**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.

**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 music 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 moments between stimuli. Participants also came to the lab between scans four times. In each of these sessions, participants listened to the stimuli in lab and completed a serious 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). All songs were extended to be 5 minutes long (using *Audacity*) by repeating a chorus or verse where needed. During the training period, participants listened to half of the songs via the online music player (4 songs, one of each type). The sets of songs 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 stimulus, 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 stimulus 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, participants heard a pair of 2 sec clips taken from the stimuli. These clips did not contain any lyrics. One clip was taken from the stimulus 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 processing**

Imaging was conducted 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 preprocessing was performed using automatic analysis (aa)…