Absolute pitch is generally considered to reflect a rare musical endowment; however, its characteristics are puzzling and its genesis is unclear. We describe two experiments in which native speakers of tone languages—Mandarin and Vietnamese—were found to display a remarkably precise and stable form of absolute pitch in enunciating words. We further describe a third experiment in which speakers of English displayed less stability on an analogous task. Based on these findings, and considering the related literatures on critical periods in speech development, and the neurological underpinnings of lexical tone, we propose a framework for the genesis of absolute pitch. The framework assumes that absolute pitch originally evolved as a feature of speech, analogous to other features such as vowel quality, and that speakers of tone language naturally acquire this feature during the critical period for speech acquisition. We further propose that the acquisition of absolute pitch by rare individuals who speak an intonation language may be associated with a critical period of unusually long duration, so that it encompasses the age at which the child can take music lessons. We conclude that the potential to acquire absolute pitch is universally present at birth, and that it can be realized by enabling the infant to associate pitches with verbal labels during the critical period for speech acquisition.
composers and performers—such as Bach, Beethoven, Mozart, Rubenstein, Toscanini and Heifetz—were known to possess it, absolute pitch is often regarded as a mysterious and exceptional musical endowment. However, its characteristics are puzzling and its genesis is unclear. In this article, we review some of the perplexing features of absolute pitch, propose a theoretical framework that accommodates these features, and describe a set of findings that support our proposed framework.

Although absolute pitch is particularly prevalent among highly accomplished musicians, it is not necessarily accompanied by superior performance on other musical processing tasks. For example, absolute pitch possessors often make octave errors in assigning names to notes, particularly when different musical instruments are involved, and they do not necessarily outperform others in making judgments of octave register (Bachem, 1937; Carroll, 1975; Lockhead & Byrd, 1981; Miyazaki, 1988, 1989; Rakowski & Morawska-Büngeler, 1987; Takeuchi & Hulse, 1993). Furthermore, absolute pitch possessors do not necessarily outperform others in making judgments of musical interval (Burns & Campbell, 1994; Miyazaki, 1992, 1993), or on tasks involving short-term memory for pitch for which verbal labels cannot be employed as cues (Bachem, 1954; Rakowski, 1972; Rakowski & Morawska-Büngeler, 1987; Siegel, 1974).

Most interestingly, although absolute pitch is often considered to reflect an unusually strong long-term memory for pitch, people without this faculty can be shown to have surprisingly accurate long-term pitch memories under certain conditions. The tritone paradox (Deutsch, 1986, 1991, 1992, 1997) provides an example. To produce this musical illusion, two computer-generated tones are presented in succession. The tones are related by a half-octave, and are so constructed that their pitch classes are clearly defined, but their octave placement is ambiguous. When subjects are asked to determine whether such patterns ascend or descend in pitch, their judgments generally show systematic relationships to the positions of the tones along the pitch class circle. The findings with respect to this illusion show that most people possess an implicit form of absolute pitch, in that their judgments depend in an orderly way on pitch class, even though they are unable to name the tones they are judging.

It has further been found that the way the pitch class circle is oriented with respect to height is related to the language or dialect to which the individual has been exposed (Chalikia & Leinfelt, 2000; Chalikia, Miller, & Vaid, 2001; Chalikia, Norberg, & Paterakis, 2000; Chalikia, & Vaid, 1999; Dawe, Platt, & Welsh, 1998; Deutsch, 1991; Deutsch, Henthorn, & Dolson, 2004; Giangrande, 1998; Ragozzine & Deutsch, 1994) and to the pitch range of his or her speaking voice (Deutsch, Henthorn, & Dolson, 1999, 2004; Deutsch, North, & Ray, 1990). Given these findings, Deutsch (2002) hypothesized that the partial form of absolute pitch that is reflected in judgments of the tritone paradox had originally evolved to subserve speech. This hypothesis is bol-
stered by findings that the pitch range of the speaking voice is related to the individual's linguistic community rather than to physiological characteristics such as his or her height, weight, chest size, length of vocal tract, and so on (Deutsch, Henthorn, & Dolson, 2004; Dolson, 1994).

There is further evidence that people who are unable to name notes that are presented in isolation nevertheless evidence a partial form of absolute pitch. Terhardt and Ward (1982) and Terhardt and Seewann (1983) found that musicians who did not possess this faculty could nevertheless judge to a large extent whether a piece of music that they knew well was played in the correct key. Later, Halpern (1989) reported that musically untrained subjects, when asked to hum the first few notes of familiar songs on different occasions, were surprisingly consistent in their choice of pitches from one occasion to the next (see also Bergeson & Trehub, 2002). In a further study, Levitin (1994) asked subjects to choose a CD that contained a popular song that they knew well, and to reproduce the song by humming, whistling, or singing. The songs had been performed by only one musical group, and so had presumably been heard in only one key. On comparing the pitches of the first notes that the subjects produced with the equivalent ones on the CD, Levitin found that, when tested with two different songs, 44% of the subjects came within 2 semitones of the correct pitch for both songs.

Taking these findings together, we can conclude that absolute pitch is a complex and baffling phenomenon. It does not appear to have an explanation purely in terms of long-term memory for pitch (though this must be part of the picture), but involves verbal labeling also. However, since the naming of notes involves choosing between only 12 possibilities—the 12 notes within the octave—we would expect the task to be an easy one. Indeed, the task should be a trivial one for trained musicians, who spend thousands of hours reading musical scores and playing the notes they read. As a related point, most people easily remember melodies by name; however, the amount of information required to name a melody is vastly greater than is required to name a single note. The real puzzle concerning absolute pitch, therefore, is not why some people possess it, but rather why it is not universal. It is as though most people have a syndrome with respect to the labeling of pitches that is equivalent to color anomia (Geschwind & Fusillo, 1966), in which the patient can recognize colors, and can discriminate between them, but cannot associate them with verbal labels (Deutsch, 1988; 2002; Levitin, 1994, 1996).

**Absolute Pitch in Relation to Speech and Language**

The verbal labeling of pitches necessarily involves speech and language, so in searching for a framework in which to place absolute pitch, we can
consider further evidence that it is tied to linguistic processing. One body of evidence concerns the neuroanatomical correlates of this faculty. Schlaug, Jancke, Huang, and Steinmetz (1995) were the first to document that musicians with absolute pitch tend to exhibit an unusual form of brain structure. In most right-handers, the planum temporale, which is critically involved in speech processing, is larger in the left than in the right hemisphere. Schlaug et al. observed that this leftward asymmetry was greater among musicians with absolute pitch than among those who did not possess this faculty. This finding indicates that absolute pitch is subserved, at least in part, by brain regions that underly speech processing (see also Keenan, Thangaraj, Halpern, & Schlaug, 2001; Schlaug, 2003; and Zatorre, Perry, Beckett, Westbury, & Evans, 1998).

Another body of evidence concerns an intriguing parallel between the critical periods involved in the acquisition of speech and language on the one hand, and the acquisition of absolute pitch on the other. In his influential book, Lennenberg (1967) pointed out that adults and young children acquire a second language in qualitatively different ways. Following puberty, such acquisition is self-conscious and labored; and even after years of experience a second language that is acquired in adulthood is spoken with a “foreign accent,” and frequently with grammatical errors. Lennenberg therefore proposed that a critical period, which extends to puberty, is involved in the acquisition of speech and language.

Lennenberg’s argument has received strong support from several lines of evidence (Doupe & Kuhl, 1999; Johnson & Newport, 1989; Newport, 1990; Newport, Bavelier & Neville, 2001). For example, children who had been socially isolated early in life and were later placed in a regular environment were found to be unable to acquire normal language (Curtiss, 1977; Lane, 1976). Studies of recovery of speech following brain injury have also pointed to a critical period for speech acquisition: The prognosis for recovery has been shown to be most positive if the injury occurred before age 6, less positive if it occurred between ages 6 and 8, and very poor following puberty (Bates, 1992; Dennis & Whitaker, 1976; Duchowney et al., 1996; Varyha-Khadem, Carr, Isaacs, Brett, Adams, & Mishkin, 1997; Woods, 1983). Studies of second language acquisition have confirmed this picture. Individuals who were first exposed to a second language in infancy or early childhood were found to be most proficient in that language. Proficiency was found to decline with increasing age of initial exposure to the second language, beginning at ages 4 to 6, and continuing until adulthood, when it was found to plateau out (Johnson & Newport, 1989; Newport, 1990; Newport, Bavelier, & Neville, 2001; Oyama, 1976).

Acquisition of absolute pitch, in relation to age at onset of musical training, presents a very similar picture, and this similarity is particularly striking in terms of the time frame involved. Although some degree of absolute
pitch can be acquired in adulthood, this occurs only through extensive and laborious training (Brady, 1970; Cuddy, 1968). In contrast, when young children acquire absolute pitch, they generally do so automatically and unconsciously, without specific training on pitch-naming tasks. In addition, absolute pitch that is acquired in adulthood does not have the same ease and proficiency as absolute pitch acquired early in life (Takeuchi & Hulse, 1993; Ward, 1999).

Furthermore, there is considerable evidence that the prevalence of absolute pitch is inversely related to the age at onset of musical training (Bachem, 1955; Miyazaki, 1988; Profita & Bidder, 1988; Sergeant, 1969). In a survey of musicians and music students, Baharloo, Johnston, Service, Gitschier, and Freimer (1998) found that 40% of those who began musical training before age 4 reported that they possessed this faculty, compared with 27% of those who began training between ages 4 and 6; 8% of those who began training between ages 6 and 9; 4% who began training between ages 9 and 12; and 2.7% who began training after age 12. The striking correspondence between the timetables for acquisition of absolute pitch on the one hand, and speech and language on the other, suggests that these different capacities may be subserved by a common brain mechanism. Although critical periods for the development of other functions have been documented, such as the development of ocular dominance columns in the visual cortex of cats (Hubel & Wiesel, 1970), imprinting in ducks (Hess, 1973), and the development of auditory localization in barn owls (Knudsen, 1988) no other critical periods have been described that show a similar correspondence with speech and language in terms of time frame (see also Trout, 2003).

The argument for a linkage between absolute pitch and speech becomes even stronger when we consider the evidence from tone languages, such as Mandarin, Cantonese, Thai, and Vietnamese. In these languages, words take on arbitrarily different lexical meanings depending on the tones in which they are enunciated. Lexical tones are defined both by their pitch heights (“registers”) and also by their pitch contours. In Mandarin, for example, the word “ma” means “mother” when spoken in the first tone, “hemp” in the second tone, “horse” in the third tone, and a reproach in the fourth tone. So when a speaker of Mandarin hears “ma” in the first tone, and attributes to it the meaning “mother,” he or she is associating a particular pitch (or combination of pitches) with a verbal label. Analogously, when a person with absolute pitch hears the note G♯, and identifies it as “G♯,” he or she is also associating a particular pitch with a verbal label.

1. At a simple level of description, in Mandarin the first tone is characterized as high and level; the second tone as mid-high and rising; the third tone as low, initially falling and then rising; and the fourth tone as high and falling.
At the neurological level, there is strong evidence that the brain structures underlying the processing of lexical tone overlap with those underlying the processing of phonemes in speech. The communication of prosody and emotion has been found to be a nondominant hemisphere function, for speakers of both tone and intonation languages (Edmondson, Chan, Seibert, & Ross, 1987; Gorelick & Ross, 1987; Hughes, Chan, & Su, 1983; Ross, 1981; Tucker, Watson, & Heilman, 1977). In contrast, the processing of lexical tone has been found to be a dominant hemisphere function. For example, identification of tones has been observed to be impaired in aphasics with left-sided brain damage who are speakers of Thai (Gandour & Dardarananda, 1983; Gandour, Ponglorpisit, Khunadorn, Dechongkit, Boongird, Boonklam, & Potisuk, 1992), Mandarin (Naeser & Chan, 1980; Packard, 1986) and East Norwegian (Moen & Sundet, 1996). In accordance with these findings on brain-damaged individuals, normal Thai speakers have been reported to exhibit a right-ear advantage in dichotic listening to Thai tones when these were presented as words, though not when the same pitch patterns were presented as hums (Van Lancker & Fromkin, 1973). A third line of evidence comes from a study on normal subjects using positron emission tomography (PET). When discriminating pitch patterns in the form of Thai words, Thai subjects showed activation of the left frontal operculum (a region near Broca’s area). However, when these subjects were presented with the same Thai words which had been low-pass filtered, the same pattern of brain activation did not occur (Gandour, Wong, & Hutchins, 1998).

These three lines of evidence, taken together, indicate that when tone language speakers perceive or produce pitches or pitch contours that signal meaningful words in their language, circuitry in the dominant hemisphere is involved. Given the evidence on critical periods for speech acquisition, it appears reasonable to assume that the development of such circuitry occurs very early in life, during the period in which infants acquire other features of speech, such as vowels and consonants (Doupe & Kuhl, 1999; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Lalonde, 1988). So we can hypothesize that if pitches and pitch contours are associated with meaningful words in infancy, these are later processed by the dominant hemisphere and are associated with words and verbal labeling. However, in the absence of such an early association, this brain circuitry is much less likely to develop.

The question then arises as to which features of pitch are critical to conveying lexical meaning in tone language. If these features were purely relational, then the present discussion would be irrelevant to the genesis of absolute pitch. If, however, absolute pitch were employed to signal lexical meaning, then we would have the beginnings of an explanation as to why speakers of intonation languages, such as English, find absolute pitch so
difficult to acquire in adulthood. The study reported here was carried out as a test of the hypothesis that absolute pitch is indeed treated by tone language speakers as a critical feature of speech. The hypothesis entails that tone language speakers would evidence absolute pitch in speech processing, and that the memory representations of the pitches of speech sounds would be qualitatively different for speakers of tone and intonation languages.

**Experiment 1**

**METHOD**

**Subjects**

Seven native speakers of Vietnamese served as subjects in this experiment and were paid for their services. These were two men and five women, and they ranged in age from 27 to 56 years (mean age, 46.3 years). They had all been born and had grown up in Vietnam, and all spoke Vietnamese as their primary language. They had been living in the United States for periods ranging from a few months to 17 years. The subjects had received minimal or no musical training.

**Procedure**

The subjects were tested individually in two sessions, which were held on different days. In each session, the subject was seated before a microphone and was handed a list of 10 Vietnamese words to read out, at a rate of roughly one word every 2 s. The words in the list were chosen so that they spanned the range of tones in Vietnamese speech.

**Apparatus**

Speech was recorded onto DAT tape at a sampling rate of 44.1 kHz, using a Nakamichi microphone and a Panasonic SV-3700 Professional Digital Audio Tape Deck. The recorded samples were then transferred to a NeXT machine, where they were stored and analyzed.

**Analysis Procedure**

The speech samples were recorded into computer memory at a sampling rate of 44.1 kHz. The sound files were then converted to a sampling rate of 11.025 kHz, and were low-pass filtered, with cutoff frequencies of 1300 Hz for the female speakers and 650 Hz for the male speakers. Pitch (F₀) estimates were then obtained at 5-ms intervals, using a procedure derived from Rabiner and Schaffer (1978) with additional signal processing by one of us (M.D.). Then for each word, the pitch estimates were averaged along the musical scale; that is, along a log-frequency continuum, thus producing an average pitch for each word. (If a speaker produced a word that resulted in fewer than 10 pitch estimates, the word was discarded from the analysis of this speaker’s readings. Five out of the 70 comparisons were thus discarded.) Then for each speaker, the difference was calculated between the average pitches for each word as it was produced on the different days, and the signed differences were averaged across the words in the list.

2. In the Rabiner and Schaffer (1978) algorithm, six pitch detectors operate in parallel, and a decision matrix is then used to determine the best $F_0$ estimate.
RESULTS

Table 1 displays the numbers of subjects whose pitch difference scores fell in each 0.25-semitone bin. As can be seen, all subjects produced pitch difference scores of less than 1.1 semitone, and two of the seven subjects produced pitch difference scores of less than 0.25 semitone. The subjects must therefore have been referring to stable and precise absolute pitch templates in enunciating the list of words.

Experiment 2

The purpose of Experiment 2 was twofold. The first was to test the generality of the findings from Vietnamese subjects to speakers of a different tone language. To this end, Mandarin was chosen as the language to study. The second purpose was to explore the extent to which the pitch differences found in Experiment 1 in enunciating the same words on different days—albeit that these differences were very small—would have reflected the limitations on the precision and stability of the subjects’ absolute pitch templates, compared with other factors. To this end, a more elaborate experimental design was constructed. As before, each subject participated in two sessions, which were held on different days. However, in each session the subject read out the word list twice in succession, with the readings separated by intervals of roughly 20 s. The question addressed was whether the pitch consistency found enunciating the same word list on two occasions would be greater for readings that occurred in immediate succession, compared with readings that occurred on different days. It was reasoned that a lack of difference between these two types of comparison would provide further evidence that these speakers were invoking an absolute pitch template in enunciating the words.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Vietnamese Speakers: Pitch Difference Scores Produced from Reading a List of Vietnamese Words on Different Occasions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Difference in Semitones</td>
</tr>
<tr>
<td></td>
<td>0-0.25</td>
</tr>
<tr>
<td>Across sessions: Day 1 vs. Day 2</td>
<td>2</td>
</tr>
</tbody>
</table>

Note—The table displays the numbers of subjects whose pitch difference scores fell in each 0.25-semitone bin.
METHOD

Subjects

Fifteen native speakers of Mandarin served as subjects in this experiment and were paid for their services. These were seven men and eight women, and they ranged in age from 22 to 40 years (mean age, 27.2 years). They had all been born and grown up in the People’s Republic of China, and all spoke Mandarin as their primary language. The subjects had been living in the United States for periods ranging from several months to 6 years. With one exception, the subjects had received minimal or no musical training.

Procedure

The procedure was essentially the same as in Experiment 1, with two exceptions. First, a list of 12 Mandarin words was used, which consisted of three words in each of the four Mandarin tones, so ordered that the same tone did not occur twice in succession.

As a second difference in procedure, each subject read out the word list twice in each session, with readings separated by a time interval of roughly 20 s. Four difference scores were then obtained: between the first readings on Day 1 and Day 2; between the second readings on Day 1 and Day 2; between the first and second readings on Day 1; and between the first and second readings on Day 2.

Apparatus and Analysis Procedure

The apparatus was identical to that used in Experiment 1. The analysis procedure was also identical, except that four difference scores were now obtained for each subject. (If a subject produced a word that resulted in fewer than 10 pitch estimates, then all four comparisons involving this word were discarded from the analysis of this subject’s readings. Ten of the 180 sets of comparisons were thus discarded.)

RESULTS

Table 2 displays, for each comparison, the numbers of subjects whose pitch difference scores fell in each 0.25-semitone bin. As can be seen, remarkable consistencies were again obtained. In particular, for all comparisons, 1/3 of the subjects produced difference scores of less than 0.25 semitone. Two independent statistical analyses were performed on the difference scores, to compare the pitch consistencies on immediate repetition (within sessions) with those across days (across sessions). For the first analysis, the difference scores derived from the first and second readings on Day 1 (within sessions) were compared with those derived from the first readings on Day 1 and Day 2 (across sessions). For the second analysis, the difference scores derived from the first and second readings on Day 2 (within sessions) were compared with those derived from the second readings on Day 1 and Day 2 (across sessions). For both analyses, the difference between the within sessions and across sessions difference scores did not approach significance ($F < 1$ in both cases). This result is as expected from the hypothesis that the subjects were referring to precise and stable absolute pitch templates in enunciating the words. It also leads to the surmise that,
although the pitch differences found on comparing the readings from different days were remarkably small, they nevertheless underestimated the precision of the subjects’ absolute pitch templates.

**Experiment 3**

Experiments 1 and 2 had examined the performance of tone language speakers only, and this raises the issue of whether speakers of an intonation language would display the same pitch consistency on an analogous task. Experiment 3 was designed to examine this issue. The design of this experiment was the same as for Experiment 2, except that native speakers of English were employed as subjects, and a list of 12 English words was used for the readings.

**METHOD**

**Subjects**

Fourteen native speakers of English served as subjects in the experiment and were paid for their services. They were five men and nine women, and they ranged in age from 18 to 43 years (mean age, 25.2 years). They had all been born and had grown up in the continental United States, and none had a parent, grandparent, or sibling who spoke a tone or pitch-stress language. Most of the subjects had received minimal or no musical training.

**Procedure**

The subjects were tested using the identical procedure as used for the speakers of Mandarin.

---

3. In pitch-stress languages such as Japanese, the pitch patterns of certain words convey lexical meaning. It is unknown whether absolute pitch levels are important in conveying lexical information in such languages.
Apparatus and Analysis Procedure

The apparatus and analysis procedure were identical to those in Experiment 2. (As in Experiment 2, if a subject produced a word that resulted in fewer than 10 pitch estimates, then all four comparisons involving this word were discarded from the analysis of this subject’s readings. Twenty-seven out of the 168 sets of comparisons were discarded on this basis.)

RESULTS

Table 3 displays, for each comparison, the numbers of subjects whose pitch-difference scores fell in each 0.25-semitone bin. Two independent statistical analyses were again performed on the difference scores, to compare the pitch consistencies on immediate repetition (within sessions) with those across days (across sessions). It was found that, in contrast with the results obtained from the Mandarin speakers, the pitch consistencies obtained within sessions were significantly greater than those obtained across sessions. The difference scores derived from the first and second readings on Day 1 (within sessions) were significantly lower than those derived from the first readings on Day 1 and Day 2 (across sessions), $F(1, 13) = 9.792, p < .01$. Further, the difference scores derived from the first and second readings on Day 2 (within sessions) were significantly lower than those derived from the second readings on Day 1 and Day 2 (across sessions), $F(1, 13) = 16.376, p = .001$. The finding that English speakers, in contrast to Mandarin speakers, showed the expected increase in pitch consistency for readings that occurred in immediate succession reinforces the conclusion that the two groups of speakers were representing the absolute pitches of the words in qualitatively different ways.

Further statistical analyses confirmed this picture. On comparing the Mandarin with the English speakers in terms of consistency within sessions, no significant difference was obtained ($F < 1$). In contrast, there was

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Speakers: Pitch Difference Scores Produced from Reading a List of English Words on Different Occasions</td>
</tr>
<tr>
<td>Difference in Semitones</td>
</tr>
<tr>
<td>Across sessions</td>
</tr>
<tr>
<td>1st reading, Day 1 vs. Day 2</td>
</tr>
<tr>
<td>2nd reading, Day 1 vs. Day 2</td>
</tr>
<tr>
<td>Within sessions</td>
</tr>
<tr>
<td>1st vs. 2nd reading, Day 1</td>
</tr>
<tr>
<td>1st vs. 2nd reading, Day 2</td>
</tr>
</tbody>
</table>

Note—The table displays, for each comparison, the numbers of subjects whose pitch difference scores fell in each 0.25-semitone bin.
a significant difference between the Mandarin and English speakers in terms of consistency across sessions, $F(1, 27) = 7.923; p < .01)$. In other words, the Mandarin and English speakers showed roughly the same degree of pitch consistency in enunciating their word lists twice in immediate succession, but the Mandarin speakers were significantly more consistent than the English speakers in enunciating their word lists on different days. Thus the Mandarin and English speakers performed differently on this reading task, both qualitatively and quantitatively, with the English speakers showing less pitch consistency across days.

**Discussion**

The findings from these three experiments are in accordance with the hypothesis that speakers of tone language employ absolute pitch as a feature of speech and that they refer to precise and stable absolute pitch templates in enunciating words. The finding that the English speakers were significantly less consistent than the Mandarin speakers in enunciating the same words on different days, whereas these two groups of speakers showed roughly the same degree of pitch consistency on immediate repetition, indicates that these two linguistic groups were processing the absolute pitch levels of speech in qualitatively different ways. This conclusion is in accordance with the evidence from the neuropsychological literature showing that, whereas pitch patterns are processed for intonation purposes primarily by the nondominant hemisphere (Edmondson et al., 1987; Gorelick & Ross, 1987; Hughes et al., 1983; Ross, 1981; Tucker et al., 1977) the processing of lexical tone primarily involves the dominant hemisphere (Gandour & Dardarananda, 1983; Gandour et al., 1992; Naeser & Chan, 1980; Packard, 1986, Moen & Sundet, 1996; Gandour et al., 1998; Van Lancker & Fromkin, 1973). The present results strongly indicate that absolute pitch is involved in such lexical processing.

The present argument that absolute pitch is employed by tone language speakers as a feature of speech carries with it the strong hypothesis that different individuals of the same sex who speak in the same dialect (and who had been exposed to the same pattern of dialects earlier in life) should match up in terms of the absolute pitch levels with which they enunciate words. Informal work has indicated that such across-speaker consistencies do indeed exist, and a formal study is currently underway to examine this issue.

Another interesting aspect of the present findings is that the English speakers, although not as consistent as the Mandarin speakers in the pitches with which they enunciated the same words on different days, nevertheless showed considerable consistency on this measure. This finding may be related to those described in the Introduction showing that speakers of intonation languages, who do not possess absolute pitch as conventionally defined, can nevertheless
be shown to possess a latent form of absolute pitch under certain conditions (Deutsch, 1986, 1991; Halpern, 1989; Levitin, 1994; Terhardt & Seewann, 1983; Terhardt & Ward, 1982). We may hypothesize that in the present study also, the English speakers evidenced a latent form of absolute pitch. We may further hypothesize that the qualitative difference exhibited by the Mandarin and English speakers in terms of consistency as a function of time reflects the fact that the Mandarin speakers had been given the opportunity to associate pitches with verbal labels in infancy. The finding by Saffran and Griepentrog (2001) that 8-month-old infants were able to perform a perceptual learning task that necessitated referring to the absolute pitches of tones provides strong evidence that the ability to acquire absolute pitch is indeed present in infancy (see also McMullen & Saffran, 2004).

The present study does not address the question of whether absolute pitch, when acquired as a feature of speech, later generalizes to music. We may surmise that absolute pitch for music is acquired by speakers of tone language as though it were a feature of a second language. If this were so, then, given the literature on second language acquisition (Johnson & Newport, 1989; Newport, 1990; Newport et al., 2001), we should expect that speakers of tone language would acquire absolute pitch for music most proficiently in early childhood, and that such proficiency would decline with increasing age at onset of musical training, and level off at around puberty. In other words, tone language speakers should also show a critical period effect for the acquisition of absolute pitch for musical tones. However, on the “second language” hypothesis, we should expect that the prevalence of absolute pitch among tone language speakers should be higher overall than for speakers of intonation language; in other words, that the function relating the prevalence of absolute pitch to the onset of musical training should be displaced upward for the tone language speakers.

Given this line of reasoning, how do we explain the rare cases of absolute pitch among individuals who had not been exposed to tone language? Deutsch (2002) hypothesized that for such individuals, the critical period for acquiring absolute pitch is of unusually long duration, so that it extends to the age at which the child can begin to take music lessons. Such a predisposition for an extended critical period could be genetically determined (Baharloo et al., 1998; Profita & Bidder, 1988) and could also be associated with an unusual form of brain organization (Keenan et al., 2001; Schlaug, 2003; Schlaug et al., 1995; Zatorre et al., 1998).

The present findings also relate to those of an interesting study by Gregersen, Kowalsky, Kohn, and Marvin (1999), who conducted a survey of the prevalence of absolute pitch on students in U.S. music conservatories and in university and college music programs. A higher prevalence of absolute pitch was reported among those students who described their ethnic background as Asian than among the other students. It is difficult to make direct comparison, however, between the findings of Gregersen et
al. and those of the present study, since the subjects in the former study were all musically trained, and the ages at which they began musical training were not given. Furthermore, the subjects’ linguistic backgrounds in the former study were not given, and we may surmise that an unknown number of the Asian students would have been exposed to tone language, or to pitch-stress language, at least early in life. To our knowledge, no study that controls for linguistic factors has documented a higher prevalence of absolute pitch among individuals of Asian origin.

Finally, the present work has implications for the issues of modularity in the processing of speech and music (McMullen & Saffran, 2004; Patel, 2003; Peretz & Coltheart, 2003; Peretz, Gagnon, Hébert, & Macoir, 2004) and of the evolutionary origins of these two forms of communication (Wallin, Merker, & Brown, 2000). It is reasonable to assume that speech and music are each subserved by a large number of modules, some of which subserve one of these faculties only, while others subserve both faculties.

Our proposed framework suggests that the brain circuitry underlying absolute pitch originally evolved to subserve speech, and that it is now involved in processing absolute pitch for both speech and music. Reasoning along these lines, we can surmise that the capacity to process certain other features of music also may have evolved to subserve speech; taking as one example the processing of rapid frequency transitions, for which common circuitry may well be involved for both domains (Samson & Ehrle, 2003). A number of principles of auditory processing apply to both speech and music (Bregman, 1990; Deutsch, 1999; Deutsch & Feroe, 1981; Lerdahl & Jackendoff, 1983), and we may assume that common brain circuitries are involved here also. On the other hand, we can also assume that certain capacities are unique to speech alone, and certain others to music alone. For example, the communication of lexical meaning is clearly unique to language (except for the special case of “program music”). We may further surmise that the processing of harmonic relationships originally evolved to subserve music, particularly choral singing, which has been argued to have a strong evolutionary value (Wallin et al., 2000). The view taken here, then, is that there is no single solution to the question of modularity; but rather that some aspects of speech and music involve circuitry that is unique to one domain or to the other, but that other aspects of speech and music, including absolute pitch, involve circuitry that is common to both domains.4

4. The findings from Experiments 1 and 2 in this article were first presented at the 138th meeting of the Acoustical Society of America (Deutsch, Henthorn, & Dolson, 1999). We are grateful to Vincent Hsieh, San Hsieh, Leonard Zhang, Karin Liu, Quyen Doan, Lynsey Doan, Phi Nguyen, and Larry McClure for their contributions in different phases of the study, and to Elizabeth Bates and Elissa Newport for valuable discussions concerning critical periods for language development.
References


