Brain and Language 187 (2018) 104–114

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Distinct brain areas process novel and repeating tone sequences

Brannon Green[a,*], Iiro P. Jääskeläinen[b,d], Mikko Sams[b], Josef P. Rauschecker[a,b,c,*]

[a] Laboratory of Integrative Neuroscience and Cognition, Interdisciplinary Program in Neuroscience, Georgetown University Medical Center, 3970 Reservoir Road NW, New Research Building-WP19, Washington, DC 20007, USA
[b] Brain and Mind Laboratory, Department of Neuroscience and Biomedical Engineering, School of Science, Aalto University, 00076 AALTO Espoo, Finland
[c] Institute for Advanced Study, TUM, Munich-Garching, 80333 Munich, Germany
[d] AMI Centre, Aalto NeuroImaging, Aalto University, Finland

ARTICLE INFO

Keywords:
Dual-stream model
Auditory sequence processing
Audiomotor integration
Learning and memory
Encoding
Chunking
Auditory objects
Basal Ganglia
Medial temporal lobe
Prefrontal cortex

ABSTRACT

The auditory dorsal stream has been implicated in sensorimotor integration and concatenation of sequential sound events, both being important for processing of speech and music. The auditory ventral stream, by contrast, is characterized as subserving sound identification and recognition. We studied the respective roles of the dorsal and ventral streams, including recruitment of basal ganglia and medial temporal lobe structures, in the processing of tone sequence elements. A sequence was presented incrementally across several runs during functional magnetic resonance imaging in humans, and we compared activation by sequence elements when heard for the first time ("novel") versus when the elements were repeating ("familiar"). Our results show a shift in tone-sequence-dependent activation from posterior-dorsal cortical areas and the basal ganglia during the processing of less familiar sequence elements towards anterior and ventral cortical areas and the medial temporal lobe after the encoding of highly familiar sequence elements into identifiable auditory objects.

1. Introduction

Hearing a story and listening to music both require the integration of auditory information over time. Novel and familiar sequences are perceived such that as each new bit of information is processed, it is coupled with previous knowledge, while the listener continues to take in new information. Representations are formed based on previous experience, and the listener can solidify these representations with each repetition, while also predicting how the remaining sequence may progress. When a series of sounds is repeated in a specific order, the sounds become associated with each other as a sequence over time, even after few presentations. Perceiving a meaningful series of sounds as an entity requires the learning and simultaneous storage of these ordered sequences, but the neural mechanisms for auditory sequence learning and storage in the brain are largely unknown (see Rauschecker, 2014).

Natural speech and music also consist of long motor sequences as they are produced. In speech, the movement of articulators in the vocal apparatus – such as the vocal folds, tongue, lips, teeth and throat – creates the sounds. In music, it is the coordinated movement of the fingers, limbs, mouth, or body that create sounds by manipulating the instrument of choice, including the singing voice. Speech and music are both learned and optimized through this coupling of the auditory and motor systems (Chen, Penhune, & Zatorre, 2008; Leaver, Van Lare, Zielinski, Halpern, & Rauschecker, 2009), which results in the formation of "internal models" (Rauschecker & Scott, 2009), as proposed more generally by control theory (Wolpert, Ghahramani, & Jordan, 1995). The dorsal auditory stream has been implicated in sensorimotor integration and concatenation of sequential events and thus in the implementation of internal sensorimotor models (Rauschecker, 2011).

The dorsal auditory stream connects posterior (caudal) areas of auditory belt cortex to areas in the inferior parietal lobule (IPL) and to premotor and prefrontal cortices. Among premotor regions, the supplementary motor area (SMA) has long been considered crucial for the coding of motor sequences (Roland, Larsen, Lassen, & Skinho, 1980). In addition, subcortical nuclei such as the basal ganglia (BG), which project to premotor areas including the SMA via the thalamus, have been shown to encode sequential behavior (Marsden, 1982). Though traditionally, based on work in nonhuman primates, both SMA and BG have been considered motor nuclei (DeLong & Georgopoulos, 1981; Strick, 1986), they have also been linked to the learning of newly associated sound sequences in more recent human studies (Zatorre, Chen, & Penhune, 2007; Leaver et al., 2009; Coull, Vidal, & Burle, 2016). Thus, auditory sequence processing by the BG and associated

* Corresponding authors at: Laboratory of Integrative Neuroscience and Cognition, Interdisciplinary Program in Neuroscience, 3970 Reservoir Road NW, New Research Building-WP19, Washington, DC 20007, USA (J.P. Rauschecker).
E-mail addresses: bmg49@georgetown.edu (B. Green), rauschej@georgetown.edu (J.P. Rauschecker).

https://doi.org/10.1016/j.bandl.2018.09.006
Received 5 April 2017; Received in revised form 3 October 2017; Accepted 23 September 2018
Available online 29 September 2018
0093-934X/

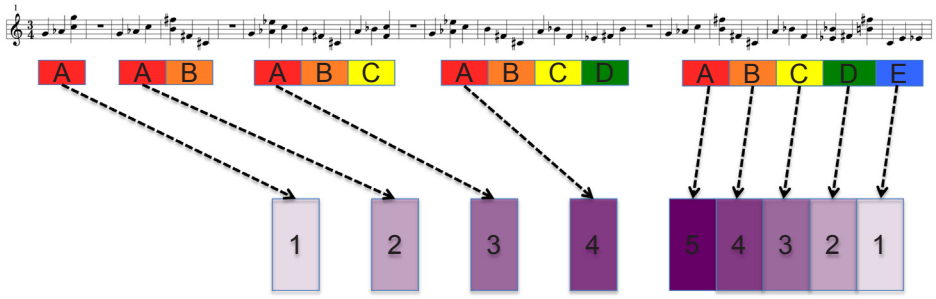

Fig. 1. Progressively presented musical sequence consisting of five sub-sequence chunks. A target sequence, ABCDE, is presented in a progressively increasing fashion such that on the first trial the first sub-sequence (A) is presented, and on each subsequent trial an additional subsequence is added to the previously heard sequences. On trial 5 the target sequence is played in its entirety with segment A having been presented 5 times and segment E being presented for the first time, thus establishing the repetition-based gradient of novelty and familiarity within a single sequence. Observing segment A across trials also establishes the gradient of novelty and familiarity.

"premotor" structures seems to be part of more generalized sequencing abilities normally epitomized by (but not limited to) the motor system (Janata & Grafton, 2003; Zatorre et al., 2007; Rauschecker, 2014).

The ventral auditory stream processes auditory objects in a hierarchical fashion, with activity proceeding anteriorly as object complexity increases (Leaver & Rauschecker, 2010). Auditory objects, analogous to visual objects, are sounds with specific auditory signatures making them easy to identify and categorize (Griffiths & Warren, 2004; Zatorre, Bouffard, & Belin, 2004; Leaver & Rauschecker, 2010; Giordano, McAdams, Zatorre, Kriegeskorte, & Belin, 2013). This pattern of hierarchical processing has also been shown to reflect the timescale over which auditory stimuli are presented and to which certain areas are tuned with different areas responding to different lengths of temporal receptive windows (Lerner, Honey, Silbert, & Hasson, 2011; Jääskeläinen et al., 2011). As the temporal receptive window increases, brain activation progresses hierarchically from primary sensory areas to higher cortical regions. As shown by Leaver et al. (2009), this "chunking" results in movement of the cortical activation to more rostral frontal regions. The same phenomenon has been described by sequence learning studies in monkeys (Pasupathy & Miller, 2005).

In the dorsal auditory stream, temporal hierarchies are also considered important for the processing of sequences in spoken language, for instance, at the level of words and sentences (Bornkessel-Schlesewsky, Schlesewsky, Small, & Rauschecker, 2015). More specifically, activations for phonemes and words migrate rostrally from middle superior temporal gyrus to anterior superior temporal gyrus, respectively (DeWitt & Rauschecker, 2012).

The current study utilized a progressively presented repeated tone sequence to create a gradient from novelty to familiarity. Tone sequences were used for increased experimental control, while preserving the essence of auditory sequence perception. Previous work has identified statistical learning of both words (Saffran, Aslin, & Newport, 1996) and tone sequences (Saffran, Johnson, Aslin, & Newport, 1999) via manipulation in transition probabilities between repeated, consecutive stimuli. The sequences in the current study were atonal in that the tones did not conform to a specific key signature to avoid triggering learned musical grammar representations. Using functional magnetic resonance imaging (fMRI) we measured blood oxygen level dependent (BOLD) activity associated with novel (few repetitions) and familiar (many repetitions) segments of the tone sequence. We hypothesized that segregated networks of activity would be associated with novel and highly familiar sequences within this paradigm. Novel sequences should activate sequence-encoding structures in the auditory dorsal stream, which should include subcortical "motor" nuclei such as the basal ganglia. We expected familiar sequences to ultimately activate object-representations in the auditory ventral stream and potentially recruit medial temporal lobe structures to facilitate the unique identification of a sequence as an auditory object and its retrieval from long-term memory. In order to control for attentional effects, the subjects performed either auditory or visual attention tasks with behavioral outcomes orthogonal to the variables of interest, namely the statistical learning of tone sequences based on repetition.

2. Materials and methods

2.1. Subjects

22 volunteers (17 male; 20 right handed; ages 21–39 years) were recruited from Aalto University and University of Helsinki, Finland. All participants were informed in writing of the procedures involved. 18 participants were native Finnish speakers, one native speaker of Russian, and three native speakers of Spanish. All participants spoke English as a second language with high proficiency and reported normal hearing. Four subjects were subsequently excluded due to scanning errors.

2.2. Stimuli

Auditory stimuli consisted of four 30-second tone sequences presented at 180 beats per minute (bpm). A random number generator (www.random.org) was used to create random, atonal sequences using the standard western musical notes from C_4 (Middle C) to C_5. Sequences were then constructed using GarageBand Software (Apple, Inc.), divided into 30 one-second segments, and edited with Audacity: Free Audio Editor and Recorder (RRID: SCR_007198). Visual stimuli, which were used to control for attention effects, consisted of 12 grey-scale Gabor patches of uniformly incremental spatial frequency that were created using a custom script in MATLAB (RRID: SCR_001622). Auditory and visual stimuli were combined into multimodal videos using VideoPad video editor (NCH Software Pty Ltd).

2.2.1. Stimulus presentation

One-second auditory sequence segments were presented progressively so that on the first trial only Segment 1 (S1) was heard; on the second trial S1 was followed by S2 as a continuous 2-second sequence; and on the third trial S1, S2, and S3 were presented as a continuous 3-second sequence and so on (Fig. 1). This progressive presentation continued through the 30th and final trial where the participants were presented with the 30 segments as one continuous 30-second sequence. Each subject was presented with four different runs, except for those subjects with missing runs due to scanning errors.

Each trial was randomly assigned between 1 and 4 auditory events intended to focus the attention of the subjects. These events were single tones exactly 7 half steps above a coincident note from the base sequence discussed above (interval). Additionally, the visual presentation of each trial included a Gabor patch in the center of the visible field for visual fixation. This patch changed spatial frequency (width of grey bands) between 1 and 4 times per trial. As described in the next section, both the auditory and visual attention stimuli were presented orthogonal to the auditory sequence stimulus of interest.

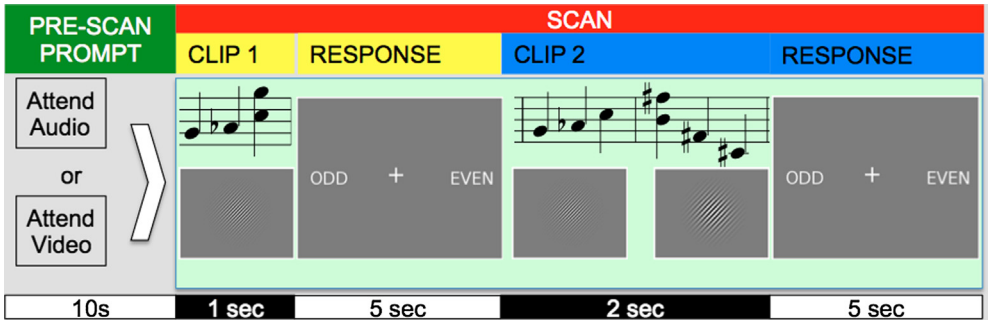

Fig. 2. Experimental design and schematic of stimulus. Each run consisted of a 10-second pre-scan period where subjects were prompted to attend and complete either an auditory or a visual attention task. During the scanning period, the subjects were presented with both auditory (musical sequences) and, as a control task, visual (Gabor filters) stimuli during 30 trials of increasing length, where the length of the nth trial was n seconds. Between trials was a 5-second response period during which subjects indicated whether or not the number of target events in either the auditory or visual attention task (max = 4) was odd or even. Target events that subjects detected were harmonizing chords (auditory task) or changes in the spatial frequency of the Gabor filter (visual task). These tasks were intended to hold the attention of subjects. Task-relevant stimuli did not contain sequence components, and both tasks were orthogonal to the stimulus variables of interest.

2.3. Behavioral tasks

While in the scanner, all participants were presented with all four of the progressive tone sequences, but were asked to perform either an auditory or visual attention task orthogonal to the sequence learning (Fig. 2). The auditory attention task consisted of counting the number of coincident musical events in each trial (harmonizing chord). Each trial was followed by a 5-second response period where the participants were prompted to indicate if the number of harmonic chords was odd or even. The visual attention task was identical to the auditory task except that participants were prompted to count the number of changes to the spatial frequency of the Gabor patches presented in each trial and indicate whether the number of detected changes was odd or even. There were between 1 and 4 total events for either task in any trial except those trials whose length was the limiting factor. While stimulus length increased with each repetition, the possible number of incidents was capped at four per trial. As such, the representations of repeated stimuli are expected to be strengthened by repetition despite low-frequency, random, incidental events. Events for either task were randomly presented during any given trial, contained no sequences, and were orthogonal to the sequence variables of interest being presented. That is to say, the variables of interest (novelty and familiarity) were defined by the number of repetitions of the unattended sound sequence, and the behavioral tasks performed in the scanner were used to assure that attention was either focused on the additional auditory task targets or visual task targets.

The novel design in the current study was created with the intention of observing a gradient of repetition-based novelty and familiarity within a musical sequence that, to the best of our knowledge, had not been studied before. This design reversed the location of novel and familiar stimulus sequences, familiar first and novel last, in each trial, thus reducing the effect of stimulus-specific adaptation in the analyses of highly repeated familiar sequences. The nature of this stimulus is biased heavily toward novel, low-repetition events, as all sequences are at some point novel, while only some sequences progress to a high level of familiarity. As such, we approached the analysis of novelty and familiarity in both a linear single-trial format and a more abstract across-trials format, finding results consistent across the methods. Since this relatively complex progressively presented design has its limitations, cautionary measures were taken to make reliable interpretations of the results. The auditory and visual tasks served as attentional controls and a baseline. Additionally, variables of interest were tested with different analysis approaches, i.e., repetition-based familiarity was observed across trials (early trials had fewer repetitions and late trials had more repetitions), and within the final trial (the first sequence segment had been presented 30 times while the last sequence segment had only been presented once).

2.4. MRI protocol

Images were obtained using a 3T MAGNETOM Skyra whole-brain body scanner (Siemens Healthcare, Erlangen, Germany), using a standard 32-channel head-neck coil. A continuous acquisition paradigm was used in which functional EPIs were collected consecutively throughout the experiment. Functional runs consisted of 325 volumes each (TR, 2 s; TE, 24 ms; voxel size, $3.4 \times 3.4 \times 3.0\,mm^3$).

2.5. fMRI analysis

All analysis was carried out using FEAT in FSL (RRID: SCR_002823). Functional images from each run were motion-corrected, spatially smoothed using a three-dimensional $2.5 \times 2.5 \times 2.5\,mm^3$ Gaussian filter, and high-pass filtered to remove low-frequency events. Preprocessed functional images were coregistered with their corresponding high-resolution MPRAGE anatomical images and interpolated into MNI152 standard space. Brain areas were identified using the Harvard-Oxford Cortical and Subcortical Structural Atlases distributed with FSL.

Parametrically weighted GLM analyses were conducted to assess the relationship between either sequence novelty or familiarity with the BOLD-fMRI signal. Weights of the GLM predictors were adjusted to reflect novelty or familiarity as a function of the number of times a stimulus had been repeated. Z statistic (Gaussianized T) images were corrected for multiple comparisons using a cluster threshold of $Z > 2.3$ at $p = 0.05$.

All analyses subsequently discussed reflect the contrast of Auditory > Visual conditions at the group level. Subjects were presented with 4 runs in the scanner (out of the 8 being used across subjects throughout the experiment). All run conditions for each subject contained a progressively presented sequence, auditory task targets, and visual task targets. Subjects were either prompted to attend to the auditory task targets or the visual task targets and then all runs of these attention conditions were averaged across subjects. They were then contrasted in the group analysis to show areas of activation that are greater for the auditory task condition than the visual task condition. This served both to control for possible attention effects, provide for a signal baseline for analysis, and to account for any attentional gating relative to the variable of interest, which was the unattended auditory sound sequence.

2.5.1. Final trial analysis

The final 30-second trial of each run was analyzed with either a 2 or 6-second non-overlapping sliding window. These 2 and 6-second windows roughly conform to the temporal receptive windows described by Lerner et al. (2011) that correspond to words and sentences, respectively. Preliminary analyses included 2-, 4-, and 6-second time windows. Preliminary data did not show a distinction between 2- and 4-

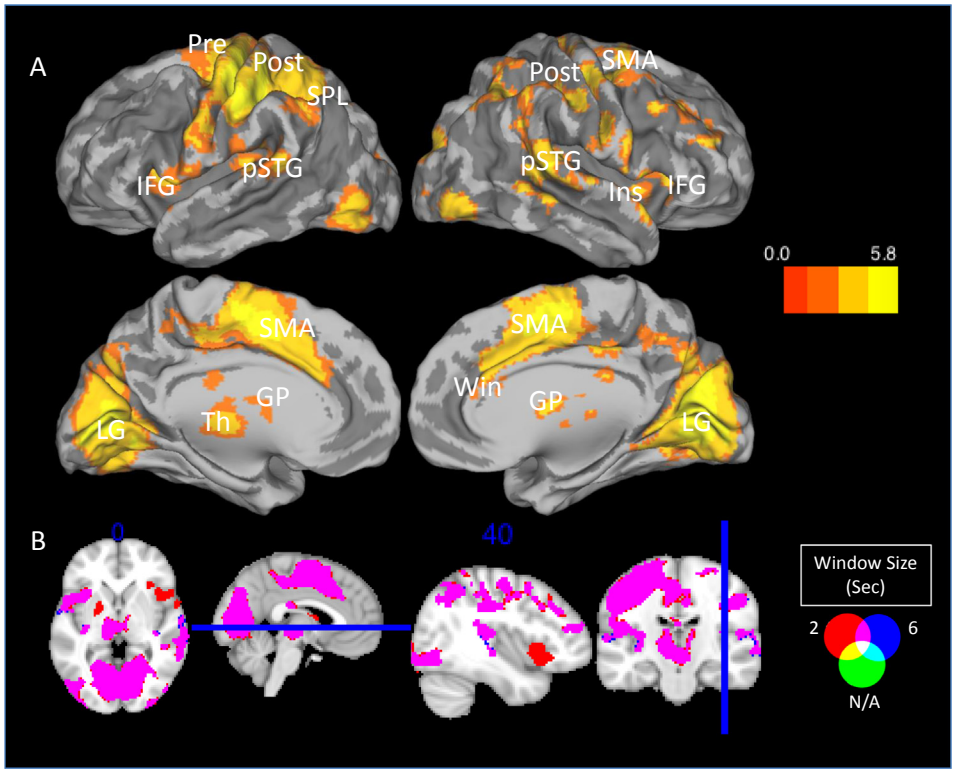

Fig. 3. Parametric analysis of novelty across the final 30-second trial. Cortical projections of the 2-second temporal window analysis (A) and slice projections of 2- and 6-second temporal windows analysis (B), shown in red and blue, respectively, (overlap in magenta) were used to analyze the final 30-second trial of the stimulus. Active clusters at both 2- and 6-second temporal windows included precentral gyrus (Pre), postcentral gyrus (Post), superior parietal lobule (SPL), posterior superior temporal gyrus (pSTG), supplementary motor area (SMA), lingual gyrus (LG), and the thalamus (Th). Active clusters unique to the 2-second temporal window include the globus pallidus (GB), inferior frontal gyrus (IFG), and the insula (Ins). Z-statistic images were thresholded at $Z > 2.3$ at $p < 0.05$ (corr) significance.

second windows, while a difference was shown between 2- and 6-second windows. GLM predictors were adjusted to reflect the average number of repetitions for sequences contained within the window in a linear fashion for the following analyses: 2-second temporal window analysis for novelty (−1.0, −0.818, −0.636, −0.455, −0.273, −0.091, 0.091, 0.273, 0.455, 0.636, 0.818, 1.0), 6-second temporal window analysis for novelty (−1.0, −0.333, 0.333, 1.0), 6-second temporal window analysis for familiarity (1.0, 0.333, −0.333, −1.0). Results for the 2-second temporal window analysis for familiarity were not significant. While there may be the potential for sampling bias based on differences in available data points, the lack of response for 2-second windows in the familiarity analysis suggests that the variables of interest have a driving role in these activation differences.

2.5.2. Intermediate sequence repetition analysis

A single 6-second sub-sequence was traced across trials. This allowed for the observation of increasing degrees of repetition-defined familiarity. This sequence began at 6 s post-stimulus onset in each trial (to account for hemodynamic onset effects), was first presented in its entirety in trial 12, and had its final presentation in trial 30 for a total of 19 repetitions. This identical sequence was compared with itself between early and late trials to identify the role of repetition on brain activations. BOLD fMRI signal associated with this sequence was averaged across trials into 5 groups according to the number of repetitions. Groups 1 through 5 averaged 2.5, 6.5, 10, 13.5, and 17.5 repetitions, respectively. GLM predictors were adjusted to reflect increased repetition in a linear fashion from group 1 to group 5 over 19 repetitions (−2.0, −1.0, 0.0, 1.0, 2.0), from group 1 to group 4 over 13.5 repetitions (−1.0, −0.333, 0.333, 1.0), from group 2 to group 5 over 13.5 repetitions (−1.0, −0.333, 0.333, 1.0), and from group 3 to group 5 over 10 repetitions (−1.0, 0.0, 1.0). Results comparing parametric increase from groups 1 to 2, 1 to 3, and 4 to 5 are not reported here.

3. Results

Our results are first presented as a comparison of parametrically increasing neural activation in response to both novel and familiar stimuli in the final trial of a progressively presented tone sequence. These results are presented in the context of temporal receptive window tuning, at either short (2-second) or long (6-second) timescales for active brain areas. Next we present results from observing unique subsequences as they were repeatedly presented across trials to investigate repetition-dependent increases in activation.

3.1. Final trial analysis

As described above, subjects heard a progressively presented tone sequence during the experiment, where the first trial was a one-second sequence. Each trial repeated the previously heard sequence, and an additional one-second sequence was added. Thus, the first trial was one second and the 30th (final) trial was 30 s long. This final trial stimulus contained a gradient of repetition-based novelty and familiarity. The first sub-sequence of the trial had been presented 30 times and was thus highly familiar, while the final sub-sequence of the trial had been presented only once, and was thus highly novel. By assigning parametric weights to each sub-sequence we were able to identify regions that showed increased activity with either more repetitions (tuned to familiarity) or fewer repetitions (tuned to novelty). Additionally, these parametric regressors were fit to sliding temporal windows of varying sizes (2 or 6 s), in order to assess the temporal receptive window characteristics of areas exhibiting either familiarity or novelty tuning.

Parametrically increasing neural responses associated with novel stimuli in the final trial were widespread at both temporal window durations. Four significant clusters ($p < 0.05$; corrected) were identified to show parametrically increasing activity for novel stimuli at the 2-second timescale (Fig. 3A). Peak maxima for these clusters were located in postcentral gyrus ($Z = 5.96$), thalamus ($Z = 3.77$), frontal operculum/inferior frontal gyrus ($Z = 3.93$), and the temporal pole ($Z = 3.56$). Non-peak local maxima included activations in globus

Table 1
Peak coordinates of local maxima for significant clusters in final trial analyses.

Analysis	Cluster	Z	X	Y	Z	Areas
2 s novel	4	5.96	−38	−30	44	Postcentral gyrus
		5.81	−34	−24	48	Precentral gyrus
		5.8	−38	−30	52	Postcentral gyrus
		5.77	−38	−20	56	Precentral gyrus
		5.75	−4	−2	50	Supplementary motor cortex, anterior cingulate
		5.62	−16	−76	4	Intracalcarine cortex, lingual gyrus
	3	3.77	−4	−20	0	Thalamus
		3.58	4	0	10	Right thalamus
		3.32	−18	0	2	Left pallidum, left putamen
		3.15	12	−16	18	Right thalamus
		3.06	−12	−26	14	Left thalamus
	2	3.92	−46	12	2	Frontal operculum, inferior frontal gyrus
		3.6	−32	18	6	Insula, frontal operculum
		3.3	−52	2	6	Central opercular cortex, precentral gyrus
		3.2	−60	2	8	Precentral gyrus, central opercular cortex
		2.84	−50	12	−12	Temporal pole
		2.77	−34	0	8	Insula
	1	3.56	54	16	−18	Temporal pole
		3.42	40	18	−12	Frontal orbital cortex, insular cortex
		3.37	60	10	24	Precentral gyrus, inferior frontal gyrus
		3.25	42	12	−4	Insula
		3.13	50	14	−2	Frontal operculum cortex, inferior frontal gyrus
		3.06	32	26	0	Insula, frontal orbital cortex
6 s novel	2	5.88	−38	−30	44	Postcentral gyrus, superior parietal lobule
		5.74	−38	−30	52	Postcentral gyrus
		5.7	−34	−24	48	Precentral gyrus, postcentral gyrus
		5.69	−38	−20	56	Precentral gyrus
		5.69	−4	−2	50	Supplementary motor cortex, cingulate gyrus
		5.46	−16	−76	4	Intracalcarine cortex, lingual gyrus
	1	3.66	−4	−20	0	Left thalamus
		3.04	12	−16	18	Right thalamus
		2.96	−8	−16	16	Left thalamus
		2.92	−12	−26	14	Left thalamus
		2.86	−6	−28	−2	Brain stem, left thalamus
6 s familiar	1	3.33	2	56	−18	Frontal pole, frontal medial cortex
		3.29	4	26	0	Subcallosal cortex, cingulate gyrus
		3.09	12	32	−2	White matter
		3.05	−2	32	−8	Subcallosal cortex, cingulate gyrus
		3.05	−4	34	−16	Frontal medial cortex, paracingulate gyrus
		3.03	−2	2	−10	Left cerebral cortex

pallidus and putamen (Z = 3.32), and in insular cortex (Z = 3.6). Two significant clusters ($p < 0.05$, corrected) were identified at the 6-second timescale with peak maxima in the postcentral gyrus/superior parietal lobule (Z = 5.88) and the thalamus (Z = 3.66). Sub-peak local maxima were also found in the precentral gyrus, supplementary motor cortex (SMA), and lingual gyrus (Fig. 3B). It is important to note here that, while the Harvard-Oxford brain atlas identified the peak cluster activation as lingual gyrus, the cluster represented additional activation across occipital visual cortices. We did not have any a-priori hypotheses for such visual area activation.

Activation at both time scales encompassed many brain areas (Table 1) and reflected widespread responses to novel stimuli with significant overlap between the two. Fig. 3B shows overlapping clusters in magenta and clusters unique to the 2-second time window in red. This analysis shows that while most of the active clusters are the same for the 2-second and 6-second receptive windows, the 2-second window has unique clusters in the basal ganglia, inferior frontal gyrus, and the anterior insula, suggesting a critical role for these structures in early sequence encoding.

Parametrically increasing neural responses associated with familiar stimuli in the final trial were also observed. One significant cluster ($p < 0.05$, corrected) was identified with peak maxima in the medial prefrontal cortex (Z = 3.33; Fig. 4; Table 1). Non-peak maxima were also observed in the anterior cingulate gyrus (Z = 3.29). This significant cluster was only observed at the 6-second time window and not at the 2-second time window, despite potential overlap in the regressor contrast weights for the volumes being analyzed, suggesting longer temporal receptive windows for brain areas tuned to familiar stimuli. These areas were located primarily in ventral medial premotor areas.

3.2. Intermediate sequence repetition analysis

The final trial was successful in creating a novelty-familiarity gradient within a single sequence where each sub-sequence had been presented an increasing number of times. Another novelty-familiarity gradient was observable in the same sub-sequence on subsequent trials. To that end we did an inter-trial analysis by tracing a single 6-second sub-sequence across multiple trial repetitions. Trials were binned into 5 groups averaging 2.5, 6.5, 10, 13.5, and 17.5 repetitions of the target sequence, respectively.

Analyses probing parametric increases in activity across different ranges of repetitions revealed a progression through distinct networks of cortical areas (Fig. 5; Table 2). Two significant clusters ($p < 0.05$, corrected) showing increased activity across 10 repetitions (Fig. 5A) were found with peak maxima in the pre/postcentral gyrus (Z = 4.56) and superior parietal lobule (Z = 3.81). Three significant clusters ($p < 0.05$, corrected) showing increasing activity across 13.5 repetitions (Fig. 5B) were found with peak maxima in the pre/postcentral gyrus (Z = 4.77 and Z = 4.38), and the fusiform/parahippocampal cortices (Z = 3.79). Non-peak local maxima also included the hippocampus (Z = 3.35) and amygdala (Z = 3.26). Two significant clusters ($p < 0.05$, corrected) showing increasing activity across the full range of presentations (15 repetitions on average, from group 1 to group 5) were found (Fig. 5C) with peak maxima in temporal pole (Z = 4.58) and medial prefrontal cortex/cingulate gyrus (Z = 4.98). Non-peak local maxima included parahippocampal gyrus (Z = 4.5). Comparisons of parametrically increasing activation from group 1 to 2 and from group 1 to 3 were not significant, but from group 1 to group 4 showed a similar activation pattern as for 1 to 5 (data not shown).

The results of the 17.5-repetition analysis mirror very closely the results of the final trial analysis for familiarity at the 6-second timescale, but the results for 13.5 and 10 repetitions differ significantly. Fig. 6 depicts an overlay of significant clusters at 10 (green), 13.5 (blue), and 17.5 (red) repetitions in both axial (Fig. 6A) and sagittal (Fig. 6B) slices. Active clusters in the 10-repetition range reside solely in the pre/postcentral gyrus and superior parietal lobule, where they overlap with some active clusters in the 13.5-repetition range (overlap in cyan). The 13.5-repetition range also has active clusters along the parahippocampal gyrus and temporal pole, where it overlaps with active clusters in the 17.5-repetition range (magenta). The 17.5-repetition range additionally has unique active clusters in the medial prefrontal cortex, anterior cingulate, and aspects of the temporal pole. These results show a clear dorsal/posterior to ventral/anterior progression from dorsal parietal areas and the central sulcus for low repetitions, to the medial temporal lobe for moderate repetitions, and proceeding to the ventral medial prefrontal cortex for high repetitions.

Finally, as shown in Fig. 7, behavioral results indicate that the subjects' level of attention did not diminish towards the end of the presentation of the progressively long stimulus sequences. Specifically, percent correct in task responses did not increase from trial 1 to trial 30.

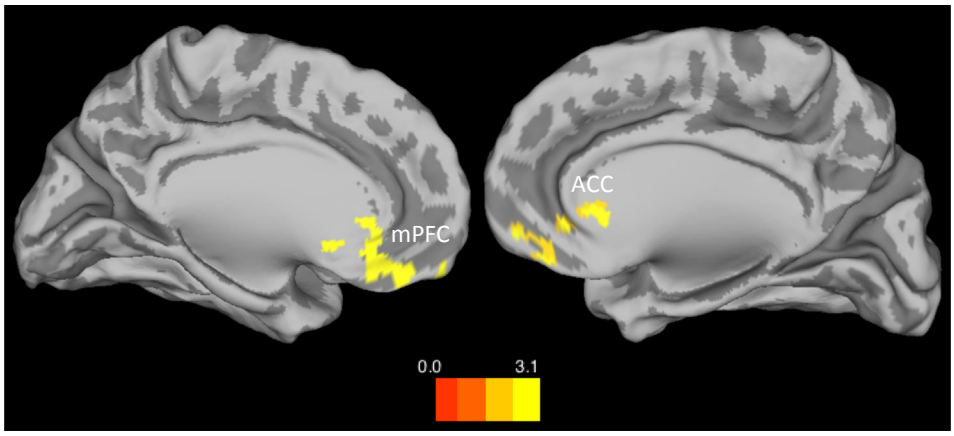

Fig. 4. Areas showing a parametrically increased activation to repeated stimuli. Six-second temporal window analysis revealed localized increases in activation across the final 30-second trial as the target sequence was repeated. Active clusters included medial prefrontal cortex (mPFC) and anterior cingulate cortex (ACC). Z-statistic images were thresholded at $Z > 2.3$ at $p < 0.05$ (corr) significance.

4. Discussion

The current study investigated the neural mechanisms underlying processing of novel and repeated auditory sequences. Our experiment utilized the progressive presentation of a tone sequence to create a gradient of novelty and familiarity. This gradient was observed within the sequence in multiple dimensions, e.g. linearly in the final 30-second trials by comparing the initial notes in the sequence that had been repeated multiple times with the last notes that had been repeated few times (Fig. 4), as well as by contrasting novel vs. repeated parts of the sequence across multiple trials (Fig. 6). We hypothesized that auditory dorsal-stream structures and basal ganglia would play a role in the encoding of novel sequences, while the perception of familiar sequences, presumably as one or more ever-evolving auditory objects, would tend to include parts of the auditory ventral stream and ventral prefrontal areas as a function of increasing familiarity.

4.1. Encoding of novel sequences

4.1.1. Dorsal auditory pathway

Listening to novel sequences revealed a broad network of active brain areas at both the 2- and 6-second timescales. This network included clusters in posterior superior temporal cortex, inferior parietal lobule, somatosensory, motor, and premotor cortices, as well as supplementary motor and dorsal prefrontal cortex (Fig. 3). These are all major hubs of the dorsal auditory pathway, and their activation is consistent with our first hypothesis.

4.1.2. Basal ganglia and inferior frontal gyrus

The basal ganglia and inferior frontal gyrus were active only during the 2-second timescale analysis of novel sequences. This suggests that these areas have shorter temporal receptive windows (Lerner et al., 2011) and are tuned to short, novel sequences, as they are first being presented for encoding. This role for the basal ganglia and inferior frontal gyrus as processors of novel sound sequences has been shown previously in work from our lab (Leaver et al., 2009). Basal ganglia

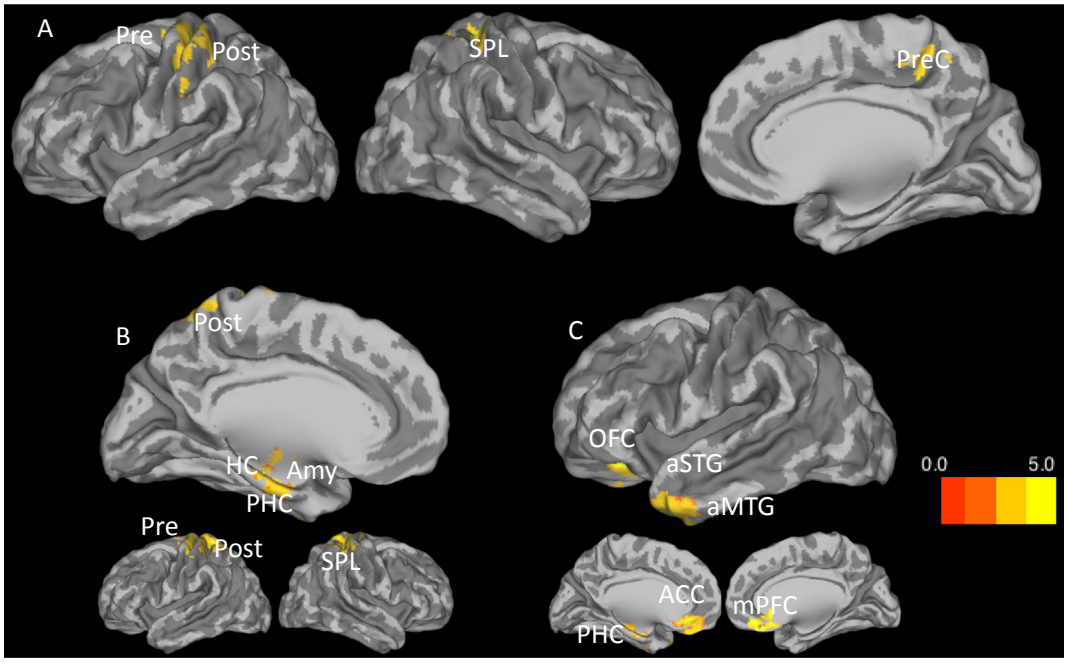

Fig. 5. Areas showing a parametric increase in activation. (A) Final 10, (B) final 13.5, and (C) all 17.5 repetitions of the target sequence. Target sequence was traced across trials for a total of 21 presentations and analyzed in trial groups resulting in 3 distinct patterns of parametrically increasing activity. Active clusters across 10 repetitions included SPL, Pre, and Post. Active clusters across 13.5 repetitions included the same as for 10 repetitions plus parahippocampal gyrus (PHC), hippocampus (HC) and the amygdala (Amy). Active clusters across 17.5 repetitions included the PHC, similar to 13.5 repetitions, but also the ACC, mPFC, anterior STG, anterior middle temporal gyrus (aMTG), and orbitofrontal cortex (OFC). Z-statistic images were thresholded at $Z > 2.3$ (corresponding to $p < 0.05$ significance).

Table 2
Peak coordinates of local maxima for significant clusters in intermediate sequence repetition analysis.

Analysis	Cluster	Z	X	Y	Z	Areas
10 Reps	2	4.56	−38	−24	54	Precentral gyrus, postcentral gyrus
		4.17	−20	−30	50	White matter
		4.13	−24	−22	64	Precentral gyrus
		4.04	−42	−24	64	Postcentral gyrus, precentral gyrus
		4.02	−38	−24	64	Postcentral gyrus, precentral gyrus
	1	3.81	24	−36	64	Postcentral gyrus, superior parietal lobule
		3.64	14	−52	62	Superior parietal lobule, precuneus cortex, postcentral gyrus
		3.51	20	−46	54	Superior parietal lobule, postcentral gyrus
		3.16	16	−36	48	Postcentral gyrus, precentral gyrus, precuneus, cingulate gyrus
		3.08	12	−36	58	Postcentral gyrus,
		2.97	32	−46	62	Superior parietal lobule
13.5 Reps	3	4.77	−24	−24	68	Precentral gyrus, postcentral gyrus
		4.58	−22	−44	68	Superior parietal lobule, postcentral gyrus
		4.49	−32	−28	66	Postcentral gyrus, precentral gyrus
		3.93	−10	−56	66	Superior parietal lobule, precuneus, lateral occipital cortex
		3.87	−30	−14	68	Precentral gyrus, superior frontal gyrus
	2	3.79	−38	−20	−24	Temporal fusiform cortex, parahippocampal gyrus
		3.74	−12	−10	−24	Anterior parahippocampal gyrus
		3.5	−18	−14	−26	Anterior parahippocampal gyrus
		3.35	−24	−22	−18	Left hippocampus
		3.26	−18	−6	−16	Left amygdala
	1	4.38	30	−28	66	Postcentral gyrus, precentral gyrus
		3.88	34	−36	66	Postcentral gyrus, superior parietal lobule
		3.71	18	−42	72	Postcentral gyrus, superior parietal lobule
		3.44	22	−46	60	Superior parietal lobule, postcentral gyrus
		3.44	40	−32	64	Postcentral gyrus
17.5 Reps	2	4.58	−38	8	−38	Temporal pole
		4.52	−48	12	−36	Temporal pole
		4.5	−38	−16	−24	Temporal fusiform cortex, anterior parahippocampal gyrus,
		3.97	−24	−22	−18	Left hippocampus
		3.94	−38	20	−36	Temporal pole
	1	4.98	−10	36	−14	Medial frontal cortex, paracingulate gyrus
		4.88	−20	32	−14	Frontal orbital cortex, frontal pole
		4.44	0	38	−26	Frontal medial cortex
		3.9	−36	42	−12	Frontal pole, frontal orbital cortex
		3.66	6	26	−6	Subcallosal cortex
		3.61	4	44	−16	Frontal medial cortex

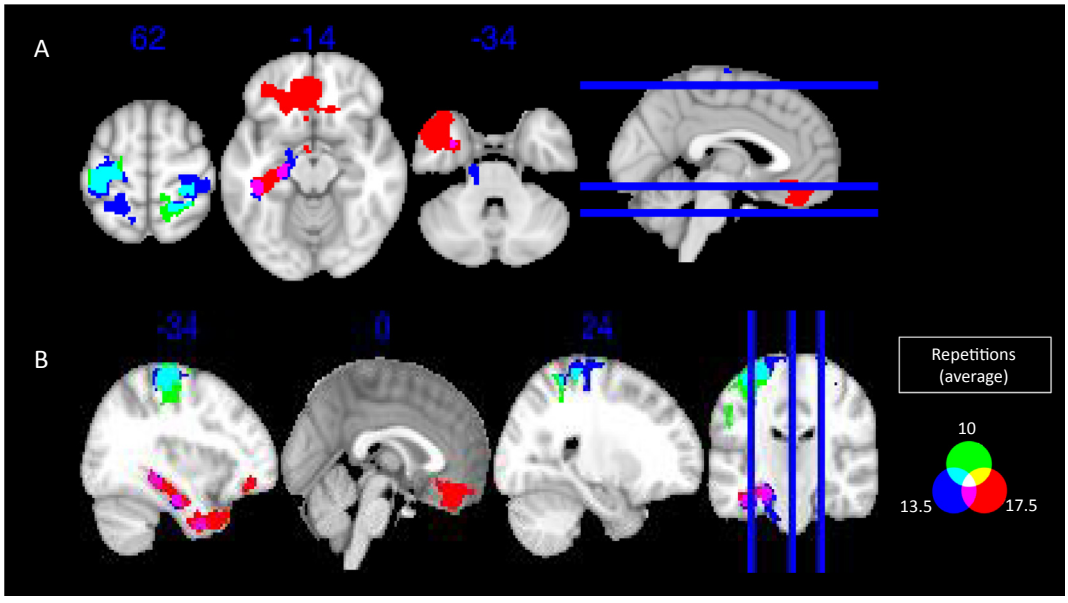

Fig. 6. Intermediate chunk familiarity in axial (A) and sagittal (B) slice images. Overlaid images of parametrically increasing responses are shown to the final 10 (green), final 13.5 (blue), and all 17.5 (red) repetitions. Activation maps from Fig. 3 are overlaid on slice images to further demonstrate the progression of activation across brain areas. Z-statistic images were thresholded at $Z > 2.3$ ($p < 0.05$ significance). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

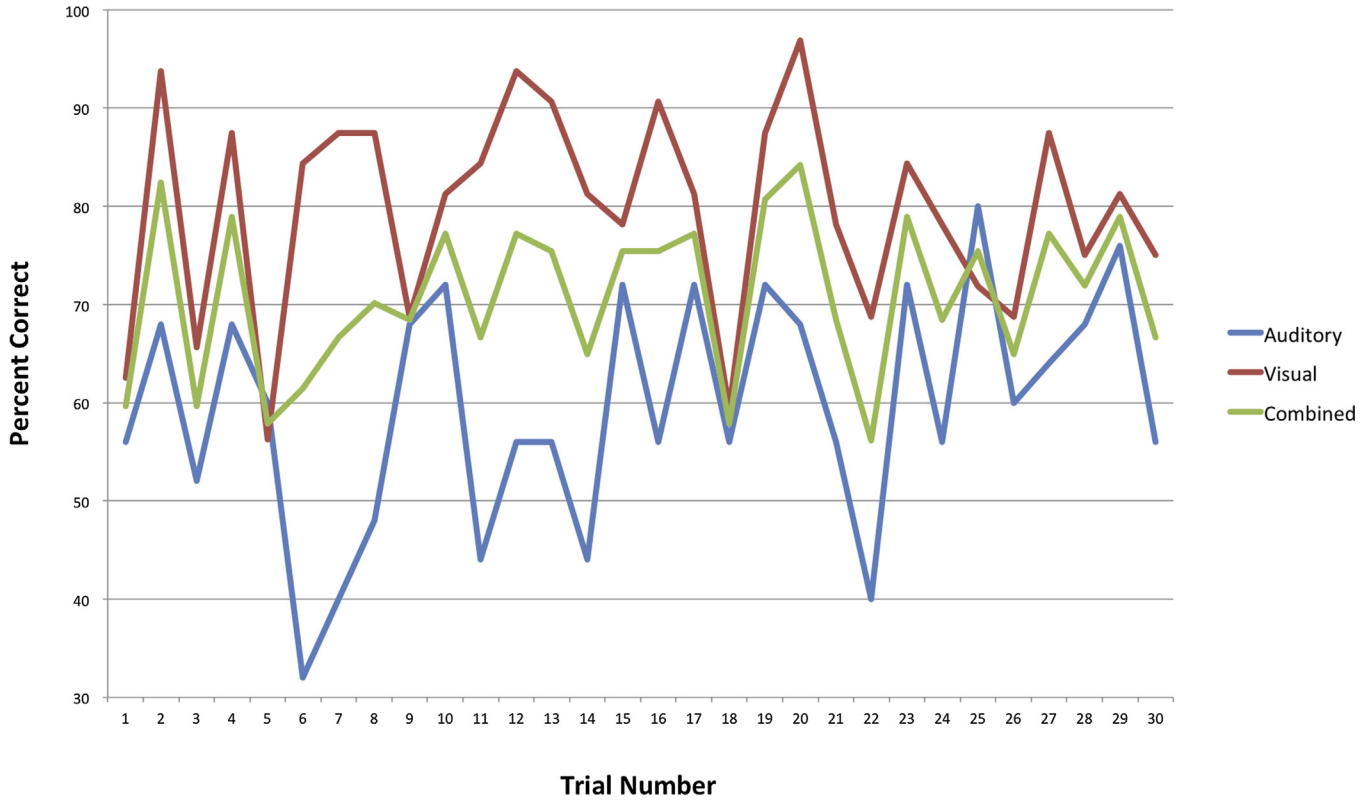

Fig. 7. Percent correct responses per trial for auditory task, visual task, and both tasks combined. Orthogonal behavioral tasks were used to control for attention. Fluctuations in percent correct are apparent for different tasks at various time points, but there is no observable increase or decrease in performance as a function of increasing number of trials.

activity has also been shown to arise during the early stages of sequence learning in monkeys (Pasupathy & Miller, 2005) as well as in other human sequence learning paradigms (Bornstein & Daw, 2012), which attests to a role for the basal ganglia in the concatenation of previously unrelated events (Hikosaka et al., 1999).

4.1.3. Sequencing in motor and sensory systems

The current study provides further evidence that, when presented with novel auditory sequences, the auditory system may borrow the sequencing capabilities of the motor system (Schubotz, 2007; Leaver et al., 2009; Doya, 1999; Rizzolatti & Luppino, 2001), and especially the basal ganglia, for the concatenation of novel, repeated sounds into longer sequences. In fact, so-called "motor" structures like the supplementary and pre-supplementary motor areas (SMA and pre-SMA) should perhaps be computationally redefined as sequencing areas that lend their capabilities to the motor system, which has a most obvious need for such sequencing mechanisms (Leaver et al., 2009; Rauschecker, 2014; for a recent review see Cona & Semenza, 2017). The SMA proper is typically active throughout sequence execution (Picard & Strick, 2003; Matsuzaka, Aizawa, & Tanji, 1992), and can be activated by auditory stimuli with increasing temporal coherence when compared to scrambled words or sentences (Lerner et al., 2011). Correspondingly, SMA has also been shown to be sensitive to disorder in the auditory stimulus (Nastase, Iacovella, & Hasson, 2014). Pre-SMA, in contrast, is transiently active at motor sequence initiation (Matsuzaka et al., 1992) and is associated with more complex motor action plans (Alario, Chainay, Lehericy, & Cohen, 2006), including chunks of sequence items (Sakai et al., 1999; Kennerley, Sakai, & Rushworth, 2004). Other sensory systems have similar needs for concatenation of sequential events; the visual system, for example, has to stitch or paste "snapshots" of visual scenes together between saccadic eye movements. Of all the sensory systems, the auditory system, however, has perhaps the most special need for such a mechanism of sequence processing, because most auditory events consist of complex acoustic patterns and occur as temporal sequences under various time scales (Lima, Krishnan, & Scott, 2016; Barascud, Pearce, Griffiths, Friston, & Chait, 2016).

4.1.4. Processing in lingual gyrus and visual cortex

We did not expect nor hypothesize activations in the visual areas, yet a peak activation in medial lingual gyrus (as well as additional lateral occipital activations) was elicited by novel sequences (Fig. 3). Auditory occipital activations (AOAs) have been reported previously and tend to occur in the blind (Weeks et al., 2000; Cate et al., 2009; Gougoux, Zatorre, Lassonde, Voss, & Lepore, 2005), although they can also occur in the sighted when listening to complex auditory scenes (Janata, Tillmann, & Bharucha, 2002), when completing cross-modal tasks (Feng, Störmer, Martinez, McDonald, & Hillyard, 2014), or listening to task-irrelevant sounds (McDonald, Störmer, Martinez, Feng, & Hillyard, 2013). It is also possible that higher activation in lingual gyrus to novel sequences reflects familiarity-related suppression, which has been linked to predictive coding (for comprehensive discussion on neural mechanisms underlying repetition suppression and enhancement, see Segaert, Weber, de Lange, Petersson, & Hagoort, 2013). Of particular interest here is that the lingual gyrus has been implicated as showing repetition suppression in cases where the original stimulus was attended to, and repetition enhancement when the original stimulus was ignored (Vuilleumier, Schwartz, Duhoux, & Dolan, & Driver, 2005). This suggests that the role of lingual gyrus is labile, context-dependent, and that the activations in the current study may be due to one or all of these factors when listening to novel sound sequences. Further study is required to fully understand the role of lingual gyrus and associated visual cortex in sequence processing more generally.

4.2. Processing and storage of highly familiar sequences

Repeated presentation of sound sequences resulted in increasing

activation in the medial temporal lobe, ventral medial prefrontal cortex, both sides of the central sulcus, and superior parietal lobule, but only at the 6-second timescale (Figs. 4–6). This suggests that areas responsible for long-term storage, recognition, and representation of sound sequences could be tuned to longer sequences that have already been encoded and concatenated into "chunks" (Leaver et al., 2009), thus forming higher-order object representations or chunks of chunks (Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015).

4.2.1. Medial temporal lobe

Signal in the medial temporal lobe, including the hippocampus, parahippocampal gyrus, and the amygdala, showed parametrically increasing activation as a specific sequence was repeatedly presented. Medial temporal lobe structures are associated with the formation of long-term memories of visual objects and their recognition (Bussey & Saksida, 2007; Eichenbaum, Sauvage, Fortin, Komorowski, & Lipton, 2012). Whether this is true for auditory memories as well is largely unknown (Munoz-Lopez et al., 2010; Muñoz, Mishkin, & Saunders, 2009), but the present results suggest that at least some traces of highly familiar melodies are retained in structures of the medial temporal lobe, consistent with Penfield's reports of patients recalling their mother's voice when he stimulated their temporal lobe (Penfield, 1975). However, these auditory memory traces may include knowledge *about* the melody, e.g. its name ("twinkle, twinkle, little star"), its composer, or its lyrics, or knowledge associated with the melody in other ways, e.g. the place where it was first or last heard, rather than the melody itself (Rauschecker, 2014). On the other hand, studies of implicit sequence learning have shown that patients with medial temporal lobe lesions have impaired declarative knowledge of the sequence (e.g. melody recognition), while procedural knowledge of the sequence (e.g. the ability to sing a melody or play it on the piano) is conserved (Schapiro, Kustner, & Turk-Browne, 2012; Schapiro et al., 2014; Valtonen, Gregory, Landau, & McCloskey, 2014; Rose, Haider, Salari, & Büchel, 2011).

Medial temporal lobe and basal ganglia have also been shown to play complementary roles in statistical and sequence learning such that the basal ganglia activate early in learning, while the medial temporal lobe activates later (Bornstein & Daw, 2012; Davis & Staddon, 1990; Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003). The medial temporal lobe has also been shown to be involved in the detection of patterns and regularities in auditory stimuli (Barascud, 2016; Geiser, Walker, & Bendor, 2014) Medial temporal lobe lesions result in temporally graded retrograde amnesia, such that older memories are less impaired than more recent memories (Squire, 2009). This suggests that while the medial temporal lobe is critical in the formation of long-term memories, the responsibility for actual long-term storage is passed to other areas of the cerebral cortex (Ullman, 2004).

4.2.2. Frontal cortex: Anterior shift in processing

The current study partially replicates the findings of previous work (Leaver et al., 2009) showing a caudal-to-rostral shift in prefrontal cortex activation as stimuli progress from novel to familiar (see also Christoff & Gabrieli, 2000). While the activation maps are not direct reflections of each other, the pattern of activation progressing from the basal ganglia to increasingly frontal areas in the medial and inferior frontal cortex suggests that long-term representations of sound sequences may indeed reside in frontal regions. Increased activations in the medial prefrontal cortex were also observed for increased repetitions of target sequences. It is possible that this area, therefore, plays an important role in the long-term representation of sound sequences, consistent with its well-known function (together with orbitofrontal cortex) in high-level cognitive and emotional processes (Öngür & Price, 2000).

4.3. Limitations of the present design

The behavioral tasks in our paradigm were orthogonal to the sequence variables of interest and did not reflect retention of the sequences themselves. Novelty and familiarity of target sequences were operationally defined by low and high sequence repetition, respectively. Because of the implicit nature of our sequence we could not conduct such behavioral experiments prior to image collection, as it would have revealed the nature of the experiment. While the lack of direct behavioral evidence may be a limitation of our study, we did observe parametrically increasing brain activations based on the number of stimulus presentations, which can serve as a substitute and putative neural correlate of the subjects' learning effect. These behavioral data from the orthogonal task, although not directly related to our hypotheses, do show that percent correct in task responses does not increase from trial 1 to trial 30 (Fig. 7). This suggests that there are no systematic fluctuations in task difficulty or attention across the length of each run of the experiment.

5. Conclusions

Our results indicate that dorsal-stream auditory cortical areas, such as posterior superior temporal gyrus, superior/inferior parietal lobes, and SMA, process novel sound sequences. The basal ganglia, inferior frontal gyrus, and insula were uniquely active for novel tone sequences at shorter timescales, but not longer ones. This suggests, as we have argued previously, that the sequence-encoding capabilities of the basal ganglia and their ability to concatenate short, novel sequences into chunks (in conjunction with associated areas in frontal cortex) are not restricted to motor skills. Instead, our results are consistent with the idea that the same mechanisms apply to the coding of sequences in various sensory domains, and different aspects of specific sequences are stored in different brain regions. These findings reflect areas subject to repetition suppression effects where novel stimuli elicited broad activations across multiple networks of brain areas with activity decreasing over time. These broad activations tend to reflect areas involved in attention and initial encoding (Segaert et al., 2013). By contrast, our results also show networks of areas that exhibit repetition enhancement. This is suggestive of unstable neural representations being strengthened through repetition and the formation of novel networks (Henson, Shallice, & Dolan, 2000). This process is likely facilitated through implicit statistical learning of transition probabilities (Saffran et al., 1996, 1999) between these novel sequences, and by preparing them for downstream processing in other areas. Our study identified the ventral-stream auditory cortical areas near the temporal pole, and in medial prefrontal cortex, as having this pattern of activation, responding more to familiar sequences at longer timescales in conjunction with the medial temporal lobe. This suggests that these areas play a role in storing sequence information (Dehaene, 2015; Rauschecker, 2014) as more complex representations.

We therefore conclude that the (implicit) encoding, reproduction, or continuation of a sequence by means of detection and prediction of sequence elements resides in the dorsal stream, whereas (explicit) declarative information relevant for the identification and recognition of complex "auditory objects" may be found in the ventral stream (Zatorre, et al., 2007; Rauschecker & Scott, 2009). Medial temporal lobe structures have already been folded into the dual-stream models of vision (Kravitz, Saleem, Baker, & Mishkin, 2011; Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013) and memory (Ullman, 2004; Eichenbaum et al., 2012). The results in the current study parallel these models and further expand the scope of the dual-stream model of audition to include these memory systems.

Significance statement

This study investigates the brain areas that subserve the encoding

and chunking of sound sequences and their eventual storage as auditory objects—a problem that needs to be solved for language processing at various levels. The initial encoding of novel sequences depends on the basal ganglia and the auditory dorsal stream. Parametrically increasing, repetition–dependent activity in prefrontal cortex and in the medial temporal lobe suggests that both of these structures play significant, but different, roles in the storage of sound sequences once they become familiar.

Acknowledgements

Funding for this study was provided by the National Science Foundation (PIRE-OISE-0730255), aivoAALTO and Academy of Finland (FiDiPRO No. 11837; CoE No. 129670; No. 276643). The authors wish to thank Marita Kattelus, Jussi Alho, and Dmitry Smirnov for their assistance with fMRI acquisition, and Juha Salmitaival for assistance with data analysis.

Author contributions

B.G., M.S., I.P.J., and J.P.R. designed research; B.G. performed research; B.G. analyzed data; B.G., M.S., I.P.J., J.P.R. wrote the paper.

Conflict of interest

The authors declare no competing financial interests.

References

Alario, F. X., Chainay, H., Lehericy, S., & Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research, 1076*, 129–143.
Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences USA, 113*, E616–E625.
Bornkessel-Schlesewsky, I., Schlesewsky, M., Small, S. L., & Rauschecker, J. P. (2015). Neurobiological roots of language in primate audition: Common computational properties. *Trends in Cognitive Science, 19*, 1–9.
Bornstein, A. M., & Daw, N. D. (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience, 35*, 1011–1023.
Bussey, T. J., & Saksida, L. M. (2007). Memory, perception, and the ventral visual-perirhinal-hippocampal stream: Thinking outside of the boxes. *Hippocampus, 908*, 898–908.
Cate, A. D., Herron, T. J., Yund, E. W., Stecker, G. C., Rinne, T., Kang, X., Petkov, C. I., Disbrow, E. A., & Woods, D. L. (2009). Auditory attention activates peripheral visual cortex. *PloS One, 4* e4645.
Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: Brain network for auditory synchronization is modulated by rhyuthm complexity and musical training. *Journal of Cognitive Neuroscience, 20*, 226–239.
Christoff, K., & Gabrieli, J. D. E. (2000). Frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology, 28*, 168–186.
Cona, G., & Semenza, C. (2017). Supplementary motor area as key structure for domain-general sequence processing: A unified account. *Neuroscience & Biobehavioral Reviews, 72*, 28–42.
Coull, J. T., Vidal, F., & Burle, B. (2016). When to act, or not to act: That's the SMA's question. *Current Opinion in Behavioral Sciences, 8*, 14–21.
Davis, D. G. S., & Staddon, J. E. R. (1990). Memory for reward in probabilistic choice: Markovian and non-markovian properties. *Behaviour, 114*, 37–64.
Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron, 88*, 2–19.
DeLong, M. R., & Georgopoulos, A. P. (1981). Motor functions of the basal ganglia. In V. B. Brooks (Ed.). *Handbook of Physiology. Section I: The Nervous System. Vol. 11: Motor Control* (pp. 1017–1061). Bethesda: American Physiological Society.
DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences of the United States of America, 109*, E505–E514.
Doya, K. (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Networks, 12*, 961–974.
Eichenbaum, H., Sauvage, M., Fortin, N., Komorowski, R., & Lipton, P. (2012). Towards a functional organization of episodic memory in the medial temporal lobe. *Neuroscience & Biobehavioral Reviews, 36*, 1597–1608.
Feng, W., Störmer, V. S., Martinez, A., McDonald, J. J., & Hillyard, S. A. (2014). Sounds activate visual cortex and improve visual discrimination. *Journal of Neuroscience, 34*, 9817–9824.
Geiser, E., Walker, K. M. M., & Bendor, D. (2014). Global timing: A conceptual framework to investigate the neural basis of rhythm perception in humans and non-human species. *Frontiers in Psychology, 5*, 159.
Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N., & Belin, P. (2013). Abstract encoding of auditory objects in cortical activity patterns. *Cerebral Cortex, 23*, 2025–2037.
Gougoux, F., Zatorre, R. J., Lassonde, M., Voss, P., & Lepore, F. (2005). A functional neuroimaging study of sound localization: Visual cortex activity predicts performance in early-blind individuals. *PLoS Biology, 3*, e27.
Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience, 5*, 887–892.
Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science, 287*, 1269–1272.
Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, X., Nakamura, K., ... Doya, K. (1999). Parallel neural networks for learning sequential procedures. *Trends in Neuroscience, 22*(10), 464–471.
Janata, P., & Grafton, S. T. (2003). Swinging in the brain: Shared neural substrates for behaviors related to sequencing and music. *Nature Neuroscience, 6*, 682–687.
Janata, P., Tillmann, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruis domain-general attention and working memory circuits. *Cognitive, Affective, & Behavioral Neuroscience, 2*, 121–140.
Jääskeläinen, I. P., Ahveninen, J., Andermann, M. L., Belliveau, J. W., Raij, T., & Sams, M. (2011). Short-term plasticity as a neural mechanism supporting memory and attentional functions. *Brain Research, 1422*, 68–81.
Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review, 110*, 316–339.
Kennerley, S. W., Sakai, K., & Rushworth, M. F. (2004). Organization of action sequences and the role of the pre-SMA. *Journal of Neurophysiology, 91*, 978–993.
Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience, 12*, 217–230.
Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Science, 17*, 26–49.
Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *Journal of Neuroscience, 30*, 7604–7612.
Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *Journal of Neuroscience, 29*, 2477–2485.
Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience, 31*, 2906–2915.
Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends in Neurosciences, 39* 572 542.
Marsden, C. D. (1982). The mysterious motor function of the basal ganglia: The Robert Wartenberg lecture. *Neurology, 32*, 514–539.
Matsuzaka, Y., Aizawa, H., & Tanji, J. (1992). A motor area rostral to the supplementary motor area (presupplementary motor area) in the monkey: Neuronal activity during a learned motor task. *Journal of Neurophysiology, 68*, 653–662.
McDonald, J. J., Störmer, V. S., Martinez, A., Feng, W., & Hillyard, S. A. (2013). Salient sounds activate human visual cortex automatically. *Journal of Neuroscience, 33*, 9194–9201.
Muñoz-Lopez, M. M., Mohedano-Moriano, A., & Insausti, R. (2010). Anatomical pathways for auditory memory in primates. *Frontiers in Neuroanatomy, 4*, 129.
Muñoz, M., Mishkin, M., & Saunders, R. C. (2009). Resection of the medial temporal lobe disconnects the rostral superior temporal gyrus from some of its projection targets in the frontal lobe and thalamus. *Cerebral Cortex, 19*, 2114–2130.
Nastase, S., Iacovella, V., & Hasson, U. (2014). Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Human Brain Mapping, 35*, 1111–1128.
Öngür, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex, 10*, 206–219.
Pasupathy, A., & Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature, 1138*, 873–876.
Penfield, W. (1975). *The mystery of the mind: A critical study of consciousness and the human brain.* Princeton: Princeton University Press21–27.
Picard, N., & Strick, P. L. (2003). Activation of the supplementary motor area (SMA) during performance of visually guided movements. *Cerebral Cortex, 13*, 977–986.
Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research, 271*, 16–25.
Rauschecker, J. P. (2014). Is there a tape recorder in your head? How the brain stores and retrieves musical melodies. *Frontiers in Systems Neuroscience, 8*, 1–6.
Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience Reviews, 12*, 718–724.
Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron, 31*, 889–901.
Roland, P. E., Larsen, B., Lassen, N. A., & Skinho, E. (1980). Supplementary motor area and other cortical areas in organization of voluntary movements in man. *Journal of Neurophysiology, 43*, 118–136.
Rose, M., Haider, H., Salari, N., & Büchel, C. (2011). Functional dissociation of hippocampal mechanism during implicit learning based on the domain of associations. *Journal of Neuroscience, 31*, 13739–13745.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.
Sakai, K., Hikosaka, O., Miyauchi, S., Sasaki, Y., Fujimaki, N., & Putz, B. (1999).

Presupplementary motor area activation during sequence learning reflects visuo-motor association. *Journal of Neuroscience, 19* RC1(1–6).
Schapiro, A. (2014). The necessity of medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience, 26*, 1736–1747.
Schapiro, A., Kustner, L., & Turk-Browne, N. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology, 22*, 1622–1627.
Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *TRENDS in Cognitive Sciences, 11*, 211–218.
Segaert, K., Weber, K., de Lange, F. P., Petersson, K. M., & Hagoort, P. (2013). The suppression of repetition enhancement: A review of fMRI studies. *Neuropsychologia, 51*, 59–66.
Strick, P. L. (1986). The organization of thalamic inputs to the "premotor" areas. *Progress in Brain Research, 64*, 99–109.
Squire, L. R. (2009). The legacy of patient H.M. For neuroscience. *Neuron, 61*, 6–9.
Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231–270.
Valtonen, J., Gregory, E., Landau, B., & McCloskey, M. (2014). New learning of music after bilateral medial temporal lobe damage: Evidence from an amnesic patient. *Frontiers in Human Neuroscience, 8*, 1–13.
Vuilleumier, P., Schwartz, S., Duhoux, S., Dolan, R. J., & Driver, J. (2005). Selective attention modulates neural substrates of repetition priming and "implicit" visual memory: Suppressions and enhancements revealed by fMRI. *Journal of Cognitive Neuroscience, 17*, 1245–1260.
Weeks, R., Horwitz, B., Aziz-Sultan, A., Tian, B., Wessinger, C. M., Cohen, L., ... Rauschecker, J. P. (2000). A positron emission tomographic study of auditory localization in the congenitally blind. *Journal of Neuroscience, 20*(7), 2664–2672.
Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science, 269*, 1880–1882.
Zatorre, R. J., Bouffard, M., & Belin, P. (2004). Sensitivity to auditory object features in human temporal neocortex. *Journal of Neuroscience, 24*, 3637–3642.
Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*, 547–558.