Encoding of Acoustic Features in Listened and Imagined MEG Responses to Melodies

Maryam Maghsoudi Shaghaghi (maryam00@umd.edu)
Department of Electrical and Computer Engineering, University of Maryland
College Park, MD 20742, United States

Mohsen Rezaeizadeh (mohsenr@terpmail.umd.edu)
Department of Electrical and Computer Engineering, University of Maryland
College Park, MD 20742, United States

Jonathon Z. Simon (jzsimon@umd.edu)
Department of Electrical and Computer Engineering, and Department of Biology, University of Maryland
College Park, MD 20742, United States

Shihab A. Shamma (sas@umd.edu)
Department of Electrical and Computer Engineering, University of Maryland
College Park, MD 20742, United States

Abstract

Musical imagery is the internal re-creation of music without external auditory input. While numerous studies have investigated the neural correlates of musical listening and imagery, fewer have explored the encoding of acoustic features during the latter. In this study, we employ Multivariate Temporal Response functions (mTRFs) to examine how melodic features, such as note onsets, envelope and envelope onsets, are encoded in magnetoencephalography (MEG) responses during both musical listening and imagery. Our analysis reveals that note onsets and envelope onsets significantly predict MEG responses in both listening and imagery conditions. Notably, correlations between these acoustic features and neural activity are evident at both group and individual levels. Furthermore, prediction correlation topographies show increased correlation values in channels located above the temporal lobe during both listening and imagery tasks.

Keywords: Musical imagery, auditory encoding, temporal response functions, magnetoencephalography

Introduction

Musical imagery is an endogenous voluntary act of hearing music in our mind without any auditory sensory input (Kosslyn et al., 2001). Understanding the underlying brain activity during musical imagery tasks can reveal the very basic levels of auditory processing in the brain. Also, studying the brain in imagery tasks holds many applications in Brain-Computer Interfaces (BCIs) for tasks such as speech recognition based on imagined speech.

Previous studies have shown that auditory imagery tasks share several brain regions with auditory listening tasks (Zatorre et al., 1996; Halpern & Zatorre, 1999; Halpern et al, 2004; Zhang et al., 2017), but less research has been done to study the temporal neural processing during auditory imagination. One challenge in investigating imagery tasks lies in precisely timing participants' onset of imagination. However, utilizing melodic stimuli with inherent temporal structures and engaging highly trained musicians adept at synchronizing with rhythmic stimuli can mitigate this issue. Marion et al. (2021) have used the same idea and have shown that melodic expectations are encoded in neural signals during listening and imagery.

In this study, we extend this line of inquiry to understand how the brain processes and encodes musical information. We use melodic stimuli and highly-trained musicians in a magnetoencephalography task, which is a high-temporal-resolution neuroimaging technique. We investigate how the acoustic features of the stimulus are encoded in imagery and listening responses.

Materials and Methods

Experimental Procedure

15 healthy and professional musicians (9 males) with self-reported normal hearing participated in the experiment. The experiment was approved by Institutional Review Boards of University of Maryland. Written, informed consent was obtained from participants prior to recording, and monetary compensation was given.

MEG data was recorded in whole head KIT (Kanazawa Institute of Technology) system, with 157 axial gradiometers at 1 kHz sampling rate with an online 500 Hz low pass filter, and a 60 Hz notch filter.

The experiment included two melodic stimuli from a MIDI corpus of Bach chorales. Each stimuli was repeated in 10 listening and 10 imagery trials in a randomized order, giving a total of 40 trials per recording. The participants were provided with the stimuli a few days before the recording session to practice and prepare for the imagery task. To make sure that the participants performed the imagery task with high temporal precision, a visual clock-shaped metronome with 120 bpm downbeat was presented on the screen. The metronome flashed every 2 seconds marking the start of each clock.

Preprocessing

Saturating and dead channels were removed. A zero-phase third-order Butterworth filter was used to filter the MEG responses from 0.1 Hz to 8 Hz. The Denoising Source Separation (DSS) technique was also used to denoise the responses by keeping its first seven components (Cheveigne & Simon, 2007). Data was z-scored and downsampled to 100 Hz. To remove the visual artifact caused by the flash onsets, notes closer than 500 ms from the flash onset were removed.

Encoding Analysis

Multivariate Temporal Response Functions (mTRF): We used the mTRF toolbox (Crosse et al., 2016) to linearly map the stimulus features into MEG responses.

$$r(t,n) = (s * \omega_n)(t) + \varepsilon(t,n) \tag{1}$$

$r(t,n)$ is the MEG response at time t at channel n, ω is the kernel we are looking to estimate, s is the stimulus feature, and $\varepsilon(t,n)$ is the estimation residual to be minimized. $\varepsilon(t,n)$ has been minimized using ordinary least square method along with a regularization parameter that was determined using leave-one-out cross-validation.

Three acoustic features of the stimuli were used to estimate the TRFs: 1) Note Onsets: A vector of zeros and ones where the ones occur at the onset of each note. 2) Envelope: Acoustic envelope of the stimulus. 3) Envelope Onsets: Half-wave rectified envelope's derivative across time.

Forward TRF models are trained on each channel independently. To evaluate the encoding of the above stimulus features in the responses, the MEG response at each channel is predicted using the trained TRF, and the linear correlation between the predicted MEG and the actual MEG is calculated. To assess the significance of these correlation values, the order of trials is shuffled so that there are no matching stimulus-response pairs during the training. These null model TRFs are then trained to minimize their prediction error with the shuffled MEG responses. A gain distribution is derived by subtracting the prediction correlation values of the null model from the TRF model. A control distribution is made by subtracting the null model correlation values from different repetitions of itself. The Wilcoxon sum-rank test is used to evaluate whether the correlation gain values are higher than the control distribution correlation values.

Results

The encoding of stimulus features in listening and imagery conditions is evaluated by calculating the TRF prediction correlation of the forward model and the null model. The null model is calculated across 20 repetitions of shuffling the order of trials. To create the control distribution the correlation values in each shuffling are subtracted from the other 19 repetitions. These gain and control distributions are calculated for each participant separately meaning that the forward model correlation values of each participant are only subtracted from its own null correlation values. The group-level histograms of correlation gain values in control and gain distributions are shown in Fig. 1 and 2, panels (a) and (d). Each point on the histogram indicates a participant and a trial.

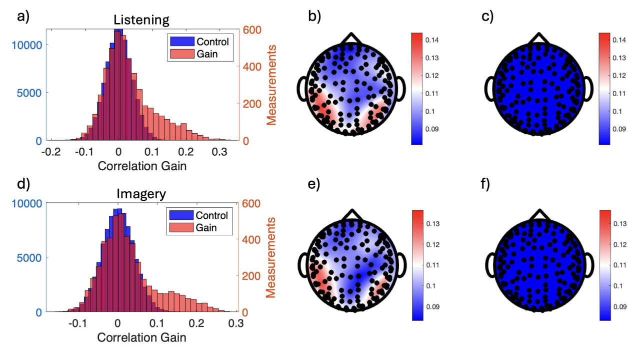

Figure 1: Melody onsets encoding during musical listening (top), and imagery (bottom). a,d) Correlation gain for control (blue) and gain (red) distributions during musical listening (a) and imagery (d). b,e) Topographies of correlation values of the onset model during musical listening (b) and imagery (e). c,f) Topographies of the null model correlation values during listening (c) and imagery (f).

The gain distribution is significantly above the control for two of melodic features, Onsets (Fig. 1) and Envelope Onsets (Fig. 2) with p-value $<<$ 0.001. This means that a linear function (TRF) convolved with the note onsets or the envelope onsets can predict the MEG response with accuracies above the chance level, and this holds for both listening and imagery conditions.

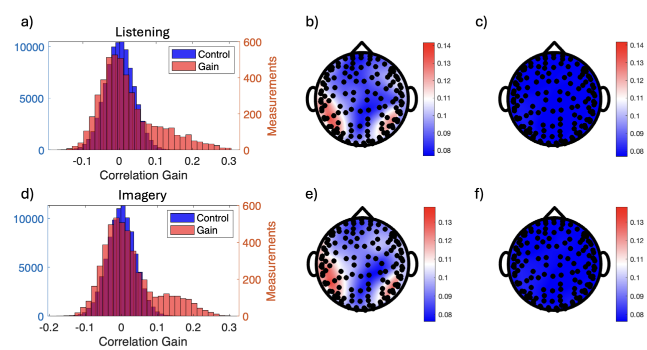

Figure 2: Melody envelope onsets encoding during musical listening (top), and imagery (bottom). a,d) Correlation gain for control (blue) and gain (red) distributions during musical listening (a) and imagery (d). b,e) Topographies of linear correlation values of the onset model during musical listening (b) and imagery (e). c,f) Topographies of the null model correlation values during listening (c) and imagery (f).

Panels (b) and (e) in figures 1 and 2 show that the prediction correlation topographies are higher above the temporal lobe. This means the channels located above the temporal lobe better encode the acoustic information of melodies. The correlation topography patterns are similar during listening (top panels) and imagination (bottom panels).

In addition to group-level observations, the prediction correlation values for each participant are shown in Fig. 3. At the single participant level, in both listening and imagery conditions, the TRF model prediction correlation values are above the ones for the null model (p-value $<<$ 0.005).

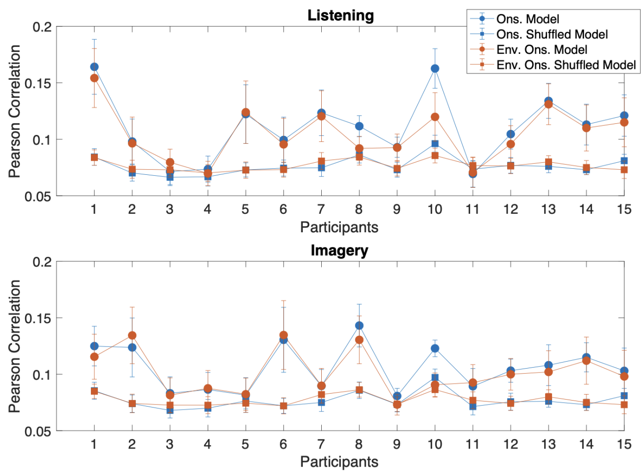

Figure 3: Onsets (Blue) and Envelope (Red) Encoding Across Participants. The x-axis shows the participants and the y-axis shows the prediction correlation value averaged over the channels and trials. Circle-shaped and square-shaped markers relate to the forward model and the null model respectively.

Acknowledgments

This work was supported by the NIH grant R01-DC019394, NSF grant SMA 1734892, NIH R01 DC005779 on Auditory Plasticity, grant from AFOSR on Decoding of Imagined Speech and Music, and NSF grant on Implicit Learning of Music.

References

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Frontiers in human neuroscience, 10, 604.

de Cheveigné, A., & Simon, J. Z. (2008). Denoising based on spatial filtering. Journal of neuroscience methods, 171(2), 331-339.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. Cerebral cortex, 9(7), 697-704.

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia, 42(9), 1281-1292.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. Nature reviews neuroscience, 2(9), 635-642.

Marion, G., Di Liberto, G. M., & Shamma, S. A. (2021). The music of silence: part I: responses to musical imagery encode melodic expectations and acoustics. Journal of Neuroscience, 41(35), 7435-7448.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: a PET investigation of musical imagery and perception. Journal of cognitive neuroscience, 8(1), 29-46.

Zhang, Y., Chen, G., Wen, H., Lu, K. H., & Liu, Z. (2017). Musical imagery involves Wernicke's area in bilateral and anticorrelated network interactions in musicians. Scientific reports, 7(1), 17066.