Tracking Attention to Multi-Stream Music with the Auditory Steady State Response

Benjamin V. Shapiro

Western University

250 782 817

Author Note

Benjamin V. Shapiro, Hsc Neuroscience, Schulich School of Medicine and Dentistry

Acknowledgements go to supervisors Christina Vanden Bosch der Nederlanden, Avital Sternin, and Jessica Grahn, to the participants, and also to the Grahn Lab for their advice and equipment. This research was supported by NSERC.

Abstract

There is an active debate in the field of auditory neuroscience over whether the processing of music and language share similar neural resources or if they rely on distinct modular processes. A potential avenue to resolve this debate is an investigation into of the processing of vocal music, which contains both musical and linguistic information. Previous work has found that attention plays a fundamental role in mediating the processing of auditory information, so in order to begin this investigation, it was important to first develop a method of tracking attention to musical stimuli. Thus, the present study used electroencephalography (EEG) to attempt to track participants’ attention to either instrumental or vocal streams of music. This was done by modulating the amplitude of the streams at 30 Hz and 40 Hz respectively and recording the auditory steady state response (ASSR) to the stimuli. This study attempted to discover if the relative frequencies of the ASSR could be used to track attention to instrumental or lyrical streams and if attention to these streams could be voluntarily controlled. It was hypothesized that attention to the musical streams would be able to be controlled by the participant and tracked by the ASSR. However, the results did not find any ASSR and no differences in ASSR frequency power between any of the attention conditions (undirected, instrumental-directed, and vocal-directed). This suggests that the ASSR does not work for either the type of complex stimuli used or the type specific experimental design used.

Tracking Attention to Multi-Stream Music with the Auditory Steady State Response

The mechanism behind the human capacity to perceive and extract meaning from music has been of great interest to researchers in the cognitive sciences over the past few decades. Much of the research in this area has looked into how music processing and language processing interact. This enterprise has produced some debate within the field, and thus two opposing conceptions of the mechanism have emerged. One side promotes the view that language and music utilize shared neural resources and brain networks from their interpretation of the results of some imaging and behavioural studies (Koelsch & Friederici, 2003). The other side turns to neuropsychological case studies which cause them to endorse a conception that involves separate neural substrates for language and music that are specific to each domain (Peretz & Coltheart, 2003). This dissonance between the two viewpoints could be a result of either flawed methodology or incomplete theories. It is very likely that both of these stances are partly correct and are not quite as mutually exclusive as they seem. One of the biggest barriers to making clear conclusions in this space is the fact that musical and linguistic stimuli differ greatly in their context, secondary variability (accent, style, pragmatics), and sonic properties (Gordon et al., 2010).

To address this problem, stimuli which have both musical and linguistic attributes are useful. Luckily, this type of stimulus exists as song lyrics, and so the perception of human vocals in music should be studied. In vocal music, one acoustic signal carries both linguistic and musical information. It makes sense to begin this line of inquiry by focussing on a construct at the top level of perception - attention. Understanding the nuances of how attention can and cannot influence the perception of these auditory stimuli will create a foundation for further research into the mechanism behind the processing of them. In order to study the how attention interacts with auditory perception, however, it is first necessary to develop a method of tracking attention to auditory streams (especially with regards to lyrical stimuli). For this reason, this paper will attempt to investigate the role of attention in the perception of instrumental and lyrical streams during the presentation of multi-stream music.

The human brain has the remarkable capacity to segregate between competing auditory streams and bind auditory information within a stream (Bregman, 1990). With regards to speech, the ability to apply one’s attention to a single voice among a set of voices and background noise has been dubbed the “cocktail party effect” (Heine & Guski, 1991). It has been established that this process takes place on multiple levels in the brain, with a significant top-down influence of attention on what streams are segregated or bound and which auditory information is ignored (Carlyon, Cusack, Foxton, & Robertson, 2001). This capacity has been studied extensively, but first was looked at in the context of competing tones and later moved towards competing speech streams (Anstis, Saida, 1985; Arons, 1992; Darwin & Carlyon, 1995). Much of this research has its roots in research done in Gestalt psychology at the onset of the twentieth century on the organizational principles of the visual domain, but has been adapted for audition (Handel, 1989). While the visual system uses attributes like similarity, proximity, continuity, and closure to segregate and bind objects in the world, the auditory system seems to use attributes like frequency, timbre, spatial position, and rhythm to do this same task (Bregman, 1990). While the majority of the work on this has been done with speech or simple tones, music perception very likely works in a similar way, as music contains all of these same sonic characteristics. Research on tracking auditory stream segregation with lyrical and instrumental streams has not yet been done. The present study attempted to track an individual’s attention to lyrical and instrumental streams after segregation.

Electroencephalography (EEG) has proved to be a useful technique in the area of sensory neuroscience. EEG involves placing electrodes on the surface of the scalp in order to measure the combined activity of hundreds of thousands of neurons firing synchronously (Gevins et al, 1995). It has been used extensively in research on the “cocktail party effect”, where it has been found that it is possible to decipher from the EEG signal which speech stream a person is attending to. It gains its utility in this space by providing high temporal resolution and access to the cortical areas where much of speech processing faculties are housed. By extracting the temporal amplitude envelope of competing speech streams, Rimmele et al. (2015) used EEG to track the phase-locking of the auditory cortex to low frequency information in the amplitude envelope of the speech. They found that in unattended speech, the precise structure of the audio information was not processed by the auditory cortex to the degree of attending speech. Their results suggested that attention may control processing of the fine structural information of speech. From this, a question arises based on whether the attentional control of the fine processing of auditory information will influence that perception of music as much as it influences the perception of speech. Music perception appears to rely less heavily on the precise low frequency information, so it might be easier to perceive musical stimuli without directly attending to it.

While the method of amplitude envelope tracking seems promising for monitoring the processing of auditory stimuli through the auditory system, it may not prove as useful for music as it does for speech due to the differences in their fine temporal properties. Other EEG methods of studying auditory processing in the brain have been used, such as measuring the amplitude and latency of event-related potentials to transient deviant auditory stimuli (Treder et al., 2014), or measuring the continuous auditory steady state response (ASSR) to amplitude modulated stimuli (Skosnik, Krishnan, & O’Donnell, 2007), both of which then use that information to make inferences about perception. Due to the continuous nature of musical stimuli, the ASSR method appears to be the most appropriate for a first step investigation like the present study.

The ASSR is an EEG phenomenon which is evoked in response to periodic auditory stimuli. It is the result of phase and frequency synchronization in the temporal cortex to the stimuli. This entrainment has been shown to be significantly influenced by attention, with findings showing attended ASSR’s having much larger amplitudes than unattended ones (Naatanen, Gaillard, & Mantysalo, 1978). Stimuli can be modified to produce an ASSR by modulating the amplitude of the sound, with the optimal ASSR being elicited by stimuli with their amplitudes modulated at frequencies around 40 Hz (Skosnik, Krishnan, & O’Donnell, 2007).

The ASSR has been elicited before in response to various types of stimuli, including music, but has not previously been looked at with regards to lyrics. Accordingly, it has not previously been used to discriminate a listeners’ attention to various musical streams. Bigand, McAdams, and Foret (2010) looked at listeners’ ability to attend to multiple melodic lines at once, but they did not investigate this in the context of lyrics. Based on their results, they proposed an “integration model” of musical attention, positing that separate melodic lines are integrated into one by the auditory system (as opposed to switching attention or having one line in the foreground and one in the background). They used an error-detection behavioural task to try to characterize attention, but the present study endeavored to use neuroimaging for a more direct representation.

The present study endeavored to find out if the ASSR could be used to track attention to competing musical streams (specifically to lyrical and instrumental streams). Additionally, it was of interest to the study if attention to competing musical streams could be voluntarily controlled. It was predicted that firstly, attention to musical streams would be able to be tracked using the ASSR and secondly, changes in attention would be represented in the EEG data as relative power differences in the amplitude-modulated frequencies. If correct, this prediction would imply that the auditory system segregates vocal and instrumental streams and that attention to these streams can be voluntarily controlled and measured.

# Methods

## Participants

3 participants volunteered to participate in the study (2 female), ranging in age from 18 to 22, with a mean age of 20.33 (*SD*=2.08). All participants reported normal hearing and none had any known neurological abnormalities. All of the participants were right-handed. Musical experience was not a controlled factor; 2 participants reported having formal music training, however none others reported currently playing an instrument regularly (more than once a week). All participants were predominantly exposed to Western tonal music. All participants gave consent to participate and all study forms were approved by Western University’s Research Ethics Board (REB).

## Stimuli

Stimuli consisted of 46 thirty-second fragments taken from seven popular recordings (see Table 1). These songs were selected on the basis of comparable length, relative popularity, and sufficient rhythmic and melodic differences between the vocal and instrumental streams. In order to index what stream participants were attending to at a given time, the amplitude of each of the vocal tracks for each song fragment was modulated at 40 Hz and the instrumental tracks had their amplitudes modulated at 30 Hz. This was done to elicit simultaneous differential steady state auditory responses for each stream. The audio files were filtered and segmented using the Logic Pro X software program (https://www.apple.com/ca/logic-pro/) and the amplitude was modulated using the NumPy package in Python (http://www.numpy.org/). All stimuli were equated both within tracks and between songs for root-mean-square (RMS) amplitude (a good approximation for perceived loudness). The instructions and fixation cross were displayed on a CRT computer monitor and stimuli were played through Sennheiser HD280 professional headphones set at 60 dB. Stimuli numbers 2, 3, and 4 had relatively consistent instrumental motifs throughout the songs, whereas stimuli numbers 1, 5, 6, and 7 were more musically variable with regards to the instrumentation. Though all of the stimuli utilized a common pop-rock arrangement (guitar, drums, bass guitar, vocals), S1, S2, S3, S4, S6, and S7 had drums featuring prominently, whereas S5 had sparse drums which functioned in the background as more of a supplement to the other instrumentation. The stimuli varied on a number of different facets, but these differences were in deliberately selected in an effort to make the listening experience as naturalistic as possible. All fragments were in the time signature of 4/4. The stimuli were not repeated between the directed listening conditions (each of which contained 23 randomized fragments), but the undirected listening condition contained 23 randomized fragments that drew from both of the directed conditions. The musical fragments were presented in randomized sequences with a short break in between each. The total duration of the experiment amounted to about 45 minutes (excluding cap fitting time).

## Procedure

Participants first were played short segments of each of the seven songs and asked to indicate their prior familiarity with each song. Then, they filled out a demographic questionnaire. Once they finished the survey, participants were fitted with the EEG system and underwent a signal quality check to ensure that the clean data was being collected. Stimuli were then presented using a custom MatLab script using Psychtoolbox (Kleiner et al., 2007). They were instructed to sit as still as possible and listen to 23 song fragments with no other instructions. This was the undirected listening block. Following the undirected listening block, they were instructed to attend to the instruments while listening to 23 randomly chosen song fragments for the instrumental-directed listening block. After that, they were instructed to attend to the vocals while listening to the remaining 23 song fragments for the vocal-directed listening block.

## EEG Recording and Data Preprocessing

The data were recorded using a BioSemi ActiveTwo system (BioSemi Inc., Amsterdam, The Netherlands) with 32 amplifier-embedded Ag/AgCl electrodes arranged according to the 10–20 system (Jasper, 1958), with the addition of two mastoid electrodes and four facial electrodes (horizontal and vertical EOG). The data from the face electrodes were used to control for facial and eye movement artifacts in preprocessing. The offsets of the active electrodes were kept below 20 uV at the start of the measurement and EEG was sampled at 1024 Hz.

The analyses were performed in MATLAB (Mathworks, Nantick USA) using EEGLab and ERPLab packages (Delorme and Makeig, 2004; Markley et al., 2011). After importing the data, it was bandpass filtered from 2 to 55 Hz to remove line and muscle noise. The EEG data was segmented into -4 to 30 s epochs of each condition (undirected listening, instrumental-directed attention, vocal-directed attention), and baseline correction with a pre-stimulus baseline of -4 to 0 s was applied. Eye movements, muscle tension, and other artifacts were corrected using Independent Component Analysis.

**Results**

EEG signal power was assessed using a Fast Fourier Transform (FFT) to compute a power spectrum (in μV2) for each segment of EEG data. Log10 power was used for statistical analyses of spectral EEG data. A one-way ANOVA was performed on the pooled epochs from all the participants between the three conditions at the electrode where the attention-controlled ASSR has been previously found, F4 (Carlyon et al., 2001; Reyes et al., 2004). The tests were two-tailed and *p*<0.05 was used as a criterion. A comparison of the mean power spectra for each frequency from the pooled epochs from all of the participants at electrode location F4 is shown in Fig. 1. ANOVA revealed no statistically significant differences in power between the groups at 30 Hz (F(2,67) = 1.65, *p* = .20) and at 40 Hz (F(2,67) = 1.40, *p* = .25). Fig. 2. shows the spectral power for all channels.

**Discussion**

The results of the present study provide interesting data for understanding the nuances of the ASSR. There are a few reasons why an ASSR may not have been elicited by the amplitude-modulated musical stimuli that were presented to participants in the study.

Lack of novelty could be one of the possible reasons that an ASSR was not elicited in the present study. It has been suggested previously that novelty may me more important than attention for the ASSR. Linden, Picton, Hamel, and Campbell (1987) investigated attentional modulation of the ASSR in an active oddball paradigm where target stimuli that had their amplitudes modulated at one frequency were to be discriminated from standard stimuli that had their amplitudes modulated at another frequency. When the targets were 40 Hz and the standards were 20 Hz, the amplitude of the ASSR was enhanced for targets. Conversely, though, when the frequencies were reversed, no ASSR was found. These results elucidate the complex nature of the ASSR. The present study utilized only what would be deemed “standard” stimuli in the paper by Linden and colleagues, so their results may be related to the findings of this paper.

Another potential factor that could have contributed to the lack of an ASSR is the fact that both instrumental and lyrical stimuli are highly complex sonic signals. One study found that the signal-to-noise ratio of an amplitude modulated stimulus (i.e. the amplitude modulated frequency to all other frequencies) correlated with the strength of the elicited ASSR in the EEG signal (Roß, Borgmann, & Draganova, 2000). The large collection of various frequencies in the different audio clips that were used could have masked the ASSR or cancelled out the effects of it. Bohórquez and Özdamar (2008) found that the ASSR arises as a result of the superposition of overlapping auditory brainstem and middle latency responses. The maximal response at 40 Hz was posited to be a product of the superposition of the resonating Pb and Pa waves of the middle latency response. To many interacting harmonic frequencies may cause a disruption in the alignment of the frequencies of these responses.

Methodological factors could have also played a part in the findings or the present study. The participant number was relatively low due to contextual factors like time constraints and technical difficulties. There is some individual variability with regards to the strength of the ASSR and it has even been suggested that some individuals do not exhibit an ASSR whatsoever in response to amplitude-modulated stimuli (Rass et al., 2010). Owing to this variability and the low sample size of the present study, it is hard to discern if the participants happened to have low receptiveness to the ASSR. Also, the volume at which the stimuli were presented had to be limited out of courtesy for the patrons of the laboratory rooms surrounding the testing room. It has been suggested that for some types of stimuli and some individuals, the ASSR responds best at high volumes (Yeung & Wong, 2007). Therefore, the volume of the presentation could have been too quiet to elicit a response large enough to be found in the EEG signal.

In order to properly interpret the meaning of the results found in the present study, further research programs should aim to investigate the nuances of the ASSR with regards to musical stimuli, especially ones that contain vocals and lyrics. One potential avenue of research is to test the ASSR using a spectrum of auditory stimuli ranging from sinusoidal tones to complex music in order to investigate how it operates at different spots along the spectrum. This same concept could be applied with speech stimuli on a spectrum from spoken to sung. Once these nuances are teased out, the present experiment could be redesigned with them in mind.

Another research project that could further this line of inquiry is to attempt to track attention using the EEG analysis method known as amplitude envelope tracking. This method has been used in the study of the cocktail party effect to decode attentional selection from EEG (Sullivan et al., 2015). It involves aligning the phase of the low frequency information in the audio signal with the time-domain EEG signal. Since it has been shown to be useful for speech stimuli, this method is another potential avenue for tracking attention to musical stimuli.

It is recommended that future researchers continue to develop a method of tracking attention to auditory stimuli (and more specifically, vocal music). Once a viable method of tracking attention is developed, the mechanism and nuance by which attention is controlled and divided when listening to this music can begin to be uncovered.

References

Anstis, S. M., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human*. Retrieved from http://psycnet.apa.org/journals/xhp/11/3/257/  
  
Arons, B. (1992). A review of the cocktail party effect. *Journal of the American Voice I/O Society*, *12*(7), 35–50.  
  
Bohórquez, J., & Ozdamar, O. (2008). Generation of the 40-Hz auditory steady-state response (ASSR) explained using convolution. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *119*(11), 2598–2607.  
  
Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.  
  
Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology. Human Perception and Performance*, *27*(1), 115–127.

Darwin, C.J. and Carlyon, R.P. (1995). Auditory Grouping. In The.. (n.d.). Retrieved from http://www.utdallas.edu/~assmann/hcs6367/darwin\_carlyon95.pdf

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.  
  
Gevins, A., Leong, H., Smith, M. E., Le, J., & Du, R. (1995). Mapping cognitive brain function with modern high-resolution electroencephalography. *Trends in Neurosciences*, *18*(10), 429–436.  
  
Gordon, R. L., Schön, D., Magne, C., Astésano, C., & Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PloS One*, *5*(3), e9889.  
  
Handel, S. (1989). Listening. *An Introduction to the Perception of Auditory Events, Cambridge, MA*.  
  
Heine, W.-D., & Guski, R. (1991). Listening: the perception of auditory events? *Ecological Psychology: A Publication of the International Society for Ecological Psychology*, *3*(3), 263–275.  
  
Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., & Others. (2007). What’s new in Psychtoolbox-3. *Perception*, *36*(14), 1.  
  
Koelsch, S., & Friederici, A. D. (2003). Toward the neural basis of processing structure in music. Comparative results of different neurophysiological investigation methods. *Annals of the New York Academy of Sciences*, *999*, 15–28.  
  
Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.  
  
Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329.  
  
O’sullivan, J. A., Power, A. J., & Mesgarani, N. (2014). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral*. Retrieved from https://academic.oup.com/cercor/article-abstract/25/7/1697/457492  
  
Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691.  
  
Rass, O., Krishnan, G., Brenner, C. A., Hetrick, W. P., Merrill, C. C., Shekhar, A., & O’Donnell, B. F. (2010). Auditory steady state response in bipolar disorder: relation to clinical state, cognitive performance, medication status, and substance disorders. *Bipolar Disorders*, *12*(8), 793–803.  
  
Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi’s of everyday life: the structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*(6), 1236–1256.  
  
Reyes, S. A., Salvi, R. J., Burkard, R. F., Coad, M. L., Wack, D. S., Galantowicz, P. J., & Lockwood, A. H. (2004). PET imaging of the 40 Hz auditory steady state response. *Hearing Research*, *194*(1-2), 73–80.  
  
Rimmele, J. M., Zion Golumbic, E., Schröger, E., & Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *68*, 144–154.  
  
Roß, B., Borgmann, C., Draganova, R., Roberts, L. E., & Pantev, C. (2000). A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *The Journal of the Acoustical Society of America*, *108*(2), 679–691.  
  
Skosnik, P. D., Krishnan, G. P., & O’Donnell, B. F. (2007). The effect of selective attention on the gamma-band auditory steady-state response. *Neuroscience Letters*, *420*(3), 223–228.  
  
Yeung, K. N. K., & Wong, L. L. N. (2007). Prediction of hearing thresholds: Comparison of cortical evoked response audiometry and auditory steady state response audiometry techniques: Predicción de umbrales auditivos. Comparación entre las técnicas de Audiometría por Respuestas Evocadas Corticales y Audiometría por Respuestas Auditivas de Estado Estable. *International Journal of Audiology*, *46*(1), 17–25.

Tables

Table 1

Stimuli Sources

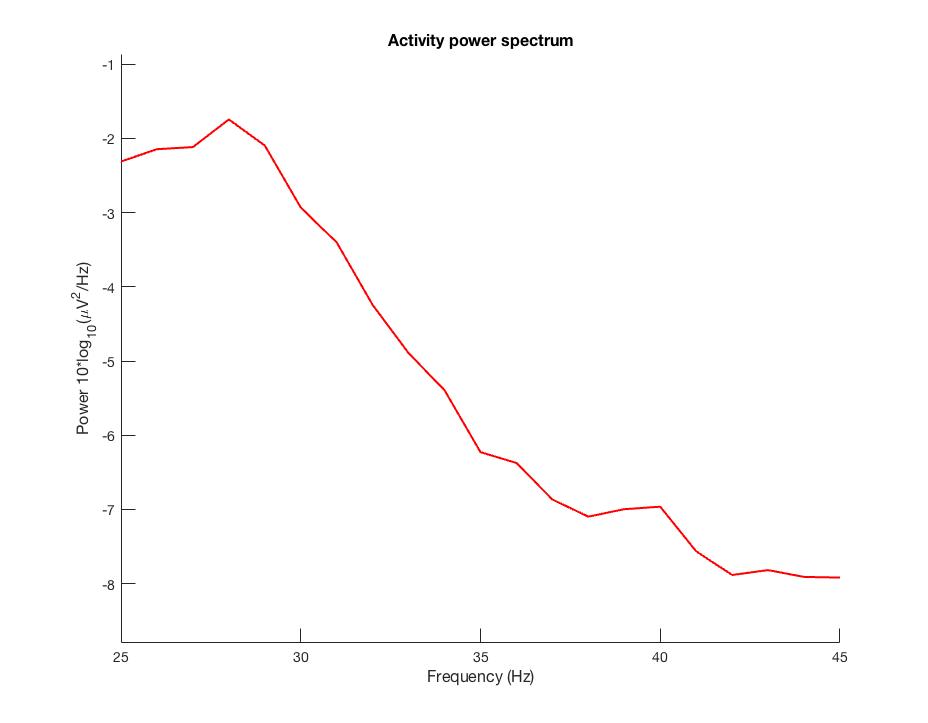
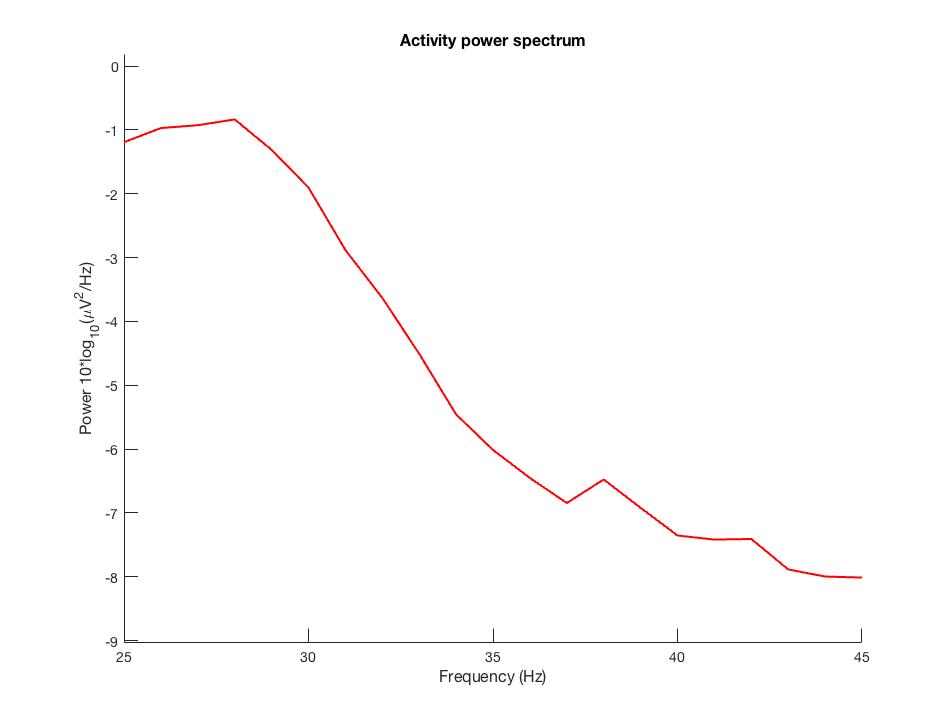
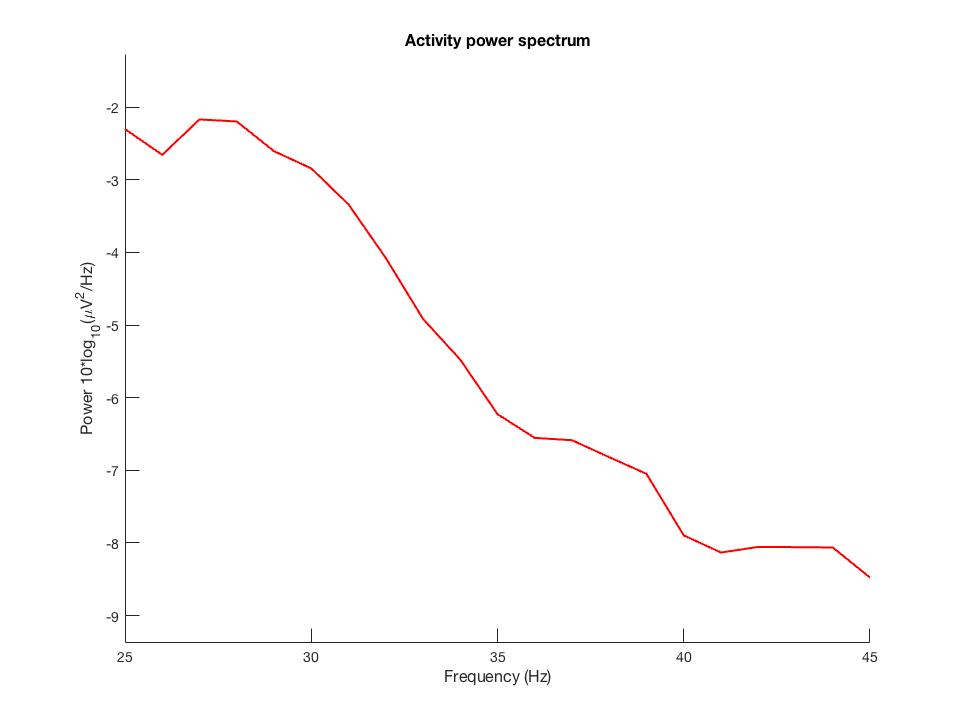
|  |  |  |
| --- | --- | --- |
| Stimulus | Title | Performer |
| S1 | Arabella | Arctic Monkeys |
| S2 | Shape of You | Ed Sheeran |
| S3 | You Make Loving Fun | Fleetwood Mac |
| S4 | Pumped Up Kicks | Foster The People |
| S5 | I’m Yours | Jason Mraz |
| S6 | Uptown Funk | Mark Ronson |
| S7 | Mr. Brightside | The Killers |

Note: Recordings from which the thirty-second fragments that were used as stimuli were made.

Figures

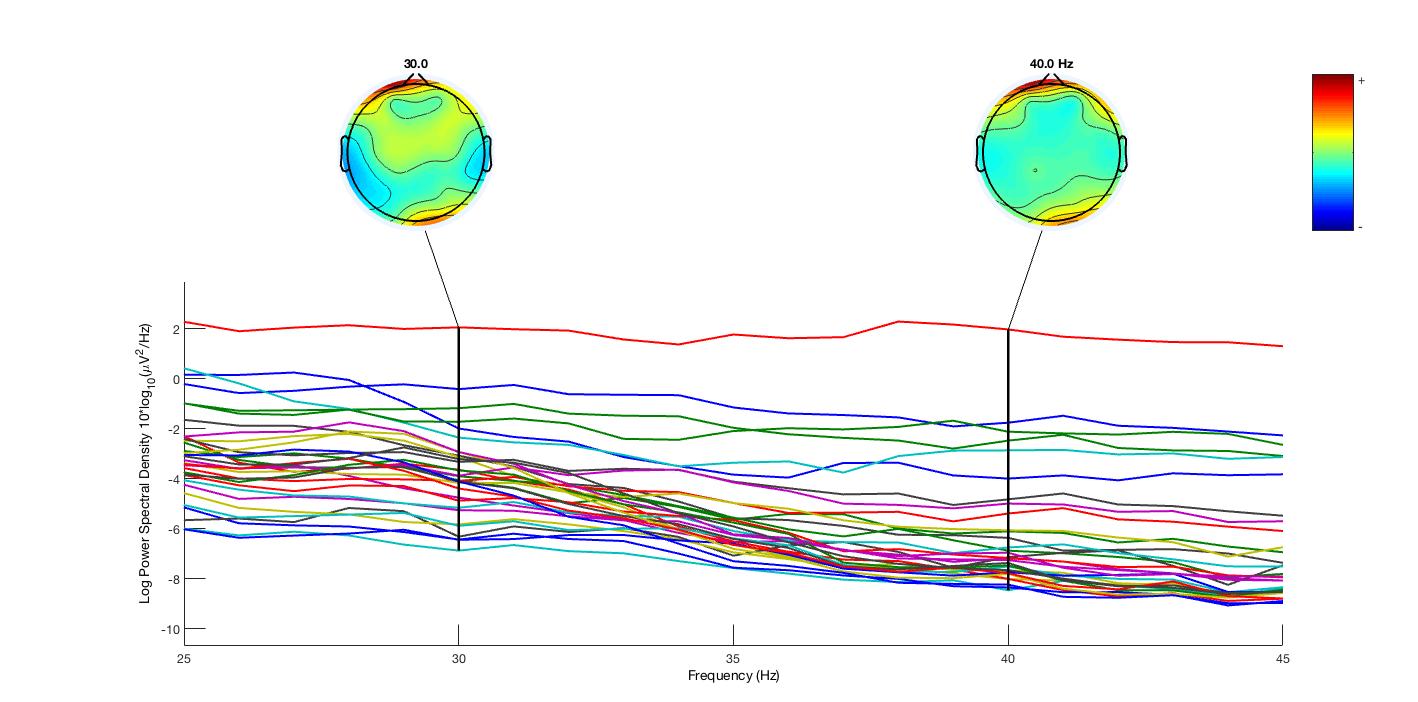
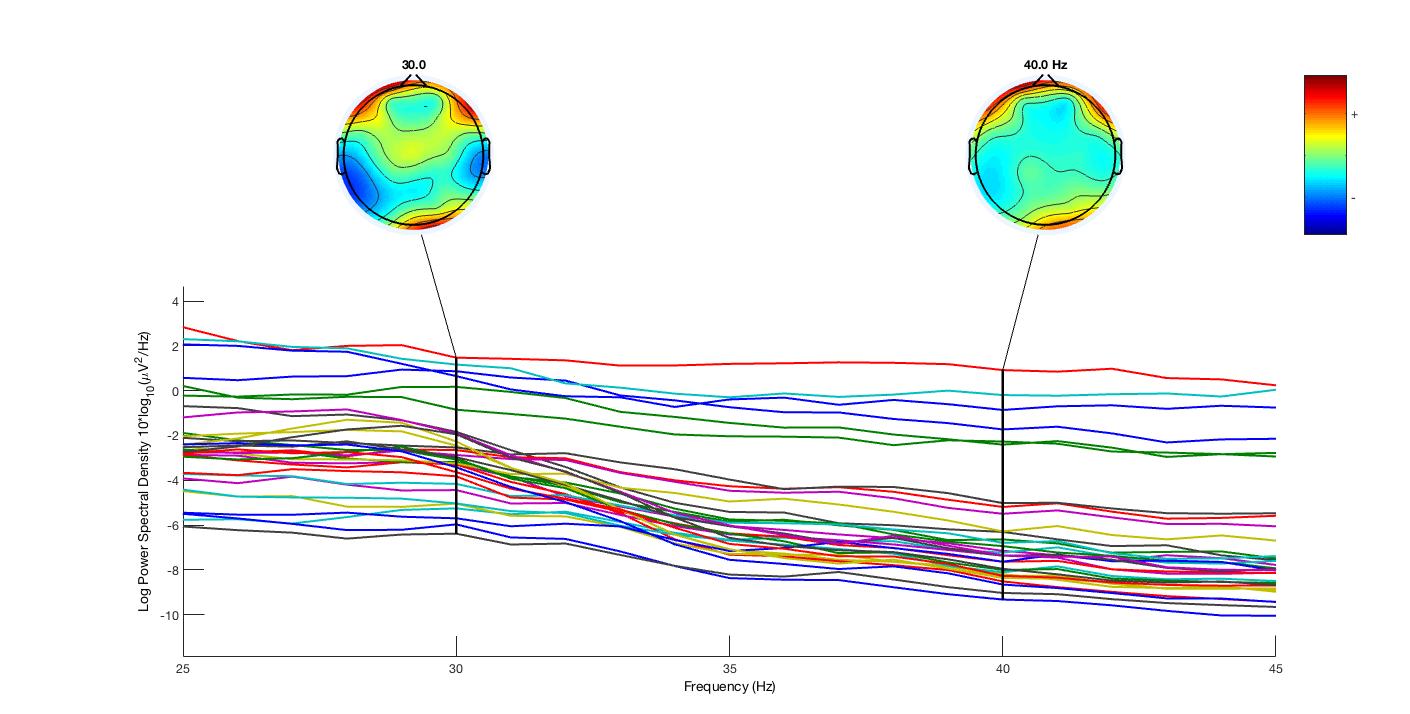
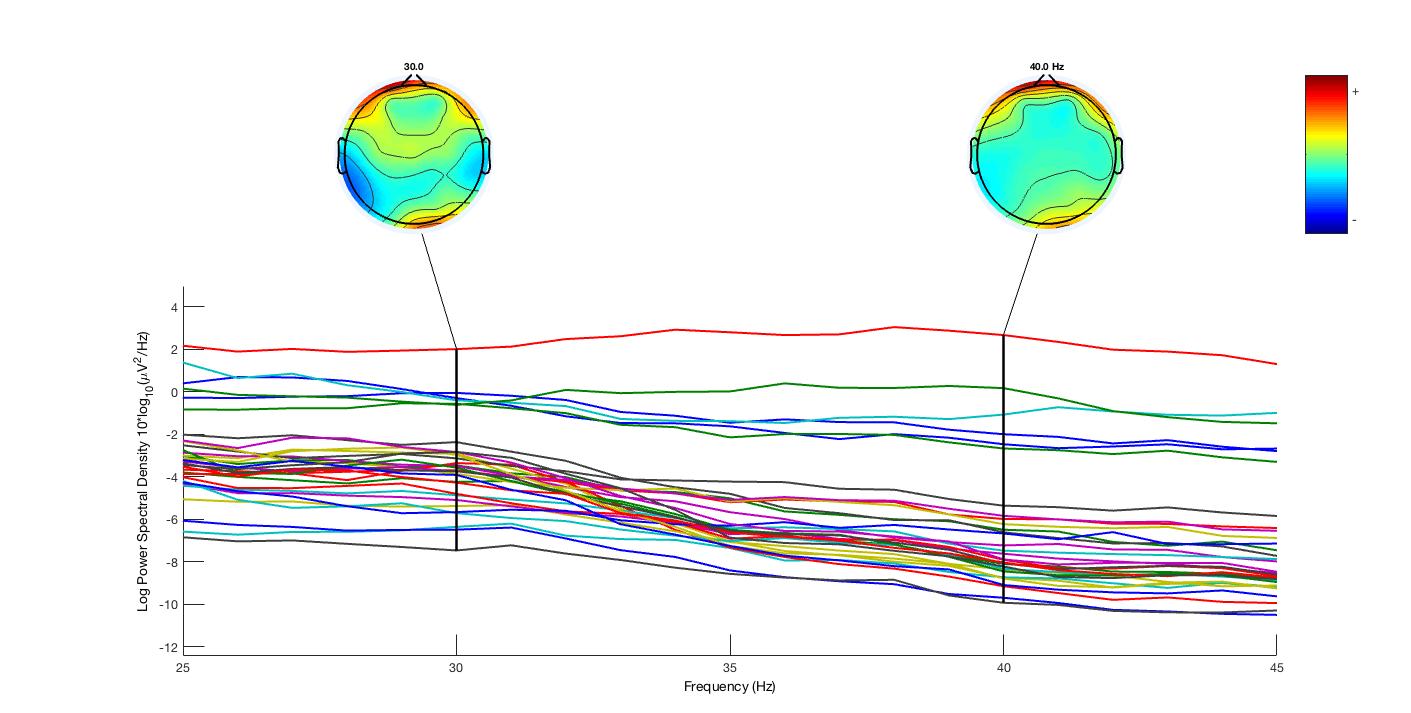
b)

a)



c)

Figure 1. Grand averaged power spectra for (a) undirected, (b) instrumental-directed, and (c) vocal-directed conditions at electrode site F4. Thirty hertz spectral power was largest for the instrumental-directed condition compared to the others and forty hertz spectral power was largest for the vocal-direction condition compared to the others.



b)

c)

a)

Figure 2. Grand averaged power spectra for (a) undirected, (b) instrumental-directed, and (c) vocal-directed conditions for each electrode channel and topographic plots at thirty hertz and forty hertz. Frequency (Hz) is on the x-axis and spectral power is on the y-axis. No visible ASSR spike is evident at either of the two frequencies in any of the three conditions.