Classifying imagined speech and rhythms using correlated component timecourses

Avital Sternin1,2, Bobby Stojanoski1, Adrian M. Owen1,2, Jessica A. Grahn1,2

1. Brain and Mind Institute
2. University of Western Ontario, Psychology Department

Western Interdisciplinary Research Building

London, Ontario, Canada N6A5B7

519 661 2111 x86057

[avital.sternin@uwo.ca](mailto:avital.sternin@uwo.ca)

Abstract

(up to 200 words)

Keywords: (up to 5 in alphabetical order)

Imagining auditory stimuli is a common occurrence as people will imagine stimuli like music or speech in the absence of an external stimulus. The relationship between the perception and imagination of auditory stimuli has previously been investigated showing significant spatial overlap between the areas responsible for these two tasks (Herholz, Halpern, & Zatorre, 2012; Kosslyn, Ganis, & Thompson, 2001).

Martin et al. (2017) investigated how encoding of auditory features occur during perception and imagery of music using ECoG. They showed that neural tuning to spectral features of the auditory stimuli occurs during imagery (in the absence of perceived sound) and occurs in the same cortical sites as during perception. However, there is lack of evidence for how perception and imagery processes occur over time and whether people are listening to and imagining these stimuli in the same way. We aim to investigate these questions using EEG.

SOMETHING HERE ABOUT SPEECH/RHYTHM CONNECTION

Deciphering the raw EEG data is a difficult task because the electrical signals recorded at the scalp include a mixture of neural activity and noise (e.g., cardiac artifacts, eye blinks, movement, and electrical environmental noise), significant covariation between electrodes adds redundancy to the data, and some electrode sites do not capture the dynamic activity of the brain as well as others. Methods such as principal component analysis (PCA) and independent component analysis (ICA) provide a means to recover meaningful information from within the data.

Recently, a new technique has emerged. Correlated component analysis (rCA) is computationally similar to (PCA) in that it computes an eigenvalue decomposition of covariance data, but where (rCA) differs is in the source of the covariance; (rCA) operates on the pooled within-subject cross covariance

and pooled between-subjects cross covariance

where

calculates the cross-covariance between participant *k* and participant *l* across all electrodes *x* at time *t*. The eigenvectors *wi* of the cross-covariance matrix *Rw-1Rb,* with the largest eigenvalues λ*i* calculated as (*Rw-1Rb)wi =* λ*i wi* are the components that maximize Pearson's correlation between subjects in the data. Like the component ranking of PCA based on explained variance, components found using rCA are rank-ordered by the magnitude of their correlation. The time courses and accompanying spatial weights of these correlated components represent patterns of evoked neural activity which are maximally correlated across all participants.

In this experiment, participants listened to and imagined short, repeated rhythms composed of tones or words to induce different rhythmic brain states. We investigated whether we could decode a user's EEG signal to determine which rhythm or speech phrase they were listening to or imagining. We used correlated components analysis to investigate our data. The components represent information that similarly tracks the stimulus across multiple recordings and may carry information that is specific to each stimulus that will allow us to classify which rhythm or speech pattern a participant is listening to or imagining.

**Materials and Methods**

**Participants**

Twelve participants (4 male), aged 20-27, with normal hearing and no history of brain injury took part in this study. Nine participants had formal musical training (1-15 years), and six of the participants played instruments regularly at the time of data collection.

**Stimuli**

Eight different stimuli were used during this experiment: four rhythmic patterns and four rhythm-matched speech phrases (see Figure 1). The audio of the stimuli can be found in the supplementary materials.

Each trial consisted of a perception and an imagination component. Four metronome clicks cued the participant to the tempo of the trial (120 bpm/2 Hz). After four clicks, the stimulus began to play while the metronome continued. After 12 seconds, the stimulus stopped playing, the metronome continued, and participants were asked to imagine the stimulus for another 12 seconds.

Rhythmic pattern stimuli were played in a random order during the first half of the experiment and speech phrase stimuli were played in a random order during the second half. Rhythm always preceded speech so that participants would not inadvertently imagine the speech patterns during the rhythm block. Each stimulus was heard and imagined 12 times throughout the experiment.

**Equipment and Procedure**

During the experiment, participants were seated in an audiometric room (Eckel model CL-13) and the EEG data were collected using a BioSemi Active-Two system with 64+2 EEG channels. Horizontal and vertical EOG channels were used to record eye movements. EEG was sampled at 512 Hz. A Cedrus StimTracker was used to ensure minimal delay (<0.05ms) between the presentation of the stimulus to the participant and the marking of the stimulus onset in the data. The experiment was programmed and presented using PsychToolbox run in MATLAB 2014a. A computer monitor displayed the instructions and speakers played the stimuli at a comfortable volume for each participant. While listening and imagining, participants were presented with a fixation cross and asked to keep their gaze steady. The volume of the stimuli was kept constant across participants.

**Analysis**

**Preprocessing.** EEG pre-processing was done using EEGLab in Matlab2014a.Data were filtered between 0.5 Hz and 30 Hz, downsampled to 256 Hz, epoched to remove break-periods, and submitted to an ICA analysis. ICA components corresponding to artefacts (eyeblinks, heart rate, etc.) were manually removed.The remaining components were used to reconstruct the EEG data.

**Correlational Component Analysis (rCA).** Pooled within and between-subjects covariances were computed separately for each of the eight stimuli conditions. Only the top component extracted from each condition (i.e., the spatial weights and time course which maximized Pearson's correlation in the group-aggregate data) was considered for further analysis here.

Although components *i* = 2...*n* undoubtedly encompass various aspects of the experience of listening and imagining, the top component reflects some neural processes that are most common across subjects and provides an optimal starting point to evaluate the utility of ISC as a means of capturing the different brain states during each condition.

**Inter-subject correlation and classification.** To assess the reliability of the correlated component at the single-subject level, time-resolved ISCs were computed by back projecting the component vectors w*i* into the original subject data to derive a component time course for each participant. This was done for each stimulus condition. With each per-subject time course, a measure of ISC encompassing the entire duration of the clip was computed first to quantify the magnitude of the correlation between each individual participant and the group and establish a distribution of synchronization in healthy controls; this is similar to the ISC analysis computed by (Naci, Sinai, & Owen, 2015).

To generate the distribution of ISCs, Pearson's correlations were calculated between all possible pairs of subjects using a sliding window technique. A sliding window of 0.5 seconds with a 60% overlap was used to generate a correlation coefficient between pairs at two-second intervals over the course of the audio. This yielded 57 correlation coefficients for each of the subject comparisons.The correlations computed for all subject pairs were then standardized using a Fisher's Z transformation and averaged at each time point to produce a mean ISC time course for each of the stimuli conditions. The time courses of the components were then used to classify which stimulus a participant was listening to or imagining.

rCA was calculated 100 times.

ISC time courses were submitted to a binary classification algorithm that compared all possible combinations of two conditions. In the classifier, 99 rCA time-course trials of the two conditions to be compared were used as training data leaving 1 trial of each condition to be classified as test data. This classification was done 100 times?

**Statistical analyses.** Non-parametric permutation statistics were used to test the significance of the group-averaged component time course for each of the stimuli conditions. Null distributions of correlation coefficients were created by iteratively phase-shifting the computed component time course for each participant and computing mean of the pairwise correlations with the rest of the group for each of the 152 time windows \citep{theiler1992testing}. We did this 1000 times to generate a null distribution of potential correlation values. The upper 5% of each null distribution was used as the significance threshold for each time point. Significance levels were adjusted for multiple comparisons using a false discovery rate (FDR) correction.

The number of significant two-second time windows was then compared between the intact and scrambled conditions using a Chi-square test of proportion with an alpha level set to 0.05.

**Results**

Within type (speech or rhythm) and modality (perception or imagination) binary classification of the four conditions occurred at 100% accuracy.

Within type but across modality (e.g. training on speech perception, testing on speech imagination) binary classification of the four conditions occurred at chance.

Classification of perception and imagination within stimulus type (e.g. training and testing on speech perception and imagination condition 1 only) occurred at 100%.

Within modality but across type (e.g. training on rhythm perception, testing on speech perception) results were varied. Binary classification of *some* pairs occurred at above chance levels.

**Discussion**

Using a correlated components analysis (rCA) we were able to classify our stimuli at a statistically significant level within type (speech or rhythm) and modality (perception or imagination) indicating that both perceived and imagined stimuli induce differentiable changes in the EEG signal. Using the time courses of correlated components to classify our stimuli indicates that the differences across stimuli are maintained across people. In other words, the brains of participants are doing similar things as they listen to or imagine the same rhythmic and speech stimuli.

When we trained and tested our classifier on the same type of stimuli (rhythm/speech) within the same modality (perception/imagery) the classifier was able to differentiate the stimuli at 100\% accuracy. When we trained and tested our classifier across modality (e.g. trained on speech perception and tested on speech imagination) our classifier was differentiating pairs of stimuli at chance levels. This lack of classification accuracy indicates that the time courses of the correlated components in perception and imagination are too similar for the classifier to be able to tell them apart. This supports previous work showing similarities between the brain's perception and imagination processes (CITE).

When we trained and tested our classifier across stimulus type (e.g. trained on rhythm perception and tested on speech perception or vice versa) the results were varied. Some pairs of stimuli could be differentiated at above chance levels. The pattern of sounds in the rhythm and speech stimuli were the same but were presented either using tones in the rhythm condition or using words in the speech condition. The similarities of this pattern of sounds between the two types of stimuli may be reflected in the time courses of the correlated components and the cross-type classification is being driven by the inherent rhythmic patterns of the stimuli. Although not all stimulus pairs could be classified at statistically significant levels the fact that some could be classified is an indication that perhaps the stimuli themselves are not suited to this type of classification and that this result is not due to a deficit in analysis technique.

mixed

scrambled

We can classify within a modality (perception/imagination) - synchrony can be driven by internally generated sources.

Compare rhythm/speech.

What are the characteristics of the stim that are classified well/poorly? Not time signature.

Takeaway: rCA is a simple, yet powerful tool that can be used to classify imagined brain states. The rhythmicity of speech stimuli can be used to produce differentiable brain states introducing a new technique that could be used to drive a BCI?

**Funding**: This work was supported by…

AS: NSERC (CGSM)

BS:

JAG:

AMO:

**Acknowledgements**

Sebastian Stober for assistance in the study design.

**References**

Herholz, S., Halpern, A., & Zatorre, R. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, *24*(6), 1382–97. https://doi.org/10.1162/jocn\_a\_00216

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural Foundations of Imagery. *Nature Reviews Neuroscience*, *2*(9), 635–642. https://doi.org/10.1038/35090055

Martin, S., Mikutta, C., Leonard, M. K., Hungate, D., Koelsch, S., Chang, E. F., … Pasley, B. N. (2017). Neural Encoding of Auditory Features during Music Perception and Imagery: Insight into the Brain of a Piano Player, 1–32. https://doi.org/10.1101/106617

Naci, L., Sinai, L., & Owen, A. M. (2015). NeuroImage Detecting and interpreting conscious experiences in behaviorally non-responsive patients. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2015.11.059

Tables

Captions to figures

Figures (individual files) .tif or .eps

A single column is 86 mm wide; a double column is 180 mm wide. Ideally figures should fit either a single or a double column.

Supplementary Material (separate from main document file)

Audio files