Congenital amusics use a secondary pitch mechanism to identify lexical tones

Oliver Bones[a,*], Patrick C.M. Wong[b,*]

[a] Acoustics Research Centre, University of Salford, Salford M5 4WT, UK
[b] Department of Linguistics & Modern Languages and Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong SAR

ARTICLE INFO

ABSTRACT

Amusia is a pitch perception disorder associated with deficits in processing and production of both musical and lexical tones, which previous reports have suggested may be constrained to fine-grained pitch judgements. In the present study speakers of tone-languages, in which lexical tones are used to convey meaning, identified words present in chimera stimuli containing conflicting pitch-cues in the temporal fine-structure and temporal envelope, and which therefore conveyed two distinct utterances. Amusics were found to be more likely than controls to judge the word according to the envelope pitch-cues. This demonstrates that amusia is not associated with fine-grained pitch judgements alone, and is consistent with there being two distinct pitch mechanisms and with amusics having an atypical reliance on a secondary mechanism based upon envelope cues.

1. Introduction

Congenital amusia, commonly referred to as "tone deafness", is a life-long disorder affecting processing and production of pitch. Pitch can be defined as that perception required for the discrimination of musical notes (Plack and Oxenham, 2005), and as such is fundamental to the perception and enjoyment of music. Sounds that produce pitch are mostly those that contain harmonic tones. These "complex tones" consist of a number of sinusoidal tones ("harmonics") that are all integer multiples of a common low tone; the fundamental frequency ($F0$). The repetition rate of the $F0$, which is usually the same as the overall repetition rate of the complex tone (the temporal "envelope"), is the acoustic feature that is most related to pitch.

The mechanical qualities of the cochlea's basilar membrane are such that it behaves like a bank of bandpass "auditory filters" (Fletcher, 1940). If a single harmonic falls within the bandwidth of an auditory filter (see Glasberg and Moore, 1990) the harmonic is said to be "resolved". Auditory nerve (AN) fibres innervating the output of the auditory filter fire at a particular phase of the resolved harmonic (Brugge et al., 1969; Russell and Sellick, 1978), such that the temporal fine-structure (TFS) is coded in the resulting inter-spike intervals. When more than one harmonic falls within the bandwidth of an auditory filter they are said to be "unresolved" (above approximately the 8th to 10th harmonic; Bernstein and Oxenham, 2003; Plomp, 1964). Interactions between unresolved harmonics within the auditory filter produce complex AN firing patterns. Pitch judgements are much easier for sounds containing resolved harmonics and which therefore produce TFS cues, compared to sounds containing unresolved harmonics and which therefore do not (Bernstein and Oxenham, 2003; Houtsma and Smurzynski, 1990). An autocorrelation-like analysis of TFS information is fundamental to most models of pitch perception, whereby the dominant period in inter-spike intervals determines pitch (e.g. Licklider, 1951; Meddis and Hewitt, 1991; Moore, 2003; but see Oxenham et al., 2011).

A recent study by Cousineau et al. (2015) tested whether congenital amusia ("amusia" hereafter) is associated with a deficit in processing TFS cues. They found no evidence for an effect of amusia on discrimination of interaural phase differences (IPD), suggesting that peripheral representation of TFS remained intact: these localization cues depend upon processing TFS at the lateral and medial superior olivary complexes respectively (Grothe et al., 2010; McAlpine and Grothe, 2003), early stages of the brainstem at which binaural interaction first occurs. Amusics and controls also performed equally well when discriminating the pitch of complex tones containing only unresolved harmonics (and which therefore did not contain TFS pitch-cues). However, amusics were poorer at discriminating complex tones containing resolved harmonics. Any mechanism for pitch-extraction is likely to reside beyond the superior olive, at or after the lateral lemniscus or the inferior colliculus (Gockel et al., 2011), suggesting that any TFS processing deficit associated with impaired pitch perception in amusia must reside at or central of this stage of the auditory brainstem. Consistent with this, a recent study has failed to find evidence of

impaired pitch encoding in the rostral auditory brainstem as indicated by the frequency-following response (Liu et al., 2014, although see also Lehmann et al., 2015).

One interpretation of these results is that amusia only affects fine-grained pitch judgements (Peretz et al., 2002): the absence of TFS cues in the unresolved condition may have added a level of noise to the pitch mechanism beyond that resulting from the effect of amusia. However, recent work suggests that amusia affects pitch discrimination of pure tones at frequencies at which it assumed place rather than temporal cues are used, and at which fine-grained pitch discrimination is therefore not possible (Whiteford and Oxenham, 2017). Alternatively, the pitch discrimination results reported by Cousineau et al. (2015) are consistent with there being two distinct temporal pitch mechanisms using TFS and envelope cues. Although performance in pitch discrimination is typically better with complex tones containing resolved harmonics, when complex tones contain only unresolved harmonics performance improves with increasing number of harmonics (Houtsma and Smurzynski, 1990), presumably because pitch perception in this case is dependent on envelope rate and the representation of the $F0$ in the envelope produced by interactions on the basilar membrane is enhanced with a greater number of harmonics (Moore and Moore, 2003). It may therefore be that amusia is a deficit that affects a primary pitch mechanism that processes TFS cues, but not a secondary mechanism that processes envelope cues. The aim of the current study is to investigate this question of whether amusia affects a primary pitch mechanism dependent upon TFS cues but not pitch discrimination based on envelope cues. This is approached by assessing the effects of amusia when both TFS and envelope pitch-cues are available but provide conflicting information, in a task that does not require fine-grained pitch discrimination judgements.

Smith et al. (2002) introduced a method for assessing the relative perceptual importance of envelope and TFS cues, by synthesizing "chimera" stimuli containing the envelope of one stimulus and the TFS of another. They demonstrated that for English language speech-speech chimeras transmitting distinct utterances in the envelope and the TFS, participants heard words contained in the envelope information approximately 80% of the time when using eight or more frequency channels to create the chimera. However, participants nearly always heard the melody information contained in TFS when listening to melody-melody chimeras, unless created using 32 or more frequency channels. Unlike English, tone-languages such as Mandarin use changes in pitch to convey meaning. Mandarin uses four lexical tones: the syllable /ma/ spoken with a high-level, rising, falling-rising, and falling tone means "mother", "hemp", "horse", and "scold" respectively. Using the method described by Smith et al. (2002), Xu and Pfingst (2003) found that Mandarin Chinese speakers identified the pitch contour and therefore the utterance in the TFS 85–90% of the time when using Mandarin speech-speech chimeras consisting of a single syllable but two lexical tones transmitted in the envelope and the TFS, demonstrating that for normal-hearing Mandarin speakers the lexical tone is typically perceived using TFS cues.

Although detection of French and English pitch-cues for speech intonation has been found to remain intact in amusics (Ayotte et al., 2002), the pitch intervals used in English intonation (typically between approximately 5–12 semitones) are larger than those used in both Western music (typically 1 or 2 semitones between consecutive notes) and in Mandarin (typically between approximately 2–5 semitones; Chao, 1948). Non-tone-language speaking amusics are less able to discriminate Mandarin lexical tones (Nguyen et al., 2009), and tone-language speaking amusics perform more poorly at lexical tone identification (Nan et al., 2010) and speech intelligibility tasks (Liu et al., 2015) indicating that whilst amusia most commonly manifests as a deficit in music perception in non-tone-language speakers, it is not a music-specific condition.

The current study tests whether native Mandarin-speaking amusics use envelope cues to determine lexical pitch and therefore identify

Table 1
Participant information.

	Participant	MBEA Pitch	Sex	Age	TONI-4	Years education
Controls	1	80	F	24	110	16
	2	70	M	23	104	15
	3	88	M	22	112	15
	4	84	F	18	108	13
	5	74	F	23	104	15
	6	83	F	20	110	15
	7	78	F	21	100	15
	8	74	F	19	100	13
	9	88	M	20	119	13
	10	79	M	19	100	14
	11	74	M	20	90	15
	12	77	F	19	102	13
	13	77	M	20	100	13
	14	80	M	19	104	13
	15	82	F	21	104	15
	16	74	F	21	106	15
	17	88	M	19	110	13
	18	80	F	19	106	13
	19	89	M	20	125	13
	20	79	F	18	108	13
	21	76	F	19	110	13
	22	81	F	18	106	13
	23	81	M	20	108	14
	24	83	F	19	94	13
Amusics	25	65	M	20	110	14
	26	52	F	20	100	13
	27	60	F	19	102	13
	28	65	M	25	106	17
	29	65	F	20	104	13
	30	64	F	21	102	14
	31	65	F	21	100	15
	32	63	F	21	104	15
	33	61	F	20	102	13
	34	65	F	21	98	13
	35	61	M	20	105	13
	36	65	F	20	98	13
	37	64	F	21	98	14
	38	62	M	23	96	15
	39	58	F	20	100	14
	40	64	F	20	108	13
	41	53	M	18	100	13
	42	64	M	19	113	13
	43	59	F	20	112	14
	44	62	M	24	102	18
	45	65	F	24	110	17
	46	63	F	21	100	16

speech even when TFS cues are available by assessing which lexical tone amusics perceive in chimera speech-speech stimuli. This would provide compelling evidence that amusia is associated with atypical TFS pitch-cue processing, and provide further evidence for there being two distinct pitch mechanisms.

2. Methods

2.1. Participants

All participants were Mandarin Chinese speakers recruited from universities in Shenzhen, Guangdong Province, China. All instructions were provided in Mandarin Chinese. Forty-six participants (see Table 1) were recruited in total (mean age 20 years, 29 females): 24 controls (14 females) and 22 amusics (15 females). The two groups were not found to be significantly different in sex ($\chi^2 = 0.48$, p = 0.49), age (mean, controls = 20, amusics = 21, SE, controls = 0.33, amusics = 0.37, $W = 345.5$, $r = -0.27$, p = 0.07), or years of formal education (mean, controls = 13.9, amusics = 14.2, SE, controls = 0.21, amusics = 0.33, $W = 287.5$, $r = -0.08$, p = 0.59). Nonverbal intelligence as measured using the TONI-4 test (Brown et al., 2010) was also not found to be

significantly different between groups (mean, controls = 106, amusics = 103, *SE,* controls = 1.49, amusics = 1.03, $t_{(40)} = -1.47, p = 0.15$).

Amusics were recruited by advertising for and screening individuals who felt they had difficulty perceiving music. None of the participants had ever received formal musical training, and all were right-handed. All participants had hearing thresholds of < 20 dBHL at 500, 1000, 1500, 2000, 3000, and 4000 Hz for both ears, and all participants were assessed for amusia using the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al., 2003). The MBEA consists of six tests: melody scale, contour and interval discrimination; rhythm discrimination; meter perception; and melody memory. Scores 2 standard deviations or more below the mean of available normative data based upon Western, non-tone-language speakers are typically considered to indicate amusia (a global composite score of 78%). The melody tests of the MBEA use melodies constructed from Western scales, and the majority of research using the MBEA has concerned non-tone-language speakers. However, more recent work concerning Chinese tone-language speakers appears to indicate the validity of this measure in the present study (e.g. Jiang et al., 2010, 2012; Liu et al., 2015; Nan et al., 2010).

In the current study, an abbreviated version of the MBEA was used, consisting of the three melody tests (Liu et al., 2015, 2010). Two standard deviations below the mean of normative data on these tests corresponds to a score of 65 out of a possible 90, or 72%.

2.2. Peripheral processing of TFS and envelope cues

All procedures were performed at the Chinese University of Hong Kong Research Institute in Shenzhen. Peripheral processing was assessed via IPD thresholds, measured using a two-alternative forced-choice paradigm in which participants were presented stimuli in two intervals, separated by 500 ms silence. Four tones were presented in each interval, separated by 20 ms silence. In one interval all four stimuli had the same phase (Φ) at both ears, whilst in the other interval the second and fourth tones contained an interaural phase shift ($\Delta\Phi$) between the two ears (Hopkins and Moore, 2010), producing a sensation of the tone moving. The task was to identify the interval in which the tones moved. The target interval was randomised between trials. Responses were collected via a bespoke MATLAB interface, and feedback was provided after each trial.

Tones had durations of 500 ms including 50 ms raised cosine ramps, and were presented at 75 dB SPL via Sennheiser HD 380 Pro headphones. All tones had a frequency of 500 Hz, amplitude modulated (AM) at 20 Hz to a depth of 100%. In one condition IPDs were created in the TFS of the AM tone, in a second condition IPDs were created in the envelope (King et al., 2014; Lacher-Fougère and Demany, 2005). At the beginning of each block $\Delta\Phi$ was set to 180°. Following two consecutive correct responses $\Delta\Phi$ was reduced; following an incorrect response $\Delta\Phi$ was increased. For the first four turn-points $\Delta\Phi$ was reduced or increased by a factor of 1.25^2; for the following eight turn-points $\Delta\Phi$ was reduced or increased by a factor of 1.25. The maximum $\Delta\Phi$ permitted was 180°. A trial ended after 12 turn-points, with the geometric mean of $\Delta\Phi$ at the eight smaller turn-points recorded as the threshold for the block. Participants each performed two blocks of AM tones with $\Delta\Phi$ in the TFS and two blocks of $\Delta\Phi$ in the envelope, with the mean of the two thresholds from the two blocks recorded as the final estimated threshold.

2.3. Pitch perception

Pitch perception using TFS cues was assessed via frequency discrimination limens (FDL), using a three-alternative forced-choice paradigm. Participants were presented tones in three intervals, separated by 500 ms. Two of the intervals contained a 600 Hz (*f*) pure tone, one of the intervals contained a pure tone with frequency *f*+*Δf*. The task was to identify the interval containing the tone with frequency *f* +*Δf*, which was randomized between trials. Responses were collected

via a bespoke MATLAB interface, and feedback was provided after each trial.

Tones had durations 500 ms including 50 ms raised-cosine ramps, and were presented at 75 dB SPL via Sennheiser HD 380 Pro headphones. At the start of each block Δf was set at *f*/2. After two consecutive correct responses Δf was reduced, after one incorrect response Δf was increased. For the first four turn-points Δf was reduced or increased by a factor of 2. For the eight subsequent turn-points Δf was reduced or increased by a factor of $\sqrt{2}$. A trial ended after 12 turn-points, with the geometric mean of Δf at the eight smaller turn-points recorded as the threshold for the block. Participants each performed three blocks, with the mean of the three thresholds from the three blocks recorded as the FDL.

2.4. Speech-speech chimera

Initially a native male and a native female speaker each provided ten spoken samples of each of the four Mandarin lexical tones using the Mandarin Chinese syllables /ma/ and /yi/. All speech samples were then normalized to 70 dB SPL, and to duration of 500 ms using Praat. These were then assessed by a native Mandarin speaker, who identified the most natural sounding example of each of the 8 speech tokens for each speaker. Chimeras were created using the method described by Smith et al. (2002). The method uses the Hilbert transform to decompose the signal into TFS and envelope components across a number of frequency channels. In the present study 16 filters were spaced in equal steps along a cochlear frequency map (Greenwood, 1990) between 80–17640 Hz, with a filter transition of 25% of the bandwidth of the narrowest filter. The upper frequency is higher than has been reported using this method previously, due to being defined as 0.8 x the Nyquist frequency and the sampling rate of the stimuli being 44,100 Hz.

Each token (e.g. /ma/, male speaker, Tone 1) was combined with each of the other tokens consisting of the same syllable and spoken by the same speaker (e.g. /ma/, male speaker, Tone 2) in both configurations e.g. Tone 1 TFS + Tone 2 Envelope; Tone 1 Envelope +Tone 2 TFS. This resulted in 96 chimeras (12 combinations of tones × 2 configurations × 2 syllables × 2 speakers).

The 96 chimeras were presented via Sennheiser HD 380 Pro headphones in a randomized sequence, using ePrime. After each presentation, participants were instructed to select which word they had heard from the four possible options. Words were represented by the corresponding Mandarin character, accompanied by a bar above the word indicating the direction of the tone contour (see Fig. 1).

Using the Hilbert transform to decompose a signal into TFS and envelope is an imperfect method. When a large number of channels is used (e.g. over 32), envelope cues may be introduced to the TFS component by filter ringing artefacts (Wang et al., 2011; Zeng et al., 2004). Envelope cues may also be reconstructed within the auditory system at the output of auditory filters (Ghitza, 2001; Heinz and Swaminathan, 2009; Schimmel and Atlas, 2005). However, this study tests the hypothesis that amusics use envelope cues to perceive pitch even when TFS cues are available, and as such a significant difference between the response of controls and amusics remains informative.
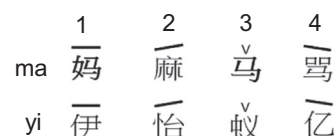
	1	2	3	4
ma	妈	麻	马	骂
yi	伊	怡	蚁	亿

Fig. 1. Each of the four possible responses for the two Mandarin syllables.

A

Interaural Phase Difference

B

Frequency Discrimination Limen
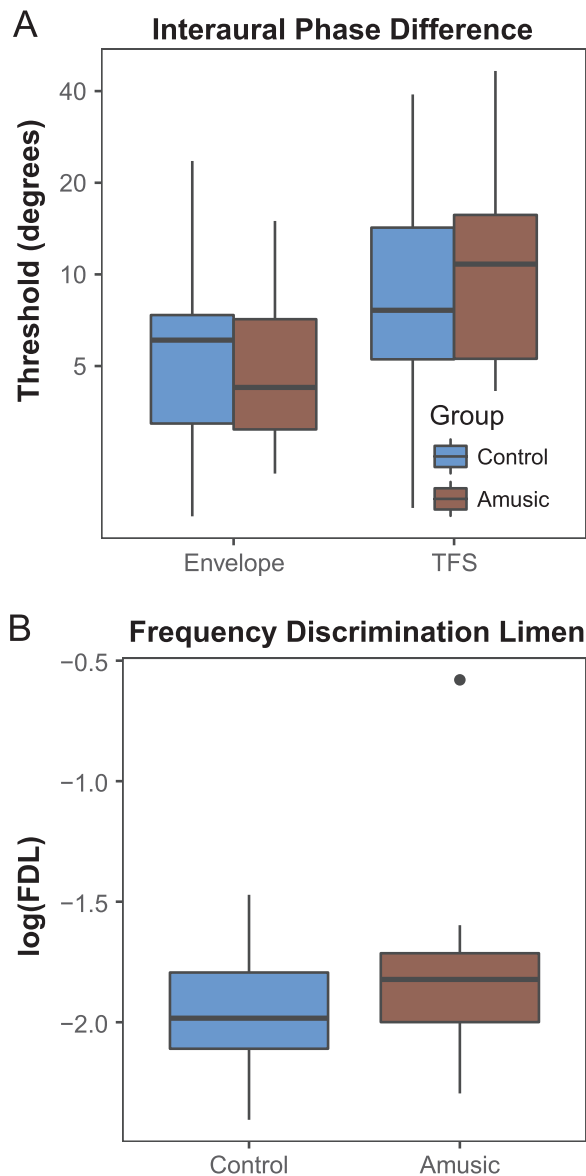

Fig. 2. Interaural phase-difference thresholds (*A*) and frequency discrimination limens (*B*) for controls and amusics.

3. Results

3.1. Peripheral processing and pitch perception

IPD threshold data (Fig. 2A) from both controls and amusics failed a Shapiro-Wilk test for normality in both the TFS (controls, $W = 0.80$, $p < 0.001$, amusics, $W = 0.83$, $p = 0.001$) and the envelope condition (controls, $W = 0.75$, $p < 0.001$, amusics, $W = 0.85$, $p = 0.003$). Consistent with results reported by Cousineau et al. (2015), there was no evidence for a deficit in peripheral processing of TFS in amusics (IPD thresholds: mean $= 13.68°$, $SE = 2.28°$) compared to controls (mean $= 11.68°$, $SE = 2.11°$) in a Wilcoxon rank sum test ($r = -0.13$, $p = 0.38$); nor was there evidence for a deficit in peripheral processing of envelope cues (controls, mean $= 6.83°$, $SE = 1.10°$, amusics, mean $= 5.68°$, $SE = 0.77°$; $r = -0.13$, $p = 0.38$).

FDL data (Fig. 2B) from amusics failed a Shapiro-Wilk test ($W = 0.80$, $p < 0.001$). Pitch discrimination in amusics (mean $= -1.83$, $SE = 0.08$) was poorer than controls (mean $= -1.96$, $SE = 0.04$) but marginally non-significant in a one-tailed Wilcoxon rank sum test p ($r = -0.28$, $p = 0.06$).
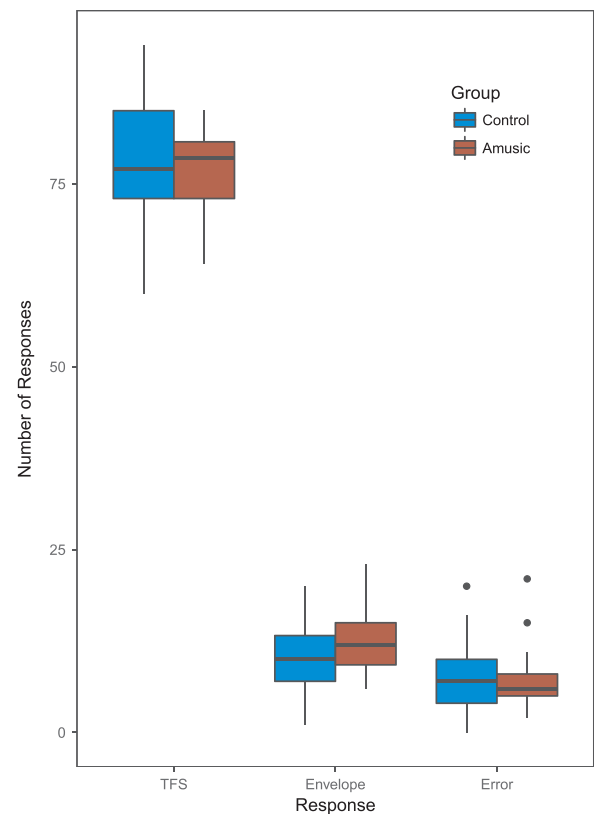

Fig. 3. The number of responses corresponding to temporal fine-structure pitch-cues, envelope pitch-cues, and errors.

3.2. Perception of speech tokens using TFS and envelope cues

Both controls (mean $= 77.96$, $SE = 1.71$) and amusics (mean $= 76.22$, $SE = 1.24$) made responses corresponding to TFS cues more often than to envelope cues (Fig. 3; Controls, mean $= 10.50$, $SE = 0.96$; Amusics, mean $= 12.59$, $SE = 0.93$). Responses which corresponded to neither cue were recorded as errors (Controls, mean $= 7.54$, $SE = 1.00$; Amusics, mean $= 7.18$, $SE = 0.90$).

Since response was a categorical forced-choice variable, an ANOVA was deemed an unsuitable method of analysis (Jaeger, 2008). Instead, a multinomial logit regression model was used to determine whether responses could be predicted by participant group. All responses were entered into the regression, rather than averaging across subjects (Jaeger, 2008, p. 439). The model demonstrated a significant effect of group on the likelihood of making a response corresponding to envelope over TFS (log-odds coefficient $B = 0.20$, $SE = 0.09$, $p = 0.03$); expressed as an odds ratio (e^B) the likelihood was greater by 1.2 for amusics compared to controls. The likelihood of making an erroneous response over a TFS response ($B = 0.02$, $SE = 0.11$, $p = 0.82$) or an envelope response ($B = 0.23$, $SE = 0.13$, $p = 0.1$) was not affected by group.

Number of responses corresponding to Envelope and TFS are plotted as a function of MBEA score in Fig. 4. Number of responses corresponding to envelope decreased with MBEA score, although the correlation was marginally non-significant ($r_{s(44)} = -0.24$, $p = 0.05$). The correlation between responses corresponding to TFS and MBEA score was also not significant ($r_{s(44)} = 0.18$, $p = 0.18$). In order to explore the relation between cue and sensitivity to pitch further, a second multinomial logit regression model was used with MBEA scores as the predictor variable. Since in this case the predictor variable was continuous, the model coefficients can be interpreted as the expected change in the logit outcome relative to the reference response for a unit increase in MBEA score. The model demonstrated a significant effect of MBEA
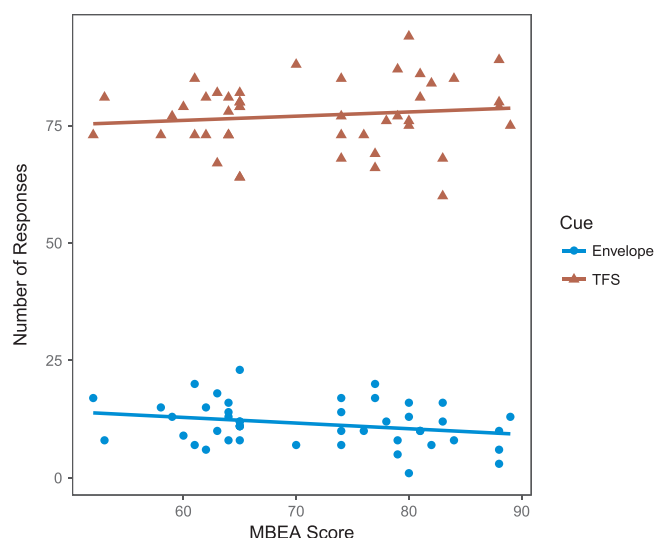
Fig. 4. Number of responses corresponding to temporal fine-structure and envelope pitch-cues as a function of MBEA Pitch score for all participants.

score on the likelihood of making a response corresponding to TFS over envelope ($B = 0.01$, $SE = 0.005$, $p = 0.01$; $e^B = 1.01$) but not over making an erroneous response ($B = -0.003$, $SE = 0.005$, $p = 0.59$). MBEA score also had a significant effect on the likelihood of making an erroneous response over a response corresponding to envelope ($B = 0.01$, $SE = 0.007$, $p = 0.04$; $e^B = 1.01$).

4. Discussion

4.1. Amusics are more likely than controls to use envelope pitch-cues to identify lexical tones

In the present study, when different TFS and envelope lexical tone pitch-cues were available, controls used TFS cues approximately 81% of the time and envelope cues 11% of the time. Using the same 16 channel method to create speech-speech chimera reported here, Xu and Pfingst (2003) found similar results, with normal hearing Mandarin speakers using TFS cues 85% of the time and envelope cues 9% of the time. Wang et al. (2011) reported normal hearing Mandarin speakers to use TFS cues 91% of the time and envelope cues 7%. The higher percentage of listeners identifying lexical tone using TFS in the latter study may be due to the authors pooling data across conditions using fewer channels (see Smith et al., 2002).

In the study by Wang et al. (2011) participants with moderate, moderate-to-severe, and severe sensorineural hearing loss used TFS cues 71%, 58%, and 38% of the time and envelope cues 21%, 31%, and 45% of the time respectively. This is consistent with numerous accounts of sensorineural hearing loss being associated with a deficit in the processing of TFS (e.g. Hopkins and Moore, 2007; Hopkins et al., 2008), and can be accounted for in part by hearing impaired listeners having broader auditory filters (e.g. Glasberg and Moore, 1986) due to a loss of outer hair cell function and therefore a loss of frequency selectivity in the basilar membrane response; fewer TFS cues are available at the level of the AN due to fewer resolved harmonics. Use of envelope cues in speech detection has been shown to remain intact in sensorineural hearing impairment (e.g. Turner et al., 1995), and the results of the study by Wang et al. (2011) suggest that where access to TFS pitch-cues is degraded listeners make more use of envelope cues.

In the present study amusics used TFS cues to determine lexical tone 73% of the time and envelope cues 14% of time, and the results of the model with MBEA score as a predictor variable indicate that robustness of pitch perception predicts use of TFS cues to identify lexical tone across all participants. This is consistent with a pitch mechanism reliant

upon TFS remaining the primary mechanism in amusics, and with amusics performing better at pitch discrimination when TFS cues are available than when not, despite performing poorly compared to controls (Cousineau et al., 2015). Consistent with Cousineau et al. (2015), and with Liu et al. (2014) who found no evidence for the representation of pitch information being degraded at the level of the auditory brainstem in amusics, no evidence was found in the present study for a deficit in the peripheral representation of TFS in amusics (Fig. 2A). However, despite the apparent integrity of the peripheral representation of TFS, the results of the multinomial logit regression model indicate that amusics are more likely to rely upon envelope pitch-cues than controls, even when TFS pitch-cues are available. A recent study by Whiteford and Oxenham (2017) found that amusics were poorer than controls at detecting amplitude modulation at 4 and 20 Hz. The results of the current study suggest that this apparent insensitivity to variation in envelope within the context of amplitude modulation does not manifest as an insensitivity to envelope pitch-cues; on the contrary, the results reported here and by Cousineau et al. (2015) are consistent with amusia being associated with over-reliance on an inferior secondary pitch mechanism that uses envelope pitch-cues, resulting from an anomaly that resides centrally to the brainstem. Interestingly, Wang et al. (2015) found that Mandarin speakers with auditory neuropathy spectrum disorder also identified lexical tones using envelope cues significantly more often than both listeners with normal hearing and listeners with sensorineural hearing loss, and made more lexical tone identification errors.

4.2. Amusia manifests in non-fine-grained pitch judgements

It has been suggested previously that the effects of amusia may be constrained to fine-grained pitch judgements (Hyde and Peretz, 2004). Whiteford and Oxenham (2017) recently challenged this by demonstrating that amusia affects pitch discrimination even at frequencies at which pitch judgements become poorer in controls. The results of the present study provide further evidence that amusia manifests in situations other than fine-grained pitch judgements. Furthermore, since pitch judgements using envelope cues are poorer than those made using TFS cues (Bernstein and Oxenham, 2003; Houtsma and Smurzynski, 1990), an atypical reliance upon envelope cues in amusics may account for poorer performance in fine-grained pitch judgements.

4.3. Methodological considerations

As discussed, the method used here for separation of the signal into TFS and envelope components is imperfect, and may result in envelope cues remaining available in the TFS component. As such, it cannot be ruled out that the amusic group used envelope cues when making responses corresponding to TFS, and that a better method for separating TFS and envelope might reveal a larger effect of amusia on use of envelope- over TFS-cues.

Surprisingly, amusics in the current study were not found to have significantly poorer pitch discrimination than controls ($p = 0.06$), as measured by FDLs. This may be accounted for by the relatively long duration stimuli used. It has been demonstrated previously that pitch discrimination improves with tone duration, and that this effect is greater for tones containing unresolved harmonics (Plack and Carlyon, 1995). It is possible therefore that the effect of stimulus duration on a pitch discrimination is greater for a pitch mechanism using envelope cues than for one that uses TFS cues; therefore if the amusics in the current study used a secondary pitch mechanism to discriminate pitch, the effects of amusia may have been suppressed by the relatively long stimulus duration. In addition, the use of an abbreviated test for amusia may have led to participants being classified as amusic that would not have been had the full MBEA been used; this may account for the non-significant pitch discrimination result and for the small effect of group on use of envelope cues over TFS cues.

5. Conclusion

Previous research has demonstrated that amusia may be associated with a deficit in processing of TFS pitch-cues but not envelope pitch-cues. However, this could have been due to amusia being a deficit in fine-grained pitch judgements, rather than being specific to TFS pitch-cues per se. The present study demonstrates that the effects of amusia are not restricted to fine-grained pitch judgements; instead, the results indicate that amusics are more likely to identify pitch based upon envelope pitch-cues over TFS pitch-cues than are controls. The results provide further evidence for there being two distinct pitch mechanisms, and for amusia being associated with an over-reliance on an inferior, secondary pitch mechanism that uses envelope pitch-cues.

Acknowledgments

References

Ayotte, J., Peretz, I., Hyde, K., 2002. Congenital amusia: a group study of adults afflicted with a music-specific disorder. Brain 125, 238–251.

Bernstein, J.G., Oxenham, A.J., 2003. Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? J. Acoust. Soc. Am. 113, 3323.

Brown, L., Sherbenou, R.J., Johnsen, S.K., 2010. TONI-4, Test of Nonverbal Intelligence. In: Pro-ed.

Brugge, J.F., Anderson, D.J., Hind, J.E., Rose, J.E., 1969. Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. J. Neurophysiol. 32, 386–401.

Chao, Y.R., 1948. Mandarin Primer. Harvard University Press, Cambridge.

Cousineau, M., Oxenham, A.J., Peretz, I., 2015. Congenital amusia: a cognitive disorder limited to resolved harmonics and with no peripheral basis. Neuropsychologia 66, 293–301.

Fletcher, H., 1940. Auditory patterns. Rev. Mod. Phys. 12, 47.

Ghitza, O., 2001. On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. J. Acoust. Soc. Am. 110, 1628.

Glasberg, B.R., Moore, B.C.J., 1986. Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. J. Acoust. Soc. Am. 79, 1020–1033.

Glasberg, B.R., Moore, B.C.J., 1990. Derivation of auditory filter shapes from notched-noise data. Hear. Res. 47, 103–138.

Gockel, H.E., Carlyon, R.P., Plack, C.J., 2011. Combination of spectral and binaurally created harmonics in a common central pitch processor. J. Assoc. Res. Otolaryngol. 12, 253–260.

Greenwood, D.D., 1990. A cochlear frequency-position function for several species—29 years later. J. Acoust. Soc. Am. 87, 2592–2605.

Grothe, B., Pecka, M., McAlpine, D., 2010. Mechanisms of sound localization in mammals. Physiol. Rev. 90, 983–1012.

Heinz, M.G., Swaminathan, J., 2009. Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech. J. Assoc. Res Otolaryngol. 10, 407–423.

Hopkins, K., Moore, B.C., 2007. Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. J. Acoust. Soc. Am. 122, 1055–1068.

Hopkins, K., Moore, B.C., 2010. Development of a fast method for measuring sensitivity to temporal fine structure information at low frequencies. Int. J. Audiol. 49, 940–946.

Hopkins, K., Moore, B.C., Stone, M.A., 2008. Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. J. Acoust. Soc. Am. 123, 1140–1153.

Houtsma, A.J.M., Smurzynski, J., 1990. Pitch identification and discrimination for complex tones with many harmonics. J. Acoust. Soc. Am. 87, 304–310.

Hyde, K.L., Peretz, I., 2004. Brains that are out of tune but in time. Psychol. Sci. 15, 356–360.

Jaeger, T.F., 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards Logit Mixed Models. J. Mem. Lang. 59, 434–446.

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.

Jiang, C., Hamm, J.P., Lim, V.K., Kirk, I.J., Yang, Y., 2012. Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. Mem. Cogn. 40, 1109–1121.

King, A., Hopkins, K., Plack, C.J., 2014. The effects of age and hearing loss on interaural phase difference discrimination. J. Acoust. Soc. Am. 135, 342–351.

Lacher-Fougère, S., Demany, L., 2005. Consequences of cochlear damage for the detection of interaural phase differences. J. Acoust. Soc. Am. 118, 2519–2526.

Lehmann, A., Skoe, E., Moreau, P., Peretz, I., Kraus, N., 2015. Impairments in musical abilities reflected in the auditory brainstem: evidence from congenital amusia. Eur. J. Neurosci. 42, 1644–1650.

Licklider, J., 1951. A duplex theory of pitch perception. Experientia 7.

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

Liu, F., Maggu, A.R., Lau, J.C., Wong, P.C., 2014. Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. Front Hum. Neurosci. 8, 1029.

Liu, F., Patel, A.D., Fourcin, A., Stewart, L., 2010. Intonation processing in congenital amusia: discrimination, identification and imitation. Brain 133, 1682–1693.

McAlpine, D., Grothe, B., 2003. Sound localization and delay lines – do mammals fit the model? Trends Neurosci. 26, 347–350.

Meddis, R., Hewitt, M.J., 1991. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: pitch identification. J. Acoust. Soc. Am. 89.

Moore, B.C.J., 2003. An Introduction to the Psychology Of Hearing, 5th ed. Academic Press, Amsterdam London.

Moore, G.A., Moore, B.C.J., 2003. Perception of the low pitch of frequency-shifted complexes. J. Acoust. Soc. Am. 113, 977.

Nan, Y., Sun, Y., Peretz, I., 2010. Congenital amusia in speakers of a tone language: association with lexical tone agnosia. Brain 133, 2635–2642.

Nguyen, S., Tillmann, B., Gosselin, N., Peretz, I., 2009. Tonal language processing in congenital amusia. Ann. New Y. Acad. Sci. 1169, 490–493.

Oxenham, A.J., Micheyl, C., Keebler, M.V., Loper, A., Santurette, S., 2011. Pitch perception beyond the traditional existence region of pitch. Proc. Natl. Acad. Sci. USA 108, 7629–7634.

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.

Peretz, I., Champod, S., Hyde, K., 2003. Varieties of musical disorders: the Montreal battery of evaluation of amusia. Ann. New Y. Acad. Sci. 999, 58–75.

Plack, C.J., Carlyon, R.P., 1995. Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics. J. Acoust. Soc. Am. 98, 1355.

Plack, C.J., Oxenham, A.J., 2005. Overview: the present and future of pitch. In: Plack, C.J., Oxenham, A.J., Fay, R.R. (Eds.), Pitch: Neural Coding and Perception. Springer, New York London.

Plomp, R., 1964. The ear as a frequency analyzer. J. Acoust. Soc. Am. 41, 1526–1533.

Russell, I.J., Sellick, P.M., 1978. Intracellular studies of hair cells in the mammalian cochlea. J. Physiol. 338, 179–206.

Schimmel, S.M., Atlas, L.E., 2005. Coherent envelope detection for modulation filtering of speech. ICASSP 221–224.

Smith, Z.M., Delgutte, B., Oxenham, A.J., 2002. Chimaeric sounds reveal dichotomies in auditory perception. Nature 416, 87–90.

Turner, C.W., Souza, P.E., Forget, L.N., 1995. Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. J. Acoust. Soc. Am. 97, 2568–2576.

Wang, S., Dong, R., Liu, D., Wang, Y., Liu, B., Zhang, L., Xu, L., 2015. The role of temporal envelope and fine structure in mandarin lexical tone perception in auditory neuropathy spectrum disorder. PLoS One 10, e0129710.

Wang, S., Xu, L., Mannell, R., 2011. Relative contributions of temporal envelope and fine structure cues to lexical tone recognition in hearing-impaired listeners. J. Assoc. Res Otolaryngol. 12, 783–794.

Whiteford, K.L., Oxenham, A.J., 2017. Auditory deficits in amusia extend beyond poor pitch perception. Neuropsychologia 99, 213–224.

Xu, L., Pfingst, B.E., 2003. Relative importance of temporal envelope and fine structure in lexical-tone perception (L). J. Acoust. Soc. Am. 114, 3024.

Zeng, F.-G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y.-Y., Chen, H., 2004. On the dichotomy in auditory perception between temporal envelope and fine structure cues (L). J. Acoust. Soc. Am. 116, 1351.