doi:10.1093/brain/awq089 Brain 2010: 133; 1682–1693 | 1682

BRAIN
A JOURNAL OF NEUROLOGY

Intonation processing in congenital amusia: discrimination, identification and imitation

Fang Liu,[1] Aniruddh D. Patel,[2] Adrian Fourcin[3] and Lauren Stewart[1]

1 Department of Psychology, Goldsmiths, University of London, London, UK
2 The Neurosciences Institute, San Diego, CA, USA
3 Department of Speech, Hearing and Phonetic Sciences, University College London, London, UK

Correspondence to: Fang Liu,
Department of Psychology, Goldsmiths,
University of London, New Cross, London,
SE14 6NW, UK
E-mail: f.liu@gold.ac.uk

This study investigated whether congenital amusia, a neuro-developmental disorder of musical perception, also has implications for speech intonation processing. In total, 16 British amusics and 16 matched controls completed five intonation perception tasks and two pitch threshold tasks. Compared with controls, amusics showed impaired performance on discrimination, identification and imitation of statements and questions that were characterized primarily by pitch direction differences in the final word. This intonation-processing deficit in amusia was largely associated with a psychophysical pitch direction discrimination deficit. These findings suggest that amusia impacts upon one's language abilities in subtle ways, and support previous evidence that pitch processing in language and music involves shared mechanisms.

Keywords: congenital amusia; intonation processing; statement–question discrimination/identification/imitation; pitch threshold; pitch change/direction

Abbreviations: F_0 = fundamental frequency; MBEA = Montreal Battery of Evaluation of Amusia; NART = National Adult Reading Test

Introduction

Around 4% of the general population have been estimated to be 'tune deaf', being unable either to sing in tune or detect an out-of-tune note in a melody (Kalmus and Fry, 1980). This disorder, termed 'congenital amusia' (amusia hereafter; Peretz *et al.*, 2002) has been investigated in relation to its phenomenology, neural underpinnings and genetic origins (Patel, 2008; Peretz, 2008; Stewart, 2008). Behavioural studies have shown that individuals with amusia (amusics hereafter) exhibit impaired music production and perception abilities, having difficulties singing in tune, dancing or tapping along with music, detecting anomalous pitches in both familiar and unfamiliar melodies, judging dissonance in musical excerpts and recognizing and memorizing melodies without lyrics (Ayotte *et al.*, 2002; Dalla Bella and Peretz, 2003; Dalla Bella *et al.*, 2009). Its core deficit has been proposed to relate to fine-grained pitch discrimination (Peretz *et al.*, 2002; Hyde and Peretz, 2004; Peretz, 2008) and pitch pattern/direction or melodic contour perception (Griffiths, 2008; Patel, 2008; Stewart, 2008). Structural imaging studies have found amusic brains to be different from neurotypical brains in subtle ways (Hyde *et al.*, 2006, 2007; Mandell *et al.*, 2007) and there is evidence that the disorder is heritable (Drayna *et al.*, 2001; Peretz *et al.*, 2007).

Although amusics rarely report problems outside the musical domain, it may be expected that these individuals would struggle with aspects of spoken language that rely on pitch-varying information. For instance, pitch is modulated in specific ways to convey different communicative meanings (e.g. word stress, focus and sentence type) in speech (Xu, 2005). However, the question of

Received November 19, 2009. Revised March 15, 2010. Accepted March 24, 2010. Advance Access publication April 23, 2010

whether amusia has implications for speech intonation perception is, as yet, unresolved. In previous studies (Ayotte *et al.*, 2002; Patel *et al.*, 2005, 2008), amusics showed no problems with focus identification and discrimination based on salient pitch accents (e.g. 'Go in *front* of the bank, I said' versus 'Go in front of the *bank*, I said'), or with statement/question identification (e.g. 'He speaks French**.**' versus 'He speaks French?'). The majority of them also had no problems with statement–question discrimination. On the basis of these reports, it has been argued that the pitch deficits seen in amusia are domain-specific (Ayotte *et al.*, 2002; Peretz and Coltheart, 2003; Peretz, 2006, 2008). However, two considerations suggest that further research is warranted. First, the intonational pitch contrasts (final pitch glides of statements versus questions) in Ayotte *et al.* (2002) and Patel *et al.* (2008) were on average 5–12 semitones (*cf.* Patel *et al.*, 2008) while amusics' average thresholds for discrimination of pitch direction have been shown to be around two semitones (Foxton *et al.*, 2004), thus making it likely that the amusics were performing at ceiling in previous studies of intonation processing. Second, the dissociation in the performance of amusics seen in Ayotte *et al.* (2002) (intact performance for speech; impaired for tone analogues) is not only explicable in terms of a genuine sparing of pitch processing in the language domain. Patel (2008) notes that the two tasks (speech and tone analogues) are not equivalent in terms of the potential they offer for a 'semantic recoding' strategy. In the speech condition, salient pitch changes can be 'tagged', according to the syllable on which this occurs, thus reducing the memory demands of the task. In the tone analogue condition, salient pitch changes are divorced from any lexical information, so comparison must be made across the whole pitch pattern. Recent findings suggest that amusics are impaired for short-term memory for pitch (Tillmann *et al.*, 2009; Williamson *et al*., 2010), making it possible that the observed dissociation relates not to domain specificity, but rather to the extent to which the two tasks rely on pitch memory.

With these considerations in mind, the present study aimed to test discrimination of pitch contours in amusia, under conditions where semantic recoding strategies would not be helpful and incorporating a range of pitch contrasts, including more subtle examples than previously used. To compare perception of pitch contours within versus outside speech, we used statement–question discrimination tasks under conditions of natural speech, gliding tone and nonsense speech analogues, and we also included two psychophysical tasks for determination of participants' thresholds for detection of pitch change (monotone versus up/down) and discrimination of pitch direction (up versus down). We hypothesized that amusics would show deficits in statement–question discrimination under all three stimulus conditions and that these deficits would be related to measured psychophysical thresholds for the discrimination of pitch direction.

In addition, the present study addressed a secondary issue, regarding a possible dissociation between perception and action in the context of speech. A recent study reported that amusics were able to imitate the correct direction of a heard pitch interval, despite their inability to report its direction (Loui *et al.*, 2008). In order to establish whether a similar dissociation may hold in the context of perception versus production of pitch contours within speech, we asked participants to identify and imitate statements and questions that differed mainly in pitch direction of the final word (down in statements and up in questions).

Materials and methods

Participants

Sixteen amusics and 16 matched controls participated in the study. All were native speakers of British English and had no self-reported neurological or psychiatric disorders. They were recruited via an online musical listening test (http://www.delosis.com/listening/home.html) consisting of the scale and rhythm subtests of the Montreal Battery of Evaluation of Amusia (MBEA; Peretz *et al*., 2003). They were invited for further on-site testing because their online scores were either in the amusic or normal range. Four MBEA subtests (scale, contour, interval and rhythm) were administered to all participants to confirm the presence or absence of amusia. Previous research has shown that amusia is characterized by poor performance on the pitch-based subtests of the MBEA (scale, contour, interval) while scores on the rhythm subtest are likely to be in the normal range for 50% of amusics (Peretz *et al*., 2003). For this reason, we calculated a composite score for the three pitch-based subtests, using 65 as a cut-off score [the sum of the cut-off scores for the three subtests in Peretz *et al*. (2003); those with composite scores at or below 65 were confirmed as amusics]. A summary of the participants' characteristics, means and standard deviations (SD) of their scores (in number of correct responses) on the four MBEA subtests, National Adult Reading Test (NART; Nelson and Willison, 1991) and digit span tests (Wechsler, 1997) are listed in Table 1. As can be seen, while amusics and controls were matched in their age, sex, handedness, musical training background, NART and digit span scores, the two groups performed significantly differently on all four MBEA subsets (detailed information and individual scores can be found in Supplementary material 1).

We additionally collected background measures on speech production (two tasks measuring pitch range and voice quality in terms of vocal fold contact phase ratio) and everyday recognition of sounds (a questionnaire including items relating to speaker identification, regional accent discrimination and identification of environmental sounds). Amusics were found to be normal in these respects (see Supplementary materials 2 and 3 for details).

Perception stimuli

Eighteen statement–question pairs were recorded by a 28-year-old female student at Goldsmiths, University of London, who spoke Southern British English with a slight London accent. The sentences ranged between three and nine syllables, with average duration of 1 s (SD 0.2 s) and average fundamental frequency (F_0) 179.5 Hz (SD 13.7 Hz). These sentence pairs were later modified using a cross-splicing technique such that each pair shared the same stem until the final word. Hence, the statement–question pairs mainly differed in the pitch contours of the final word, with those in statements going down and those in questions going up at the end. We intentionally did not manipulate the duration and amplitude of the final word in order to preserve the natural quality of the utterances. As can be seen from Table 2, compared with those in Patel *et al.* (2008), the current stimuli were spoken with faster rates and with smaller final glide sizes and rates (rate = size/duration; for detailed

Table 1 Subject characteristics

Group	Age	Sex	Handedness	Musical training	Scale	Contour	Interval	Rhythm	Pitch composite	NART	Digit span
Amusic											
μ	55.31	10F	2L	1.28	18.94	19.13	17.81	24.19	55.88	42.00	20.80
σ	9.24	6M	14R	3.24	2.79	3.05	2.32	3.43	6.73	4.80	2.86
Control											
μ	52.25	10F	2L	1.52	26.25	26.81	25.69	27.56	78.75	44.67	19.69
σ	9.14	6M	14R	3.11	2.18	1.72	2.47	1.59	5.12	4.35	4.36
t-tests											
t	0.94			−0.21	−8.26	−8.78	−9.31	−3.57	−10.82	−1.59	0.84
P	0.3535			0.8339	<0.0001	<0.0001	<0.0001	0.0018	<0.0001	0.1221	0.4059

Summary of the two groups in terms of age, sex, handedness, musical training (in years), score on Montreal Battery of Evaluation of Amusia (MBEA): scale, contour, interval and rhythm subtests (number of correct responses, out of 30), NART and digit span scores.
μ = mean; σ = standard deviation; F = female; M = male; L = left; R = right; the pitch composite score is the sum of the scores on the MBEA scale, contour and interval subtests; *t* is the statistic of the Welch two sample *t*-test (two-tailed).

Table 2 Sentences used in the intonation tasks

Sentence	Sentence rate (syl/s)		Size of final pitch glide (st)		Rate of final pitch glide (st/s)		Duration of final pitch glide (s)	
	S	Q	S	Q	S	Q	S	Q
It's a lie./?	4.7	4.4	−2.5	7.1	−11.3	40.4	0.22	0.18
This is love./?	3.9	3.5	−3.6	2.6	−17.4	20.7	0.21	0.12
He hurt his knee./?	4.3	4.2	−3.0	3.7	−16.7	21.6	0.18	0.17
The answer is no./?	5.6	5.6	−2.7	4.7	−24.3	28.3	0.11	0.17
The deal is still on./?	5.5	5.3	−3.7	4.4	−24.3	31.4	0.15	0.14
He just turned one./?	3.9	4.0	−5.1	4.0	−29.5	16.6	0.17	0.24
She looks like Anne./?	3.9	4.0	−2.4	3.4	−17.2	26.3	0.14	0.13
She changed her name./?	4.3	4.3	−4.5	5.5	−16.7	35.3	0.27	0.16
It's a menu./?	5.4	5.5	−5.4	2.9	−18.8	17.2	0.29	0.17
She looks manly./?	4.3	3.9	−6.2	4.1	−14.8	12.4	0.41	0.33
He lives in Ealing./?	5.7	5.6	−6.4	8.1	−27.1	32.6	0.24	0.25
She grew up in Ely./?	5.8	5.8	−4.0	9.6	−18.3	29.2	0.22	0.33
They were in a limo./?	6.7	6.4	−3.5	6.9	−17.5	40.7	0.20	0.17
They named her Lilly./?	5.3	5.4	−4.2	4.9	−25.5	26.5	0.17	0.18
It's from Emily./?	6.1	6.1	−4.1	5.0	−26.3	12.7	0.16	0.39
He speaks Romany./?	4.6	4.4	−3.3	2.9	−7.5	10.3	0.44	0.28
He was born in Illinois./?	6.1	5.9	−3.0	3.2	−8.0	18.9	0.37	0.17
He considers her his enemy./?	6.7	6.5	−3.7	4.8	−10.4	19.3	0.35	0.25
Mean (current study)	5.1	5.0	−4.0	4.9	−18.4	24.5	0.24	0.21
SD (current study)	0.9	1.0	1.2	1.9	6.6	9.3	0.10	0.08
Mean (Patel *et al.*, 2008)	4.2	4.2	−7.4	12.3	−37.8	73.5	0.22	0.18
SD (Patel *et al.*, 2008)	0.8	0.8	2.0	1.9	11.9	23.9	0.11	0.07

S = statement; Q = question; syl = syllable; st = semitone.

definitions and measurement methods, see Patel *et al.*, 2008). It is worth mentioning that the relatively small pitch changes in the current stimuli were not deliberately elicited and that the speaker was instructed to read the sentences naturally as if she was talking to somebody, asking questions and making statements.

Using the technique developed in Patel *et al.* (1998) for converting intonation patterns of spoken utterances to tone analogues, we first generated the discrete tone analogue of each syllable in every sentence using Praat (Boersma, 2001). The newly created sound was the sum of the fundamental frequency at the median F_0 of the original syllable [= $(\max F_0 + \min F_0)/2$] plus its seven odd harmonics (of the same amplitude and with sine phase), and was sampled at 44 100 Hz. The tone analogue had the same duration as the original syllable rhyme, since people tend to base their judgements of syllable duration on the rhyme (Goedemans, 1998). An 8 ms onset and offset taper was later applied to the tone to adjust the rise/decay time. The tone analogues of all the syllables in each sentence were then combined together, preserving the silent gaps as in the original spoken utterances, to form a discrete-tone sequence, which had the same temporal pattern as the original sentence. The pitch of each tone in the sequence

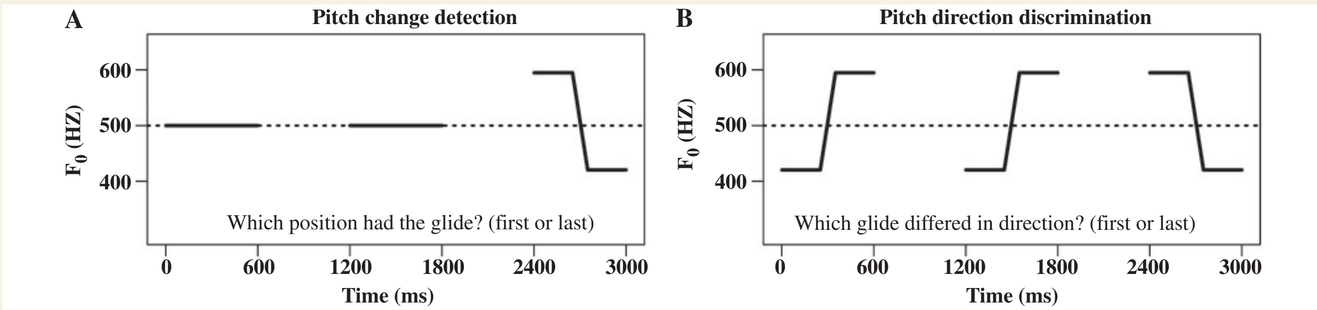

Figure 1 Illustrations of the pitch threshold tasks: (**A**) pitch change detection and (**B**) pitch direction discrimination. The dotted line represents the 500 Hz base frequency, and the solid lines represent the auditory stimuli (discrete or gliding tones). The tones and the silences between them were 600 ms in duration, and the excursion within the gliding tones was 100 ms in duration.

was then changed from a discrete pitch to a gliding pitch that exactly followed the original F_0 contour of the syllable rhyme from which the tone had been constructed. This resulted in a gliding tone sequence that mirrored the pitch and temporal patterns of the original utterance.

Nonsense speech analogues were made from sequences of [li], which was originally the second syllable in 'Ely' (as in 'She grew up in Ely' uttered by the same female speaker). The duration of [li] was lengthened or shortened by Praat to match that of its corresponding gliding tone. These [li] sounds were then combined with silences to make up a sequence that was comparable with the natural speech and gliding tone stimuli in terms of both pitch and rhythm patterns (sound examples of the different conditions can be found at http://www.gold.ac.uk/music-mind-brain/speechproject/).

The peak amplitudes of the three sets of discrimination stimuli were normalized to maximum in Praat to roughly equate perceived loudness. During testing, the participants were also allowed to adjust the output level of the external sound card (Edirol UA-4FX USB Audio Capture) to a comfortable loudness level.

The identification and imitation tasks employed the same set of statements and questions as in the statement–question discrimination task in natural speech.

Procedure

Five intonation perception tasks and two pitch threshold tasks were administered to the participants in fixed order (statement–question discrimination in natural speech and gliding tones; nonsense speech, statement–question identification and imitation, pitch change detection and pitch direction discrimination), so as to keep the possible carry-over effects the same across participants. The experiments were conducted in a sound-attenuated chamber. Participants completed written consent forms to participate in the research, which was reviewed and approved by the Goldsmiths, University of London Ethics Committee.

Statement–question discrimination

Each of the 18 statement–question pairs appeared in both 'same' (randomly-selected nine pairs as statement–statement and the other nine pairs as question–question) and 'different' (randomly-selected nine pairs as statement–question and the other nine pairs as question–statement) configuration. This resulted in 36 pairs in total in the testing session with 750 ms interstimulus interval. Two more trials, which were distinct from the test stimuli, were included in the practice session to acquaint participants with the testing procedure and sound materials. The task was run using Praat, and the 36 statement–question pairs were pseudo-randomized so that they appeared in the same order for all participants. Participants listened to one pair at a time and informed the experimenter whether the two stimuli were 'same' or 'different'. The experimenter then made the response on behalf of the participant. Following the response, another pair was played two seconds later.

The gliding tone and nonsense speech analogues, each including two practice trials, were arranged and presented to the participants in the same way as their natural speech counterparts. Participants were not informed that these stimuli were based on the statements and questions tested above.

Statement–question identification and imitation

In the identification task, the 36 statements and questions were pseudo-randomized and presented one at a time in a series to the participants, who were required to report whether they heard a statement or a question.

In the imitation task, the 36 statements and questions were presented again, in the same way as in the previous identification task. Participants were required to imitate the pitch patterns of these utterances as closely as possible. They were also reminded that even though they may have a different way of producing statements and questions compared with the model, they should try to mirror exactly the way the model phrased the utterances, using a comfortable pitch range and loudness level. The experimenter controlled the pace by playing the sentences one at a time. Participants imitated each sentence immediately following its presentation. The imitations were recorded in the same fashion as in the two speech production tasks (see Supplementary material 2 for details).

Pitch threshold tasks

In the pitch change detection task (Fig. 1A), participants heard three sounds, each 600 ms in duration. Two sounds were steady-state pure tones of 500 Hz; the other was a pure tone containing a pitch glide (250 ms steady-state onset, 100 ms excursion and 250 ms steady-state offset), logarithmically centred on 500 Hz. The participant was required to identify the sound containing the pitch glide as the 'odd-one-out', which always appeared as the first or last of the three-sound

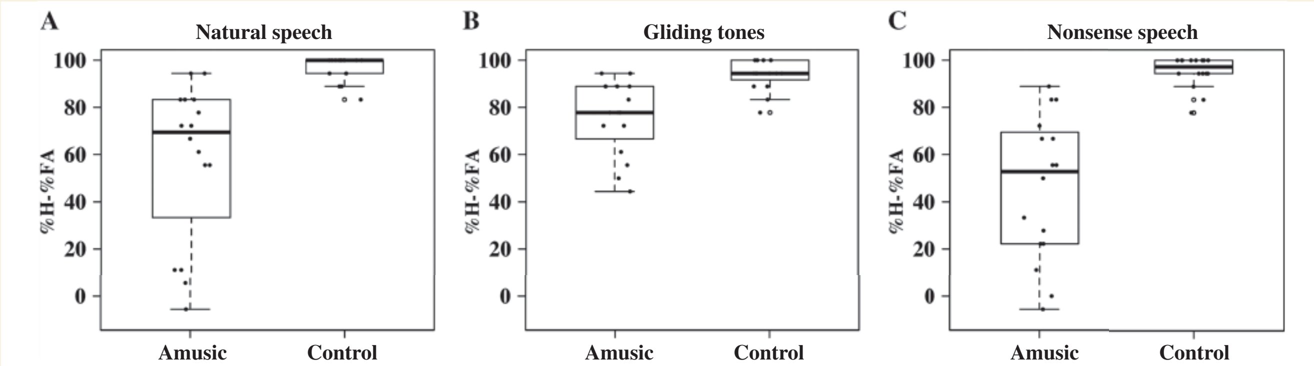

Figure 2 Boxplots of the 'percentage of hits – percentage of false alarms' (%H – %FA) scores of the amusic and control participants on the three discrimination tasks: (**A**) natural speech, (**B**) gliding tones and (**C**) nonsense speech. These boxplots contain the extreme of the lower whisker, the lower 'hinge', the median, the upper 'hinge' and the extreme of the upper whisker. The two 'hinges' are the first and third quartile, and the whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. The black dots denote individual measurements of the 16 participants in each group, with those arranged at the same level horizontally having the same value. The data points that lie beyond the extremes of the whiskers are outliers, which are further denoted by small open circles.

sequence. Participants indicated the position of the 'odd-one-out' using a verbal response: 'first' or 'last'.

Like the pitch change detection task, the pitch direction discrimination task (Fig. 1B) also required the participant to identify the 'odd-one-out' of three sounds centred on 500 Hz. Here, all sounds contained pitch glides: two in one direction; the other ('odd-one-out') in the opposite direction. Again, the target appeared in the first or last position.

Adaptive-tracking was used, with a '2 down, 1 up' staircase method and a variable change in step size. Starting with a default excursion of six semitones, the initial step size was one semitone, reducing to 0.1 semitones after four reversals and 0.02 semitones after eight reversals. The session was terminated after 14 track reversals and the threshold was calculated as the mean excursion value of the target glide for the last six reversals.

Statistical analyses

Statistical analyses were conducted in R (R Development Core Team, 2009). Since most of the data deviated significantly from a normal distribution (Shapiro–Wilk normality tests), non-parametric tests were performed on all data. Specifically, Wilcoxon rank sum tests (two-sided) were used for between-groups comparisons, Wilcoxon signed rank tests (two-sided) for within-group between-tasks comparisons, and Kendall's τ (one-sided) for correlation analyses. We do not report correlation analyses for the control participants because ceiling effects in this group would make interpretation of such analyses problematic.

Results

Statement–question discrimination

Performance on the three discrimination tasks was calculated in terms of 'percentage of hits–percentage of false alarms' (individual scores can be seen in Supplementary material 4). A hit was defined as a different pair judged as different, and a false alarm was defined as a same pair judged as different. 'Percentage of hits – percentage of false alarms' was the difference between percentage of hits (across 18 trials) and percentage of false alarms (across 18 trials). Figure 2 shows boxplots of individual 'percentage of hits–percentage of false alarms' scores on the three discrimination tasks from the two groups.

Wilcoxon rank sum tests revealed that the amusic group performed significantly worse than the control group on all three discrimination tasks [natural speech: amusics mean (SD): 57.6 (33.3), controls: 96.9 (5.4), $W = 9.5$, $P < 0.0001$; gliding tones: amusics: 76.0 (15.8), controls: 93.7 (6.4), $W = 30$, $P = 0.0002$; nonsense speech: amusics: 45.8 (30.5), controls: 95.1 (6.7), $W = 5.5$, $P < 0.0001$].

Pair-wise comparisons were done to examine performance differences across the three discrimination tasks for both amusics and controls. Amusics performed significantly better on gliding tones discrimination than on natural speech discrimination [mean difference (SD): 18.4 (29.7), Wilcoxon signed rank test $V = 25.5$, $P = 0.03$], which was in turn significantly better than nonsense speech discrimination [mean difference (SD): 11.8 (20.0), $V = 89$, $P = 0.02$]. Amusics' performance on gliding tones discrimination was also significantly better than nonsense speech discrimination [mean difference (SD): 30.2 (24.7), $V = 118$, $P = 0.001$]. There was no significant difference in performance across the three different types of discrimination tasks for controls (all $P > 0.05$), which may reflect a ceiling effect.

Analysis of the errors made by amusics on natural speech discrimination revealed that there were significantly more misses than false alarms (84.4 versus 15.6% of the total errors, $\chi^2 = 33.99$, df = 15, $P = 0.003$). A similar pattern was also seen in amusics' performance on discrimination of gliding tones (misses: 85.5%; false alarms: 14.5%; $\chi^2 = 26.85$, df = 15, $P = 0.03$) and nonsense speech (misses: 84.0%; false alarms: 16.0%; $\chi^2 = 57.11$, df = 15, $P < 0.0001$). These findings confirm that errors mainly arose

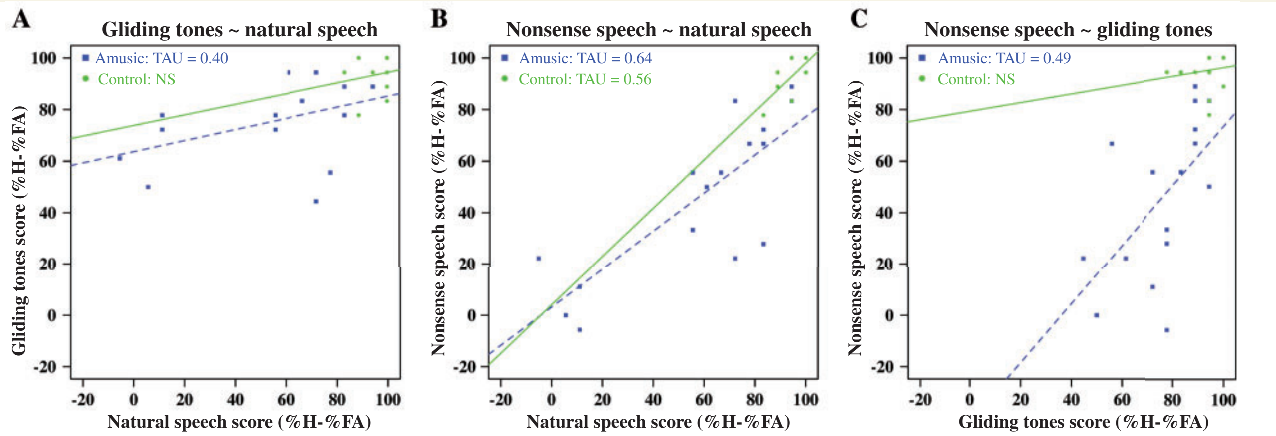

Figure 3 Scatter plots across the three discrimination tasks: (**A**) gliding tones scores against natural speech scores, (**B**) nonsense speech scores against natural speech scores and (**C**) nonsense speech scores against gliding tones scores. Regression lines were based on linear regressions of paired tasks for each group. %H – %FA = percentage of hits – percentage of false alarms; TAU = Kendall's τ; NS = non-significant.

because amusics failed to detect subtle differences between spoken utterances (Patel *et al.*, 2008).

Correlation analyses (Fig. 3) across the three discrimination tasks revealed a significant positive association between amusics' performance on all three tasks (natural speech and gliding tones: $z = 2.03$, $P = 0.02$, $\tau = 0.40$; nonsense speech and gliding tones: $z = 2.52$, $P = 0.01$, $\tau = 0.49$; natural speech and nonsense speech: $z = 3.33$, $P = 0.0004$, $\tau = 0.64$).

In order to examine what might have caused amusics' discrimination inaccuracy in the three tasks, we categorized trials into those which were correctly versus incorrectly discriminated by the amusic participants. We then compared final glide sizes, rates and durations between correct and incorrect trials for statements and questions of all three types of stimuli. Wilcoxon rank sum tests revealed that for natural speech and nonsense speech stimuli, incorrect trials had significantly smaller glide sizes than correct trials in statements (natural speech: $W = 14\,479.5$, $P = 0.003$; nonsense speech: $W = 17\,519.5$, $P = 0.01$), but not in questions (natural speech: $W = 18\,630$, $P = 0.66$; nonsense speech: $W = 21\,864$, $P = 0.36$). For gliding tones, incorrect trials had significantly smaller glide sizes than correct trials in both statements ($W = 8648$, $P = 0.001$) and questions ($W = 9360$, $P = 0.01$). There was no statistically significant difference in glide rate across incorrect versus correct trials in either statements or questions for all three types of stimuli. The effect of glide duration was only seen in gliding tones questions, with incorrect trials having significantly smaller glide durations than correct trials ($W = 7136$, $P < 0.0001$).

Logistic regressions of the counts of correct and incorrect responses by amusics on the length of the sentences (number of syllables) revealed no significant length effect for natural speech ($\chi^2 = 0.97$, df = 1, $P = 0.33$) or nonsense speech ($\chi^2 = 0.34$, df = 1, $P = 0.56$) discrimination. For gliding tones stimuli, the longer the sentence, the better was the discrimination ($\chi^2 = 13.01$, df = 1, $P = 0.0003$). These results demonstrate that amusics' discrimination inaccuracy is unlikely to have been caused by memory deficiency.

Statement–question identification and imitation

Performance on the identification task was calculated in terms of the percentage of correct responses across the 36 trials. Figure 4A shows boxplots of individual performance from the two groups (Supplementary material 4). Wilcoxon rank sum test revealed significant group difference for the statement–question identification task, with amusics showing significantly worse performance than controls [amusics mean (SD): 71.7 (10.3), controls: 92.4 (10.9), $W = 238.5$, $P < 0.0001$].

The imitation task explicitly required the participants to mirror exactly the pitch patterns of the model sentences, thus imitations were scored as correct only if they shared the same sign in glide size as the models (i.e. statements imitated as statements would have negative glide sizes, and questions imitated as questions with final rises would have positive glide sizes; see Supplementary material 5 for details), yielding percentage of correct responses scores for each participant (Fig. 4B). Wilcoxon rank sum test showed that controls did significantly better than amusics on imitating the pitch patterns of the model sentences [amusics mean (SD): 86.3 (12.6), controls: 97.7 (2.9), $W = 212$, $P = 0.001$]. Interestingly, we observed that although some participants imitated questions incorrectly (with a negative, as opposed to positive final glide size), they nonetheless produced a pitch pattern that conformed to other legitimate ways of producing a question in British English (Grabe, 2002), suggesting that these participants were sometimes only able to imitate the question by asking it in their own way, rather than mimicking the pitch pattern of the model. However, this was as likely to occur in controls as in amusics: for controls, 4% of questions (11 out of 288) were incorrectly imitated; 27% of which (three out of 11) were imitated using a legitimate question intonation; for amusics, 21% of questions (61 out of 288) were imitated incorrectly, of which 31% (19 out of 61) were imitated using a legitimate question intonation (see Supplementary material 5 for details).

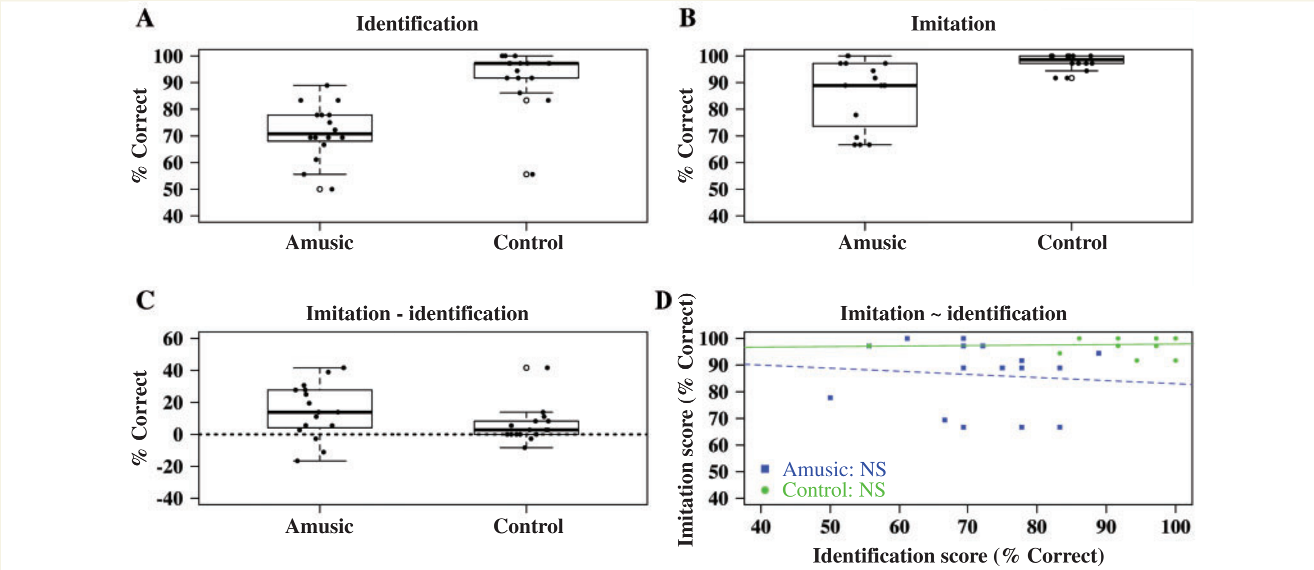

Figure 4 Boxplots and the scatter plot of the percentage of correct responses (%Correct) scores of the amusic and control participants on the identification and imitation tasks: (**A**) boxplots of identification scores, (**B**) boxplots of imitation scores, (**C**) boxplots of imitation performance minus identification performance and (**D**) the scatter plot of the identification and imitation scores, where the regression lines were based on the linear regressions of the paired scores. NS = non-significant.

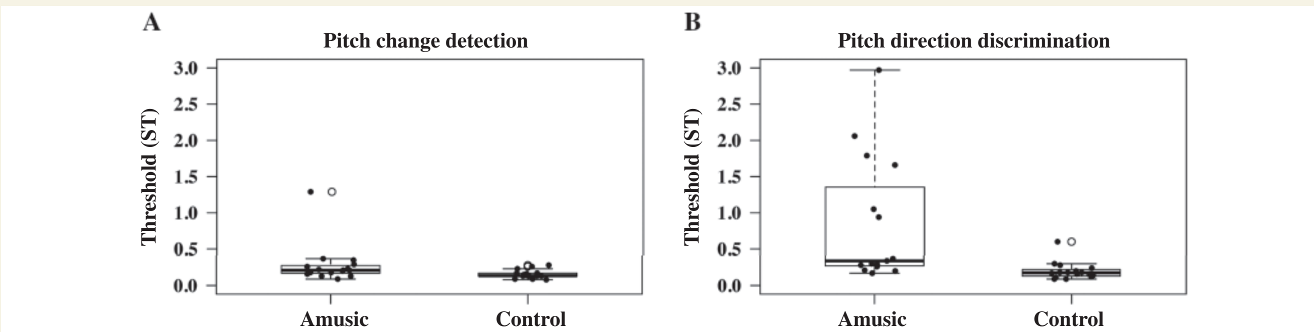

Figure 5 Boxplots of pitch thresholds (in semitones) of the amusic and control participants in the two psychophysical tasks: (**A**) pitch change detection and (**B**) pitch direction discrimination.

As shown in Fig. 4C, both groups (but particularly the amusic group) performed significantly better on the imitation task compared with the identification task [Wilcoxon signed rank test: amusics' mean difference (SD): 14.6 (16.9), $V = 120.5$, $P = 0.007$; controls: 5.4 (11.1), $V = 58$, $P = 0.03$], constituting a pattern of identification–imitation dissociation at the group level: better imitation than identification of statements and questions. As indicated in Fig. 4D, there was no significant correlation between identification and imitation scores for either group [(Kendall's τ (one-sided): correlation between amusics' identification and imitation scores: $z = -0.88$, $P = 0.81$, $\tau = -0.18$; controls: $z = 0.51$, $P = 0.30$, $\tau = 0.11$].

Pitch threshold tasks

Figure 5 shows the boxplots of individual pitch thresholds in the two psychophysical tasks from the amusic and control groups (Supplementary material 4). As found in previous studies (e.g. Foxton *et al.*, 2004), there was substantial overlap in pitch thresholds between the two groups in both tasks, with some amusics having thresholds within the control range.

Wilcoxon rank sum tests suggested that amusics as a group had significantly higher thresholds than controls in both pitch threshold tasks. The group difference for the pitch direction discrimination task was more pronounced [amusics' mean (SD): 0.86 (0.87); controls: 0.20 (0.12), $W = 212$, $P = 0.0003$] than that for the pitch change detection task [amusics: 0.28 (0.28); controls: 0.15 (0.06), $W = 193.5$, $P = 0.01$].

Correlations between intonation performance and pitch thresholds

Correlation analyses were conducted between participants' performance on the five intonation tasks and their pitch thresholds.

Figure 6 shows the corresponding scatter plots between participants' 'percentage of hits–percentage of false alarms' scores on the three discrimination tasks and their pitch thresholds.

For amusics, pitch change detection thresholds were negatively correlated with their performance on gliding tones discrimination (with marginal significance as shown in Fig. 6B: when the outlier was included: $z = -1.97$, $P = 0.02$, $\tau = -0.38$; when the outlier was removed: $z = -1.42$, $P = 0.08$, $\tau = -0.29$), but not with their performance on natural speech ($z = -0.73$, $P = 0.23$, $\tau = -0.14$) or nonsense speech ($z = -1.22$, $P = 0.11$, $\tau = -0.23$) discrimination. In contrast, their pitch direction discrimination thresholds were negatively correlated with their performance on all three discrimination tasks (the smaller the threshold, the better the discrimination performance; pitch direction and natural speech: $z = -1.80$, $P = 0.04$, $\tau = -0.35$; pitch direction and gliding tones: $z = -2.87$, $P = 0.002$, $\tau = -0.57$; pitch direction and nonsense speech: $z = -2.54$, $P = 0.01$, $\tau = -0.50$). This suggests that amusics' poor performance on the statement–question discrimination tasks is linked to a psychophysical pitch direction discrimination deficit.

Figure 7 shows scatter plots of participants' percentage of correct responses scores on the statement–question identification and imitation tasks against their pitch thresholds. For amusics, there was a negative association between their imitation scores and pitch direction discrimination thresholds (the smaller the thresholds, the better the imitation; $z = -1.71$, $P = 0.04$, $\tau = -0.34$), but no significant relationship was found for other three pairs of tasks (identification and pitch change detection: $z = -1.14$, $P = 0.13$, $\tau = -0.22$; imitation and pitch change detection: $z = 0.28$, $P = 0.61$, $\tau = 0.05$; identification and pitch direction discrimination: $z = -1.25$, $P = 0.11$, $\tau = -0.25$).

Discussion

Congenital amusia is not domain-specific

The extent to which amusia can be considered domain-specific is an actively debated question. While amusics have been shown to have deficits in other domains (Douglas and Bilkey, 2007; Thompson, 2007; Jones *et al.*, 2009*a*, *b*), conflicting results have been reported for their speech intonation perception abilities. Ayotte *et al.* (2002) found that amusics performed as well as controls on both identifying and discriminating the focus (shift in location) and sentence type information (statement versus question) of spoken utterances. However, they did significantly worse than controls on discrimination of the discrete tone analogues of these focus-shift and statement–question pairs. Similarly, in Patel *et al.* (2005), amusics achieved good performance on discrimination of focus-shift pairs in natural speech, but did poorly on both discrete and gliding tone analogues. These results suggested that the pitch processing deficits in amusia were domain-specific such that pitch processing within speech is preserved. However, as noted by Patel (2008), any dissociation in performance between natural speech and tone analogues in the context of focus-shift utterances may be accounted for by the different potential these stimuli offer for 'semantic recoding'. In other words, while salient pitch changes in a linguistic utterance can be 'tagged' to a particular syllable, the same pitch change in a delexicalized context must be encoded as part of the entire pitch pattern, which may exceed memory limits in amusics (Tillmann *et al.*, 2009; Williamson *et al*., 2010). In order to compare discrimination of pitch changes within, versus outside, speech, it is thus necessary to use statement–question discrimination, where the pitch patterns are identical until the final word such that the memory load for the speech and tone analogues is equivalent for both these conditions. Using precisely this approach, Patel *et al.* (2008) found that 30% of amusics exhibited worse performance on discrimination of statements and questions in natural speech than discrete tone analogues, pointing to speech intonation perception deficits in a minority of amusics. Given that the stimuli used in Patel *et al.* (2008) had relatively large pitch excursions, this raised the possibility that speech intonation perception deficits would be evident in the majority of amusics if tested using stimuli that included smaller excursions within natural speech.

Thus the present study used relatively small but ecologically valid intonational pitch contrasts, and examined statement–question perception abilities of 16 British amusics and 16 matched controls. Compared with controls, amusics demonstrated impaired ability to discriminate between statements and questions that differed mainly in the direction of final pitch glides, regardless of whether these stimuli were natural speech or analogues made from gliding tones and nonsense speech. These findings extend those of Patel *et al.* (2008), where only 30% of amusics had difficulties in distinguishing between natural speech statements and questions, and suggest that, for the majority of amusics, intonation perception is not as robust as controls. Acoustic analysis of the final pitch glides in correct versus incorrect trials revealed that, in most cases, incorrectly discriminated trials had significantly smaller final glide sizes than correctly discriminated trials. Furthermore, a significant negative correlation was observed between amusics' performance on statement–question discrimination and their psychophysical pitch direction discrimination thresholds (the better the discrimination, the smaller the thresholds). Finally, amusics also showed impaired ability to identify and imitate statements and questions in natural speech, and their imitation performance was also negatively correlated with their pitch direction discrimination thresholds. These results indicate that under laboratory conditions, it is possible to observe cases of speech intonation perception deficits in amusics when the sizes of intonational pitch contrasts are small, providing evidence that amusia involves a pitch direction discrimination deficit that is domain-general rather than specific to the musical domain.

It is worth noting that previous authors have acknowledged that the deficit may be 'music-relevant', rather than 'music-specific', owing to the smaller pitch changes utilized in music, as opposed to speech (Peretz and Hyde, 2003; Patel, 2008). However, the present study revealed deficits in the perception of speech intonation using real-world natural speech stimuli. Thus, even if natural speech in everyday situations is rarely problematic for amusics (presumably due to the presence of additional cues to meaning),

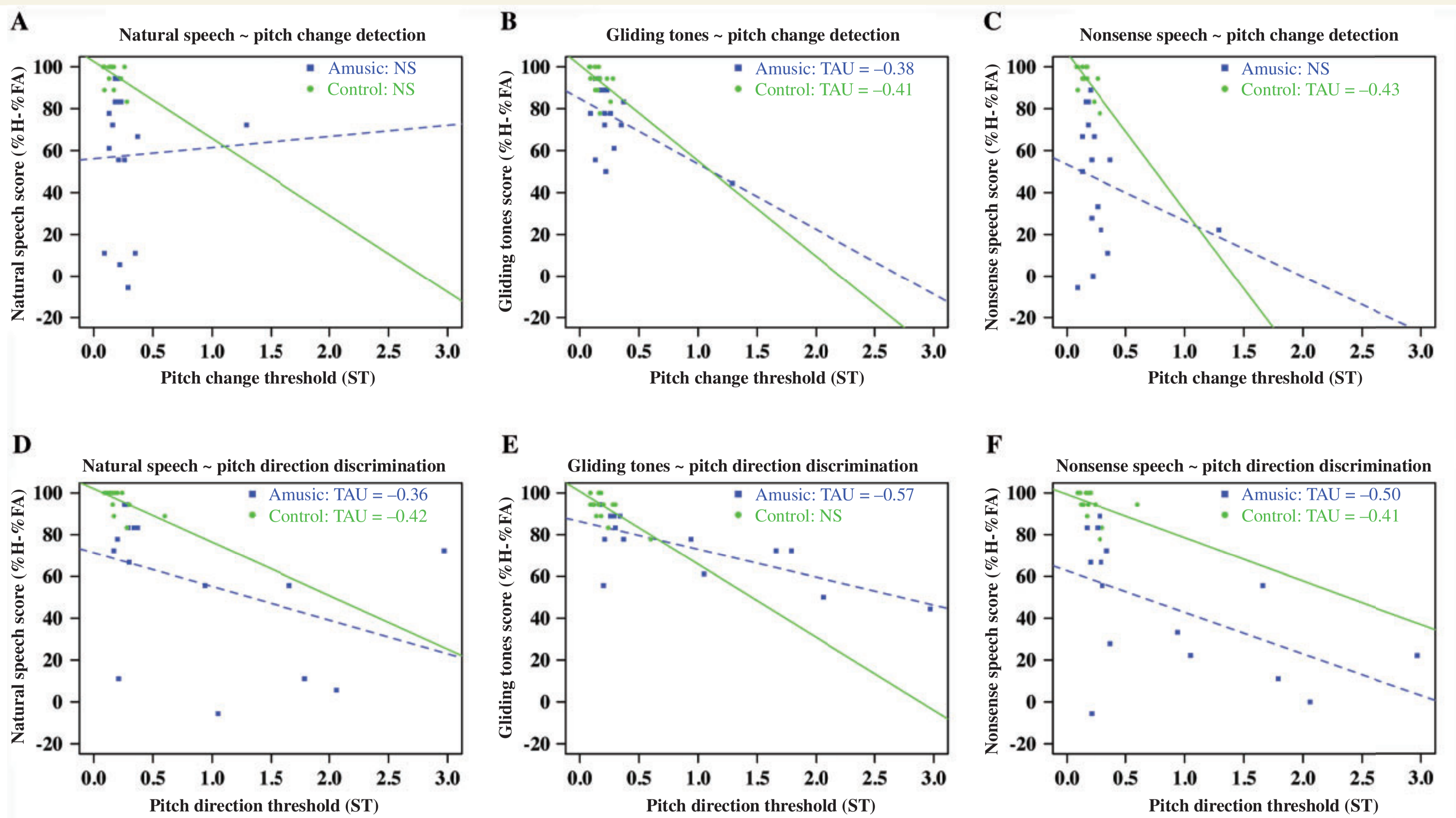

Figure 6 Scatter plots of discrimination scores against pitch thresholds: (**A**) natural speech scores against pitch change detection thresholds, (**B**) gliding tones scores against pitch change detection thresholds, (**C**) nonsense speech scores against pitch change detection thresholds, (**D**) natural speech scores against pitch direction discrimination thresholds, (**E**) gliding tones scores against pitch direction discrimination thresholds and (**F**) nonsense speech scores against pitch direction discrimination thresholds. Regression lines were based on linear regressions of paired tasks for each group. ST = semitone; TAU = Kendall's τ; NS = non-significant; %H – %FA = percentage of hits – percentage of false alarms.

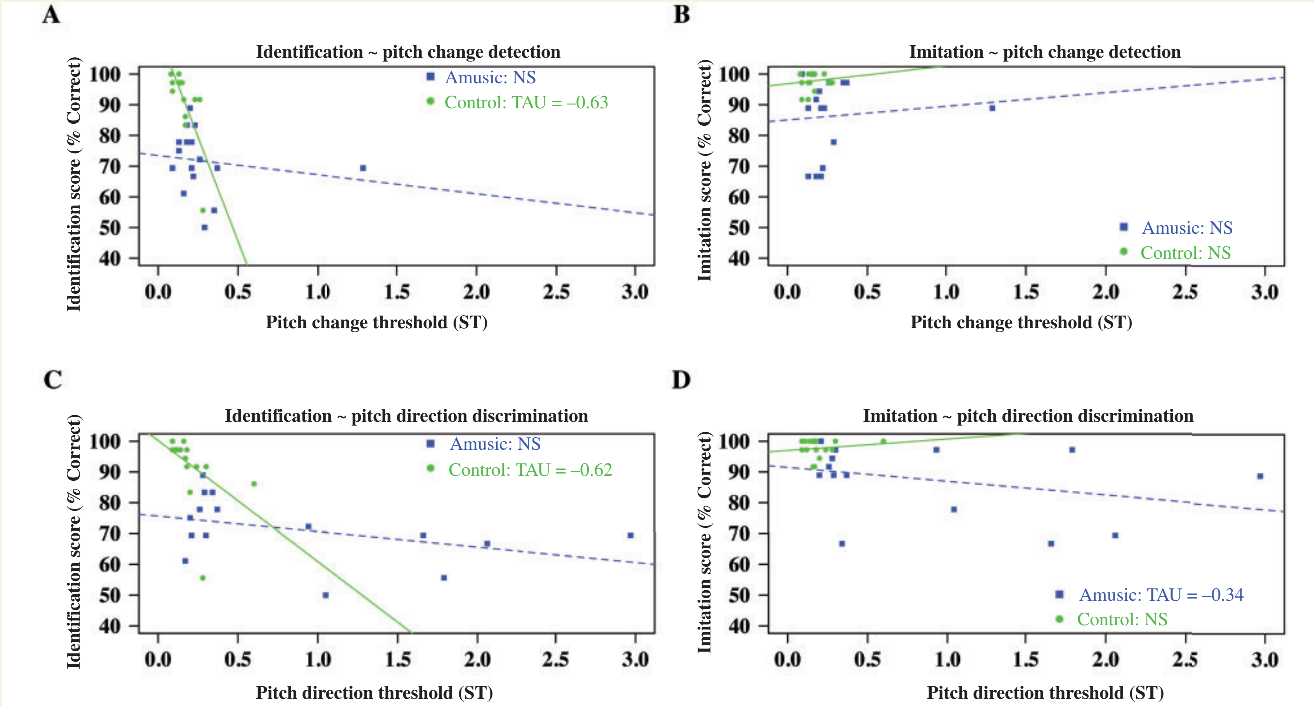

Figure 7 Scatter plots of identification and imitation scores against pitch thresholds: (**A**) identification scores against pitch change detection thresholds, (**B**) imitation scores against pitch change detection thresholds, (**C**) identification scores against pitch direction discrimination thresholds and (**D**) imitation scores against pitch direction discrimination thresholds. Regression lines were based on linear regressions of paired tasks for each group. ST = semitone; TAU = Kendall's τ; NS = non-significant; %Correct = percentage of correct responses.

the present study demonstrates that pitch deficits in amusia can be behaviourally relevant for both speech and music. It is also interesting to note that data from our language questionnaire (Supplementary material 3) suggest that the percentage of people who reported difficulties in speaker identification and recognition of regional/foreign accents was bigger in the amusic group than in the control group. We hypothesize that carefully-designed experimental studies involving subtle differences in speaker voices and speech accents may reveal deficits in amusic participants.

As with previous studies of amusia, there was significant variability in performance across the amusic group, with one quarter of participants behaving like controls in the discrimination tasks (Supplementary material 4). However, it is not simply the case that these participants are less severely affected in their musical perception, since no significant correlation was found between amusics' performance on intonation tasks and their scores on MBEA subtests (Kendall's τ: $P > 0.05$). It remains an issue for future research to determine why a small subgroup of amusics nevertheless performs well on tests of pitch direction and intonation discrimination.

Identification–imitation dissociation in congenital amusia

While amusics performed worse than controls on both identification and imitation in the present study, both groups—but particularly the amusic group—showed superior performance for imitation compared with identification. This is somewhat similar to the results reported in Loui *et al.* (2008), although the dissociation reported in that study (production of a pitch interval better than perception, measured by imitation and labelling, respectively) held for amusics alone. While the dissociation reported for amusics in Loui *et al.* (2008) follows the criteria for a 'classical dissociation' (task A performed in the normal range, task B performed below the normal range), the pattern of dissociation found in the current study for the amusic group follows the criteria for a 'strong dissociation', in which neither task is performed at a normal level, but task A (imitation) was performed very much better than task B (Shallice, 1988, p. 227–8). While this pattern held for amusics at the group level, some individuals showed the reverse of this, producing incorrect imitations that, nonetheless, conformed to legitimate ways of asking a question in British English (Grabe, 2002), suggesting that these individuals had correctly discerned that utterance as a statement or question but were unable to mimic the utterance in another's style. There was some evidence of this in two control participants as well but when prompted they were able to adjust their motor programmes, while the same was not true of the amusic participants. British English provides a particularly interesting model for addressing the extent to which individuals can modify their own intonation patterns to mimic those of another, given the potential for questions to be realized in several different ways. In future studies it would be important to investigate the relationship explicitly between pitch patterns used in spontaneous and imitated speech, in order to examine the extent to which amusics can match pitch patterns that may not be common to their action repertoire, thus informing discussions on the coupling between perception and action (Bosshardt *et al*., 1997; Braun *et al*., 2006; Loui *et al*., 2008; Over and Gattis, 2010).

Glide rate versus glide size/duration

In Patel *et al*. (2008), acoustic analyses of correctly versus incorrectly discriminated speech pairs by the 30% of British amusics who showed statement–question discrimination deficits indicated that final glide rates were slower in incorrectly versus correctly discriminated trials. However, this pattern did not hold in a separate group of French amusics performing a similar task in their native language. Analysis of the current data also found no rate difference for correctly versus incorrectly discriminated final glides by our amusic participants. Rather, their errors were associated with reduced final glide sizes in natural and nonsense speech statements and in gliding tones statements and questions. Furthermore, in gliding tones, incorrectly discriminated questions had significantly smaller glide durations than correctly discriminated ones. The discrepancy in the role of glide rate versus glide size/duration in amusics' discrimination inaccuracy may be due to different pitch contrasts involved as in Patel *et al*. (2008) and the current study. When final glides are large in size (on average 5–12 semitones) as in Patel *et al*. (2008), glide rate may play an important role in one's discrimination of pitch directions: the faster the rate, the better the discrimination. When final glide sizes are relatively small as in the current study (on average 4–5 semitones), glide rate may play a less important role than glide size/duration in one's discrimination of pitch directions. It will be of interest to investigate this issue further in future studies.

Pitch thresholds in speech versus non-speech contexts

In the current study, the amusic group showed significantly higher thresholds than the control group in both pitch threshold tasks. However, most amusics achieved thresholds of less than one semitone for both detection of pitch change and discrimination of pitch direction. Given that the smallest glide size in our intonation tasks was 2.4 semitones, which exceeded all but one amusic's pitch direction discrimination threshold, one may wonder why the amusic group still encountered problems in intonation perception. One possibility is that people have different pitch thresholds in speech versus non-speech contexts. As reviewed by Moore (2008), F_0 contrasts that convey linguistic meaning are much larger than psychophysically obtained pitch thresholds. However, despite the likely absolute differences in threshold for pitch changes in speech versus non-speech, the relationship between thresholds across the two domains can be seen in the present study, since in most cases, the trials which were incorrectly discriminated by amusics had significantly smaller glide sizes than those which were correctly discriminated. Furthermore, amusics' pitch direction discrimination thresholds were negatively correlated with their performance on the three statement–question discrimination tasks: the smaller the pitch threshold, the better the performance on intonation perception.

Conclusion

In summary, despite reporting normal speech communication abilities in language questionnaires, the amusic group in the current study showed a statement–question discrimination deficit when exposed to small pitch direction contrasts at the sentence-final position. Compared with controls, the amusic group also showed impaired performance on statement–question identification and imitation. Except for statement–question identification, amusics' performance on intonation tasks appeared related to a psychophysical pitch direction discrimination deficit. These findings indicate that congenital amusia is not a music-specific disorder, and support previous evidence that the processing of pitch within and outside language may share common mechanisms.

Acknowledgements

We owe our thanks to Guy Peryer for assistance on collecting language questionnaires; to Yi Xu for suggestions of using the /li–li–li/ stimuli and language questionnaires; to Sukhbinder Kumar for development of the psychophysical procedures; and to Bruno Repp for insightful comments on an earlier version of the manuscript. We also thank two anonymous reviewers for helpful comments.

Funding

This work was supported by the Economic and Social Research Council (grant numbers RES-061-25-0155 to L.S., PTA-026-27-2480 to F.L.); and by Neurosciences Research Foundation as part of its programme on music and the brain at The Neurosciences Institute, where A.D.P. is the Esther J. Burnham Senior Fellow.

Supplementary material

Supplementary material is available at *Brain* online.

References

Ayotte J, Peretz I, Hyde K. Congenital amusia: a group study of adults afflicted with a music-specific disorder. Brain 2002; 125: 238–51.

Boersma P. Praat: a system for doing phonetics by computer. Glot Int 2001; 5: 341–5.

Bosshardt HG, Sappok C, Knipschild M, Hölscher C. Spontaneous imitation of fundamental frequency and speech rate by nonstutterers and stutterers. J Psycholinguist Res 1997; 26: 425–48.

Braun B, Kochanski G, Grabe E, Rosner BS. Evidence for attractors in English intonation. J Acoust Soc Am 2006; 119: 4006–15.

Dalla Bella S, Peretz I. Congenital amusia interferes with the ability to synchronize with music. Ann NY Acad Sci 2003; 999: 166–9.

Dalla Bella S, Giguère J, Peretz I. Singing in congenital amusia. J Acoust Soc Am 2009; 126: 414–24.

Douglas K, Bilkey D. Amusia is associated with deficits in spatial processing. Nat Neurosci 2007; 10: 915–21.

Drayna D, Manichaikul A, de Lange M, Snieder H, Spector T. Genetic correlates of musical pitch recognition in humans. Science 2001; 291: 1969–72.

Foxton JM, Dean JL, Gee R, Peretz I, Griffiths TD. Characterization of deficits in pitch perception underlying 'tone deafness'. Brain 2004; 127: 801–10.

Goedemans R. Weightless segments – a phonetic and phonological study concerning the metrical irrelevance of syllable onsets [LOT dissertation series 9]. The Hague: Holland Academic Graphics; 1998.

Grabe E. Variation adds to prosodic typology. In: Bel B, Marlien I, editors. Proceedings of Speech Prosody 2002. Aix-en-Provence: Laboratoire Parole et Langage; 2002. p. 127–32.

Griffiths T. Tone deafness: a model complex cortical phenotype. Clin Med 2008; 8: 592–5.

Hyde K, Peretz I. Brains that are out of tune but in time. Psychol Sci 2004; 15: 356–60.

Hyde K, Zatorre R, Griffiths T, Lerch J, Peretz I. Morphometry of the amusic brain: a two-site study. Brain 2006; 129: 2562–70.

Hyde K, Lerch J, Zatorre R, Griffiths T, Evans A, Peretz I. Cortical thickness in congenital amusia: when less is better than more. J Neurosci 2007; 27: 13028–32.

Jones J, Lucker J, Zalewski C, Brewer C, Drayna D. Phonological processing in adults with deficits in musical pitch recognition. J Commun Disord 2009a; 42: 226–34.

Jones J, Zalewski C, Brewer C, Lucker J, Drayna D. Widespread auditory deficits in tune deafness. Ear Hear 2009b; 30: 63–72.

Kalmus H, Fry D. On tune deafness (dysmelodia): frequency, development, genetics and musical background. Ann Hum Genet 1980; 43: 369–82.

Loui P, Guenther F, Mathys C, Schlaug G. Action-perception mismatch in tone-deafness. Curr Biol 2008; 18: R331–2.

Mandell J, Schulze K, Schlaug G. Congenital amusia: an auditory-motor feedback disorder? Restor Neurol Neurosci 2007; 25: 323–34.

Moore BCJ. Basic auditory processes involved in the analysis of speech sounds. Philos Trans R Soc Lond B Biol Sci 2008; 363: 947–63.

Nelson HE, Willison J. National adult reading test manual. 2nd edn. Windsor, UK: NFER-Nelson; 1991.

Over H, Gattis M. Verbal imitation is based on intention understanding. Cogn Dev 2010; 25: 46–55.

Patel AD. Music, language, and the brain. New York: Oxford University Press; 2008.

Patel AD, Foxton JM, Griffiths TD. Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. Brain Cogn 2005; 59: 310–3.

Patel AD, Peretz I, Tramo M, Labreque R. Processing prosodic and musical patterns: a neuropsychological investigation. Brain Lang 1998; 61: 123–44.

Patel AD, Wong M, Foxton J, Lochy A, Peretz I. Speech intonation perception deficits in musical tone deafness (congenital amusia). Music Percept 2008; 25: 357–68.

Peretz I. The nature of music from a biological perspective. Cognition 2006; 100: 1–32.

Peretz I. Musical disorders: from behavior to genes. Curr Dir Psychol Sci 2008; 17: 329–33.

Peretz I, Coltheart M. Modularity of music processing. Nat Neurosci 2003; 6: 688–91.

Peretz I, Hyde K. What is specific to music perception? Insights from congenital amusia. Trends Cognit Sci 2003; 7: 362–7.

Peretz I, Champod S, Hyde K. Varieties of musical disorders: the Montreal Battery of Evaluation of Amusia. Ann NY Acad Sci 2003; 999: 58–75.

Peretz I, Cummings S, Dubé MP. The genetics of congenital amusia (or tone-deafness): a family aggregation study. Am J of Hum Genet 2007; 81: 582–8.

Peretz I, Ayotte J, Zatorre R, Mehler J, Ahad P, Penhune V, et al. Congenital amusia: a disorder of fine-grained pitch discrimination. Neuron 2002; 33: 185–91.

R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; ISBN 3-900051-07-0; URL http://www.R-project.org (14 December 2009, date last accessed); 2009.

Shallice T. From neuropsychology to mental structure. Cambridge, UK: Cambridge University Press; 1988.

Stewart L. Fractionating the musical mind: insights from congenital amusia. Curr Opin Neurobiol 2008; 18: 127–30.

Thompson WF. Exploring variants of amusia: tone deafness, rhythm impairment, and intonation insensitivity. In: Schubert E, Buckley K, Eliott R, Koboroff B, Chen J, Stevens C, editors. Proceedings of the International Conference on Music Communication Science. Sydney: HCSNet; 2007. p. 159–63.

Tillmann B, Schulze K, Foxton JM. Congenital amusia: a short-term memory deficit for non-verbal, but not verbal sounds. Brain Cogn 2009; 71: 259–264.

Wechsler D. Wechsler adult intelligence scale-III (WAIS-III). San Antonio (TX): The Psychological Corporation; 1997.

Williamson V, McDonald C, Deutsch D, Griffiths T, Stewart L. Faster decline of pitch memory over time in congenital amusia. Adv Cogn Psychol 2010; 6: 15–22.

Xu Y. Speech melody as articulatorily implemented communicative functions. Speech Commun 2005; 46: 220–51.