

Classifying song and speech: effects of focal temporal lesions and musical disorder

Julia Merrill, Marc Bangert, Daniela Sammler & Angela D. Friederici

To cite this article: Julia Merrill, Marc Bangert, Daniela Sammler & Angela D. Friederici (2016) Classifying song and speech: effects of focal temporal lesions and musical disorder, *Neurocase*, 22:6, 496-504, DOI: [10.1080/13554794.2016.1237660](https://doi.org/10.1080/13554794.2016.1237660)

To link to this article: <http://dx.doi.org/10.1080/13554794.2016.1237660>

View supplementary material

Published online: 11 Oct 2016.

Submit your article to this journal

Article views: 74

View related articles

View Crossmark data

Classifying song and speech: effects of focal temporal lesions and musical disorder

Julia Merrill^{a,b,c}, Marc Bangert^{a,d}, Daniela Sammler^{a,e} and Angela D. Friederici^a

^aDepartment of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; ^bMusic Department, Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany; ^cInstitute of Music, University of Kassel, Kassel, Germany; ^dInstitute of Musicians' Medicine, Dresden University of Music, Dresden, Germany; ^eOtto Hahn Group "Neural Bases of Intonation in Speech and Music", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ABSTRACT

Song and speech represent two auditory categories the brain usually classifies fairly easily. Functionally, this classification ability may depend to a great extent on characteristic features of pitch patterns present in song melody and speech prosody. Anatomically, the temporal lobe (TL) has been discussed as playing a prominent role in the processing of both. Here we tested individuals with congenital amusia and patients with unilateral left and right TL lesions in their ability to categorize song and speech. In a forced-choice paradigm, specifically designed auditory stimuli representing sung, spoken and "ambiguous" stimuli (being perceived as "halfway between" song and speech), had to be classified as either "song" or "speech". Congenital amusics and TL patients, contrary to controls, exhibited a surprising bias to classifying the ambiguous stimuli as "song" despite their apparent deficit to correctly process features typical for song. This response bias possibly reflects a strategy where, based on available context information (here: forced choice for either speech or song), classification of non-processable items may be achieved through elimination of processable classes. This speech-based strategy masks the pitch processing deficit in congenital amusics and TL lesion patients.

ARTICLE HISTORY

Received 3 March 2016
Accepted 12 September 2016

KEYWORDS

Song; speech; amusia; lesion; temporal lobe

1. Introduction

Song and speech are the two major forms of structured human vocalization; most of the time they are easy to distinguish, even though they share many acoustic features: song as well as speech consist of sequential acoustic patterns which show orderly variations of pitch (intonation), stress (duration, loudness, and pitch), and rhythm of elements. Following the understanding of pitch as a percept, i.e., a tone complex that contains harmonically related frequencies, pitch sequences are integrated over time to form melodies in song and speech (in speech, the respective term is "prosody"). At this basic level of comparing fundamental acoustic parameters, song melody and the prosodic aspect of speech are in close correspondence.

On the other hand, melody as a feature of singing and prosody as a feature of speech exhibit pronounced differences. At the phenomenological level, sung melodies are "quantized", i.e., (i) their pitches usually have discrete relations at $n \sqrt[2]{2}$ (semitones in Western music) and (ii) their rhythm typically shows discrete onsets at integer multiples of the underlying metric beat or its subdivisions. In speech, this "quantized" quality can be found in special cases such as recited poetry, where the rhythmic/metric timing is divided into two or three beats per measure, but a discreteness of pitch is missing: speech shows continuously changing (gliding) pitch over the course of a sentence (cf. Patel, Wong, Foxton, Lochy, & Peretz, 2008; Zatorre & Baum, 2012). It can be assumed that human perception makes use of these features (besides others such as the situational context) to classify song and speech in everyday life.

Under some circumstances, however, the listener might fail to correctly classify song and speech. For example, certain art forms exploit this confusion when a performer on stage uses speech which is actually "composed" in a musical form (e.g., in a *Melodrama*). These conditions interfere with the listener's capabilities to label these vocalizations as being sung or spoken. This, on the other hand, demonstrates that song and speech might not be as discrete as assumed but might rather constitute a continuum, the ambiguous center of which can be prone to individual or contextual interpretation. Here, categorization might rely on more strategic processes that draw, for example, on the current listening context (e.g., opera house vs. parliament) or the participant's listening history, while typical exemplars of speech and song might be easy to classify based on the acoustic parameters of the stimulus itself. The present study seeks to measure stimulus-based auditory and individual strategic classification by testing participants with different auditory processing abilities and by using ambiguous vocalizations, i.e., stimuli that are able to trigger both associations – song and speech.

Two neurocognitive preconditions might play a role in the process of classifying speech and song based on auditory features: Firstly, on the systemic level, intact processing mechanisms of melody and prosody, constituting a fundamental difference between speech and song, and secondly, on the neural level, an intact temporal lobe (TL) as an important anatomical structure for melody and prosody processing. It is tempting to assume that the latter simply is the substrate necessary to accomplish the former; however, as will be outlined later, impairment in one of

these aspects is not always strictly tied to an impairment in the other. To investigate the necessity of intact melody/prosody processing and the role of the TL for song-speech classification, the current study examined two groups of participants in addition to normal controls: (i) participants with a known deficit in processing (pitch) musical melody, congenital amusics and (ii) patients with lesions in the temporal cortex.

Congenital amusia (also known under the outdated label “tone deafness”) is a lifelong impairment of music perception, in spite of normal audiograms, education, cognitive ability, and memory (Ayotte, Peretz, & Hyde, 2002; Peretz et al., 2013; Stewart, 2011) that affects about 2.5% of the population (for a recent description of the phenomenon, see Tillmann, Albouy, & Caclin, 2015). Case studies on congenital amusia have shown a variety of music processing deficits with a core deficit in pitch perception and pitch memory (cf. Ayotte et al., 2002; Peretz et al., 2002, 2013; Williamson & Stewart, 2010). As these two features are not unique to music, amusics have been tested on the speech domain as well. Overall, findings point to unimpaired speech processing, however, with heterogeneous results on linguistic tone and intonation processing (Ayotte et al., 2002; Hutchins, Gosselin, & Peretz, 2010; Jiang, Hamm, Lim, Kirk, & Yang, 2010; Liu, Patel, Fourcin, & Stewart, 2010; Liu, Xub, Patel, Francart, & Jiang, 2012; Patel, Foxton, & Griffiths, 2005; Patel et al., 2008). Some authors argue that amusics’ pitch awareness deficit seems to be most noticeable in music due to usually smaller intervals (ranging down to a semitone) compared to speech (Foxton, Dean, Gee, Peretz, & Griffiths, 2004; Hyde & Peretz, 2004; Liu et al., 2010).

The TL has been repeatedly linked to the classification process of speech and song in the literature. Along with an ongoing discussion on the existence of specific music processing regions in the brain (compared to speech-related areas), some recent studies have proposed regions in the TL (bilaterally) to be sensitive to musical sounds, for example, melodies compared to sentences (Rogalsky, Rong, Saberi, & Hickok, 2011) and musical structure versus other high-level linguistic representations (Fedorenko, McDermott, Norman-Haignere, & Kanwisher, 2012). Angulo-Perkins et al. (2014) reported the anterior superior temporal gyrus (STG) and the planum polare to be particularly involved in processing musical stimuli compared to other complex sounds, for example, speech, implying that this region plays an important role in processing music-related acoustic parameters. Brain imaging studies focusing on song and speech perception in particular reported overlapping activation in temporal areas, with the (right) anterior STG being important for song in comparison to speech perception (Callan et al., 2006; Schön et al., 2010). Similarly, in the special case when actual speech is mis-perceived as song, the bilateral anterior STG, in addition to the right mid-posterior STG, was observed to be active (Tierney, Dick, Deutsch, & Sereno, 2012). Furthermore, it has been shown that damage to the TL in some cases can lead to music processing deficits (Ayotte, Peretz, Rousseau, Bard, & Bojanowski, 2000; Liégeois-Chauvel, Peretz, Babai, Laguitton, & Chauvel, 1998; Peretz, 1990, 1996; Peretz et al., 1994; Schuppert, Münte, Wieringa, & Altenmüller, 2000). Based on these combined results, it can be hypothesized that the TL might be also a core area for classifying song

and speech. Thus, we also tested patients with focal unilateral lesions in the right or left TL.

To our knowledge, the current study is the first to introduce vocalizations specifically designed to be perceptually ambiguous, in comparison to distinct speech and song recordings. These hybrid stimuli (referred to as ambiguous stimuli (AMB) in the following) were perceived as “halfway between” song and speech as cross-validated by a large participant sample in a pilot rating study. This particular 50:50 ambiguity exploits the defining feature of the ambiguous sounds, namely, that they are validated to equally likely evoke the responses “song” and “speech”, respectively, without any defining acoustical cues inherent in the sound itself. Combining the ambiguous stimulus subset with a forced-choice paradigm (requiring participants to cognitively choose to perceive each of these stimuli as sung or spoken) allows for an investigation of whether pitch and music perception abilities and/or TL lesions influence the categorization of song and speech based on their rating strategy.

We hypothesize that

- (1) congenital amusics and TL lesion patients will be able to categorize unambiguously sung and spoken stimuli;
- (2) for perceptually balanced ambiguous stimuli, however, their functional and/or structural impairment (congenital amusia, TL lesion) will affect their rating strategy in a way that will skew the outcome in an experimental forced-choice (Is it song or speech?) scenario;
- (3) this imbalance will manifest as a statistical bias toward labeling the ambiguous stimuli as speech, as a consequence of a failure to detect the song-like components of the stimulus while easily extracting the speech-like aspects of it.

In other words, for (1) clear and unmanipulated stimuli, we expect the null hypothesis to be confirmed (no performance difference between experimental groups and controls). However, for ambiguous stimuli, we test for the alternative hypothesis in (2) undirected (difference between experimental groups and controls) and (3) directed manner (experimental groups show bias toward “speech” responses). In addition, from the considerations earlier, we also expect to be able to draw conclusions on the importance of music processing deficits (amusia) and the role of the TL in processing sung and spoken utterances.

2. Methods

2.1. Participants

The present study tested song and speech perception in two groups, TL lesion patients ($N = 7$; 2 female) and congenital amusics ($N = 5$; 2 female). A control group of 12 healthy, musically untrained (i.e., musical experience did not exceed basic school education) control subjects was recruited, matched by gender (4 female), age (median = 55.5; interquartile range (IQR) = 50–64), handedness (evaluated with the Edinburgh Inventory; Oldfield, 1971), and school education (median = 10; IQR = 10–12 years). For details see Tables 1–3. The group matching was confirmed by Mann–Whitney U-tests,

Table 1. Personal and neuropsychological data and description of the lesions of the patients.

Code	Left-sided TL lesion patients			Right-sided TL lesion patients			
	TL 1	TL 2	TL 3	TL 4	TL 5	TL 6	TL 7
Age	59	46	51	67	67	57	50
Gender	M	F	F	M	M	M	M
Handedness	R	R	R	L	R	R	R
School education	12	10	12	12	12	10	10
Etiology	IS	IS	HE	IS	IS	IS	CH
Lesion onset (yrs.; mon.)	7;10	4;7	6;4	5;6	1;10	1;9	8;10
TL lesion (BA)	21,22	21,22,38	20,21,22, 38	21,22,38	21,22,42	21,22,38, 41,42	22,41,42
Additional lesions (BA)	39,40		11,28,34, 35,36, amygdala	11,47		11	39,40
Token Test	1	7	0	0	0	0	11
MMSE	25	24	29	29	29	29	24
Digit span forward/backward	13/2	1/2	98/34	95/85	28/78	88/67	0/2
MBEA scale	23	24	18	20	22	16	21
MBEA contour	22	16	17	22	18	19	25
MBEA interval	24	12	18	14	14	21	24
MBEA melodic score	69	52	53	56	54	56	70
Song–speech MT (% song)	60	42	56	50	92	92	62
Song–speech MP (% song)	82	68	40	80	92	96	78
Song–speech rating (% song)	71	55	48	65	92	94	70

Gender: F = female, M = male. Handedness is indicated according to the Edinburgh Inventory (Oldfield, 1971). R = right-handed, L = left-handed. School education is indicated in years. Etiology: IS = ischemic stroke, CH = cerebral hemorrhage, HE = herpes encephalitis. The average time between lesion onset and the experimental session is indicated in years; months. Lesions are listed in Brodmann Areas (BA) that were partly damaged. The extent of the lesions was determined in MRI scans by a neurologist. The severity of language comprehension deficits is indicated by the number of mistakes in the Token Test: no/very mild disorder (0–6); mild (7–21); medium (22–49); severe (>40). Mini-Mental State Examination (MMSE): mild cognitive impairment (<26); moderate (<18); severe (<10). Verbal STM and WM were tested with the digit span test forward and backward; results (age corrected) are indicated as percentages (digit span forward/backward); values below 16 indicate a deficient memory performance. The scores achieved in the Montreal Battery of Evaluation of Amusia (MBEA) are depicted as number of correct responses. For the MBEA Melodic Score, bold font indicates scores under the cutoff of 65 points, i.e., marks a deficient performance. Song–speech ratings are indicated in percentages of song ratings, listed for the monotonous (MT), the mimicked prosody (MP), and the mean ratings.

revealing no difference in age (congenital amusics vs. controls: $U = 103.5$, $Z = -.476$, $p = .634$; TL vs. controls: $U = 118.5$, $Z = -.127$, $p = .899$), handedness (laterality index; congenital amusics vs. controls: $U = 38$, $p = .823$; TL vs. controls: $U = 62.5$, $p = .975$), and education (congenital amusics vs. controls: $U = 42$, $Z = -.365$, $p = .715$; TL vs. controls: $U = 117$, $Z = -.293$, $p = .770$). All participants reported to be musically untrained and gave written informed consent before testing in accordance with the Declaration of Helsinki.

2.1.1. Congenital amusics

The Montreal Battery of Evaluation of Amusia (MBEA; Peretz, Champod, & Hyde, 2003) was administered in order to assess their individual music processing capabilities. Only the first three, melody-related tasks, of the battery were administered because the main experiment focused exclusively on melody (rhythm was identical between conditions). Correct responses were summed across the three subtests, and a cutoff score of 65, which represents 72.2% of the total score, was chosen to classify the participants as being amusic (Liu et al., 2010, 2012; Peretz et al., 2003). All congenital amusics scored below the 65 cutoff score (Median = 58; IQR = 50.5–61; see Table 2). The control group scored 75 points (median; IQR = 72–79), indicating unimpaired music processing, in contrast to the congenital amusics ($U = 15$, $Z = -3.172$, $p = .002$).

2.1.2. Patients with temporal lobe lesions

TL lesion patients were selected with respect to their lesion site only. The group comprised three patients with focal lesions in the left and four patients with focal lesions in the right TL with different etiologies (post-lesional delay of 5;6 (median; IQR = 1;10–7;10) years; months). Lesions

encompassed the superior temporal gyrus and sulcus (STG/STS), and the temporal pole (Brodmann area (BA) 38) with the core lesion overlay in the anterior TL (see Figure 1 and Table 1). The MBEA was also administered to assess music processing deficits in the TL patients group. The group scored below the cutoff score (median = 56; IQR = 53–69) and can be classified as having acquired amusia in comparison to the healthy control group ($U = 29.5$, $Z = -3.433$, $p = .001$), even though two patients scored slightly above the cutoff with 69 and 70 points (a left- and right-sided patient). Congenital amusics and TL patients did not differ in their MBEA performance ($U = 30.5$, $Z = -.326$, $p = .744$).

2.1.3. Neuropsychological testing

Language comprehension deficits were assessed by means of the Token Test, a subtest of the Aachen Aphasia Test (Huber, Poeck, Weniger, & Willmes, 1993). At the time of testing, language comprehension was normal in all congenital amusics and all but two TL patients, the latter showing mild impairment. Mann–Whitney U-tests for independent samples did not reveal significant differences between the TL patients and controls ($U = 57.5$, $Z = -.898$, $p = .369$), or the congenital amusics and controls ($U = 83.5$, $Z = -1.490$, $p = .136$).

All patients scored 24 or higher in the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975) licensing their inclusion in the present study (<26 indicating a mild cognitive impairment). Two outliers in the control group (C8 and C12) reached scores slightly under the cutoff (23 and 22 of 24, respectively). As they showed otherwise normal performance (see Figure 3), their exclusion did not appear to be justified. Neither the performance between TL patients and

Table 2. Personal and neuropsychological data and description of the congenital amusics.

Code	Congenital amusics				
	CA 1	CA 2	CA 3	CA 4	CA 5
Age	57	67	51	67	49
Gender	M	M	F	M	F
Handedness	R	A	R	R	A
School education	12	12	10	10	10
Token Test	0	0	0	0	0
MMSE	27	29	29	27	30
Digit span forward/ backward	98/80	98/85	95/34	28/27	95/52
MBEA scale	18	18	16	23	20
MBEA contour	18	21	17	21	23
MBEA interval	16	19	16	15	20
MBEA melodic score	52	58	49	59	63
Song–speech MT (% song)	62	80	66	90	94
Song–speech MP (% song)	72	82	100	98	94
Song–speech rating (% song)	67	81	83	94	94

Gender: F = female, M = male. Handedness is indicated according to the Edinburgh Inventory (Oldfield, 1971). R = right-handed, L = left-handed. School education is indicated in years. Etiology: IS = ischemic stroke, CH = cerebral hemorrhage, HE = herpes encephalitis. The average time between lesion onset and the experimental session is indicated in years; months. Lesions are listed in Brodmann Areas (BA) that were partly damaged. The extent of the lesions was determined in MRI scans by a neurologist. The severity of language comprehension deficits is indicated by the number of mistakes in the Token Test: no/very mild disorder (0–6); mild (7–21); medium (22–49); severe (>40). Mini-Mental State Examination (MMSE): mild cognitive impairment (<26); moderate (<18); severe (<10). Verbal STM and WM were tested with the digit span test forward and backward; results (age corrected) are indicated as percentages (digit span forward/backward); values below 16 indicate a deficient memory performance. The scores achieved in the Montreal Battery of Evaluation of Amusia (MBEA) are depicted as number of correct responses. For the MBEA Melodic Score, bold font indicates scores under the cutoff of 65 points, i.e., marks a deficient performance. Song–speech ratings are indicated in percentages of song ratings, listed for the monotonous (MT), the mimicked prosody (MP) and the mean ratings.

controls ($U = 57$, $Z = .888$, $p = .374$) nor between congenital amusics and controls ($U = 91$, $Z = -.294$, $p = .769$) revealed significant differences.

Short-term (STM) and working memory (WM) abilities were tested using the forward and backward digit span (WAIS-III; German adaption: WIE; Aster, Neubauer, & Horn, 2006; converted according to the normative values of their age group). Despite notable outliers in the TL patient group, Mann–Whitney U-tests for independent samples did not reveal significant differences between the TL patients and controls (forward: $U = 54$, $Z = .172$, $p = .172$ /backward: $U = 67.5$, $Z = -.212$, $p = .832$), or the congenital amusics and controls ($U = 94$, $Z = -1.490$, $p = .136$ / $U = 96.5$, $Z = -1.215$, $p = .224$). Accordingly, differences in behavioral performance cannot be attributed to a different STM or WM capacity of TL patients, congenital amusics, and controls.

All participants, healthy controls, congenital amusics, and TL patients had normal hearing as tested with the HTTS Audiometry (2008) by SAX GmbH (<http://sax-gmbh.de/htts/httsmain.htm>). HTTS is a program for performing a hearing test (audiometry) on a multimedia PC using a logarithmic frequency scale. Pitches were presented in a randomized order and both ears were tested independently. Eight frequencies between 250 and 10,000 Hz were tested, each sinus tone increased by 0.5 dB four times per second, starting 10 dB below the set value that represents the general auditory threshold.

2.2. Stimuli

For all stimuli, simple German sentences were used (Kotz & Paulmann, 2007) exhibiting fixed grammatical structure

Table 3. Personal and neuropsychological data and description of the control group.

Code	Controls											
	C 1	C 2	C 3	C 4	C 5	C 6	C 7	C 8	C 9	C 10	C 11	C 12
Age	59	46	51	65	44	60	51	61	69	52	65	50
Gender	M	F	F	M	F	M	M	M	M	F	M	F
Handedness	R	R	R	L	R	R	R	R	R	R	R	R
School education	12	10	12	12	10	10	10	12	12	10	10	10
Token Test	0	2	0	0	1	0	0	0	0	0	2	5
MMSE	29	30	29	27	29	29	30	23	27	29	28	22
Digit Span forward/ backward	48/5	53/2	53/52	53/27	75/12	67/38	100/71	67/67	76/53	53/71	76/13	53/52
MBEA Scale	26	28	29	26	28	24	30	29	26	26	26	28
MBEA Contour	22	24	25	26	23	24	26	29	24	24	23	24
MBEA Interval	23	26	25	24	23	21	23	23	22	23	23	27
MBEA Melodic Score	71	78	79	76	74	69	79	81	72	73	72	79
Song–Speech MT (% song)	28	52	80	58	20	28	20	22	18	34	40	72
Song–Speech MP (% song)	36	74	52	76	22	50	70	40	62	32	88	80
Song–Speech Rating (% song)	32	63	66	67	21	39	45	31	40	33	64	76

Gender: F = female, M = male. Handedness is indicated according to the Edinburgh Inventory (Oldfield, 1971). R = right-handed, L = left-handed. School education is indicated in years. Etiology: IS = ischemic stroke, CH = cerebral hemorrhage, HE = herpes encephalitis. The average time between lesion onset and the experimental session is indicated in years; months. Lesions are listed in Brodmann Areas (BA) that were partly damaged. The extent of the lesions was determined in MRI scans by a neurologist. The severity of language comprehension deficits is indicated by the number of mistakes in the Token Test: no/very mild disorder (0–6); mild (7–21); medium (22–49); severe (>40). Mini-Mental State Examination (MMSE): mild cognitive impairment (<26); moderate (<18); severe (<10). Verbal STM and WM were tested with the digit span test forward and backward; results (age corrected) are indicated as percentages (digit span forward/backward); values below 16 indicate a deficient memory performance. The scores achieved in the Montreal Battery of Evaluation of Amusia (MBEA) are depicted as number of correct responses. For the MBEA Melodic Score, bold font indicates scores under the cutoff of 65 points, i.e., marks a deficient performance. Song–Speech ratings are indicated in percentages of song ratings, listed for the monotonous (MT), the mimicked prosody (MP) and the mean ratings.

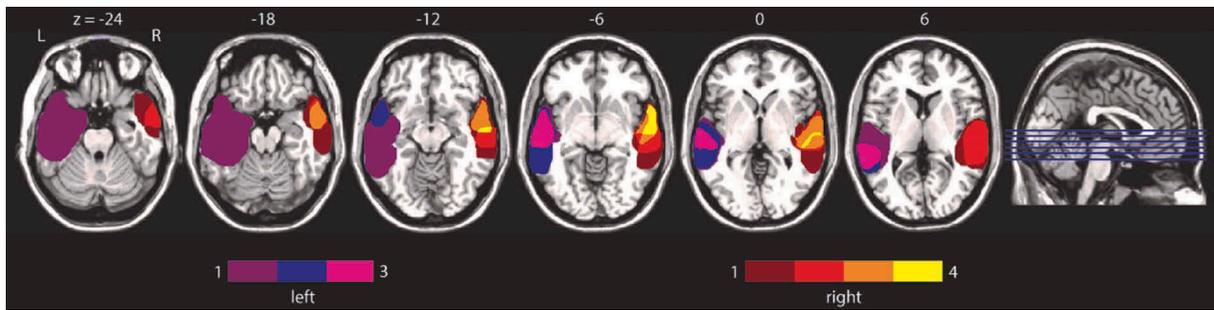


Figure 1. Overlay of the temporal lobe lesions of all lesion patients in MNI space. Lesions are color-coded: each color represents one individual. Left-sided TL lesions blue/pink ($N = 3$), right-sided TL lesions red/yellow ($N = 4$). Far right: visualization of the z-coordinates from which the six depicted slices were taken. (For MR scans of the individual patients, see Supplementary Material.)

(pronoun, auxiliary, article, noun, past participle) and length of seven syllables, recorded from a male and a female voice. Rhythmically, a ternary pattern was applicable to all stimuli, irrespective of whether they were sung or spoken (musically, a triple meter; lyrically, a dactyl).

During the recording session, the voice artists were instructed to produce a wide variety of speech and song styles based on melodies pre-composed according to interval transitions typical of Western tonal music (Figure 2). To obtain rather unambiguously sung stimuli, melodies were composed as a classical cadence or consisted of fifths (Figure 2(d,c)). To obtain a basis for ambiguous stimuli, we asked the singers to produce (i) utterances with monotonous pitch or (ii) with a pitch contour that was composed to resemble spoken prosody (Figure 2(a,b)). These melodies had been labeled beforehand to better instruct the singers on producing a wide range of vocal utterances. Which melody was actually perceived as song or speech was pending until the pre-test (and the current experiment). All recordings

were digitally normalized and re-sampled to exactly three seconds in duration using MATLAB 7.0 (The MathWorks, Inc., Natick, MA, USA).

In a pilot study, 62 participants rated the resulting pool of 674 stimuli on a 10-point Likert scale ranging from 0 for speech to 9 for song. The rating resulted less in a stimulus set along a song–speech continuum, but rather in three groups of stimuli, clustering around the end points and the center of the Likert scale. Therefore, 40 stimuli with consistent ratings across participants were selected from the pre-evaluation for the experiment: 10 clearly spoken (SPK; mean 0.66, $SD \leq 1.4$), 10 clearly sung (SNG; mean 7.7, $SD \leq 0.8$) and 20 ambiguous (AMB; mean 4.62, $SD \leq 2.3$) stimuli. The AMB stimuli consisted of two subsets, of which 10 exhibited a “monotonous” pitch contour and 10 a “mimicked prosodic” pitch contour (MT and MP; corresponding to the melodic prototype classes shown in Figure 2(a,b)). Both groups of stimuli, despite their difference in pitch contour, were hybrid in nature, i.e., no tendency toward song and speech was inherent in the stimuli themselves and not perceived by the participants in the pre-test. During the main experiment, each stimulus was presented five times in a randomized order, totaling 200 stimulus presentations.

Er hat die Spie-le ge - spielt.

a MT

b MP

c SNG

d SNG

e SPK

Er hat den Bogen gespannt.

Figure 2. Stimulus material. All stimuli generated for the stimulus pool were based on one of four melodic prototypes, respectively: (a) a monotonous pitch contour = MT, (b) a melody that mimics a neutrally spoken prosody = MP, (c) a typical classical music cadence, (d) a melody consisting of fifths, both served as basis for the sung (SNG) stimuli, and (e) neutrally spoken sentences (SPK). Rhythm and meter was ternary for all stimuli. Conditions MT and MP served as basis for ambiguous stimuli.

2.3. Procedure

Participants sat in front of a computer and were presented with the stimuli via headphones. For presentation and recording of responses, a custom-developed Flash® animation (Adobe® Systems Software Ireland Ltd.) was used. After each stimulus presentation, participants were asked to decide whether the stimulus was “song” or “speech” by a 2-alternative forced-choice button decision. In a self-paced paradigm, participants were instructed to wait with the button press until a prompt occurred on the screen after the stimulus presentation. After the button press, the next stimulus presentation started automatically after one second with a cross symbol for visual fixation on the screen. The participants were instructed to take all the time they needed to make their decision and that there were no correct or incorrect answers. However, the actual auditory presentation of a stimulus was only delivered once during a single trial and could not be repeated while the subjects made their decision. A training session with 10

examples assured the participants' understanding of the experimental procedure. The experiment itself took about 18 min to complete.

Of interest for the statistical analysis was (i) whether participants were able to correctly respond to clearly sung and spoken stimuli and (ii) whether they showed a bias during categorization of ambiguous stimuli (e.g., a tendency to rate them more toward song or more toward speech under forced choice). Mann–Whitney U-tests for independent samples were chosen to statistically evaluate the differences between the groups. Global alpha level for significance was set at 0.05. All tests were corrected for multiple comparisons according to Bonferroni. All statistics were done with PASW Statistics 18.0.

3. Results

All participants were able to correctly identify the clearly sung (SNG) and spoken stimuli (SPK), as indicated by 100% (median; IQR = 97.5%–100%) correct responses. Broken down into the individual groups, the control group scored 100% (98–100) for SNG and 100% (97.5–100) for SPK; congenital amusics 100% (98–100) for SNG and 100% (97–100) for SPK; TL patients 100% (97.5–100) for SNG and 100% (97–100) for SPK. No differences between the groups were found, neither for the clearly spoken, nor for the clearly sung stimuli, revealed by Mann–Whitney U-tests: congenital amusics vs. control group for SPK: $U = 43$, $Z = -.263$, $p = .793$ and for SNG: $U = 106.5$, $Z = -.215$, $p = .830$; TL vs. control group for SPK: $U = 119.5$, $Z = -.052$, $p = .959$ and for SNG: $U = 68.5$, $Z = -.166$, $p = .868$.

For the AMB stimuli, the control group showed a wide variety of individual response biases, i.e., eight participants rated over half of the AMB stimuli toward “speech” while six participants over half toward “song” (see Figure 3). These

mixed preferences became obvious by applying a one-sample t -test with a test value of 50 which revealed no tendency in the average controls' rating ($t(11) = -.367$, $p = .721$). In contrast, both the TL patient and the amusia group rated the stimuli significantly higher than the 50% midpoint, indicating a rating bias of these groups toward “song” (TL: one-sample t -test (0.5): $t(6) = 3.172$, $p = .019$; congenital amusics: one-sample t -test (0.5): $t(4) = 6.768$, $p = .002$).

To statistically compare the AMB stimuli ratings between groups, Mann–Whitney U-tests were applied because of the mutual independence of the samples. The congenital amusic group differed significantly from the control group in their rating results ($U = 79.5$, $Z = -3.008$, $p = .003$, $r = .729$), the same was true for the TL patients group ($U = 93$, $Z = -2.282$, $p = .022$, $r = .524$). Conversely, ratings did not differ between congenital amusics and TL ($U = 37$, $Z = -1.390$, $p = .164$), or between left and right TL patients ($U = 8$, $Z = -1.414$, $p = .157$). As Mann–Whitney U-tests were performed pairwise on all three permutations of the three groups, a Bonferroni correction factor 2 was applied (for independence of two out of the three permutations respectively). The test between right and left TL patients serves a separate hypothesis and is uncorrected. A lack of relationship between rating bias and years after lesion onset was suggested by a nonsignificant Kendall rank correlation ($p = .293$, $r = .333$).

Regarding the two subsets of AMB stimuli, the control group rated the “monotonous” (MT; median = 31%; IQR = 21–57%) stimuli significantly less toward song than the “mimicked prosody” (MP; median = 57%; IQR = 37–75.5%) stimuli ($Z = -2.239$, $p = .025$) according to a Wilcoxon signed rank test for paired samples. This differential rating bias was not objectified in the TL and congenital amusic groups (TL: $Z = -1.682$, $p = .093$; congenital amusics: $Z = -1.826$, $p = .068$).

4. Discussion

Individuals with lesions in the TL and congenital amusia were tested on their song and speech classification abilities using ambiguous stimuli and clear sung and spoken stimuli. For clear stimuli, performance of either experimental group was comparable to the control group, thus supporting hypothesis (1). When rating ambiguous stimuli, congenital amusics and TL patients showed a significant bias toward responding “song”. This finding is simultaneously both expected and surprising: While a behavioral bias is clearly present and in line with hypothesis (2), the direction of this bias is exactly opposite to what was predicted based on a review of the relevant literature (hypothesis (3)). These results will be discussed in turn.

4.1. Unambiguous stimuli (SNG and SPK)

Despite TL lesions and evident pitch processing deficits, the forced-choice paradigm did not reveal obvious problems in the experimental groups in classifying unambiguous stimuli as sung or spoken, as expected. The nearly perfect performance can most likely be explained by a ceiling effect. When looking at the TL patients only, a unilateral lesion (as in the tested lesion patients) is probably not sufficient to disrupt the classification of unambiguous song and speech, due to bilateral TL

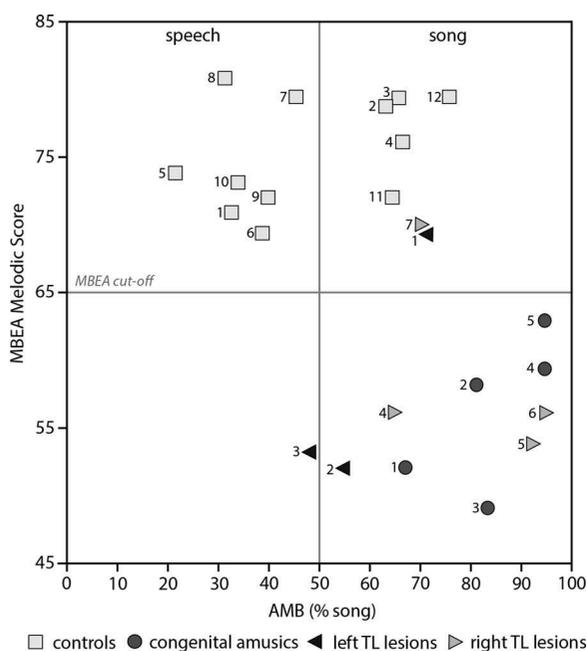


Figure 3. Performance of all participants in the MBEA (Y-axis, score, MBEA Melodic Score) and the song–speech rating of ambiguous (AMB) stimuli (X-axis, percent rated for “song”). Numbers on the left side of the symbols indicate the number of the participant according to Tables 1–3.

involvement in this task (i.e., prosody and pitch processing) (Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Tzourio et al., 1997; Zatorre & Belin, 2001; Zatorre, Belin, & Penhune, 2002; Zatorre & Samson, 1991). Overall, the perfect classification of the stimuli does not conflict with the TL patients' and congenital amusics' deficits in pitch perception if one takes strategic processes during forced-choice paradigms into account: Here, participants' rating behavior shows that they were able to *differentiate* the stimulus types (SNG and SPK) in context but does not give an answer if they were able to *identify* and label the stimuli themselves. The mere ability to clearly recognize only one of the categories, either song or speech, allows a perfect performance by classifying the stimuli by exclusion as, for example, "speech" and "no speech", regardless of what the other stimulus was. That means, forced-choice paradigms on clearly identifiable stimuli are not sufficient to tease out the participants' processing problems, while the listeners' rating bias of the AMB stimuli can shed light on strategic processes, for example, a deficit-related classification by exclusion as will be explained later.

4.2. Ambiguous stimuli

In the ambiguous condition, both groups, the congenital amusics and the TL patients, appeared systematically biased to let their forced classification gravitate toward "song". Healthy controls, on the other hand, showed a range of individual biases toward song or speech, resulting in a balanced rating at group level (in line with the pretest). It may be speculated that the ambiguous stimuli were classified by matching the perceived sound pattern with an internalized, experience-dependent inventory of acoustically and situationally distinct prototypes. The individually tolerable range of feature properties for each of the prototypes will certainly depend heavily on previous biographical exposure to samples from each category, along with relevant situational contexts. In a laboratory setting, when access to supporting contextual information is removed, the participants have to resort to fundamental acoustic features of the stimulus alone. Which of these properties will be weighted most saliently in the prototype is, again, likely subject to lifelong biographical tuning of the prototypes.

It might appear surprising that the patients tended to choose the song category for the ambiguous stimuli, rather than the speech category (as predicted by hypothesis (3)). They seemingly opt to make a decision in favor of something they are known to be unable to process and even possibly know themselves to be unable of. The question arises whether or not their response behavior reflects their processing deficit, and which actual cognitive strategy they use to classify song and speech. It may well be the case that the participants, due to their melodic processing deficits, really only classify into two categories: "speech" (correct feature extraction possible) and "non-speech" (everything that fails to pass through an intact feature matching process, here, including both SNG and AMB stimulus types). The instructions did not explicitly reveal information about the presence of a third, manipulated stimulus type besides the two specified as the response buttons. Since the experimental design, specifically the forced-choice

response options, implicitly suggested that the only alternative category to "speech" was "song", any participant with deficits in song processing might have been under the impression he or she was supposed to use the "song" bin to dump anything "non-speech".

The focus of previous studies on intonation processing in amusia has been on pitch patterns only (melody and prosody; see Section 1), i.e., never the amusics' tolerance of spoken stimuli in general, for example, their general recognition and classification abilities, has not been tested specifically. From the unexpected response bias in the current study, one might assume that a rating of intoned stimuli depends on the classification ability of the participants, i.e., if a stimulus is not perceived as proper speech, might lead to a certain rating strategy and result, respectively. This might call for a careful procedure in amusia studies, as the chosen behavioral task itself might introduce response biases arising from a generally skewed perception of speech and song.

The rating behavior might be explained by a (i) missing context in a laboratory situation, (ii) missing "unambiguous" acoustical features, (iii) in case of congenital amusics, an impairment in storing melodies or songs (e.g., Peretz, 1996; Peretz & Coltheart, 2003; Williamson & Stewart, 2010), and (iv) in case of the TL patients a failure to access previously stored melodies or songs; the latter as a consequence of a specific brain lesion. Indeed, nearly all of the TL patients were acquired amusics and had a lesion locus in the anterior STG, partly extending into the temporal pole (BA 38).

In recent imaging studies, the anterior TL has been reported when processing music (including song) in comparison to speech, for instance, sung folk songs vs. spoken lyrics (Callan et al., 2006), perceiving actual spoken utterances as song (Tierney et al., 2012), for musical melodies and structure (Angulo-Perkins et al., 2014; Fedorenko et al., 2012; Rogalsky et al., 2011) (for details, see Section 1). The temporal pole (BA 38) has been discussed as end point of the ventral auditory stream (e.g., Rauschecker & Scott, 2009) supporting a higher level of music processing than BA 22 (e.g., perception of song over speech and main effect for music over language processing (Schön et al., 2010), and improvising sung vs. spoken utterances (Brown, Martinez, & Parsons, 2006)). These findings give reason for proposing that the anterior TL is a significant factor in song processing, in which the patient group exhibits a deficit and has, therefore, lead to the chosen rating strategy.

Interestingly, no difference between the ratings of left- and right-sided lesion patients occurred, neither for the clear nor for the ambiguous stimuli. This observation might be an indicator that on the level of auditory processing needed to distinguish between song and speech, no hemispheric difference seems to persist; however, given the small clinical sample available for the study, any attempt to sub-differentiate into lesion loci would border on anecdotal-level observations. Therefore, the authors refrain from any implications of lesion hemisphere for the interpretation of the behavioral finding.

4.3. Considerations and outlook

It is noteworthy that controls consistently rated both AMB stimulus sets differently: the "monotonous" stimuli were

more rated toward speech and the “mimicked prosody” stimuli with a tendency toward song. These stimulus-specific rating tendencies are not, as it seems, in contradiction to the pre-study evaluation of the recordings, during which the particular subset of stimuli (subsequently used for this study as AMB) was most consistently rated as being in the very center of a 9-point scale ranging from “clear song” to “clear speech”. In the main experiment, the healthy controls did not have the option to choose ambiguity on a 9-point scale but had to make a forced choice. It can be speculated that under these forced conditions, a fallback strategy to a particular salient stimulus feature (or set of features) is employed, which differs according to the specific sub-class of AMB stimuli. At any rate, to disentangle what specific stimulus properties are used as a fallback under which conditions, is an interesting question for future studies that may be addressed using systematically modulated synthesized mock vocalizations. As we only tested small groups, further investigation is necessary on a wider population range of patients on the TL lesion spectrum as well as individuals with music processing impairments.

Taken together, temporal lobe lesions as well as impaired music perception, irrespective of whether congenital or acquired, evoke a bias to rate ambiguous stimuli as song. This bias most likely reflects the use of a specific strategy (classification by exclusion) that is indirectly related to pitch processing deficits (e.g., pitch perception or pitch memory). These results have methodological implications for studies including amusics (congenital and acquired through lesions) on the perception of prosody and melody and may point to a possible mechanism of how we classify song and speech. Apart from acoustic features, the individual listening history and memory for melody and song may influence our perception of song and speech.

Acknowledgments

We thank Sonja Kotz and Silke Paulmann for providing the sentences used to create the stimuli, Anne-Kathrin Franz for recruiting the participants and Derek Ott for neuroanatomical support on the lesion patients.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Federal Ministry of Education and Research under Grant 01GW0773 awarded to Angela D. Friederici.

References

Angulo-Perkins, A., Aube, W., Peretz, I., Barrios, F., Armony, J. L., & Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: Differences between musicians and nonmusicians. *Cortex*, *59*, 126–137. doi:10.1016/j.cortex.2014.07.013

Aster, M. V., Neubauer, A., & Horn, R., (Eds.). (2006). Wechsler Intelligenztest für Erwachsene (WIE). In *Deutschsprachige Bearbeitung und Adaptation des WAIS-III von David Wechsler*. Frankfurt a. M.: Pearson Assessment.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, *125*, 238–251. doi:10.1093/brain/awf028

Ayotte, J., Peretz, I., Rousseau, I., Bard, C., & Bojanowski, M. (2000). Patterns of music agnosia associated with middle cerebral artery infarcts. *Brain*, *123*, 1926–1938. doi:10.1093/brain/123.9.1926

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, *23*, 2791–2803. doi:10.1111/ejn.2006.23.issue-10

Callan, D. E., Tsytarev, V., Hanakawa, T., Callen, A. M., Katsuhara, M., Fukuyama, H., & Turner, R. (2006). Song and speech: Brain regions involved with perception and covert production. *NeuroImage*, *31*, 1327–1342. doi:10.1016/j.neuroimage.2006.01.036

Fedorenko, E., McDermott, J. H., Norman-Haignere, S., & Kanwisher, N. (2012). Sensitivity to musical structure in the human brain. *Journal of Neurophysiology*, *108*, 3289–3300. doi:10.1152/jn.00209.2012

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state (a practical method for grading the state of patients for the clinician). *Journal of Psychiatric Research*, *12*, 189–198. doi:10.1016/0022-3956(75)90026-6

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying “tone deafness”. *Brain*, *127*, 801–810. doi:10.1093/brain/awh105

Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1993). *Aachener aphasia test (AAT)*. Göttingen: Hogrefe.

Hutchins, S., Gosselin, N., & Peretz, I. (2010). Identification of changes along a continuum of speech intonation is impaired in congenital amusia. *Frontiers in Psychology*, *1*, 236. doi:10.3389/fpsyg.2010.00236

Hyde, K., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, *15*, 356–360. doi:10.1111/psci.2004.15.issue-5

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia*, *48*, 2630–2639. doi:10.1016/j.neuropsychologia.2010.05.009

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, *1151*, 107–118. doi:10.1016/j.brainres.2007.03.015

Liégeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain*, *121*, 1853–1867. doi:10.1093/brain/121.10.1853

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation. *Brain*, *133*, 1682–1693. doi:10.1093/brain/awq089

Liu, F., Xub, Y., Patel, A. D., Francart, T., & Jiang, C. (2012). Differential recognition of pitch patterns in discrete and gliding stimuli in congenital amusia: Evidence from Mandarin speakers. *Brain and Cognition*, *79*, 209–215. doi:10.1016/j.bandc.2012.03.008

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, *17*, 73–88. doi:10.1002/(ISSN)1097-0193

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113. doi:10.1016/0028-3932(71)90067-4

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition*, *59*, 310–313. doi:10.1016/j.bandc.2004.10.003

Patel, A. D., Wong, M., Foxton, J. M., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception*, *25*, 357–368. doi:10.1525/mp.2008.25.4.357

Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, *113*, 1185–1205. doi:10.1093/brain/113.4.1185

Peretz, I. (1996). We can lose memory for music? A case of music agnosia in a nonmusician. *Journal of Cognitive Neuroscience*, *8*, 481–496. doi:10.1162/jocn.1996.8.6.481

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, *33*, 185–191. doi:10.1016/S0896-6273(01)00580-3

- Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The montreal battery of evaluation of amusia. *Annals of the New York Academy of Sciences*, 999, 58–75. doi:10.1196/annals.1284.006
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6, 688–691. doi:10.1038/nn1083
- Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S., & BÉland, R. (2013). A novel tool for evaluating children's musical abilities across age and culture. *Frontiers in Systems Neuroscience*, 7, 30.
- Peretz, I., Kolinsky, R., Tramo, M. J., Labreque, R., Hubler, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, 117, 1283–1301. doi:10.1093/brain/117.6.1283
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12, 718–724. doi:10.1038/nn.2331
- Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. *Journal of Neuroscience*, 31, 3843–3852. doi:10.1523/JNEUROSCI.4515-10.2011
- Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J.-L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage*, 51, 450–461. doi:10.1016/j.neuroimage.2010.02.023
- Schuppert, M., Münte, T. F., Wieringa, B. M., & Altenmüller, E. (2000). Receptive amusia: Evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain*, 123, 546–559. doi:10.1093/brain/123.3.546
- Stewart, L. (2011). Characterizing congenital amusia. *The Quarterly Journal of Experimental Psychology*, 64, 625–638. doi:10.1080/17470218.2011.552730
- Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2012). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex*. doi:10.1093/cercor/bhs003
- Tillmann, B., Albouy, P., & Caclin, A. (2015). Congenital Amusias. In G. G. Celesia & G. Hickok (Eds.), *Handbook of clinical neurology* (Vol. 129, pp. 589–605). Elsevier B.V.
- Tzourio, N., Massiou, F. E., Crivello, F., Joliot, M., Renault, B., & Mazoyer, B. (1997). Functional anatomy of human auditory attention studied with PET. *NeuroImage*, 5, 63–77. doi:10.1006/nimg.1996.0252
- Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory*, 18, 657–669. doi:10.1080/09658211.2010.501339
- Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune?. *Plos Biology*, 10, e1001372. doi:10.1371/journal.pbio.1001372
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11, 946–953. doi:10.1093/cercor/11.10.946
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, 6, 37–46. doi:10.1016/S1364-6613(00)01816-7
- Zatorre, R. J., & Samson, S. (1991). Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain*, 114, 2403–2417. doi:10.1093/brain/114.6.2403