**Introduction**

Understanding how an individual experiences the world is a fundamentally difficult process. One approach to begin characterizing an individual’s experience is to quantify how similar or different it is to the experience of others in the same context. To probe the similarities between how different brains process the same stimuli we can use neuroimaging and an analysis called inter-subject synchronization (ISS). ISS is calculated as a voxel-wise correlation in fMRI activity over time between subjects and quantifies the similarity in brain activation between individuals exposed to the same stimulus. This similarity in brain activation can be interpreted as a proxy for shared experience (). One of the strengths of an ISS analysis is its reliance on stimuli that unfold over time. Therefore, naturalistic stimuli, such as movies or music, are ideally suited to an ISS analysis and they increase the ecological validity of the results by creating an experience inside the lab that more closely resembles the outside world than other traditional experimental paradigms.

Previous experiments have used a variety of stimuli to induce ISS such as movies, audio stories, and music. Using movies and stories, the researchers hypothesized that the strong ISS was driven by participants’ engagement with the stimulus and that this engagement was largely driven by how interesting participants found the plot of the movie or story (Campbell et al., 2015; Naci, Sinai, & Owen, 2017; Regev et al., 2018). However, inter-subject synchrony was also found between participants listening to classical orchestral music (Abrams et al., 2013). The music, despite not having a cohesive narrative, produced consistent and reliable patterns of activity in both cortical and subcortical brain areas previously shown to be involved in music processing. Together, these results indicate that although the presence of language and a narrative induce strong ISS, synchrony may rely on other factors as well. However, the factors that are necessary to produce consistent and reliable synchrony are yet unknown and the characterization of how ISS changes as a result of the presence of language has not been done. In this experiment we chose to tightly control our stimuli to better understand how various stimulus characteristics have an effect on synchrony.

Music is a unique stimulus category because it reliably creates *behavioural* synchronicity; people synchronize their movements as they clap or dance to music (). It is possible that the mechanisms involved in allowing people to synchronize to music are also involved in the ISS recorded using fMRI. Music is also different than other stimuli in the way that it is remembered over long periods of time. There is evidence to show that older adults with neurodegenerative disorders remember music from their youth even after other semantic memories have been forgotten (). Traditionally, the method for investigating memory for music using neuroimaging has involved a comparison between the blood-oxygen level dependent (BOLD) activation during unknown and well-known music (). One downside to this approach is that it collapses over the time dimension over which music unfolds. Using ISS takes into account the fluctuations in brain activity over time. It is possible that using an approach such as ISS may be a more sensitive method for understanding the unique processing of music and that will also allow us to understand the mechanisms behind ISS.

Although previous ISS investigations have not investigated memory or familiarity with stimuli specifically, some EEG and fMRI work has shown global ISS decreases with stimulus repetition (). This result may be explained by the brain’s adaptation response to the same stimulus over time. However, all of the stimuli have contained language (either words themselves(), or movies()). No studies to date have explored how inter-subject synchrony varies as a function of the memory for the stimuli and the presence of language or music. Spoken words and music have both been shown to induce strong ISS, but in this experiment we are interested in understanding how the combination of words and music change the quality of the ISS and whether the pattern of these changes is modified as a function of familiarity with the stimuli. To control for differences in auditory characteristics between the familiar and the unfamiliar stimuli, we employed a strict training paradigm to induce familiarity in participants who listened to initially novel stimuli that included all possible combinations of language and music.

**Methods**

**Participants**

Twenty-six neurologically healthy participants (14 female) aged 18-39 (mean=24) were recruited via posters and word of mouth at The University of Western Ontario. *Include relevant musical demographic information based on what goes into the analysis.*

**Testing procedure**

Participants completed two functional MRI scans that were separated by a stimulus training period (14-29 days; mean = 19 days). During both scans, participants passively listened to the stimuli (described below). During the training period, participants listened to the stimuli via an online player (designed in-lab) that tracked the number of times each stimulus was played. To ensure participants were listening, the player presented a simple question about the stimulus (e.g. “*were there lyrics present in the previous song?”*) at random between stimuli. Participants also came to the lab between for four sessions between scans. In each of these sessions, participants listened to the stimuli in lab and completed a series of behavioural tasks (described below).

**Stimuli.** Eight different auditory stimuli were created from songs written and recorded by a lab member between *year1-year2* in Cambridge, UK. These songs were chosen to lower the likelihood of participants being familiar with the stimuli. Two songs were kept whole (vocals & instruments), two songs had the vocals removed leaving just the instruments, two songs had the instruments removed leaving a single a capella voice, and the lyrics of two songs were recorded in-lab as spoken word (no music). The voice in all stimuli was the same. All stimuli were extended to 5 minutes long (using *Audacity*) by repeating a chorus or verse where needed. During the training period, participants listened to half of the stimuli via the online player (4 songs, one of each type). The training sets of stimuli were counterbalanced across participants.

**Behavioural tasks.** During the training period, participants came into the lab four times. Each session lasted less than one hour. Participants listened to the stimuli in-lab and completed 2-3 of the following behavioural tasks in each session.

To test whether participants were learning their training stimuli, two tests designed in-lab were used. The first, was a lyric modification task that presented participants with pairs of lyrics. Each pair consisted of a lyric taken directly from their stimuli training group and a modified version of the same lyric. Participants indicated which lyric was the correct lyric. The lyric pairs were tested for their validity before being included in this study. Before the first scan session, participants were presented with the entire set of 25 lyric pairs to obtain a baseline measurement. As participants were not familiar with the stimuli these lyrics were taken from, they were asked to indicate which lyric they believed was most likely to come from a song. In each behavioural session, participants responded to a subset of 10 lyric pairs to track learning progress. After the second scan session participants completed the full set of 25 lyric pairs again.

The second test of familiarity was a melody recognition task. After the second scan only, participants heard pairs of 2 sec clips taken from the stimuli. These clips did not contain any lyrics. One clip was taken from the stimuli training set, the other clip was from a stimulus the participant did not train on. Participants were asked to indicate which clip was most familiar to them.

Participants completed a questionnaire regarding musical abilities and training as well as a test of melodic memory and a test of beat perception taken from the Goldsmith’s Musical Sophistication Index (CITE). Finally, participants completed a musical association form where they described what each of the trained stimuli reminded them of (memories, other songs, etc.) and a lyric orientation questionnaire that measured to what degree the participant focuses on the lyrics in a song over the melodic content (D. VUVAN).

**~~Preference ratings.~~** ~~In each lab session and after the second scan, participants rated on a scale of 1-7 how much the liked the songs in their stimulus training group allowing us to track how preference changed as familiarity increased.~~

**fMRI acquisition and analyses**

Imaging was conducted at the Robarts Research Institute on a Siemens Magnetom 7 Tesla scanner with a 32-channel head coil. Functional scans were acquired with 54 slices per volume (TR = 1.25 s; TE = 20 ms; flip angle = 35°; FOV = 220 x 220 mm; voxel size = 2.5 mm3). Between functional runs within the first session only, a whole-head anatomical scan was acquired (TR = 6s; TE = 2.69 ms; FOV = 240 x 240 mm; voxel size = 0.75 mm3; 208 slices). Eight functional runs, each lasting five minutes (the length of the stimuli) were collected in both the first and second scans. The order of the eight songs was randomized in each scanning session for each participant.

Data were processed using SPM12 and automatic analysis software (AA; [www.cusacklab.org](http://www.cusacklab.org)). Data were corrected for motion and normalized to a template brain. Regressors were calculated to account for artifacts in white matter and cerebrospinal fluid. Smoothing was done with a Gaussian kernel of 8 mm FWHM (Peigneux et al., 2006). These normalization parameters were then applied to all echoplanar images.

Intersubject synchrony (ISS) was calculated for each stimulus by extracting the timecourse of every voxel and taking a correlation between each subject’s voxel timecourses and the mean timecourses from every other participant (leave-one-out method). This created a map of correlation values for each participant indicating their degree of correlation to the group. Performing a t-test on each voxel produced a spatial map of areas where the group was highly synchronized while listening to that stimulus.

The same ISS calculation was performed in specific ROIs. The Yeo 7 network parcellation (CITE) was used to identify frontoparietal areas. Two other ROIs were used as defined by neurosynth.org: motor areas (from 2565 studies), and auditory areas (from 1252 studies).

**Results**

**Behavioural Results**

Over the course of the training period, participants listened to the stimuli on average 12.8 times (SD = 4.6). The familiarity tests indicate that behaviourally, participants learned the stimuli they trained on. A 2(session: 1 & 2) x 3(type: spoken, a capella, whole) ANOVA was run on the lyric modification results. There was a main effect of session (*F*(1,138)=159.2, *p*<0.001) with results increasing from an average of 36% correct (SD=13.7) before the first scan, to an average of 82% correct (SD=9.8) after the second scan (see Figure 1). There was no main effect of stimulus type (*F*(2,138)=3.0, *p*=0.05). The melody recognition results collected after the second scan were at ceiling (mean=92%, SD=6.4) (see Figure 1). Four participants were excluded from the rest of the data analysis, because their average score across both memory tests did not reach 70%.

**fMRI results**

Initially, we investigated differences in intersubject synchronization (ISS) across stimulus type (instrumental, a capella, spoken, and whole). Pair-wise comparisons between the amount of synchrony resulting from each stimulus type showed that the largest amount of ISS was driven by the spoken stimuli, followed by the a capella stimuli, the whole stimuli, and finally the instrumental stimuli (see FIGURE). The whole brain analysis of ISS highlighted three areas of the brain that were most synchronized across individuals: auditory areas, frontoparietal areas, and motor areas. The ISS analysis was calculated again within each of these regions.

FIGURE: (bar graphs maybe?)

Whole brain: spoken > a capella > whole > instrumental

Auditory: spoken > a capella > whole > instrumental

Frontal: spoken > a capella = whole = instrumental

Motor: spoken > a capella = whole = instrumental

Within each of the ROIs pair-wise comparisons between stimuli type showed that the auditory region followed the same pattern as the whole brain results, while frontal and motor areas showed the most synchrony to spoken stimuli and no differences between the other types of stimuli.

*Will also look at whether the difference between the spoken and a capella (or spoken and instrumental?) covaries with the Lyric Orientation questionnaire results.*

Next, we investigated whether there were differences in the amount of synchrony in each ROI between sessions. T-tests between synchrony values in auditory areas showed more synchrony in session 1 than in session 2 to a capella (*t*=4.07, *p*<0.001), spoken (*t*=3.16,*p*=0.003), and whole stimuli (*t*=3.91,*p*<0.001). The amount of synchrony between the two sessions for instrumental stimuli was not different. T-tests between synchrony values in frontal areas showed more synchrony in session 1 than in session 2 to spoken stimuli only (*t*=3.47 ,*p*=0.001). There were no differences between sessions in motor areas.

The analysis described above included both trained and untrained stimuli in the second session. When t-tests were run on the trained and untrained stimuli separately we found no differences in synchrony across sessions in frontal and motor areas. In auditory areas only two differences were found in the session comparison between whole trained stimuli and in the a capella untrained stimuli. In both cases, more synchrony was present in the first session. It is likely that the loss of differences between sessions is due to half as many stimuli being included in this analysis.

T-tests across sessions were also run in the whole brain and resulted in a similar pattern with no differences seen in instrumental stimuli and more synchrony in the first session to spoken, a capella, and whole stimuli. However, we also see more synchrony to whole stimuli in the second session over the first session in the insula. This is the only area and stimulus that shows more synchrony in the second session.

**Discussion**

Two main findings:

The spoken stimuli drive more synchrony across people than any other type of stimulus. Once music is added (either as melody in a capella or as instruments and melody in whole) ISS is reduced. Too much information to follow? Previous Hasson lab work has shown strong ISS to stories. It is possible that there is a categorical difference between ISS to stories and songs and once melody is added, even to spoken song lyrics, the words are no longer processed as a story and are processed as a song.

Familiarity doesn’t really matter. If memory of a stimulus mattered, we would have expected to see differences in frontal areas. But because we only see effects in auditory areas across sessions, this is likely to do with a stimulus repetition effect rather than a memory effect. Repetition decreases synchrony – shown in previous work.