

III. Results

The Spearman correlation coefficients for the anticipatory and perceptual models for the listen condition are shown in Figure 3. Except for Participant 6, all perceptual models generated statistically significant correlations for the listen condition. This is also the case for the anticipatory models except for Participants 4 and 6. Examining the average across participants, the anticipatory models performed marginally lower than the perceptual models (p=0.1562, Wilcoxon signed rank test).

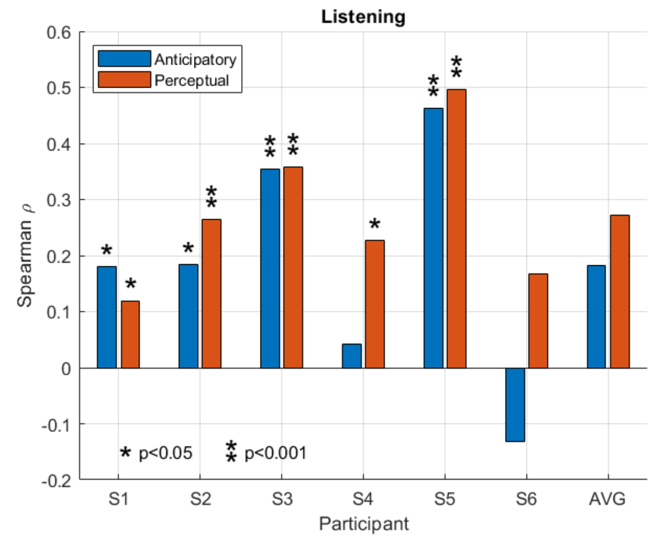

Fig. 3. Bar graph of the correlations for the anticipatory and perceptual models for the listen condition. The asterisks indicate the level of significance based on the randomization tests.

The Spearman correlation coefficients for the anticipatory and perceptual models for the imagine condition are shown in Figure 4. Only the perceptual model from Participant 4 was significant for the imagine condition.

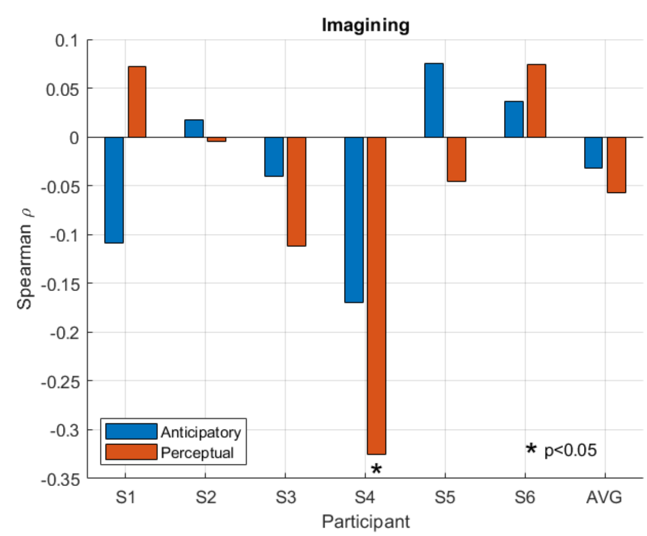

Fig. 4. Bar graph of the correlations for the anticipatory and perceptual models for the imagine condition. The asterisks indicate the level of significance based on the randomization tests.

Figure 5 shows an example reconstruction from Participant 5 for the perceptual model that generated the highest correlation for the listen condition. Figure 6 shows a topography of the model weights that associated with the reconstruction in Figure 5. As expected for such a perceptual model, the largest model weights are over the auditory cortex from 50 to 100 ms, with negligible feature contributions at 0 ms or beyond 200 ms. There is also some contribution from the posterior parietal area around 150 ms, which is consistent with prior findings examining perceived rhythm [35], and has been linked to encoding and retrieval of music/rhythmic patterns [30], [36].

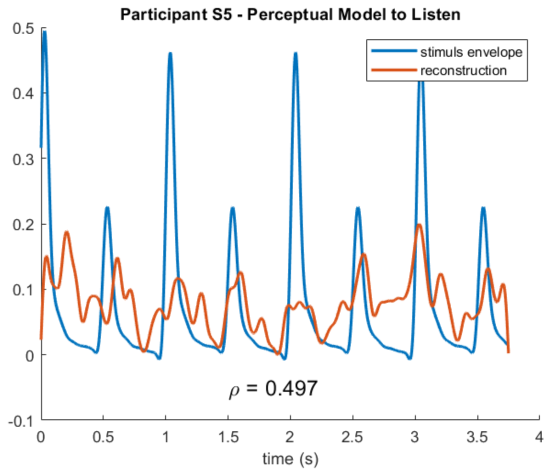

Fig. 5. An example reconstruction from Participant 5 for the perceptual model for the listen condition.

Figure 7 shows an example reconstruction from Participant 5 for the anticipatory model that generated the highest correlation for the listen condition. Figure 8 shows a topography of the model weights that associated with the reconstruction in Figure 7. For this model, the largest model weights are from -150 ms to -100 ms, also over the auditory cortex, similar to the perceptual model. There are negligible feature contributions at 0 ms or beyond -200 ms. There is also some contribution from a more anterior parietal area around -50 ms that may also span the region associated with encoding and retrieval of music/rhythmic patterns [30], [36].

IV. Discussion

As expected, the perceptual models applied to the listen data yielded the best overall performance. This performance is largely due to the fact that all participants had some electrode coverage of the superior temporal gyrus, often containing primary auditory cortex. This is exemplified by the selected model weights for this condition, e.g., Figure 6, with the relevant features being represented at time lags commensurate with typical auditory response times for ECoG [28]. The anticipatory models also exhibited comparatively high correlations for the listen condition, with feature weights primarily concentrated over the auditory cortex around 50-150 ms prior to the stimulus. Such a causal design is necessary in the pursuit of a real-time brain-computer interface application. It should be noted that one measure of the rhythm is 1000 ms in duration, with a beat occurring every 500 ms. Since the neural auditory response to a beat should resolve within 200 ms after a beat, and the anticipatory model weights are concentrated around 50-150 ms prior to the current model output, it is

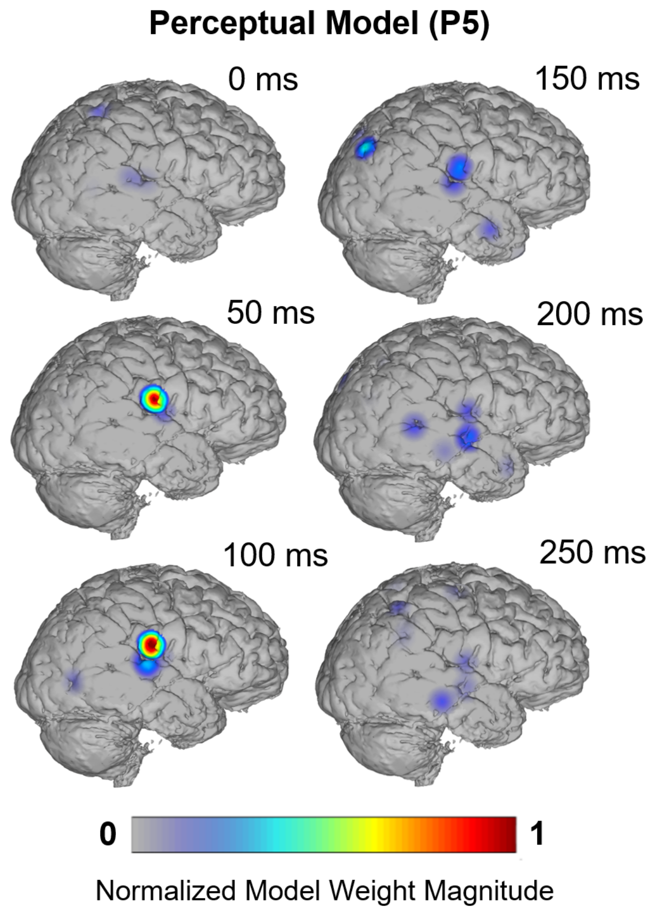

Fig. 6. Topographies of the normalized model weight magnitudes of the perceptual model for Participant 5.

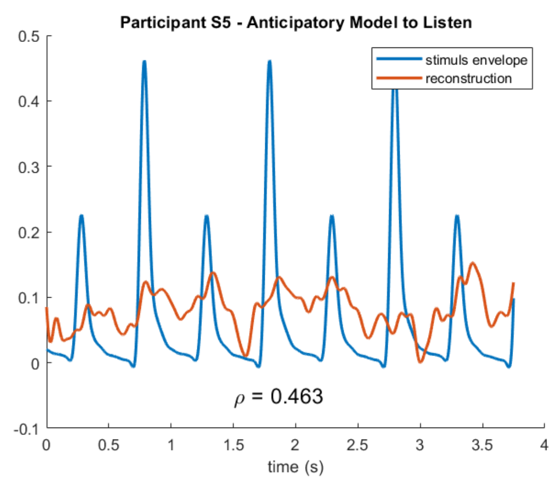

Fig. 7. An example reconstruction from Participant 5 for the anticipatory model for the listen condition.

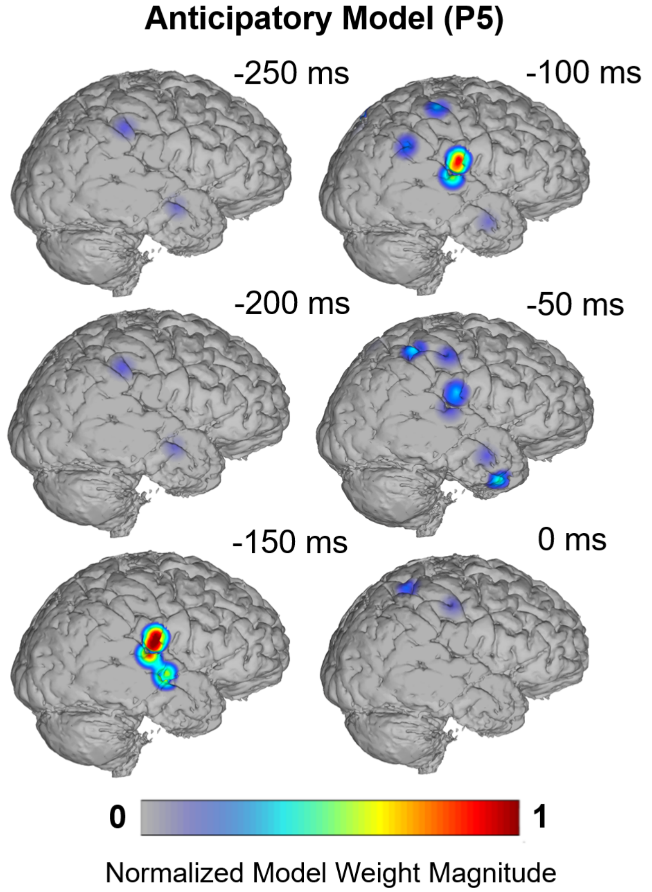

Fig. 8. Topographies of the normalized model weight magnitudes of the anticipatory model for Participant 5.

unlikely that the performance of the anticipatory model is exclusively attributed to perceptual transients. However, the bandpass filter used to isolate the broadband gamma is 300 ms in duration, which is effectively doubled with zero-phase filtering. Thus, it is possible that this temporal overlap could impact the interpretation of the results.

Only one significant correlation was observed for the imagined condition, which was obtained from the perceptual model for Participant 4. Curiously, this correlation is negative. One possible explanation is that the reconstruction may be out of phase with the correct stimulus envelope alignment. This could be due to the participant losing synchronization with the rhythm during the imagine condition. However, it is perplexing that this negative correlation (ρ = -0.325) is larger in magnitude than the participant's perceptual model in the listen condition (ρ = 0.227), where the neural synchronization should be naturally maintained from the audible stimulus. It is conceivable that this significant negative correlation is merely an aberration of the selected data since limited data were available to perform a more thorough cross validation.

The general failure of the models for the imagined condition is likely analogous to what has been observed in imagined speech decoding from intracranial signals, where models trained on overt speech tend to fail when evaluated on imagined speech [37]. However, the present protocol does not provide a sufficient duration of imagine data to conduct a more comprehensive characterization or to train a model on imagine for comparison to the models trained on listen.

V. Conclusion

This study demonstrates that it is possible to design models that can reconstruct simple perceived acoustic rhythm patterns from both causal, anticipatory and noncausal perceptual ECoG recordings made during listening of the rhythms. While the quality of the reconstructions varied across participants, the simple linear models were trained with a very limited amount of data from electrodes in suboptimal locations for this decoding application. It is expected that the reconstructions can be significantly improved with better electrode coverage, larger amounts of data, and more sophisticated decoding models. While both the anticipatory and perceptual models trained on listening data fail to reliably reconstruct the expected imagined patterns, for a proper assessment and comparison to the models trained with listening data, additional imagine data is required to train the models.

References

[1] S. Koelsch and W. A. Siebel, "Towards a neural basis of music perception," *Trends in Cognitive Sciences*, vol. 9, pp. 578–584, 2005.

[2] V. J. Schmithorst, "Separate cortical networks involved in music perception: preliminary functional mri evidence for modularity of music processing," *NeuroImage*, vol. 25, pp. 444–451, 2005.

[3] V. Alluri, P. Toiviainen, I. P. Jääskeläinen, E. Glerean, M. Sams, and E. Brattico, "Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm," *NeuroImage*, vol. 59, pp. 3677–3689, 2012.

[4] B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, and E. F. Chang, "Reconstructing speech from human auditory cortex," *PLoS Biology*, vol. 10, 2012.

[5] D. A. Moses, S. L. Metzger, J. R. Liu, G. K. Anumanchipalli, J. G. Makin, P. F. Sun, J. Chartier, M. E. Dougherty, P. M. Liu, G. M. Abrams, A. Tu-Chan, K. Ganguly, and E. F. Chang, "Neuroprosthesis for decoding speech in a paralyzed person with anarthria.," *The New England journal of medicine*, vol. 385 3, pp. 217–227, 2021.

[6] S. V. Norman-Haignere, J. Feather, D. Boebinger, P. Brunner, A. L. Ritaccio, J. H. McDermott, G. Schalk, and N. G. Kanwisher, "A neural population selective for song in human auditory cortex," *Current Biology*, vol. 32, pp. 1470–1484.e12, 2022.

[7] S. Martin, C. Mikutta, M. K. Leonard, D. Hungate, S. Koelsch, S. Shamma, E. F. Chang, J. d. R. Millán, R. T. Knight, and B. N. Pasley, "Neural Encoding of Auditory Features during Music Perception and Imagery," *Cerebral Cortex*, vol. 28, pp. 4222–4233, 10 2017.

[8] G. Schalk and E. C. Leuthardt, "Brain-computer interfaces using electrocorticographic signals," *IEEE Reviews in Biomedical Engineering*, vol. 4, pp. 140–154, 2011.

[9] T. Fujioka, L. J. Trainor, E. W. Large, and B. Ross, "Beta and gamma rhythms in human auditory cortex during musical beat processing," *Annals of the New York Academy of Sciences*, vol. 1169, 2009.

[10] J. A. Grahn and J. D. McAuley, "Neural bases of individual differences in beat perception," *NeuroImage*, vol. 47, pp. 1894–1903, 2009.

[11] S. Nozaradan, Y. Zerouali, I. Peretz, and A. Mouraux, "Capturing with eeg the neural entrainment and coupling underlying sensorimotor synchronization to the beat.," *Cerebral cortex*, vol. 25 3, pp. 736–47, 2015.

[12] S. Nozaradan, A. Mouraux, J. Jonas, S. Colnat-Coulbois, B. Rossion, and L. Maillard, "Intracerebral evidence of rhythm transform in the human auditory cortex," *Brain Structure and Function*, vol. 222, pp. 2389–2404, 2017.

[13] S. Nozaradan, P. E. Keller, B. Rossion, and A. Mouraux, "Eeg frequency-tagging and input–output comparison in rhythm perception," *Brain Topography*, vol. 31, pp. 153–160, 2018.

[14] T. Lenc, P. E. Keller, M. Varlet, and S. Nozaradan, "Neural tracking of the musical beat is enhanced by low-frequency sounds," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, pp. 8221 – 8226, 2018.

[15] T. Lenc, P. E. Keller, M. Varlet, and S. Nozaradan, "Neural and behavioral evidence for frequency-selective context effects in rhythm processing in humans," *Cerebral Cortex Communications*, vol. 1, 2020.

[16] G. M. D. Liberto, G. Marion, and S. A. Shamma, "The music of silence: Part ii: Music listening induces imagery responses," *The Journal of Neuroscience*, vol. 41, pp. 7449 – 7460, 2021.

[17] G. Marion, G. M. D. Liberto, and S. A. Shamma, "The music of silence: Part i: Responses to musical imagery encode melodic expectations and acoustics," *The Journal of Neuroscience*, vol. 41, pp. 7435 – 7448, 2021.

[18] G. Novembre and G. D. Iannetti, "Tagging the musical beat: Neural entrainment or event-related potentials?," *Proceedings of the National Academy of Sciences*, vol. 115, pp. E11002 – E11003, 2018.

[19] F. J. P. Langheim, J. H. Callicott, V. S. Mattay, J. H. Duyn, and D. R. Weinberger, "Cortical systems associated with covert music rehearsal," *NeuroImage*, vol. 16, pp. 901–908, 2002.

[20] N. Kraus, E. Skoe, A. Parbery-Clark, and R. D. Ashley, "Experience-induced malleability in neural encoding of pitch, timbre, and timing," *Annals of the New York Academy of Sciences*, vol. 1169, 2009.

[21] A. R. Halpern, R. J. Zatorre, M. A. Bouffard, and J. A. Johnson, "Behavioral and neural correlates of perceived and imagined musical timbre," *Neuropsychologia*, vol. 42, pp. 1281–1292, 2004.

[22] M. J. Henry, B. Herrmann, and J. A. Grahn, "What can we learn about beat perception by comparing brain signals and stimulus envelopes?," *PLoS ONE*, vol. 12, 2017.

[23] C. Herff, L. Diener, M. Angrick, E. M. Mugler, M. C. Tate, M. A. Goldrick, D. J. Krusienski, M. W. Slutzky, and T. Schultz, "Generating natural, intelligible speech from brain activity in motor, premotor, and inferior frontal cortices," *Frontiers in Neuroscience*, vol. 13, 2019.

[24] M. Angrick, C. Herff, E. M. Mugler, M. C. Tate, M. W. Slutzky, D. J. Krusienski, and T. Schultz, "Speech synthesis from ecog using densely connected 3d convolutional neural networks," *Journal of Neural Engineering*, vol. 16, 2018.

[25] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, pp. 493 – 498, 2018.

[26] S. A. Herff, C. Herff, A. J. Milne, G. D. Johnson, J. J. Shih, and D. J. Krusienski, "Prefrontal high gamma in ecog tags periodicity of musical rhythms in perception and imagination," *eNeuro*, vol. 7, 2020.

[27] C. Potes, A. Gunduz, P. Brunner, and G. Schalk, "Dynamics of electrocorticographic (ecog) activity in human temporal and frontal cortical areas during music listening," *NeuroImage*, vol. 61, pp. 841–848, 2012.

[28] C. Potes, P. Brunner, A. Gunduz, R. T. Knight, and G. Schalk, "Spatial and temporal relationships of electrocorticographic alpha and gamma activity during auditory processing," *NeuroImage*, vol. 97, pp. 188–195, 2014.

[29] I. Sturm, B. Blankertz, C. Potes, G. Schalk, and G. Curio, "Ecog high gamma activity reveals distinct cortical representations of lyrics passages, harmonic and timbre-related changes in a rock song," *Frontiers in Human Neuroscience*, vol. 8, 2014.

[30] Y. Ding, Y. Zhang, W. Zhou, Z. Ling, J. Huang, B. Hong, and X. Wang, "Neural correlates of music listening and recall in the human brain," *The Journal of Neuroscience*, vol. 39, pp. 8112 – 8123, 2019.

[31] L. Bellier, A. Llorens, D. Marciano, G. Schalk, P. Brunner, R. T. Knight, and B. N. Pasley, "Encoding and decoding analysis of music perception using intracranial eeg," *bioRxiv*, 2022.

[32] J. Kubanek and G. Schalk, "Neuralact: A tool to visualize electrocortical (ecog) activity on a three-dimensional model of the cortex," *Neuroinformatics*, vol. 13, pp. 167–174, 2015.

[33] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "Bci2000: a general-purpose brain-computer interface (bci) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1034–1043, 2004.

[34] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the royal statistical society series b-methodological*, vol. 58, pp. 267–288, 1996.

[35] M. H. Thaut, P. D. Trimarchi, and L. M. Parsons, "Human brain basis of musical rhythm perception: Common and distinct neural substrates for meter, tempo, and pattern," *Brain Sciences*, vol. 4, pp. 428 – 452, 2014.

[36] N. Konoike, Y. Kotozaki, H. Jeong, A. Miyazaki, K. Sakaki, T. Shinada, M. Sugiura, R. Kawashima, and K. Nakamura, "Temporal and motor representation of rhythm in fronto-parietal cortical areas: An fMRI study," *PLoS ONE*, vol. 10, 2015.

[37] P. Z. Soroush, C. Herff, S. K. Riès, J. J. Shih, T. Schultz, and D. J. Krusienski, "The nested hierarchy of overt, mouthed, and imagined speech activity evident in intracranial recordings," *NeuroImage*, vol. 269, 2022.