Original Article

Reduced functional connectivity between left frontal and right auditory cortices during speech perception in congenital amusia.

Kyle Jasmin[1,2], Fred Dick[1,3], Lauren Stewart[4], and Adam Tierney[1]

1. Department of Psychological Sciences, Birkbeck University of London
2. UCL Institute of Cognitive Neuroscience, University College London
3. Department of Experimental Psychology, University College London
4. Department of Psychology, Goldsmiths University of London

Correspondence to:

Kyle Jasmin
Department of Psychological Sciences
Birkbeck University of London
Malet Street
London WC1E7 HX
kyle.jasmin.11@ucl.ac.uk

Abstract | Congenital amusia is a condition characterized by disordered perception of and memory for pitch. Investigations of amusia's neural basis have typically used musical or tonal tasks to demonstrate reduced connectivity between frontal and auditory cortices within the right hemisphere. However, pitch is not only a musical property, but is also an important aspect of spoken language — one which amusics, presumably because of their impairment, tend to rely on less than non-amusic adults. It is an open question whether amusics' decreased reliance on pitch for speech shares the same right hemispheric neural substrate as their music-related deficits. Here we used functional magnetic resonance imaging to scan 15 individuals with amusia and 15 non-amusic controls while they performed a speech categorization task (intonational phrase boundary detection). On each trial, participants matched auditory and visual sentences by relying on pitch cues, duration cues, or both combined. We used a data-driven analysis to identify the strongest functional connectivity differences between the amusics and controls. Group differences in global functional connectivity (Control > Amusia) were strongest in four regions in lateral prefrontal cortex. Connectivity with these 'seed' regions was examined with respect to the rest of the brain. Participants with amusia showed prominent decreases in functional connectivity between language-related regions (left inferior and middle frontal gyrus/DLPFC) and pitch-related regions (right auditory and anterior insular cortex). No connectivity differences between the cue conditions were detected and no condition-by-group interactions were found, suggesting that decreased functional connectivity in amusia persisted irrespective of which acoustic cues judgments were based on. There were also no group differences in degree of task-related activation. We conclude that functional connectivity between left—as well as right—frontal cortex and pitch-related right hemisphere regions is decreased in amusia, and suggest this decreased connectivity may reflect amusics' strategic downweighting of pitch information during processing of spoken language. Our results also suggest that individual differences in auditory abilities relating to specific dimensions (such as pitch) are reflected in differential functional connectivity between regions that process those dimensions.

[H1] Introduction

Congenital amusia is a rare condition characterized by impaired perception of and memory for pitch (Peretz *et al.*, 2002). Although amusia presents as an auditory condition, auditory cortical responses (Moreau *et al.*, 2013; Norman-Haignere *et al.*, 2016) are normal, as is subcortical encoding of pitch (Liu, Maggu, *et al.*, 2015). The dominant view of amusia's neural basis is that connectivity between right inferior frontal cortex and right auditory cortex is impaired, resulting in impaired conscious access to pitch information for guiding behavior (Hyde et al., 2011; Albouy et al, 2013; Leveque et al., 2016; Zendel, et al., 2015; see Peretz, 2016 for review). While congenital amusia is believed to be innate, there is evidence that recovery is possible through training (Whiteford & Oxenham, 2018).

Although pitch is usually associated with music, it is also important for cueing categories in spoken language (de Pijper & Sanderman, 1994; Streeter, 1978) and conveying emotion in speech (Frick et al., 1985). In highly-controlled laboratory tasks in which speech perception judgments must be made based on pitch alone, only minor deficits have been observed in amusia (Liu, Jiang, *et al.*, 2015; Patel *et al.*, 2008). In naturalistic speech perception contexts, people with amusia rarely report any difficulties (Liu *et al.*, 2010). This may be because, in natural speech, pitch variation tends to co-occur with variation in other acoustic dimensions, such as duration and amplitude. Our lab has shown that in such cases where multiple redundant cues are available, English-speaking individuals with amusia tend to rely less on pitch than non-amusic controls. This suggests that they can calibrate their perception by down-weighting the cues that are less reliable for them (Jasmin *et al*., 2019).

It is unknown how decreased reliance on a particular acoustic cue during speech perception (such as pitch cues in amusia) is reflected in the brain. Previous neural studies of cue integration have focused on integration of multiple modalities, e.g. the "weighted connections" model of multisensory integration. In this model, the relative reliability of the

modalities involved with perception of a stimulus is related to differential connectivity strength (Beauchamp *et al.*, 2010; Rohe and Noppeney, 2018). For example, when participants simultaneously view and feel touches to the hand and reliability of visual and tactile perception is manipulated experimentally via introduction of noise, connection strength (effective connectivity measured with functional MRI and structural equation modeling) between unimodal and multimodal sensory areas adjusts accordingly. More concretely, when visual information is degraded, the connection strength between lateral occipital cortex (a visual area) and intraparietal sulcus (a multimodal area) decreases, and when tactile perception is made noisier, connection strength between secondary somatosensory cortex and intraparietal sulcus becomes weaker (Beauchamp et al., 2010). Similarly, effective connectivity between the (multimodal) superior temporal sulcus (STS) and visual and auditory areas has shown similar modulations during processing of audiovisual speech: connection strength between auditory cortex and the STS is weaker when noise has been introduced to the auditory speech, and conversely connection strength between visual cortex and STS is weaker if visual noise is introduced (Nath and Beauchamp, 2011).

Just as connectivity differences have been shown to reflect the precision of different dimensions during multi*sensory* integration, an analogous phenomenon may be at work within a single modality during multi*dimensional* integration. As mentioned, the acoustic speech signal carries multiple co-occurring acoustic dimensions (e.g. roughly described as pitch, duration, and amplitude), which often provide redundant cues to disambiguate a linguistic category (Patel, 2014; Winter, 2014; Jasmin *et a*l., 2019a). Individuals with typical pitch perception have learned through a lifetime of experience with speech acoustics that vocal pitch is a useful and reliable cue. By contrast, individuals with amusia, who have unreliable perception of and memory for pitch (analogous to the ‘noise’ introduced in the

multisensory integration studies cited above), would have learned that, for them, pitch is not a reliable cue for processing spoken language. Thus, by analogy to the multisensory weighting results described above, we hypothesize that amusics may exhibit decreased connectivity between language regions and pitch-related areas during speech processing.

The neural foundations of perceptual weighting in speech have thus far not been investigated in atypical individuals. Indeed, only one previous functional neuroimaging study has examined the neural processing of spoken material in people with amusia. In this study no group differences were detected in task-related activation or functional connectivity during processing of speech (whereas group differences were observed during processing of tones; Albouy *et al.*, 2018). However, this study used tasks for which pitch processing would not have been necessary (verbal memory), as well as *a priori* regions of interest derived from studies that used tonal/melodic (rather than linguistic) stimuli. It therefore may not be surprising that neural differences between amusics and controls did not emerge. It remains an open question how functional connectivity in amusic and non-amusic participants may differ during pitch-related language tasks with regions of interest selected with a whole-brain data-driven approach. Although we aim to focus on amusia here, our results may illuminate the neural basis of dimensional weighting in speech perception more generally.

As discussed, the relative reliability of senses in multisensory perception is reflected in neural connection strength—is reliability of dimensions within a sense reflected similarly? Given that amusics have unreliable perception of and memory for pitch and have low pitch cue weights during language processing (Jasmin *et al.*, 2019a), do they also exhibit correspondingly decreased functional connectivity between regions typically found to be involved in pitch processing and frontal regions associated with speech and language? Here we set out to answer these questions. We used functional magnetic resonance imaging to scan 15 individuals with amusia and 15 controls. Participants matched spoken sentences with

visually-presented ones on the basis of the position of intonational phrase boundaries. These intonation changes were conveyed differently, in three conditions: pitch cues only, duration cues only, or both these cue types together (Jasmin et al., 2019 a,b). Functional connectivity was then examined using a data-driven approach that allowed us to identify the largest group differences, without the need for regions of interest to be selected *a priori*. The benefit of this approach is that any set of regions could emerge, not only ones reported in previous literature. Crucially, task performance was matched between the groups (based on prior behavioural testing (Jasmin et al., 2019a), ensuring that any neural differences did not simply represent an inability to perform the task.

[H1] Materials and Methods

[H2] Participants

Participants, 15 individuals with amusia (10 F, age = 60.2 ± 9.4, range = 43–74) and 15 controls (10 F, age = 61.3 ± 10.4, range = 38–74), were recruited from the UK and were native British English speakers. All participants gave informed consent and ethical approval was obtained from the relevant UCL and Birkbeck ethics committees. Amusia status was obtained using the Montreal Battery for the Evaluation of Amusia (MBEA). Participants with a composite score (summing the Scale, Contour and Interval tests scores) of 65 or less were classified as having amusia (Peretz et al., 2003).

[H2] Stimuli

The stimuli were 42 compound sentences that consisted of a pre-posed subordinate clause followed by a main clause (see Fig. 1 for an example, and Jasmin et al., 2019a,b for details). There were two versions of each sentence: (1) an ‘early closure’ version, where the verb of the subordinate clause was used intransitively and the following noun was the subject of a new clause [“After Jane dusts, the dining table [is clean]”]; and (2), ‘late closure’, where the verb was transitive and took the following noun as its object, moving the phrase boundary to

a slightly later position in the sentence [“After Jane dusts the dining table, [it is clean]”]. Crucially for the task, the words in both versions of the sentence were identical from the start of the sentence until the end of the second noun (“After Jane dusts the dining table …”).

A native British English speaking male (trained as an actor) recorded early closure and late closure versions of each sentence in a sound-proofed room. The recordings were cropped such that only the portions with the same words remained, and silent pauses after phrase breaks were removed. Synthesized versions of these sentences were created with STRAIGHT voice-morphing software (Kawahara and Irino, 2005). First, the two versions of the sentence were manually time-aligned by marking corresponding ‘anchor points’ in the two recordings. Then, morphed speech was synthesized by varying the degree to which the early closure and late closure recordings contributed duration and pitch information. We synthesized pairs of stimuli in three conditions: (1) In the Pitch condition, the stimulus pair had exactly the same durational properties (that is, the length of phonemes, syllables, and words was the average between the two original recordings) but the vocal pitch indicated early or late closure at a morphing level of 80%; (2) in the Time condition, vocal pitch in the stimulus pair was identical (at 50% between both versions) but the durational characteristics indicated early or late closure at a morphing level of 80%; (3) in the Combined condition, both pitch and time cued early or late closure simultaneously at 80%. The morphed speech varied only in duration and pitch, while all other aspects of the acoustics (such as amplitude and spectral characteristics other than pitch) were the same, held constant at 50% between the two original recordings during morphing. This stimulus set is freely available (Jasmin et al., 2019b).

[H2] MRI data collection.

Subjects were scanned with a Siemens Avanto 1.5 Tesla magnetic resonance imaging scanner with a 32-channel head coil, with sounds presented via Sensimetrics S14 earbuds, padded

around the ear with NoMoCo memory foam cushions. Functional data were collected using a slow event-related design with sparse temporal sampling to allow presentation of auditory stimuli in quiet. We used an echo planar image sequence, with 40 slices, slice time 85 ms, slab tilted to capture the entire cerebrum and dorsal cerebellum, ascending sequential acquisition; 3×3×3 mm voxel size; silent stimulus and response period = 8.7s, volume acquisition time = 3.4 s, total inter-trial interval = 12.1s, flip angle = 90 degrees, bandwidth = 2298 Hz/pixel, echo time (TE) = 50ms. After collecting functional runs, a high-resolution T1-weighted structural scan was collected (MPRAGE, 176 slices, sagittal acquisition, 2x GRAPPA acceleration, 1 mm isotropic voxels, acquisition matrix = 224 × 256).

[H2] Procedure (see schematic in Fig. 1)

Each run began with three dummy scans to allow magnetic stabilization. Each trial (repetition time) lasted 12.1 seconds. The start of each trial was triggered by a pulse corresponding to the start of a volume acquisition (which acquired neural data from the previous trial, at a delay). At t=1 s into the trial, the sentence appeared on the screen; before scanning participants were instructed to read each sentence silently to themselves. At t=5 seconds (plus or minus a random 100 ms jitter) participants heard a spoken version of the first part of the sentence. At t = 7.4 seconds (plus or minus 100 ms jitter) the second version was presented. The two spoken versions contained the same words but their pitch and/or timing characteristics cued a phrase boundary that occurred earlier or later in the sentence. Following this, there were approximately 2 seconds of silence during which the participant responded with the button box, before the scanner began acquiring the next volume at t=12 s. Participants performed three blocks of 42 trials (14 each of Pitch, Time, and Combined) with 8 Rest trials interspersed within each block.

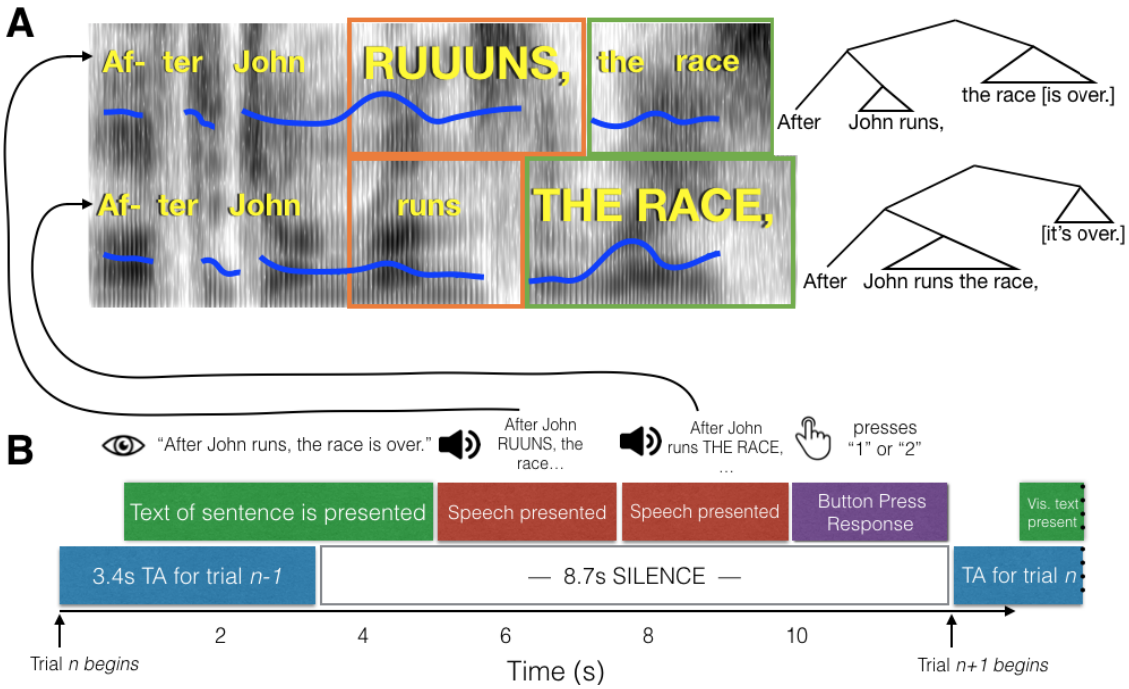

Figure 1 Caption: Schematic of experimental paradigm. (A) Example spectrograms of the early closure (top) and late closure (bottom) stimuli for the Combined condition. Fundamental frequency contours are indicated with blue lines. The relative duration of the critical words are indicated with orange and green boxes. To the right, syntactic trees for the two sentences are shown to highlight the grammatical structure indicated by the phrase boundaries. (**B)** The time course of a single trial. Participants read a text version of the sentence from the screen, which was either early or late closure. This was followed by auditory presentation of the late and early closure versions. After both recordings were played, participants chose whether the first or second recording they heard matched the visually presentence sentence better. A single whole-brain volume was acquired after the button press, timed to capture the peak of the hemodynamic response.

[H2] MRI pre-processing

Image preprocessing was performed with FreeSurfer 6.0.0 (Fischl, 2012) and AFNI-SUMA 18.1.18 (Cox, 1996). Anatomical images were registered to the third echo planar image of the first run using Freesurfer's *bbregister* and processed with FreeSurfer's automated pipeline for segmenting tissue types, generating cortical surface models, and parcellating subcortical structures. Masks of inferior colliculi were obtained by manually examining individual

subjects' anatomical images and selecting a single EPI voxel located at its centre, bilaterally. Freesurfer cortical surface models were imported to AFNI with the @SUMA_Make_Spec_FS program. Then a standard pre-processing pipeline using AFNI's *afniproc.py* program was used: all echo planar image volumes were aligned to the third repetition time of the first run using AFNI's 3DAllineate, intersected with the cortical surface with SUMA, smoothed along the surface with a 2D 6-mm-FWHM kernel, and converted to a standard mesh (std.141) for group analyses, separately for each hemisphere, where each vertex in the mesh (198812 per hemisphere) is aligned to the 'same' location in the cortex across subjects, using curvature-based morphing.

[H2] Motion

The magnitude of transient head motion was calculated from the six motion parameters obtained during image realignment and aggregated as a single variable using AFNI's @1dDiffMag to calculate a Motion Index (Berman *et al.*, 2016; Gotts *et al.*, 2012; Jasmin *et al.*, 2019c). This measure is similar to average Frame Displacement over a scan (Power *et al.*, 2012) and is in units of mm per repetition time. The difference in average motion between the groups was small (amusia group mean motion = .31mm/TR; control group mean = .28mm/TR) and amounted to 32 micrometers (~1/30th of a millimeter) per TR. The mean and distribution of motion did not differ statistically between groups (two sample *t*-test $P = 0.7$, two-tailed).

[H2] Beta series analysis of context-modulated functional connectivity

Given the previous reports (described above) of changes in connection strength between unimodal and multimodal areas in response to noise (Beauchamp, *et al.*, 2010; Nath and Beauchamp, 2011), we chose a connectivity-based analysis approach for our study. *Beta series correlation* (Rissman *et al.*, 2004) is a technique for examining functional connectivity and its modulation by task, using correlations in trial-by-trial responses. It has been shown to

be more powerful than alternatives such as generalized psycho-physiological interaction (gPPI) for event-related designs (Cisler *et al.*, 2014). In a beta series analysis, one beta weight is calculated for each trial in the experiment (rather than for each condition). All of the trial-wise betas associated with a given condition are then ordered to form a "beta series". Finally, using the beta series in the same way as a standard BOLD fMRI time series, functional connectivity (measured as Pearson correlations) is calculated between seed regions of interest and the rest of the brain. Differences in functional connectivity can then be examined by comparing groups, comparing conditions, or examining the interaction of these factors.

[H2] Obtaining trial-wise beta weights

Our experiment used a slow event-related design with a long repetition time (12.1s) and sparse temporal sampling (with volume acquisition separated by silent periods). Therefore, the time between acquisitions was long enough for the haemodynamic response to return to baseline, and each echo planar image acquisition corresponded to exactly one trial (Fig. 1). For this reason, we did not convolve the echo planar image time series with a basis function during subject-level statistical analysis (Hall et al., 1999). In the design matrix for obtaining trial-wise betas, 126 column regressors were used (one for each non-rest trial). Each column vector was of length 150 (corresponding to all trials, including rest trials) and had a single "one" in the position where the trial associated with that column occurred, while zeros were located in every other position. Polynomials up to second degree were also included in the model, on a run-wise basis, to remove the mean and any linear or quadratic trends. Fitting the trial regressors on a subject-wise basis resulted in cortical surface models of beta weights for each of the 126 trials, at each vertex on the reduced-vertex icosahedral cortical surface, with beta weights reflecting the neural response associated with that trial. As noted above, trial-wise betas were then serially ordered to form beta series for each of the three experimental conditions (Pitch, Time, and Combined) (Rissman *et al.*, 2004). In total, 90 beta series were

created for each voxel, corresponding to each of the three task conditions (Pitch, Time, Combined) from all 30 participants.

[H2] Defining seed regions of interest

Beta series analysis requires initial seed voxels, vertices, or regions to be identified, whose trial-to-trial changes in activity are then compared to those of the rest of the brain. Rather than choose *a priori* seeds derived from the literature, which used mainly musical tasks or resting state, we used a data-driven approach to search for the largest group and condition differences in functional connectivity (Berman *et al.*, 2016; Cole *et al.*, 2010; Gotts *et al.*, 2012; Jasmin *et al.*, 2019c; Meoded *et al.*, 2015; Song *et al.*, 2015; Steel *et al.*, 2016; Stoddard *et al.*, 2016; Watsky *et al.*, 2018). To do this, we first calculated the "whole-brain connectedness" (similar to 'centrality' in graph theory) of each cortical vertex (a procedure available in AFNI as the *3dTCorrMap* function). The whole-brain connectedness of a given vertex is defined as the Pearson correlation of activity within that vertex/voxel and the average signal across all neural gray matter in the rest of the brain. Mathematically, this is equivalent to calculating thousands of Pearson correlations, of a given vertex/voxel series and every other vertex/voxel series in the brain, and then taking the mean of those correlations (Cole *et al.*, 2010), then repeating the process for every individual voxel/vertex. As such, it represents the global connectedness (or 'global correlation') of a vertex/voxel.

To calculate whole-brain connectedness, first, the average of trial-wise betas in gray matter across the brain was calculated in volume space, separately for each subject and for each condition (Pitch, Time, Combined) by running first-level (subject) models. The statistical models were identical to those conducted on the cortical surface, described above, but were performed on volumetric Talairach images instead of the cortical surfaces. The reason for this choice was so that voxels in cortex and subcortex would contribute equally to our measure of global (whole-brain) connectivity. First, average gray-matter beta value was

calculated for each trial by intersecting each image in the beta series with a whole-brain gray matter mask (which excluded white matter and ventricles) and calculating the average beta value within the mask (Gotts *et al.*, 2012; Jasmin *et al.*, 2019c). Next, this gray matter average was correlated with each cortical surface vertex's beta series, separately for each subject and condition, to obtain whole-brain connectedness maps. These values were then subjected to a statistical analysis based on our 2 (Group) × 3 (Condition) experimental design. Linear mixed effects models (AFNI's *3dLME*) (Chen *et al.*, 2013) were constructed whose dependent variables were the vertex-wise whole-brain connectedness maps from each beta series. Group and Condition and their interaction were included as fixed effects. Participant was treated as a random intercept. Results of this step were corrected vertex-wise for multiple comparisons with False Discovery Rate ($q < 0.05$), separately for each test (Main Effect of Group; Main Effect of Condition; Interaction of Group by Condition) by pooling the *p*-values from both hemispheres' cortical surfaces. This False Discovery Rate threshold corresponded to uncorrected $p < 4\times10^{-6}$ for the Main Effect of Group. Four significant results (contiguous significant vertices) survived this threshold and were taken forward for the next analysis step. For the Main Effect of Condition and Interaction of Condition x Group, no results survived statistical correction FDR ($q < 0.05$).

A similar procedure was performed for subcortical structures. Beta series were obtained for each subject, structure, and experimental condition, from their standard Freesurfer subcortical parcellations by masking the EPI data within each structure and calculating the average of the voxels. Each structure's beta series was then correlated with the whole-brain gray matter beta average, separately for each condition, and the resulting values were subjected to linear mixed effects models with the same factors as above. Tests for Main Effect of Condition, of Group, and the Interaction of these factors was performed. All *p*-

values were greater than $p > 0.001$ and no results survived an FDR-correction calculated over them.

[H2] Follow-up seed-to-whole-brain testing

The first analysis step (seed definition, described above) identified which, if any, brain areas showed the largest connectivity differences between groups. However, this step is insufficient to localize the other specific regions driving this pattern. An analogy is in Analysis of Variance, where a significant omnibus test indicates a difference exists, but follow-up testing is required to determine where in the model differences exist (Gotts *et al.*, 2012). Thus, to locate the regions driving this pattern, we undertook a second step: follow-up seed-to-whole-brain testing (Cole *et al.*, 2010; Gotts *et al.*, 2012; Jasmin *et al.*, 2019c). Each seed region was examined with respect to its connectivity pattern with every cortical vertex and subcortical structure.

For each of the 90 beta series (30 subjects by three conditions), values within the seed vertices was averaged and then correlated with the beta series for every vertex in the brain. These correlations were Fisher z-transformed and used as the dependent variables in linear mixed effects models (*3dLME*) with the same fixed and random effects as above. For each of the seeds, we tested for the group difference (Amusia vs Control) in connectivity. Results were False Discovery Rate corrected to ($q < 0.05$) across all eight follow-up tests [4 seeds × 2 hemispheres] corresponding to a threshold of $p < 0.00035$. Similar, for the subcortical structures, each seed beta series was correlated with subcortical structure beta series, with resulting values subject to statistical testing. An FDR correction over all tests involving subcortex was applied. For display in figures, the data were converted from SUMA's standard mesh (std.141) to Freesurfer's standard surface (fsaverage) using AFNI's SurfToSurf program and mapping values from the closest nodes (i.e. vertices).

[H2] Analysis of activation

A standard General Linear Model comparing activation strength (rather than connectivity) was also conducted. As in the General Linear Model for obtaining beta weights, no basis function was used, and polynomials up to second degree were included in the models.

[H2] Data availability

The data that support the findings of this study are openly available in the Birkbeck repository.

[H1] Results

[H2] In-scanner Behavior

On each trial, participants read one visually presented text sentence, then heard two auditory versions of the sentence, only one of which contained an acoustically-conveyed phrase boundary in the same place as in the text sentence. Trials were scored as correct if a participant pressed the button associated with the auditory sentence that correctly matched the text sentence. Proportions of correct judgments were subjected to a repeated-measures Analysis of Variance. Overall, proportion correct across amusia and control groups was matched (main effect of Group, $F(1,84) = 0.16$, $P = 0.69$), interaction of Group by Condition ($F(2,84) = 0.374$, $P = 0.96$). This lack of interaction was predicted based on previous results obtained from a similar paradigm using out-of-scanner data but from the same participants (Jasmin et al., 2019a). There was a main effect of condition ($F(2,84) = 3.32$, $p = 0.04$. Follow-up post-hoc testing indicated that performance in the Combined condition (with pitch and duration present in combination) was more accurate than either Pitch-Alone ($T(84) = 2.3$, $p = 0.02$) or Time-Alone ($T(84) = 2.1$, $p = 0.03$), a result that was also predicted and which replicates the behavioral findings in Jasmin *et al.* (2019a).

[H2] Neuroimaging - whole-brain connectedness

As discussed in the Methods, a data-driven approach was taken to identify brain regions with the largest group- and condition-related differences in functional connectivity. Comparing

whole-brain connectedness values by group (Amusia vs. Controls) revealed four significant locations (where *z* of peak vertices > 4.61, FDR-corrected 0.05) that showed greater whole-brain connectedness for the control than amusia group (see Fig. 2). All group differences were located in the inferior frontal cortex: two left hemisphere vertices (inferior frontal gyrus *p. triangularis,* and dorsolateral prefrontal cortex); and two right hemisphere vertices (inferior frontal gyrus *p. triangularis* and *p orbitalis*). There were no areas where whole-brain connectedness differed by Condition, or showed an interaction of Group and Condition.

Figure 2: Seed locations. Inflated surfaces show the locations of False Discovery Rate-corrected group differences (Control > Amusia) in whole-brain connectivity, which were used as seeds in subsequent analyses. All four seed vertices were located in inferior frontal cortices (left inferior frontal gyrus; left DLPFC, right inferior frontal gyrus *pars triangularis*. and right inferior frontal gyrus *orbitalis*)

[H2] Follow-up seed-to-whole brain tests

Follow-up testing was conducted on the four significant regions (Control > Amusia, collapsed across the three conditions) identified above to characterize the specific cortical regions driving these group connectivity differences (Berman *et al.*, 2016; Gotts *et al.*, 2012; Jasmin *et al.*, 2018; Song *et al.*, 2015). Relative to control participants, amusic participants' left inferior frontal gyrus seed region showed particularly notable decreases in connectivity with the right posterior superior temporal and inferior parietal cortex, as well as with the right

posterior superior temporal sulcus (Fig. 3). Analysis of subcortical connectivity indicated that there was also weaker connectivity with the right nucleus accumbens.

The left dorsolateral prefrontal cortex in amusic participants showed decreased functional connectivity with the mid portions of the right superior temporal gyrus, posterior part of the right middle temporal gyrus extending into the inferior bank of the superior temporal sulcus, and the right anterior insula (Fig. 3). Several subcortical structures also showed significantly reduced (FDR-corrected) connectivity with the seed in amusics: bilateral caudate nucleus and putamen, bilateral pallidum, bilateral cerebellum, and bilateral thalamus.

The right *pars triangularis* seed showed Control > Amusic connectivity with right dorsolateral prefrontal cortex and left posterior superior temporal gyrus (Fig. 4). It also showed decreased connectivity with left nucleus accumbens. Right *pars orbitalis* showed decreased connectivity with right dorsolateral prefrontal cortex (Fig. 4). There was also decreased connectivity with the left thalamus.

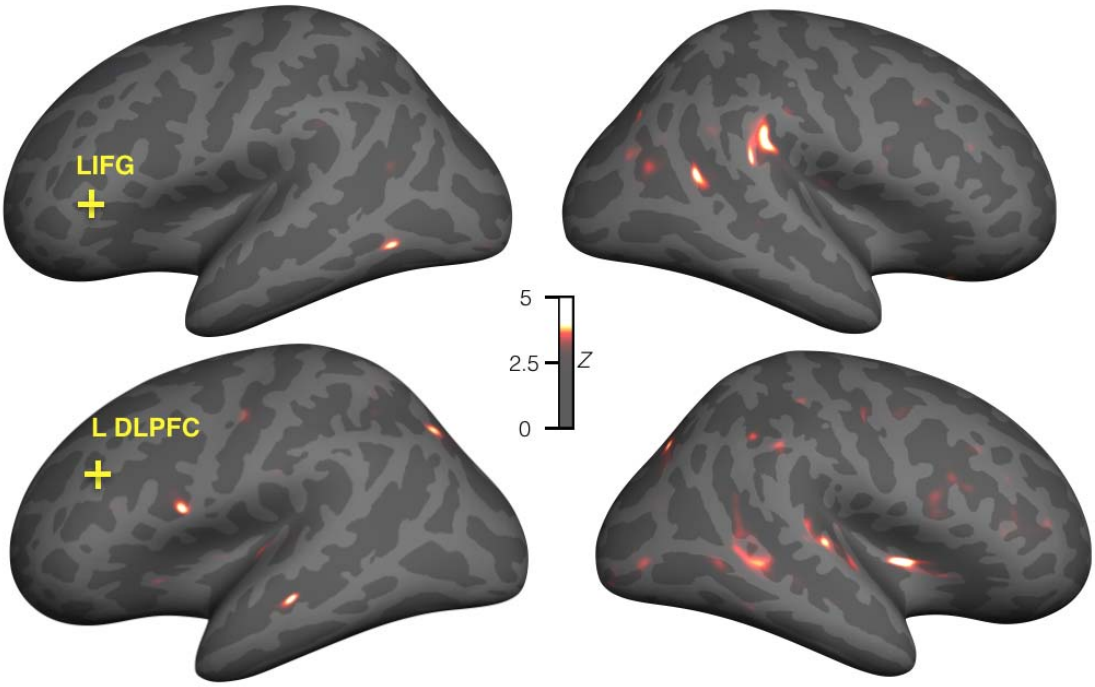

Figure 3: Individuals with amusia show decreased connectivity between two regions in left inferior frontal cortex and right auditory and pitch-processing areas during a language task. Significant group differences (control > amusia) in functional connectivity with left hemisphere seeds, False Discovery Rate corrected, minimum Z=3.57). Warm colors indicate greater connectivity in the control than amusia participants. The largest decreases in connectivity in the amusia group were located in right Heschl's gyrus, superior temporal plane and gyrus, the posterior middle temporal gyrus onto the inferior bank of the superior temporal sulcus, and anterior insula.

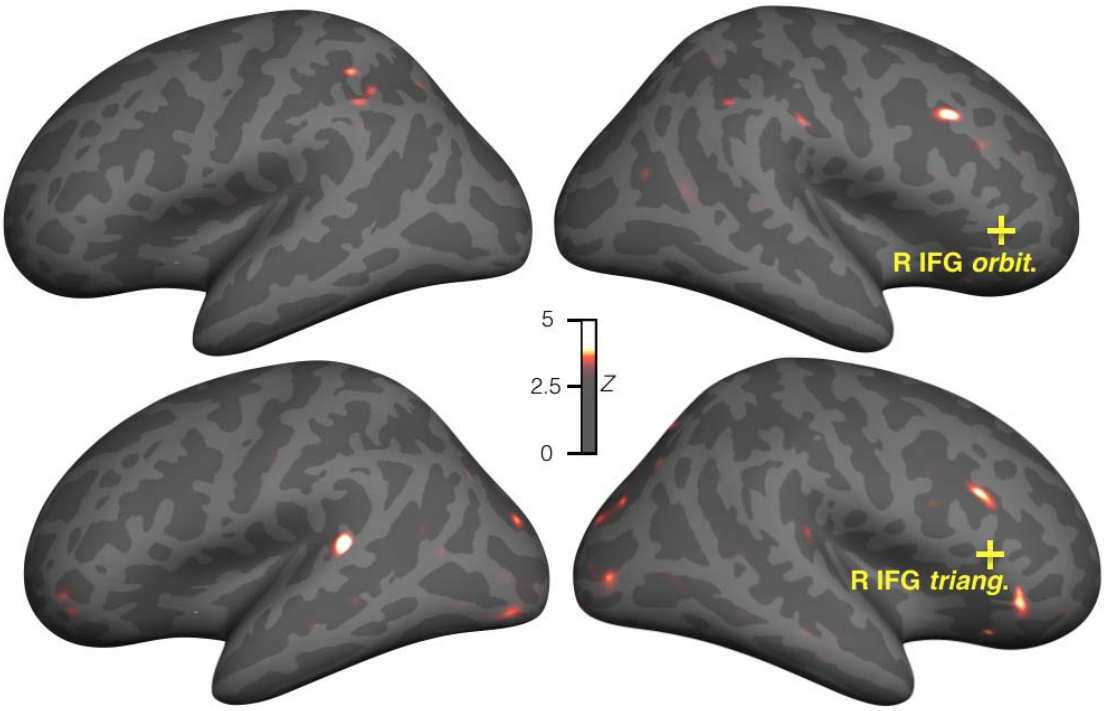

Figure 4: Individuals with amusia show decreased connectivity between right inferior frontal cortex and many other brain areas. Significant group differences (control vs amusia) in functional connectivity with right hemisphere seeds, False Discovery Rate corrected, minimum Z=3.57). Warm colors indicate greater connectivity in the control than amusia participants. Prominent decreases in connectivity with the right inferior frontal gyrus in individuals with amusia included superior temporal plane, and areas of occipital, frontal, and parietal cortex.

[H2] Activation Results

Although we were concerned with functional connectivity rather than activation, we also tested for differences in activation levels between groups and conditions. False Discovery Rate correction was used to correct for multiple comparisons across both hemispheres for each test (Group, Condition and Group X Condition). No significant differences were detected for the main effects of group and condition, nor the interaction of those factors. Results images from these analyses are available online (see Data Availability Statement for details).

[H1] Discussion

We found that individuals with amusia, who have been previously shown to rely less on pitch than controls to process spoken language (Jasmin et al. 2019a), exhibit decreased functional connectivity between left frontal areas and right hemisphere pitch-related regions. In our task, participants matched spoken sentences with visually presented sentences based on pitch, duration, or both these acoustic dimensions together. Using a data-driven approach, we identified four regions in left and right inferior frontal cortex for which the amusic group exhibited decreased functional connectivity with several other sites in frontal, temporal and occipital cortex. The most prominent of these results was decreased connectivity between left frontal regions classically implicated in language processing (left IFG and DLPFC) and right hemisphere regions —in the superior temporal gyrus and sulcus, Heschl's gyrus, and anterior insula—that have been implicated in pitch processing (Lee et al., 2011; Garcea et al., 2017; Warren et al., 2003). We suggest that this decreased connectivity between right hemisphere pitch and left hemisphere frontal cortices may relate to the unreliability of the amusics' perception of and memory for pitch. This is similar to the "weighted connections" model of multisensory integration, where a more (or less) reliable modality is given a stronger (or weaker) weight. (Beauchamp et al., 2010).

Congenital amusia is often described as a disorder related to structural and functional connectivity within the right hemisphere, particularly between right inferior frontal cortices and right posterior temporal cortex (see Peretz, 2016 for review). Consistent with this proposal, we found in the present study that right inferior frontal cortex exhibited strongly decreased functional connectivity in the amusia group, and follow-up seed testing revealed that right auditory areas were involved as well. However, we also found that sites in *left* frontal cortex also showed large decreases in connectivity in amusia, also most prominently with right hemisphere auditory areas. Our results are consistent with an account that right

hemisphere auditory areas are not only abnormally connected to right frontal areas (as observed during tonal tasks), but are less integrated with left hemisphere when processing speech and language.

Our null results for group differences in activation are consistent with prior reports that amusics and controls do not differ in pitch representations within sensory regions. For example, the extent of pitch-responsive regions within auditory cortex has been shown to be similar in participants with amusia and controls (Norman-Haignere et al. 2016). Brainstem encoding of pitch in speech and musical stimuli is similarly unimpaired in individuals with amusia (Liu et al., 2015). Moreover, in oddball EEG paradigms, amusics show similar pre-attentive mismatch negativity responses to small pitch deviants, but impaired attention-dependent P300 responses (Moreau et al. 2009; Peretz et al. 2009; Goulet et al. 2012; Moreau et al. 2013). These findings, along with the fact that amusics show intact non-volitional behavioral responses (unconscious pitch shifts) when presented with pitch-altered feedback of their own voice (Hutchins and Peretz 2013), have been interpreted as evidence that amusia is a disorder of pitch awareness rather than one of low-level pitch processing (Peretz et al. 2009), with differences in structural connectivity as one possible foundation of this putative impaired pitch awareness (Hyde et al. 2006; Loui et al. 2009; but see Chen et al. 2015).

Our interpretation of differences in functional connectivity between amusics and controls diverges somewhat from these previous approaches: we argue that down-weighting of pitch information during perceptual categorization in both speech and music is adaptive, inasmuch as amusics have learned that pitch is an unreliable source of evidence relative to other perceptual dimensions. The evidence above suggesting that encoding of pitch in the brainstem and auditory cortex and pre-attentive responses to pitch changes are unaffected in amusia can be interpreted as suggesting that the fundamental deficit in amusia may not be increased perceptual noise or decreased pitch awareness but difficulties with retention of

pitch information in memory. This interpretation is consistent with evidence suggesting that amusics have difficulty with pitch sequence processing tasks even when discrimination thresholds are accounted for (Tillmann et al. 2009), as well as the finding that delaying the time interval between standard and comparison tones exacerbates pitch discrimination impairment in individuals with amusia (Williamson et al. 2010). Moreover, the pitch awareness account of amusia cannot explain the Jasmin et al. (2019a) finding that pitch cues are downweighted only during longer-scale suprasegmental speech perception, while pitch weighting is not different between amusics and controls during shorter-scale segmental speech perception, despite pitch cues being arguably more subtle in the segmental condition. However, this finding can be explained by the pitch memory account, as the suprasegmental task requires detection of and memory for pitch patterns within a complex sequence, while the segmental task does not. Furthermore, an account of amusia which suggests that the disorder primarily stems from differences in structural connectivity cannot account for the recent finding that functional connectivity patterns do not differ between amusics and controls during a verbal memory task (Albouy et al. 2018). We suggest, therefore, that amusics neglect pitch because they have implicitly learned that their memory for pitch is unreliable, and that this down-weighting of pitch is reflected in decreased functional connectivity between right auditory areas and downstream task-relevant areas which integrate information from perceptual regions. One way to test this hypothesis would be to examine functional connectivity during perceptual categorization of consonant-vowel syllables as voiced versus unvoiced based on a pitch cue (F0 of the following vowel) and a durational cue (voice onset time). We predict, based on our previous findings (Jasmin et al. 2019a), that functional connectivity will not differ between amusics and controls on this task, a finding which would not be predicted by the pitch awareness account of amusia.

Several other future directions are suggested by our results, particularly for examining cue weighting during auditory/speech perception. In the multimodal integration studies mentioned above (Beauchamp et al., 2010; Nath and Beauchamp. 2011), reliability of two different sensory modalities was manipulated experimentally, resulting in changes in connectivity. Similarly, aspects of speech could be selectively masked with noise in order to make them less reliable, which in turn could cause corresponding changes in functional or effective connectivity. Indeed, behavioral work has indicated that when fundamental frequency (pitch) or durational aspects of speech are manipulated to be unreliable cues, categorization behavior shifts such that participants place less relative weight on the dimension that has been made less reliable (Holt & Lotto, 2006). Certain groups, such as tone language speakers, are known to have fine-grained pitch perception abilities, and tend to place greater weight on pitch even when processing speech from a second, non-tonal language that they have learned (e.g. English; Yu, et al., 2010; Zhang et al, 2010, Zhang et al., 2008; Qin et al., 2017). Given the increased reliability of their pitch perception, tone language speakers may exhibit correspondingly high connectivity strength between right hemisphere auditory regions and left hemisphere 'language regions' when pitch cues are present (more so than native non-tonal language speakers).

Figure 1 Caption: Schematic of experimental paradigm. (A) Example spectrograms of the early closure (top) and late closure (bottom) stimuli for the Combined condition. Fundamental frequency contours are indicated with blue lines. The relative duration of the critical words are indicated with orange and green boxes. To the right, syntactic trees for the two sentences are shown to highlight the grammatical structure indicated by the phrase boundaries. (**B)** The time course of a single trial. Participants read a text version of the sentence from the screen, which was either early or late closure. This was followed by auditory presentation of the late and early closure versions. After both recordings were played, participants chose whether the first or second recording they heard matched the visually presentence sentence better. A single whole-brain volume was acquired after the button press, timed to capture the peak of the hemodynamic response.

Figure 2: Seed locations. Inflated surfaces show the locations of False Discovery Rate-corrected group differences (Control > Amusia) in whole-brain connectivity, which were used as seeds in subsequent analyses. All four seed vertices were located in inferior frontal cortices (left inferior frontal gyrus; left DLPFC, right inferior frontal gyrus *pars triangularis.* and right inferior frontal gyrus *orbitalis*)

Figure 3: Individuals with amusia show decreased connectivity between two regions in left inferior frontal cortex and right auditory and pitch-processing areas during a language task. Significant group differences (control > amusia) in functional connectivity with left hemisphere seeds, False Discovery Rate corrected, minimum Z=3.57). Warm colors indicate greater connectivity in the control than amusia participants. The largest decreases in connectivity in the amusia group were located in right Heschl's gyrus, superior temporal plane and gyrus, the posterior middle temporal gyrus onto the inferior bank of the superior temporal sulcus, and anterior insula.

Figure 4: Individuals with amusia show decreased connectivity between right inferior frontal cortex and many other brain areas. Significant group differences (control vs amusia) in functional connectivity with right hemisphere seeds, False Discovery Rate corrected, minimum Z=3.57). Warm colors indicate greater connectivity in the control than amusia participants. Prominent decreases in connectivity with the right inferior frontal gyrus in individuals with amusia included superior temporal plane, and areas of occipital, frontal, and parietal cortex.

[H1] Acknowledgments

We would like to thank our study participants.

[H1] Funding

This study was funded by Seed Grant No. 109719/Z/15/Z from The Wellcome Trust to A.T.T., a Reg and Molly Buck Award from the Society for Education, Music and Psychology Research to K.J., and an Early Career Fellowship from the Leverhulme Trust (ECF-2017-151) to K.J.

[H1] Competing Interests

The authors have no competing interests to declare.

[H1] References

Albouy P, Mattout J, Bouet R, Maby E, Sanchez G, Aguera PE, Daligault S, Delpuech C, Bertrand O, Caclin A, Tillmann B. Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. Brain. 2013 Apr 23;136(5):1639-61.

Albouy P, Peretz I, Bermudez P, Zatorre RJ, Tillmann B, Caclin A. Specialized neural dynamics for verbal and tonal memory: fMRI evidence in congenital amusia. Hum. Brain Mapp. 2018; 40: 855–867.

Beauchamp MS, Pasalar S, Ro T. Neural substrates of reliability-weighted visual-tactile multisensory integration. Frontiers in Systems Neuroscience. 2010 Jun 23;4:25.

Berman RA, Gotts SJ, McAdams HM, Greenstein D, Lalonde F, Clasen L, et al. Disrupted sensorimotor and social–cognitive networks underlie symptoms in childhood-onset schizophrenia. Brain 2016; 139: 276–291.

Chen JL, Kumar S, Williamson VJ, Scholz J, Griffiths TD, Stewart L. Detection of the arcuate fasciculus in congenital amusia depends on the tractography algorithm. Frontiers in psychology. 2015 Jan 21;6:9.

Cisler JM, Bush K, Steele JS. A comparison of statistical methods for detecting context-modulated functional connectivity in fMRI. NeuroImage 2014; 84: 1042–1052.

Cole MW, Pathak S, Schneider W. Identifying the brain's most globally connected regions. NeuroImage 2010; 49: 3132–3148.

Cox RW. AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. Computers and Biomedical Research 1996; 29: 162–173.

de Pijper JR, Sanderman AA. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. The Journal of the Acoustical Society of America 1994; 96: 2037–2047.

Fischl B. FreeSurfer. NeuroImage 2012; 62: 774–781.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. Psychological Bulletin, 97(3), 412-429. http://dx.doi.org/10.1037/0033-2909.97.3.412

Garcea FE, Chernoff BL, Diamond B, Lewis W, Sims MH, Tomlinson SB, Teghipco A, Belkhir R, Gannon SB, Erickson S, Smith SO. Direct electrical stimulation in the human brain disrupts melody processing. Current Biology. 2017 Sep 11;27(17):2684-91.

Goulet GM, Moreau P, Robitaille N, Peretz I. Congenital amusia persists in the developing brain after daily music listening. PLoS One. 2012 May 11;7(5):e36860.

Gotts SJ, Simmons WK, Milbury LA, Wallace GL, Cox RW, Martin A. Fractionation of social brain circuits in autism spectrum disorders. Brain. 2012 Jul 11;135(9):2711-25.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. “Sparse” temporal sampling in auditory fMRI. Human brain mapping. 1999;7(3):213-23.

Holt LL, Lotto AJ. Speech perception as categorization. Attention, Perception, & Psychophysics. 2010 Jul 1;72(5):1218-27.

Hutchins S, Peretz I. Amusics can imitate what they cannot discriminate. Brain and language. 2012 Dec 1;123(3):234-9.

Hyde KL, Zatorre RJ, Griffiths TD, Lerch JP, Peretz I. Morphometry of the amusic brain: a two-site study. Brain. 2006 Aug 24;129(10):2562-70.

Hyde KL, Zatorre RJ, Peretz I. Functional MRI Evidence of an Abnormal Neural Network for Pitch Processing in Congenital Amusia. Cerebral Cortex 2011; 21: 292–299.

Jasmin K, Dick F, Holt LL, & Tierney A. Tailored perception: Individuals’ speech and music perception strategies fit their perceptual abilities. Journal of Experimental Psychology: General. 2019a Advance online publication. http://dx.doi.org/10.1037/xge0000688

Jasmin K, Dick F, Tierney A. The Multidimensional Battery of Prosody Perception (MBOPP). bioRxiv 2019b; doi: https://doi.org/10.1101/555102

Jasmin K, Gotts SJ, Xu Y, Liu S, Riddell CD, Ingeholm JE, Kenworthy L, Wallace GL, Braun AR, Martin A. Overt social interaction and resting state in young adult males with autism: core and contextual neural features. Brain. 2019c Jan 28;142(3):808-22.

Kawahara H, Irino T. Underlying Principles of a High-quality Speech Manipulation System STRAIGHT and Its Application to Speech Segregation. In: Speech Separation by Humans and Machines. Boston: Kluwer Academic Publishers; 2005. p. 167–180.

Lee YS, Janata P, Frost C, Hanke M, Granger R. Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. Neuroimage. 2011 Jul 1;57(1):293-300.

Leveque Y, Fauvel B, Groussard M, Caclin A, Albouy P, Platel H, et al. Altered intrinsic connectivity of the auditory cortex in congenital amusia. Journal of Neurophysiology 2016; 116: 88–97.

Liu F, Jiang C, Wang B, Xu Y, Patel AD. A music perception disorder (congenital amusia) influences speech comprehension. Neuropsychologia 2015; 66: 111–118.

Liu F, Jiang C, Wang B, Xu Y, Patel AD. A music perception disorder (congenital amusia) influences speech comprehension. Neuropsychologia 2015; 66: 111–118.

Liu F, Maggu AR, Lau JCY, Wong PCM. Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. Front. Hum. Neurosci. 2015; 8: 35.

Loui P, Alsop D, Schlaug G. Tone deafness: a new disconnection syndrome?. Journal of Neuroscience. 2009 Aug 19;29(33):10215-20.

Liu F, Patel AD, Fourcin A, Stewart L. Intonation processing in congenital amusia:

discrimination, identification and imitation. Brain 2010; 133: 1682–1693.

Meoded A, Morrissette AE, Katipally R, Schanz O, Gotts SJ, Floeter MK. Cerebro-cerebellar connectivity is increased in primary lateral sclerosis. NeuroImage: Clinical 2015; 7: 288–296.

Moreau P, Jolicœur P, Peretz I. Automatic brain responses to pitch changes in congenital amusia. Annals of the New York Academy of Sciences. 2009 Jul 1;1169(1):191-4.

Moreau P, Jolicœur P, Peretz I. Pitch discrimination without awareness in congenital amusia: Evidence from event-related potentials. Brain and Cognition 2013; 81: 337–344.

Nath AR, Beauchamp MS. Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. Journal of Neuroscience. 2011 Feb 2;31(5):1704-14.

Norman-Haignere SV, Albouy P, Caclin A, McDermott JH, Kanwisher NG, Tillmann B. Pitch-Responsive Cortical Regions in Congenital Amusia. Journal of Neuroscience 2016; 36: 2986–2994.

Patel AD. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. Hearing research. 2014 Feb 1;308:98-108.

Patel AD, Wong M, Foxton J, Lochy A, Peretz I. Speech intonation perception deficits in musical tone deafness (congenital amusia). Music Perception: An Interdisciplinary Journal 2008; 25: 357–368.

Peretz I, Brattico E, Järvenpää M, Tervaniemi M. The amusic brain: in tune, out of key, and unaware. Brain. 2009 Mar 31;132(5):1277-86.

Peretz I, Champod AS, Hyde K. Varieties of musical disorders: the Montreal Battery of Evaluation of Amusia. Annals of the New York Academy of Sciences. 2003 Nov;999(1):58-75.

Peretz I, Ayotte J, Zatorre RJ, Mehler J, Neuron PA, Penhune VB, et al. Congenital amusia: a disorder of fine-grained pitch discrimination. Neuron 2002; 33: 185–191.

Peretz I. Neurobiology of Congenital Amusia. Trends in Cognitive Sciences 2016; 20: 857–867.

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage 2012; 59: 2142–2154.

Qin Z, Chien Y, Tremblay A. Processing of word-level stress by Mandarin-speaking second language learners of English. Applied Psycholinguistics 2017; 38:541-570

Rohe T, Noppeney U. Reliability-weighted integration of audiovisual signals can be modulated by top-down attention. eNeuro. 2018 Jan;5(1).

Rissman J, Gazzaley A, D'Esposito M. Measuring functional connectivity during distinct stages of a cognitive task. NeuroImage 2004; 23: 752–763.

Song S, Gotts SJ, Dayan E, Cohen LG. Practice Structure Improves Unconscious Transitional Memories by Increasing Synchrony in a Premotor Network. Journal of Cognitive Neuroscience 2015; 27: 1503–1512.

Steel A, Song S, Bageac D, Knutson KM, Keisler A, Saad ZS, et al. Shifts in connectivity during procedural learning after motor cortex stimulation: A combined transcranial magnetic stimulation/functional magnetic resonance imaging study. CORTEX 2016; 74: 134–148.

Stoddard J, Gotts SJ, Brotman MA, Lever S, Hsu D, Zarate C, et al. Aberrant intrinsic functional connectivity within and between corticostriatal and temporal–parietal networks in adults and youth with bipolar disorder. Psychological Medicine 2016; 46: 1509–1522.

Streeter LA. Acoustic determinants of phrase boundary perception. The Journal of the Acoustical Society of America 1978; 64: 1582–1592.

Tillmann B, Schulze K, Foxton JM. Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. Brain and cognition. 2009 Dec 1;71(3):259-64.

Warren JD, Uppenkamp S, Patterson RD, Griffiths TD. Separating pitch chroma and pitch height in the human brain. Proceedings of the National Academy of Sciences. 2003 Aug 19;100(17):10038-42.

Watsky RE, Gotts SJ, Berman RA, McAdams HM, Zhou X, Greenstein D, et al. Attenuated resting-state functional connectivity in patients with childhood- and adult-onset schizophrenia. Schizophrenia Research 2018

Whiteford KL, Oxenham AJ. Learning for pitch and melody discrimination in congenital amusia. Cortex. 2018 Jun 1;103:164-78.

Williamson VJ, McDonald C, Deutsch D, Griffiths TD, Stewart L. Faster decline of pitch memory over time in congenital amusia. Advances in Cognitive Psychology. 2010;6:15.

Winter B. Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. BioEssays. 2014 Oct;36(10):960-7.

Yu V, Andruski J. A cross-language study of perception of lexical stress in English. J Psycholinguist Res 2010l; 39:323-344.

Zendel BR, Lagrois MÉ, Robitaille N, Peretz I. Attending to pitch information inhibits processing of pitch information: the curious case of amusia. Journal of Neuroscience. 2015 Mar 4;35(9):3815-24.

Zhang Y, Francis A. The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. Journal of Phonetics 2010; 38:260-271.

Zhang Y, Nissen S, Francis A. Acoustic characteristics of English lexical stress produced by

native Mandarin speakers. JASA 2008; 123:4498-4513.