Research on the relation of spoken and musical melodies in non-tone languages is much less developed than in tone languages. In one study, Arnold and Jusczyk (2002) examined spoken and sung versions of three nursery rhymes. The spoken versions were made by individuals who were unfamiliar with the musical melodies that normally accompanied them, and who read them as texts of connected prose (rather than as verse organized into couplets). To conduct cross-domain comparison, musical pitch contours were analyzed simply in terms of high and low pitch targets, in other words, maxima and minima in the song's F0 track that were aligned with stressed syllables. Speech contours were analyzed in terms of the ToBI system of analysis to identify high and low tones (cf. section 4.3.1). The authors found that pitch targets in the sung version did tend to correspond in type (high or low) to those realized on the same words in the text, though the amount of correspondence varied widely across the three nursery rhymes. Clearly, this is an area that is open for research. One idea for quantifying the degree of correspondence between linguistic intonation and musical melody in songs is to use the prosogram to identify the melodic contour of the lyrics when they are spoken, and then to compare this contour to the musical melodic contour to which the text is set in a song. It would be interesting to know if songs with melodies congruent with the spoken pitch contour have an advantage in terms of how easily they are learned or remembered (cf. Patel, 2006a).

4.5 Melodic Statistics and Melodic Contour as Key Links

The purpose of this section is to delve into two areas in which direct comparisons of musical and linguistic melody are proving fruitful. Section 4.5.1 addresses structure, and shows that a specific aspect of the statistics of speech intonation is reflected in instrumental music. Section 4.5.2 concerns processing, and provides evidence that perception of melodic contour in speech and music engages overlapping cognitive and neural machinery. This latter section also outlines a hypothesis (the "melodic contour deafness" hypothesis) that aims to account for the apparent dissociation of musical and linguistic melody perception in musically tone-deaf individuals. As we shall see, one implication of this hypothesis is that a behavioral dissociation is not the same as a neural dissociation.

4.5.1 Melodic Statistics

Analysis of intonation in terms of the syllable tones of the prosogram (section 4.3.2) opens the way to quantitative comparisons of melody in speech and music. For example, musical melodies are characterized by a number of statistical regularities that appear cross-culturally (summarized in Huron, 2006). To what extent are these regularities unique to music versus shared with speech? If some regularities are shared, what is the cognitive significance of this fact?

One well-known regularity of musical melodies is the predominance of small ("conjunct") intervals between successive pitches. Figure 4.13 shows the relative frequency of intervals of different sizes in a sample of Western music (Vos & Troost, 1989). The general pattern is similar to that seen when melodies from a range of cultures are sampled (Huron, 2006:74).

As is clear from the figure, musical melodies are dominated by small intervals. Another way of putting this is that most melodic motion is by small steps (of 2 semitones or less). The predominance of small intervals is reflected in listeners' expectations for how novel melodies will continue when stopped midstream (cf. section 4.2.4), showing that this is a perceptually relevant feature of melodies.

Why do small intervals dominate melodies? The two most commonly proffered reasons are motor and perceptual. The motor explanation suggests that small intervals are easier to produce in succession than large ones, with the voice and with most instruments (cf. Zipf, 1949:336–337). The perceptual explanation suggests that too many large pitch movements risk splitting a melody into separate perceptual streams, destroying the perceptual cohesion between successive tones (Bregman, 1990; cf. Narmour, 1990). Of course, the motor and perceptual explanations are not mutually exclusive: Indeed, they may reinforce each other. One can thus group them into a single category of "constraint-based explanations" based on the notion of intrinsic constraints on production and perception.

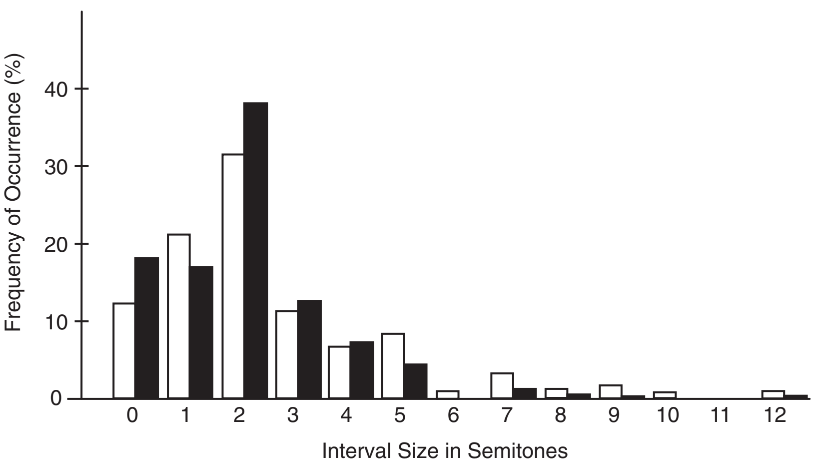

Figure 4.13 Histograms showing the relative proportion of pitch intervals of different sizes in Western melodies (classical and rock: white bars; folk, dark bars). From Vos & Troost, 1989.

An alternative to constraint-based theories is the idea is that the preference for small intervals in music arises out of experience with speech. That is, if speech melodies are dominated by small pitch movements, and if listeners absorb the statistics of spoken pitch patterns in their environment, then this might influence their proclivities in terms of shaping musical melodies. This hypothesis invokes a perceptual mechanism known as "statistical learning," which has been discussed in Chapters 2 and 3. Statistical learning refers to tracking patterns in the environment and acquiring implicit knowledge of their statistical properties, without any direct feedback. Statistical learning is discussed in more detail later in this chapter. For now, the relevant point is that statistical learning may provide one mechanism for a cross-domain influence of speech on music.

How can one test this idea? The first step is to determine if speech melodies are in fact dominated by small pitch movements. Consider once again the prosogram in Figure 4.4b. The intervals between adjacent level tones are listed in Table 4.2. (Note that that there are 18 tones in the prosogram of Figure 4.4b, but only 15 intervals appear in Table 4.2. This is because one of the prosogram tones is a glide, and intervals are computed only between immediately adjacent level tones.) One fact that is immediately apparent from this table is that most intervals are quite small, reminiscent of the pattern found in music. How representative is this pattern? Figure 4.14 shows a histogram of intervals between level tones in a corpus of 40 English and French sentences that have been analyzed using prosograms like that in Figure 4.4b.

As can be seen, intonational pitch patterns are dominated by small pitch movements. In Figure 4.14, 56% of all intervals between level tone pitches are

Table 4.2 The Size of Pitch Intervals in Semitones Between Adjacent Level Tones in a Prosogram of an English Sentence (cf. Figure 4.4b)

Interval Number	Size in Semitones
1	2.56
2	–0.01
3	–2.51
4	4.96
5	–1.72
6	–0.83
7	–1.64
8	–0.56
9	–0.21
10	0.54
11	–0.90
12	–1.02
13	0.01
14	0.47
15	–1.89

less than or equal to 2 semitones. One notable difference between the distribution of intervals in speech and music is that in musical melodies intervals of 2 semitones are more common than intervals of 1 or 0 semitones (see Figure 4.13). No such tendency is seen in speech. (In Figure 4.14, the percentage of intervals between 0-1, 1-2, and 2-3 semitones, respectively, are 35%, 21%, and 19%). This predominance of 2-semitone intervals in musical melodies likely reflects the fact that intervals of 2 semitones are more common than intervals of 1 semitone in musical scales (cf. Chapter 2). Despite this difference between music and speech, the larger point is that speech melodies, like musical melodies, are dominated by small pitch intervals. This is probably not a peculiarity of English and French; it is likely to hold for languages generally.

Finding that small intervals predominate in speech melodies is consistent with the idea that music reflects speech. However, it is also consistent with the idea that pitch patterns in speech and music are subject to the same set of motor and perceptual constraints. This illustrates a basic problem in interpreting commonalities between statistical regularities in music and speech: Such commonalities could be due to the operation of some physiological or perceptual factor that influences both speech and music, rather than to any cross-domain influence.

Hence if one is interested in the possibility of speech influencing music, some strategy is needed to eliminate the problem of common external influences. One such strategy is to look for quantifiable *differences* between the statistics of speech intonation in different cultures, and to see if these differences are reflected in instrumental music of these cultures. Because there is no way that

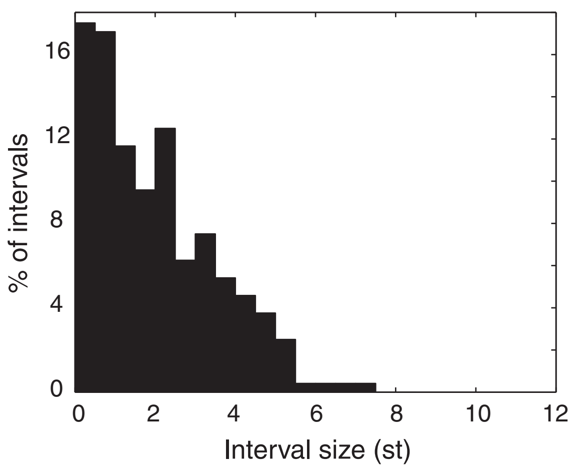

Figure 4.14 Histogram showing the relative proportion of pitch intervals of different sizes (n = 543 total) between level tones in prosograms of British English and French sentences. Histogram bin size = 0.5 st.

general motor or perceptual factors can explain differences in speech melody between languages, a reflection of linguistic differences in music makes a stronger case for cross-domain influence.

This "difference-based" approach to comparing language and music has been applied by Patel, Iversen, and Rosenberg (2006) to the speech and instrumental classical music of England and France. This study complemented an earlier study with a similar approach, which focused on rhythm (Patel & Daniele, 2003a; cf. Chapter 3, section 3.5.1). Once again, the topic of investigation was the provocative claim (made by a number of musicologists and linguists over the past 50 years) that a culture's instrumental music reflects the prosody of its native language. For example, the linguist Hall (1953) suggested a resemblance between Elgar's music and the intonation of British English.

In addressing this idea, we used the same corpus of speech and music as in our earlier work on rhythm. That is, the speech corpus consisted of short, news-like utterances read by native speakers of each language, whereas the musical corpus consisted of themes from instrumental classical music by 16 composers whose lives spanned the turn of the 20th century (such as Elgar and Debussy). Although we focused on just two cultures, a principal goal of this research was to establish methods of broad applicability in comparative speech-music studies.

We computed prosograms for all sentences in our speech corpus, resulting in representations like that in Figure 4.4b. Then, focusing on just the level pitches (which represented about 97% of pitches assigned by the prosogram algorithm, the rest being glides), we computed two measures of melodic statistics for each sentence. The first was variability in pitch height, which simply measured how widely spread the level pitches were around their mean pitch. The second was variability in pitch interval size within a sentence, in which an interval was defined as the jump in pitch between immediately successive level pitches. This measure serves to indicate whether steps between successive level pitches tend to be more uniform or more variable in size.

Initially it may seem odd to focus on pitch intervals in speech. Although the human perceptual system is quite sensitive to interval patterns in music (in which, for example, a melody can be recognized in any transposition as long as its interval pattern is preserved), music features well-defined interval categories such as minor second and perfect fifth, whereas speech does not (cf. section 4.1.1). Might the perceptual system attend to interval patterns in speech despite the lack of stable interval structure? Recent theoretical and empirical work in intonational phonology suggests that spoken pitch intervals may in fact be important in the perception of intonation, even if they do not adhere to fixed frequency ratios (Dilley, 2005; cf. Hermes, 2006). If this is the case, then one can understand why composers might have implicit knowledge of the "spoken interval statistics" of their language, which might in turn be reflected in the music that they write.

We used semitones in our measurements of pitch height and pitch intervals in speech, reflecting the perceptual scaling of intonation (Nolan, 2003).[15] We also used semitones in measuring pitch height and pitch intervals in music, working directly from music notation. The height measurements revealed no significant difference in variability between English and French speech, nor between English and French music. The results of interval measurements were more revealing. Spoken French had significantly lower pitch interval variability than spoken English, and music mirrored this pattern. That is, French musical themes had significantly lower interval variability than English themes, Put another way, as the voice moves from one syllable to the next in speech, the size of each pitch movement is more uniform in French than in English speech. Similarly, as a melody moves from one note to the next, the size of the pitch movement is more uniform in French than in English music. Figure 4.15 shows the results of the pitch interval variability measures combined with our previous findings on rhythm. Note that the measure of variability on the vertical axis is "melodic interval variability" (MIV), defined as 100 times the coefficient of variation (CV) of interval size (CV = standard deviation/mean). Scaling the CV by 100 serves to put MIV in the same general range of absolute values as nPVI.

We call Figure 4.15 an "RM space" (rhythm-melody space) plot, because it shows the position of a given language or music in a two-dimensional space with rhythm on one axis and melody on the other. One appealing thing about RM space is that the "prosodic distance" between two languages can be quantified via the length of a line connecting the points representing the mean (nPVI, MIV) of the two languages (see Patel et al., 2006 for details). In Figure 4.15, the prosodic distance between English and French speech is 27.7 RM units. In contrast, the line connecting the English and French music is only about 1/3 as long (8.5 RM units). Thus the musical difference is smaller than the linguistic one. This is not surprising, given that music is an artistic endeavor with substantial intracultural variation and (unlike speech) no a priori reason to follow rhythmic or melodic norms. What is remarkable is that despite this variation, quantitative differences emerge between the music of the two nations that reflect linguistic differences.

By what route do speech patterns find their way into music? One oft-heard proposal is that composers borrow tunes from folk music, and these tunes bear the stamp of linguistic prosody because they were written with words. This might be termed the "indirect route" from speech to music. If this is the case,

[15] Earlier work by Hermes and Van Gestel (1991) had suggested ERBs should be used in measuring pitch distances in speech. The precise choice of units for speech is unlikely to influence the results reported here. Our measure of pitch variability, the coefficient of variation (CV), is a dimensionless quantity that would allow one to measure pitch distances in speech and music in different units (ERBs vs. semitones), and still compare variability across domains with the CV as a common metric.

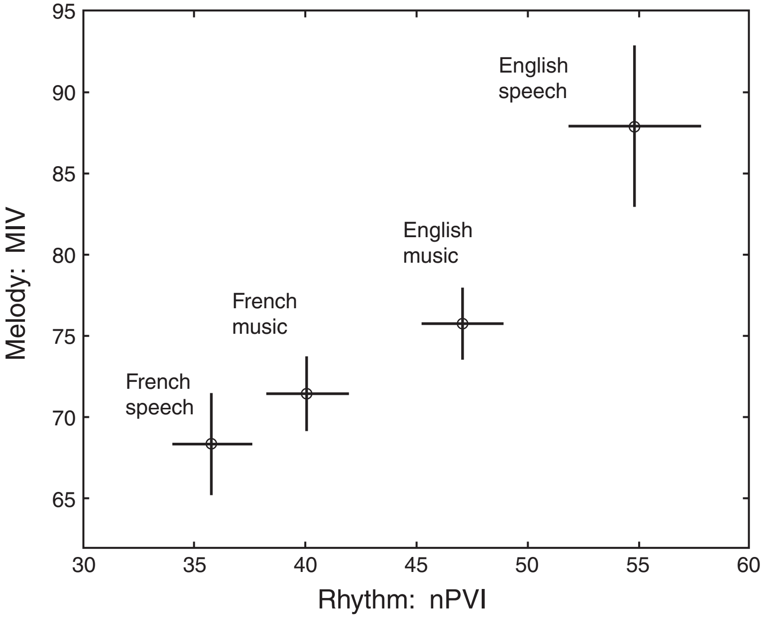

Figure 4.15 Rhythm-melody (RM) space for speech and music. Error bars show +/– 1 standard error. See text for details.

then there are no strong cognitive implications for speech-music relations, because the speech-music resemblance is simply the result of borrowing of music that has been consciously shaped to fit linguistic phrases.

We favor a different hypothesis, however. This is the idea that implicit learning of prosodic patterns in one domain (ordinary speech) influences the creation of rhythmic and tonal patterns in another domain (instrumental art music). This might be termed the "direct route" between speech and music, because contact between speech and music need not be mediated by an intermediate speech-music blend (such as songs). One advantage of the direct-route hypothesis is that it can account for the reflection of speech in music not thought to be particularly influenced by folk music (e.g., much of Debussy's and Elgar's work; Grout & Palisca, 2000).

The direct-route hypothesis centers on the notion of statistical learning of prosodic patterns in the native language. Recall that statistical learning refers to tracking patterns in the environment and acquiring implicit knowledge of their statistical properties, without any direct feedback. Statistical learning of prosodic patterns in one's native language likely begins early in life. Research on auditory development has shown that infants are adept at statistical learning of phonetic/syllabic patterns in speech and of pitch patterns in nonlinguistic tone sequences (Saffran et al., 1996, 1999) Thus it seems plausible that statistical learning of rhythmic and tonal patterns in speech would also begin in infancy, especially because infants are known to be quite sensitive to the prosodic patterns of language (Nazzi et al., 1998; Ramus, 2002b). Of course, statistical learning of tone patterns need not be confined to infancy. Adult listeners show sensitivity to the distribution of different pitches and to interval patterns in music (Oram & Cuddy, 1995; Saffran et al., 1999; Krumhansl, 2000; Krumhansl et al., 1999, 2000). Importantly, statistical learning in music can occur with atonal or culturally unfamiliar materials, meaning that it is not confined to tone patterns that follow familiar musical conventions.

It is worth emphasizing that the direct-route hypothesis does not imply that speech prosody influences musical structure in a deterministic fashion. It simply implies that composers can be influenced by their implicit knowledge of their native language's rhythm and melody. Music is an artistic medium, after all, and composers are free to do what they like. In particular, the influence of music of other cultures may override native linguistic influences on musical structure (see Patel & Daniele, 2003b; Daniele & Patel, 2004).

Two final points deserve mention. First, it remains to be explained why English speech should have a greater degree of pitch interval variability than French. One idea is that British English may use three phonologically distinct pitch levels in its intonation system, whereas French may only use two (cf. Willems, 1982; Ladd & Morton, 1997; cf. Figure 4.11 for one model of English intonation based on three distinct pitch levels). A compelling explanation, however, awaits future research. Second, our study focused on a very simple aspect of melodic statistics (pitch variability). More sophisticated analyses of melodic statistics are clearly called for. One idea is to study the alignment of pitch and duration patterns, to see if differences in the "joint accent structure" of speech melody between languages are reflected in music (cf. section 4.2.9, and Patel et al., 2006).

4.5.2 Melodic Contour Perception

Melodic contour refers to a melody's pattern of ups and downs of pitch over time, without regard to exact interval size. As discussed in section 4.2.3, melodic contour perception is important in both music and speech. Yet it is not obvious on a priori grounds whether the processing of melodic contour in the two domains is mediated by common cognitive and neural machinery. One might imagine that the lack of stable pitch interval structure in speech, combined with a speech-specific tendency for pitch declination (cf. section 4.1.1) would lead linguistic melodic contours to be processed differently from musical melodic contours. Indeed, Peretz and Coltheart (2003) have proposed a modular model of music processing in which melodic contour analysis is a domain-specific aspect of music perception, not shared with speech. Thus cross-domain studies of melodic contour perception are relevant to debates on the overlap of musical and linguistic processing in the brain.

In terms of evidence from cognitive neuroscience, reports of selective impairments of music, or "selective amusia," are of particular relevance for the cross-domain study of melodic contour processing. The two sections below discuss melodic contour processing in two populations: individuals with "acquired amusia" and those with "congenital amusia," or musical tone-deafness.

Melodic Contour Perception in Acquired Amusia

Acquired amusia refers to deficits in musical perception and/or production abilities following brain damage that are not simply due to hearing loss or some other peripheral auditory disorder (Marin & Perry, 1999; Peretz, 2006). The condition is reported relatively rarely, probably due to social factors. That is, individuals who experience dramatic changes in music perception after brain damage may not seek medical attention for this problem, or may not be directed by their physicians toward a neuropsychologist who studies music. From the standpoint of brain localization, amusia has been associated with damage to diverse regions of the brain (not just the auditory cortices), though there are a preponderance of cases with right-hemisphere damage in the published literature (Griffiths, 2002; Stewart et al., 2006).

Patel, Peretz, et al. (1998) examined intonation perception in two amusic individuals with different kinds of music perception deficits. The first participant, CN, was an "associative amusic" who was able to discriminate musical pitch and rhythm patterns but was unable to identify culturally familiar tunes, suggesting a selective difficulty in accessing stored representations for familiar tunes (Peretz, 1996). The second participant, IR, was an "apperceptive amusic" who could not discriminate musical pitch and rhythm patterns. Thus her deficit was at a lower level than that of CN.

To test the intonation perception of these individuals, Patel, Peretz, et al. (1998) used sentence pairs in which the members of the pair were lexically identical but had different intonation. Furthermore, computer editing of each sentence pair was used to ensure that the timing of syllables in the two sentences was identical and that intensity differences were minimized, so that the only salient cue for discrimination was pitch. These sentence pairs could take one of two forms: "statement-question pairs" in which one version of a sentence was a statement and the other a question (such as "He wants to buy a house next to the beach" spoken as a statement vs. a question), and "focus-shift pairs" in which the two sentences had contrastive focus on a different word (such as "Go in FRONT of the bank, I said" vs. "Go in front of the BANK, I said").

In addition to these linguistic stimuli, pairs of nonlinguistic tone sequences were also presented for discrimination. These tone sequences were created from the intonation contours of the sentence pairs, replacing each syllable of a sentence with a tone whose pitch was fixed at the Hz value midway between the maximum and minimum F0 values for that syllable (Figure 4.16). The linguistic and nonlinguistic stimuli were thus matched for overall length and pitch range.

Tone onsets occurred at the vowel onset times of corresponding syllables, and tone offsets were determined by the offset of F0 within each syllable. Thus each tone sequence had the same temporal rhythm as the syllables in the parent sentence. All tones had a complex frequency structure consisting of a fundamental and a few harmonics of decreasing amplitude. An example of a sentence pair and its melodic analogs are given in Sound Examples 4.9a–d. (Note that the sentences are in French, as the study was conducted with French-speaking amusics. Similar sound examples in English are provided later.) The rationale behind the study was that if the amusics' perceptual deficits were confined to

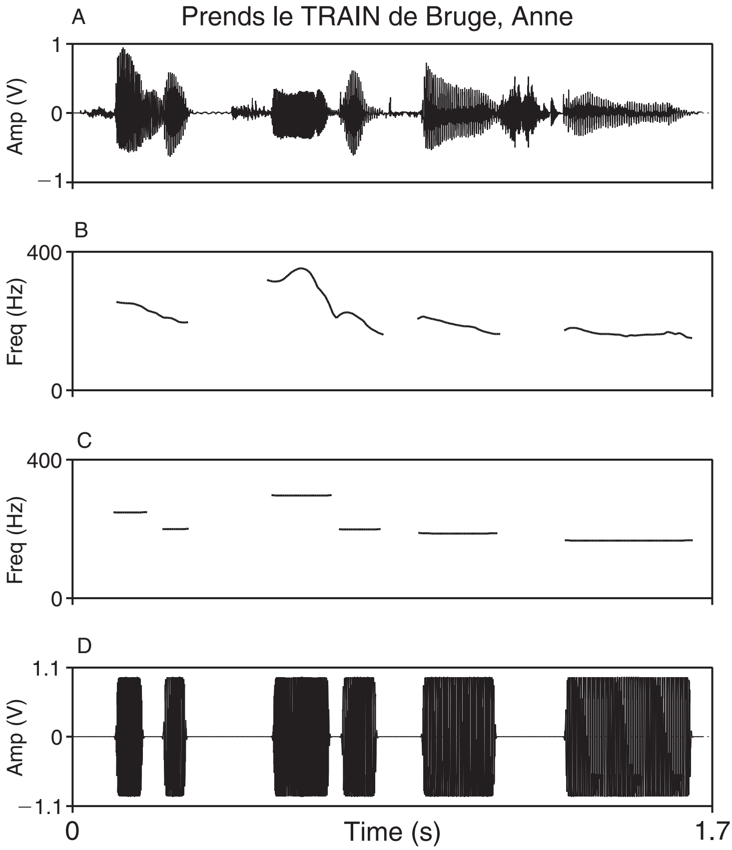

Figure 4.16 Illustration of the process of converting the intonation contour of a sentence into a discrete tone analog. The waveform of a French sentence is shown in (A) and its F0 contour in (B). In (C), the F0 of each syllable has been set to a fixed value. (D) shows the waveform of the tone analog. Adapted from Patel, Peretz, et al., 1998.

music, they should perform well on discriminating the sentences but have difficulty with the tone sequences, in other words, a dissociation between speech and nonlinguistic tone sequence processing should be observed. On the other hand, if intonation and tone-sequence processing overlap in the brain, then similar performance on the two types of sequences should be found.[16]

The results of the study supported the second conclusion. CN did well on both the linguistic and nonlinguistic sequences, whereas IR had difficulty discriminating both types of sequences. Furthermore, IR's problems could not be attributed to simple low level perceptual problems in perceiving pitch patterns, because further experiments showed that she could accurately label individual statements and questions as either "statement" or "question," and could identify which word in a given focus-shift sentence carried the main emphasis. Thus it appeared that IR's intonation *perception* was only mildly compromised, but her memory for melodic contours suffered (cf. Belleville et al., 2003). This interpretation was supported by a comparison of the lesion profiles of CN and IR. Compared to CN, IR had additional damage in left primary auditory cortex and right frontal cortex (Figure 4.17).

The latter lesion is likely to be associated with a memory-based deficit, because right frontal and temporal cortex has been implicated in pitch memory tasks in both speech and music (Zatorre et al., 1994). This suggested that both intonation and nonlinguistic tone sequences rely on common brain regions when melodic contours must be retained in working memory (cf. Semal et al., 1996).[17]

Melodic Contour Perception in Musical Tone Deafness

For those interested in comparing melody perception in music and speech, the phenomenon of musical tone-deafness (henceforth mTD) is of particular interest. Also referred to as "congenital amusia" (Peretz & Hyde, 2003), mTD

[16] Note that this study was conducted before the authors were aware of the prosogram, and thus the construction of tonal analog of intonation was based on intuitive criteria rather than on an explicit algorithm for F0 stylization. Nevertheless, the resulting pitch sequences do bear some resemblance to sequences that would be generated by a prosogram, although they lack any glides (cf. section 4.3.2).

[17] Another amusic who has been studied with a similar paradigm has also demonstrated equivalent difficulties with intonation and nonlinguistic tone sequence analogs (Nicholson et al., 2003). In this case, however, the damage was in the right parietal cortex and this may have disrupted the ability to extract pitch patterns in both domains, rather than causing a problem in remembering these patterns (cf. Griffiths et al., 1997). It is also worth noting that IR was retested on the materials of Patel et al. approximately 10 years after the original study and showed a virtually identical pattern of performance. Thus her condition is very stable (Lochy & Peretz, personal communication, 2004).

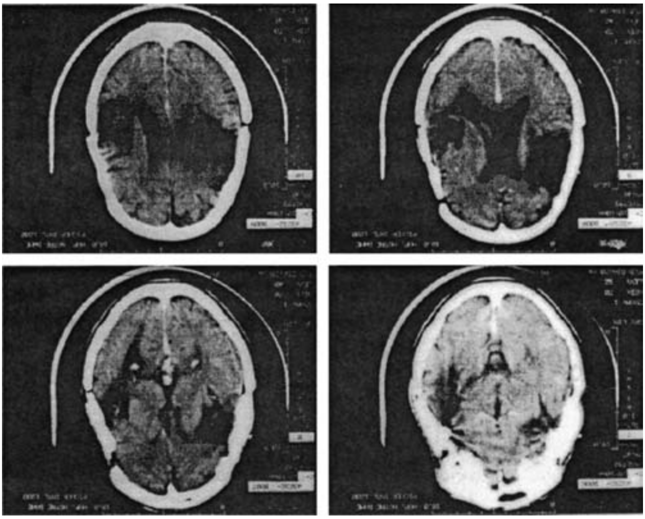

Figure 4.17 (a) Transverse CT scans of the amusic individual IR. Note that the right side of the brain is shown on the left of each image, per radiological convention. Scans proceed superiorly in 10-mm increments, in the order top left, top right, bottom left, and bottom right. The scans show bilateral temporal and right inferior frontal lobe damage (see Patel, Peretz, et al., 1998, for details).

refers to severe problems with music perception and production that cannot be attributed to hearing loss, lack of exposure to music, or any obvious nonmusical social or cognitive impairments. As mentioned in section 4.1.1, musically tone-deaf individuals (henceforth mTDIs) are typically unaware when music, including their own singing, is off-key. For example, they have difficulty detecting out-of-key notes in novel melodies, a task that most individuals (even those with no musical training) find quite easy (Ayotte et al., 2002). They also have difficulty discriminating and recognizing melodies without lyrics, even melodies that are quite common in their culture.

It is important to distinguish true mTD from the "tone deafness" label that people sometimes apply to themselves. In survey studies, about 15% of people self-define as tone deaf (Cuddy et al., 2005). Research reveals that many such people are in fact referring to their poor singing skills, even if they have a keen ear for music and show no music perception deficits (Sloboda et al., 2005). Lack of singing ability in such individuals may simply reflect a lack of training, which could presumably be ameliorated by guided practice. It is also important to distinguish mTD (which appears to have its roots in pitch perception deficits, as discussed below) from a less commonly reported disorder in which musical sounds lose their normal timbral qualities, being heard, for example, as "the banging of pots and pans" (cf. Sacks, 2007, for descriptions of such cases).

An important fact about mTD is that it is not due to lack of exposure to music, as many mTDIs report having music lessons in childhood and/or coming from musical households (Peretz & Hyde, 2003). Furthermore, research on twins suggests that there are specific genes that put one at risk for this condition (Drayna et al., 2001). Thus it makes sense that most mTDIs report having their condition for as long as they can remember. Indeed, one oft-heard story is that an mTDI first discovered his or her problem in childhood, when a choir teacher in school asked them to stop singing and simply move their lips to the music. Until that time, he or she was unaware of any musical problem.

Although the existence of mTD has been noted for over a hundred years (Allen, 1878), systematic research on this phenomenon is a relatively recent endeavor (e.g., Peretz et al., 2002; Ayotte et al., 2002; Peretz & Hyde, 2003; Foxton et al., 2004). Research in this area is poised to grow rapidly, because mTDIs are estimated to comprise about 4% of the population (Kalmus & Fry, 1980) and are thus easier to locate and work with than acquired amusics.

For the current purposes, the significance of mTD is that it manifests as a highly selective deficit for music, with no readily apparent consequences for other cognitive abilities (though see Douglas and Bilkey, 2007). Indeed, mTDIs may excel in other domains: Their ranks include numerous famous individuals, including the Nobel-prize–winning economist Milton Friedman. This makes mTD particularly attractive for the comparative study of speech and music perception. Returning to the topic at hand, what is known about musical versus spoken melodic contour perception in mTD?

In one relevant study, Ayotte et al. (2002) tested mTDIs for their ability to discriminate between sentences that differed only in intonation, and to discriminate between nonlinguistic tone sequences created from the intonation patterns of the sentences. (The tone sequences were created using the methods of Patel, Peretz, et al., 1998, described in the previous section.) In contrast to the earlier findings of Patel, Peretz, et al., which had focused on acquired amusia, Ayotte et al. found a dramatic dissociation between performance on the sentences and their nonlinguistic analogs. Specifically, mTDIs had no difficulty discriminating between sentences differing in intonation, but had substantial difficulty discriminating between the corresponding tone sequences. Controls, in contrast, performed equally well in both domains. This suggests a dissociation between melodic contour processing in speech and music, with normal speech intonation processing but impaired musical melodic contour processing.

How can the dissociation reported by Ayotte et al. (2002) be explained? Psychophysical research with mTDIs has revealed that they have deficits in detecting fine-grained pitch changes. For example Hyde and Peretz (2004) had mTDIs listen to sequences of 5 high-pitched piano tones (C6, which has a fundamental frequency of 1046 Hz), the fourth of which could differ in pitch from the others. The listeners had to indicate if the sequence contained a pitch change or not. To reach a criterion of 75% correct, mTDIs needed a pitch change of

1/2 semitone. Controls, in contrast, were nearly 100% correct for the smallest pitch change used: 1/4 semitone. Furthermore, mTDIs did not perform as well as controls until the pitch change was 2 semitones. Based on such findings, Peretz and Hyde (2003) suggested that pitch contrasts in the sentences used by Ayotte et al. (2002) were coarser than in the tone sequences, and were thus large enough to overcome the mTDIs' deficits in pitch change detection. This was a reasonable suggestion, as the process of converting F0 contours to tones used by Patel, Peretz, et al. (1998) involved converting the dynamically changing pitch within each syllable to a single pitch near the mean F0 of the syllable, thus compressing the overall amount of pitch variation in a sentence. Peretz and Hyde's suggestion also led to a testable prediction: namely, if the nonlinguistic tone sequences followed the original F0 contours exactly, then tone-deaf individuals should be able to discriminate them without difficulty. Patel, Foxton, and Griffiths (2005) tested this idea, using a population of mTDIs from the United Kingdom. The stimuli were based on an English version of the prosody-music battery, and concentrated on focus-shift pairs in language and their nonlinguistic tone analogs.

Patel, Foxton, and Griffiths (2005) made two types of nonlinguistic analogs from the focus-shift pairs. The first type was the same as used previously, in other words, each syllable was replaced by a tone of fixed pitch set to the Hz value midway between the maximum and minimum F0 values within that syllable. These were referred to as the "discrete pitch" analogs. In the other (new) type of analog, each tone's pitch exactly followed the F0 contour within the syllable, gliding up and/or down just as the F0 did: These were referred to as the "gliding pitch" analogs. In both types of sequences, tone onsets occurred at the vowel onset times of corresponding syllables, and tone offsets were determined by the offset of F0 within each syllable. Thus each tone sequence had the same temporal rhythm as the syllables in the parent sentence. All tones had a fundamental and several harmonics, giving the analogs a clarinet-like quality (Sound Examples 4.10a–f). The three types of stimuli (speech, discrete pitch analogs, and gliding pitch analogs) were presented in separate blocks, and scoring of performance was based on percentage of hits minus percentage of false alarms, in accordance with the procedure of Ayotte et al. (2002).[18] Figure 4.18 shows results on the three tasks.

Consistent with the findings of Ayotte et al. (2002), tone-deaf individuals were significantly better at discriminating sentences based on intonation than they were at discriminating discrete-pitch analogs of the intonation contours (cf. Ayotte et al., 2002, Figure 3). Surprisingly, however, tone-deaf individuals also had difficulty discriminating gliding-pitch analogs that exactly mimicked

[18] A hit was defined as a *different* configuration pair that was classified as different, whereas a false alarm was defined as a *same* configuration pair classified as different.

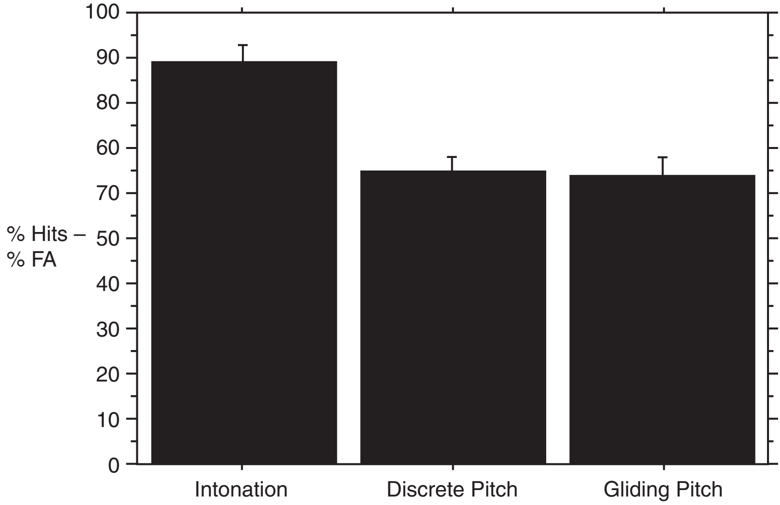

Figure 4.18 Performance of musically tone-deaf individuals on discrimination of three types of pitch patterns: intonation contours in speech, nonlinguistic analogs of intonation contours based on discrete pitches, and nonlinguistic analogs of intonation contours based on gliding pitch movements that exactly replicate the pitch pattern of intonation. The vertical axis shows percentage of hits minus percentage of false alarms. Error bars show 1 standard error. From Patel, Foxton, & Griffiths, 2005.

the intonation patterns of the sentences. In fact, performance on the discrete and gliding pitch analogs was indistinguishable, and performance on the gliding pitch analogs was significantly worse than on the sentences.

These findings are particularly striking in light of certain psychophysical findings by Foxton et al. (2004) with these same mTDIs. Foxton et al. examined thresholds for detecting a pitch change between two successively presented pure tones. In one condition ("segmented pitch-change detection"), the tones were separated by a short silent interval. In another condition ("gliding pitch-change detection"), this interval was filled by a linear frequency ramp that bridged the pitch difference between the tones. For the mTDIs in the study of Patel, Foxton, and Griffiths (2005), the threshold for pitch-change detection was significantly smaller when pitches were connected by an intervening glide than when they were separated by a silent interval (threshold = 0.21 st for gliding pitch change, versus 0.56 st for segmented pitch change, based on 75% correct discrimination).

Thus not only do mTDIs have difficulty discriminating gliding-pitch analogs of intonation, they have these difficulties *despite* the fact that they have substantially smaller thresholds for detecting pitch changes in gliding-pitch patterns than in segmented pitch patterns. It therefore appears that the relatively normal intonation perception of mTDIs cannot be explained by the idea that intonation uses coarse pitch contrasts that exceed their psychophysical thresholds for pitch-change detection.

Clearly, a hypothesis is needed that can account for the normal performance of mTDIs in discriminating spoken intonation contours versus their impaired performance in discriminating the same contours extracted from a phonetic context. In the following section, I propose one such hypothesis, based on the idea that melodic contour perception in speech and music does in fact rely on common neural circuitry.

The Melodic Contour Deafness Hypothesis

Prima facie, the finding of Patel, Foxton, and Griffiths (2005) seems to be good evidence for the modularity of contour processing (Peretz & Coltheart, 2003) because pitch contours in speech are discriminated better than the same contours extracted from speech. There is another possibility, however, as explored in this section. I call this the "melodic contour deafness hypothesis." This hypothesis proposes that mTDIs have equivalent problems in judging the *direction* of pitch change in speech and music, but that intonation perception is largely robust to this problem, whereas music perception is not. Before discussing how this hypothesis can account for the dissociations observed in the study of Patel, Foxton, and Griffiths (2005), it is worth delving into the findings that motivated the development of this hypothesis.

One such finding was the discovery that some mTDIs do in fact have problems with intonation perception in language. Specifically, Lochy et al. (2004; cf. Patel, Wong, et al., in press) tested a group of mTDIs using the sentence stimuli of Patel, Peretz, et al., (1998), and found that some had deficits in same/different discrimination of linguistic statement-question pairs. That is, when asked to tell whether two sentences were acoustically identical or not, they had difficulty when one sentence was a question and the other was a statement, a task that control participants found easy. In contrast, they did well when the two sentences were members of a focus-shift pair, in which emphasis (as signaled by pitch movement) was on a different word in each sentence (cf. two subsections back for examples of a statement-question and a focus-shift pair). The critical difference between these tasks is that the statement-question task requires discriminating the *direction* of pitch movement on the same word (up versus down), whereas the focus-shift task simply requires detecting a salient pitch movement within a sentence, because different words bear the large movement in the two members of a focus-shift pair. That is, sensitivity to the direction of pitch movement is irrelevant to the focus-shift task: As long as one can detect a pitch change, and can remember that this change happened on the same or different words in the two sentences, one can solve the task.

Another finding that motivated the melodic contour deafness hypothesis was the discovery that mTDIs have marked deficits in nonlinguistic tasks that require the perception of pitch direction. Figure 4.19 shows thresholds for 75% correct performance on three types of tasks: segmented pitch-change detection, gliding pitch-change detection, and pitch-direction judgments (Foxton et al., 2004).[19]

The first two tasks are simple pitch-change detection tasks, and were described in the previous section. The third task involves judging pitch direction: Listeners hear two pairs of pure tones and decide in which pair the pitch goes up. (Pitch always goes down in one pair and up in the other. In both pairs, tones are connected by a glide.) Note the dramatic difference between mTDIs and controls on this latter task: Their thresholds are about 20 times higher than those of controls, compared to thresholds that are 2 or 3 times worse than controls in terms of pitch-change detection.

The data of Lochy et al. (2004) and Foxton et al. (2004) suggest that mTDIs have problems with pitch-direction perception in both speech and nonlinguistic sounds. If the melodic contour deafness hypothesis is correct, then a common processing deficit for pitch direction underlies these deficits, but speech intonation perception is largely robust to this deficit, whereas music perception is not.

Why would intonation perception be robust to the deficit? One reason is that in intonation languages such as English (in which pitch does not distinguish lexical items, as it does in tone languages), the direction of a pitch change is seldom crucial to understanding. For example, if a pitch movement is used to signal focus on a word, it may matter little to a listener if the movement is upward or downward, as long as it is salient and detectable. Although the direction of pitch movement *is* important for distinguishing statements from questions, there are often redundant syntactic, semantic, or contextual cues to indicate whether an utterance is a question or not. Hence a pitch direction deficit in speech may be largely asymptomatic, revealing itself only in controlled situations in which pitch direction is crucial to the task and redundant sources of information are suppressed. A second reason that intonation perception may be robust to a pitch-direction deficit is that such a deficit is not all-or-none, but a matter of degree. Recall from Figure 4.19 that the direction deficit is defined by an elevated *threshold* for accurate detection of pitch direction. In the study of Foxton et al. (2004), mTDIs succeeded on the pitch-direction task when pitch movements were significantly above their threshold.

Hence the key question is how the size of *linguistically relevant* pitch movements in speech compares to the pitch direction thresholds of mTDIs. Consulting Figure 4.19, it can be seen that the average pitch-direction threshold for mTDIs is a little above 2 semitones. How does this compare to the size of rising or falling pitch accents in English (e.g., L+H* movements in autosegmental-metrical theory)? Existing research on intonation suggests that 2 semitones is

[19] Data presented in Figures 4.19 and 4.20 were kindly provided by Jessica Foxton. The data come from the same set of participants doing different perceptual tasks. Two outliers have been excluded from the analysis (one amusic and one control).

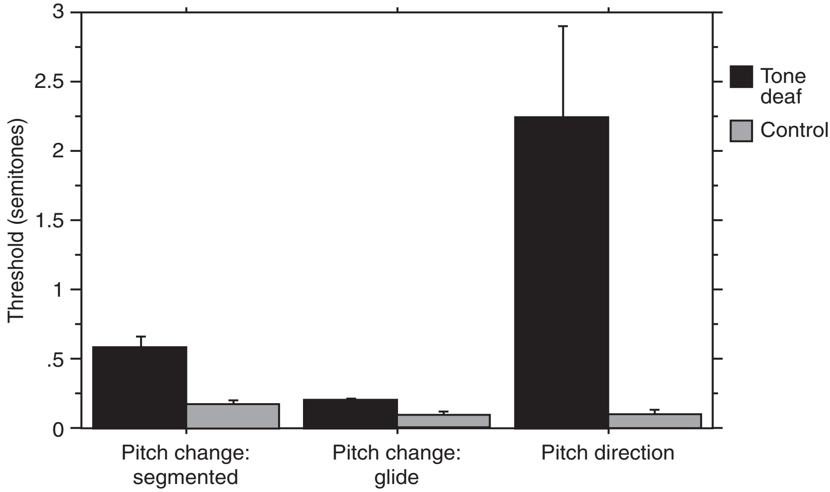

Figure 4.19 Performance of tone-deaf individuals on three kinds of pitch perception tasks: segmented pitch-change detection, gliding pitch-change detection, and pitch-direction detection. Error bars show 1 standard error. Data from the study by Foxton et al. (2004).

near the low end of the spectrum for rising or falling pitch accents in speech (e.g., Xu & Xu, 2005; Arvaniti & Garding, in press). In other words, most linguistically relevant pitch movements are likely to be in excess of the pitch-direction thresholds of mTDIs.[20] However, circumstances may arise in which an individual with a large threshold hears a linguistically relevant pitch movement below their threshold, and cannot determine its direction. Such a circumstance may explain the cases found by Lochy et al. (2004), who could not discriminate questions from statements.

Turning to music, why would music perception be so severely affected by a pitch direction deficit? As noted earlier in this chapter (section 4.5.1), musical melodies are dominated by small pitch intervals of 2 semitones or less in size. Note that this is below the average pitch-direction threshold of mTDIs. This suggests that mTDIs often cannot tell if a musical melody is going up or down in pitch. This degraded perception of musical melodic contour would make it very difficult to gauge crucial aspects of melodic structure such as motivic similarity and contrast. Furthermore, without an accurate mental representation

[20] Interestingly, in Mandarin, the rising and falling lexical tones appear to have a lower limit of about 2 semitones in connected speech (Xu, 1994, 1999), suggesting that these tones will exceed the pitch direction thresholds of most mTDIs. If this lower limit is representative for tone languages generally, it would suggest that mTD should have little consequence for speech perception in such languages under ordinary circumstances.

of melodic contour, there would be no developmental framework for learning musical intervals (Dowling, 1978). Hence mTDIs would never acquire the normal tonal schemata for music perception, which might explain why they fail to detect off-key ("sour") notes in music.

How can the melodic contour deafness hypothesis account for the dissociations observed in the study of Patel, Foxton, and Griffiths (2005)? Consider the focus-shift questions used in the speech intonation task. In such sentences, the salient pitch movement is on a different word (as noted earlier in this section). Thus, a same/different task with such sentences can be solved with a "semantic recoding strategy," in other words, by simply listening for a word with a salient pitch movement in each sentence, and then deciding if these words are the same or different. If the intonation contours are separated from their lexical context, however, as in the discrete-pitch and gliding-pitch analogs of intonation, then this strategy is no longer possible and success depends on remembering the patterns of ups and downs of pitch over time. Hence a problem in perceiving pitch direction could disrupt this task.

An important question about the melodic contour deafness hypothesis concerns its proposed neural foundation. In this regard, it is interesting to examine the relationship between mTDIs problems with pitch direction and their problems with simple pitch-change detection. Figure 4.20 shows thresholds for pitch-change detection versus pitch-direction detection for 12 mTDIs. As can be seen, the pitch-change detection thresholds do not predict pitch-direction thresholds.

This suggests an independent pitch-direction deficit. There are reasons to believe that such a deficit could arise from abnormalities in right auditory cortex. Research on patients with surgical excisions of temporal lobe regions has revealed that individuals with excisions of right secondary auditory cortex (lateral Heschl's gyrus) have pronounced deficits in judging pitch direction, even though their thresholds for simple pitch-change detection are normal. In contrast, patients with comparable excisions of left auditory cortex show no such direction deficits (Johnsrude et al., 2000). Evidence supporting a link between pitch-direction detection and melodic contour perception is the fact that both are disrupted by lesions to right auditory cortex (Johnsrude et al., 2000; Liégeois-Chauvel et al., 1998).

Neurophysiological research on animals also supports a right-hemisphere basis for coding of pitch direction (Wetzel et al., 1998), and suggests that direction sensitivity may arise from patterns of lateral inhibition among cortical neurons (Shamma et al., 1993; Ohl et al., 2000; cf. Rauschecker et al., 1998a, 1998b). Indeed, asymmetric lateral inhibition has been successfully used to give cells pitch-direction sensitivity in computational models of auditory cortex (e.g., Husain et al., 2004). In these models, direction sensitivity emerges from specific anatomical patterns of inhibitory connections between neighboring cells in tonotopic maps.

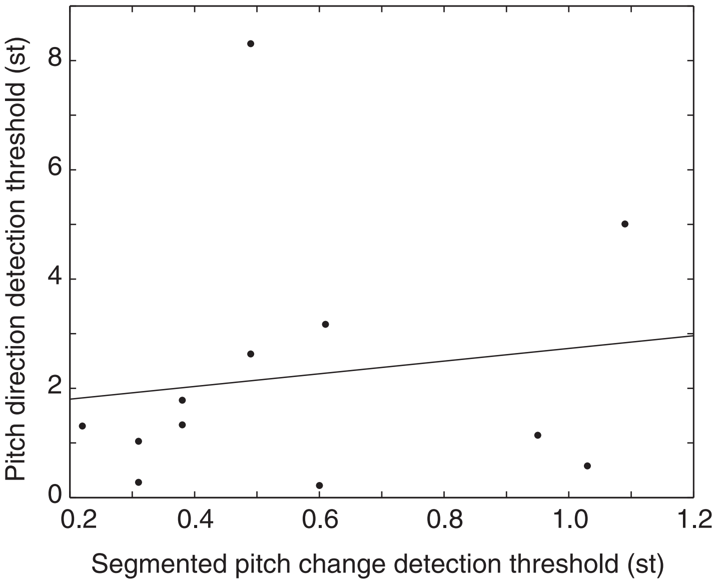

Figure 4.20 The relationship between performance on the segmented pitch-change detection task and the pitch-direction detection task for 12 musically tone-deaf individuals. The best fitting regression line is shown. The severity of the pitch-direction deficit is not predicted by the severity of the pitch-change detection deficit. Equation for regression line: Direction threshold = 1.16 * change threshold + 1.57 ($r^2 = 0.02$, $p = .65$).

Thus one plausible hypothesis for the etiology of the pitch-direction deficits in mTDIs is abnormal wiring of inhibitory connections in tonotopic maps of right secondary auditory cortex. This miswiring may occur early in development due to genetic factors (Drayna et al., 2001). Such miswiring may not be detectable using macro-imaging methods such as MRI, and may require cellular-level investigations (e.g., postmortem histology) or electrophysiological techniques (cf. Peretz et al., 2005). Interestingly, in-vivo structural brain imaging has revealed that mTDIs have a neural anomaly in the right cerebral hemisphere (Hyde et al., 2006). However, the anomaly is in a region of right inferior frontal cortex, which is thought to be involved in short-term memory for pitch patterns. The anomaly consists of a decrease in white matter tissue (i.e., the fibers that connect nerve cells rather than nerve cell bodies themselves). At the moment, it is not known whether the anomaly is inborn or if it arises in development due to anomalous connections between right frontal regions and auditory cortex, particularly because both gray and white matter patterns in the frontal cortices develop well into adolescence (Toga et al., 2006). For example, a problem in perceiving melodic contour (arising in auditory cortex) may lead to the underdevelopment of normal musical memory skills, because it