This is the author's final version of the work, as accepted for publication following peer review but without the publisher's layout or pagination.

http://researchrepository.murdoch.edu.au/3015

Pitch and time, tonality and meter:

How do musical dimensions combine?

Jon B. Prince

University of Toronto Mississauga

William F. Thompson

Macquarie University, Sydney, Australia

Mark A. Schmuckler

University of Toronto Scarborough

Address correspondence to:
Jon B. Prince
Department of Psychology
University of Toronto at Mississauga
3359 Mississauga Road N.
Mississauga, Ontario CANADA L5L 1C6
Telephone: 905-608-9691
jon.prince@utoronto.ca

Abstract

We examined how the structural attributes of tonality and meter influence musical pitch-time relations. Listeners heard a musical context followed by probe events that varied in pitch class and temporal position. Tonal and metric hierarchies contributed additively to the goodness of fit of probes, with pitch class exerting a stronger influence than temporal position (Experiment 1), even when listeners attempted to ignore pitch (Experiment 2). Speeded classification tasks confirmed this asymmetry. Temporal classification was biased by tonal stability (Experiment 3), but pitch classification was unaffected by temporal position (Experiment 4). Experiments 5 and 6 ruled out explanations based on the presence of pitch classes and temporal positions in the context, unequal stimulus quantity, and discriminability. We discuss how typical Western music biases attention towards pitch, and distinguish between dimensional discriminability and salience.

Keywords: Pitch, Time, Music, Dimensional Interactions

Introduction

A central goal of cognitive psychology is to understand the process by which listeners apprehend the complex visual and auditory stimuli with which they come into contact, and how they subsequently represent this information internally. This process must act upon the multidimensional nature of objects, and exhibit stability in the face of complex and often subtle variations that frequently occur along any number of physical dimensions simultaneously. Therefore, one of the critical questions in understanding perceptual and cognitive organization is how changes along various stimulus dimensions lead to varying percepts and correspondingly different cognitive structures. Equally important, however, is determining the inter-relations between the dimensions, and how processing of change in the attributes in one perceptual dimension may or may not be influenced by change in the attributes of a different dimension. Accordingly, a core question of perceptual and cognitive processing concerns how stimulus dimensions combine in perception.

Classic research by Garner (1974) provided probably the most widely known work on dimensional interactions in information processing, defining separable and integral stimulus dimensions. Consider a simple bidimensional stimulus. If the two dimensions are integral, the perceiver cannot tease apart the component dimensions in the stimulus and instead processes it holistically. Therefore, when presented with a task that requires selective attention to one dimension in the stimulus, any variation along an irrelevant but integral dimension will affect performance because the perceiver is unable to ignore it. Conversely, independent processing occurs if the dimensions are perceptually separable. Accordingly, variation in the irrelevant dimension will have no effect on the perception and processing of the relevant one.

This question of dimensional separability versus integrality has been explored across several domains, including vision (Garner & Felfoldy, 1970; Handel & Imai, 1972), audition (Ben-Artzi & Marks, 1999; Krumhansl & Iverson, 1992; Melara & Marks, 1990a, 1990b, 1990c; Tekman, 2002) and cross-modal perception (Amazeen & Turvey, 1996; Ben-Artzi & Marks, 1995; Marks, 1987; Melara, 1989; Melara & O'Brien, 1987). However, the strict categorization of dimensions as either integral or separable does not always fit the data. Asymmetric integrality (or separability) occurs when one dimension interferes with another, but not vice versa (Garner, 1976). There are several explanations for such situations – differences in discriminability, level of processing or physical primacy of the dimensions. Differences in discriminability appear when processing dimension A alone is easier than dimension B alone. Thus, when classifying dimension B, there is interference from irrelevant dimension A due to its lesser processing requirements (Garner & Felfoldy, 1970). Accordingly, task demands can change how integral or separable dimensions appear to be. Asymmetric interference can also occur if extraction and processing of one dimension happens at an earlier stage than another, as the earlier-processed dimension may interfere with the dimension not yet analyzed (Garner, 1974). Finally, when one dimension cannot physically exist without another (thus exerting physical primacy), asymmetric integrality often emerges. For example, when classifying a speech syllable, in order to be an auditory signal, it must have a pitch. Thus pitch can asymmetrically interfere with syllable identification (Wood, 1974).

There are reports of degrees of integrality – dimensions that are neither integral nor separable, but somewhere in between (L. B. Smith & Kemler, 1978). Developmental research suggests that this continuum may be due to an age-related ability to separate a stimulus into its component dimensions (L. B. Smith & Kemler, 1977). Even in adulthood, however, individual differences can emerge in the pattern of observed integrality or

separability of dimensions (J. D. Smith & Baron, 1981). Moreover, in some cases the strategy participants adopt to complete the task may cause differing patterns of dimensional processing (Pomerantz, 1983).

Another approach, based on signal detection theory (Green & Swets, 1966), is that dimensional interactions can occur at two different levels, perceptual and decisional (Ashby & Townsend, 1986). The perceptual process refers to the phenomenal experience of the perceiver when exposed to a given stimulus, whereas the decisional process is the rule used to select a response based on the stimulus percept. Perceptual separability occurs when one dimension fails to elicit perceptual effects on another at early processing levels. Accordingly, decisional separability holds when an irrelevant dimension does not affect the decisional rule used by participants, thus happening at later processing levels. Garner interference can happen as a result of failure of perceptual or decisional separability; furthermore, failure of either type of separability does not imply failure of the other. This distinction has been successfully applied in work on facial recognition (Thomas, 2001), as well as dissociating identification and categorization processes (Maddox, 2001; Maddox & Dodd, 2003).

Taken together, such findings suggest that the classification of dimensions as purely integral or purely separable does not adequately capture the complex nature of processes that handle combinations of perceptual dimensions. As Potts et al. (1998) suggest, establishing a list of integral and separable dimensions along with numerous qualifications as to when the rules do not apply is not necessarily the most productive venture. Instead, perhaps the question of interest should be about the circumstances that do and do not give rise to joint processing of dimensions.

Although some dimensions vary along a continuum (e.g., intensity), others have complex internal structure (e.g., pitch, color), such as prototype structure, or multidimensional similarity relations (Tversky, 1977). One unaddressed issue is whether the internal structure of a dimension affects how they combine in perception. When perceivers apprehend structure within a dimension of an object, they do not perceive continuous variation along that dimension. Rather, they perceive differences as change relative to certain set points within the dimension. For instance, although there is constant variation along the color spectrum, focal colors exist that function as perceptual reference points, such as a "good" red, compared to an "off" red (Rosch, 1973). Encoding information relative to these reference points provides an organizational framework that helps to classify objects into categories. Indeed, humans naturally sort objects into semantic categories, at optimal levels of abstraction chosen to maximize the informative value of the categories (Rosch, 1973, 1975; Rosch, Mervis, Gray, Johnson, & Boyesbraem, 1976). In the language domain, the most prototypical words in a category are the most informative perceptually, and are maximally different from other prototypes (Rosch et al., 1976). An object that exemplifies a category functions as the central prototype around which the category forms; membership is then determined based on the possession of defining features (necessary for membership) for the category, as a function of the similarity of the objects along their component dimensions (Tversky, 1977). Furthermore, an exemplar within the category is not defined by possession of a particular set of characteristic features (optional, but preferred for membership), but as a better or worse example of a category according to its similarity to the prototype along a number of dimensions (Rosch & Mervis, 1975).

But how does the structure of these component dimensions combine to form a percept? Are there conditions under which the internal structure of one dimension may affect that of another? If so, then a small change along one of an object's dimensions may have larger perceptual consequences than those due to that dimension alone. The seemingly

small change might alter the perceived internal structure of that dimension, and furthermore, a structural variation could then affect the perceived structure of another dimension. In such a manner, a shift along one of an object's dimensions may cause a cascade of perceived change, ultimately resulting in a much different percept. Thus understanding of how the internal structures of an object's dimensions integrate and subsequently contribute to perception is critical to understanding how the attributes of perceptual objects are extracted and encoded. In this discussion, perceptual "objects" are not limited to those that are static and three-dimensional, but also include auditory objects and events such as words or musical notes.

Indeed, music is an ideal arena for investigating questions about the nature of dimensional interactions and the internal organization of stimuli in memory, for several reasons. First, it is a psychologically rich stimulus that provides several dimensions, such as pitch, time, timbre and loudness, which may or may not interact. These dimensions are different from circle size and color, most noticeably in that unlike such static objects that provide their own self-contained context, musical events (or auditory objects) unfold over time. Thus the dimensions of musical events often require a preceding context to establish their internal structure.

Second, there are extensive models of stimulus structure from music theory that translate into general and quantitative psychological theories. In particular, musical pitch perception illustrates classic themes of categorization and stimulus prototypes (Krumhansl, 1990); research on musical time is relevant to the same issues (Jones, 1976). Third, within a musical context the primary dimensions of pitch and time also display a quantifiable internal hierarchical organization. Whereas color, spatial position, texture and timbre have a nebulous internal structuring (i.e., only qualitative structure), musical pitches and temporal positions are organized based on their hierarchical position when a preceding context establishes this hierarchy. Specifically, listeners perceive musical events in accordance with the culturally learned internal structures of these dimensions (i.e., tonality and meter). As a result, there can be quantitative descriptions of the psychological structure of pitch and time in music, and thus investigation of how the internal structures of the two dimensions combine in a musical context.

Pitch and time in music perception

Pitch and time are the primary dimensions of music. Accordingly, much research in music cognition has focused on these dimensions, deriving rich theoretical models of how pitch and time establish, and subsequently encode within, the hierarchical structures of tonality and meter.

When listeners perceive a hierarchical ordering of stability of pitch classes in music, the resultant perceptual structure is the musical "key" (Krumhansl, 1990). Krumhansl and Kessler (1982) used a probe-tone technique that involves presenting listeners with a key-defining musical excerpt followed by a single tone, or probe, and then obtaining "goodness of fit" ratings. Listeners rate each of the 12 pitch classes in the chromatic scale; the rating determines its level of stability. Examination of these ratings suggests three hierarchical levels of tonal stability: within-tonic triad, within-key and outside-key (see Figure 1 for a graphical representation of the tonal hierarchy). These levels correspond to category membership of exemplars – the within-tonic triad members are the most prototypical pitches for the key. Additionally, the remaining pitches fall into categories in accordance with their level of similarity to the prototypical pitches. The pattern of ratings for all the pitch classes correlates near-perfectly with frequency of usage in real music and music-theoretic accounts, perhaps suggesting that the ratings are merely a function of the frequency of

occurrence in the musical context preceding the probe tone. However, randomizing the assignment of pitch classes to relative frequencies of occurrence fails to elicit any hierarchical perception of pitch (N. A. Smith & Schmuckler, 2004). In other words, if the preceding context uses a hierarchy based on frequency of occurrence that is anything other than the standard tonal hierarchy, listeners perceive no hierarchy at all. Thus the correlation of frequency of occurrence with the tonal hierarchy in a preceding musical context cannot solely account for the observed pattern of ratings.

Similarly, there is a large body of research on the role of temporal patterns in music perception, resulting in perceptual accounts of time and meter (Jones, 1976; Palmer & Krumhansl, 1990). Sequences of durations related to each other by simple ratios (i.e., 1:2 or 1:3) often characterize temporal patterns in music. These durational ratios contribute to the perception of accented (strong) and unaccented (weak) points in time. This oscillation defines musical meter – a regular pattern of alternation between weak and strong temporal positions that serves as an organizational basis for musical information. Lerdahl and Jackendoff (1983) formalized a theory of grouping of musical events based on nested hierarchies and well-formedness rules, derived from a linguistics approach. Upon establishing a meter, or pattern of relative accents, the psychological stability of temporal positions varies according to their location in the metric hierarchy (Palmer & Krumhansl, 1990).

The combination of pitch and time

The question of if pitch and time jointly contribute to the perception of musical events has a long history. For the most part, there have been two opposing viewpoints, arguing for pitch and time as either independent or interactive dimensions.

Numerous researchers have failed to find any joint effects of pitch and time in music perception, using a variety of methodologies such as judgments of melody completion, pleasantness and similarity (Makris & Mullet, 2003; Monahan & Carterette, 1985; Palmer & Krumhansl, 1987a, 1987b; Pitt & Monahan, 1987), as well as recall, pitch change detection, matching, well-formedness ratings and modulation detection tasks (Krumhansl, 1991; K. C. Smith & Cuddy, 1989; Thompson, 1993, 1994; Thompson, Hall, & Pressing, 2001). In addition, neuroimaging data and clinical evidence of double dissociations support the idea of biologically independent neural foci specialized in processing pitch and time (Fries & Swihart, 1990; Mavlov, 1980; Peretz, 1990, 1996; Peretz & Kolinsky, 1993; Peretz et al., 1994; Schön & Besson, 2002).

But opposite findings reporting interactive relations between musical pitch and time are equally common, with judgments of expectancy, melody completion and perceived emotion (Boltz, 1989a, 1989b, 1993; Schellenberg, Krysciak, & Campbell, 2000; Schmuckler & Boltz, 1994) as well as melody recognition, detection of pitch changes, duration judgments and notation (Abe & Okada, 2004; Boltz, 1989c, 1991, 1992, 1995, 1998b; Crowder & Neath, 1995; Deutsch, 1980; Jones, Boltz, & Kidd, 1982; Jones, Johnston, & Puente, 2006; Jones, Moynihan, MacKenzie, & Puente, 2002; Jones & Ralston, 1991; Jones, Summerell, & Marshburn, 1987; Kelley & Brandt, 1984; Kidd, Boltz, & Jones, 1984; Monahan, Kendall, & Carterette, 1987). There is also some neuroimaging evidence that points toward interactive relations between musical pitch and time (Griffiths, Johnsrude, Dean, & Green, 1999; Nittono, Bito, Hayashi, Sakata, & Hori, 2000). According to Jones (1987), temporal accents provide a perceptual anchor that listeners use to form expectancies about upcoming events, even when listeners have instructions to ignore temporal structure.

There have been several attempts to reconcile these divergent findings on how pitch and time contribute to the perception of music. One suggestion is that they combine

relatively late in processing, such that tasks using lower-level "early" processing preclude joint contributions of pitch and time (Hébert & Peretz, 1997; Peretz & Kolinsky, 1993; Pitt & Monahan, 1987; Thompson et al., 2001; Tillmann & Lebrun-Guillaud, 2006). Accordingly, higher-level tasks might allow top-down influence on a judgment, such that joint expectations about pitch and time become relevant. If so, then pitch height comparison tasks that require no knowledge of musical structure (implicit or explicit) and therefore do not involve top-down processes should prevent temporal manipulations from influencing the task. However, Jones et al. (2002) report the opposite when listeners compare the pitch height of two tones separated by randomly-pitched intervening tones.

Another explanation is that local judgments foster the ability to focus selectively on one dimension, whereas more global judgments that require integration of material over longer time periods makes selective attention difficult (Jones & Boltz, 1989; Tillmann & Lebrun-Guillaud, 2006). The local/global distinction differs from the notion of top-down/bottom-up processing in that both local and global judgments may involve top-down and/or bottom-up processes. As such, these explanations are separate in that they do not necessarily refer to the same type of processing, although depending on the task requirements there may be overlap between these theories. The global/local idea aligns with work on melodic completion ratings that report joint influences of the two dimensions (Boltz, 1989b). Yet recall that other melodic completion judgments failed to find this pattern (Palmer & Krumhansl, 1987a, 1987b) and furthermore that melodic similarity judgments exhibited an attentional tradeoff between pitch and temporal structures (Monahan & Carterette, 1985).

Boltz (1998a, 1999) offers another explanation, based on the inherent structure of the melody and the degree of the participants' learning (i.e., experience with the stimuli used in the experiment). It is easy to encode pitch and temporal structure when they are coherent in a melody, thus allowing a joint encoding in a dynamic shape. Conversely, a melody with little musical coherence requires more effortful processing because there are no ordered relations between the dimensions that enable a joint encoding; as a result, selective attention becomes easier. But the dimensions have shown joint contributions when the tonal coherence of a melody was systematically varied (Deutsch, 1980). The other component of Boltz's theory concerns learning, more specifically that the perceived coherence of a melody changes as the participant gains more experience with it (Boltz, 1999). As most studies use a variety of melodies and do not explicitly test differences across presentations, it is difficult to assess if this explanation can account for many of the contradictory findings.

From this review, it is clear that categorical claims of pitch and time as being independent or interactive dimensions are not productive. There are numerous methodological differences that exist between studies, but simply observing that a different task gives a different result sheds little light on the issue. However, if determining the circumstances under which pitch and time combine jointly (or not) can reveal what causes these patterns of processing, then more general insights on the perception of music may result. Moreover, understanding how dimensional structure in a stimulus affects perception can further promote understanding of more general issues such as object perception.

Certainly, it is true that pitch and temporal structures are correlated in Western music, such that stable pitches (i.e., those occupying higher positions in the tonal hierarchy) occur at stable points in time (higher positions in the metric hierarchy). Passive exposure to this connection between pitch and temporal structures in Western music likely creates expectations for these combinations; thus violations of this principle could be detrimental to a judgment along either dimension. Indeed, Lebrun-Guillaud and Tillmann (2007) reported

that the accuracy of detecting a temporal shift improved when the tonal function of the note was strong.

The aim of the present research is to explore further how pitch and time combine in the perception of musical events. Given the variability in results across task, these experiments test subjective ratings of goodness of fit as well as objective classifications of pitch-time events. This research also adds a new component to the issue of pitch-time integration in music – testing how the organizational principles of these two dimensions (tonality and meter) contribute to this phenomenon. In addition to adding insight to the complex question of how pitch and time combine in music, examining the contribution of the tonal and metric hierarchies provides an opportunity to explore dimensional relations within the context of complex hierarchically structured categories, an unexplored issue in research on dimensional organization.

Experiment 1: Pitch-time dimensional relations in probe event ratings

The goal of Experiment 1 was to examine perceptual relations between pitch and time, as well as their structural organization of tonal and metric hierarchies in the perception of complex musical passages. Previous research has often involved goodness of fit ratings to assess the psychological stability of musical events. Such "probe-tone" ratings have been examined for tones that vary in pitch (see Krumhansl, 1990, for a review) and temporal position (e.g., Palmer & Krumhansl, 1990), but there are no studies that systematically varied the tonal (pitch) and metric (temporal) stability of musical events simultaneously. The combinations of values along these two dimensions (pitch-time events) may influence how they combine perceptually. For example, the metric stability of a note's temporal position may affect its tonal stability. Moreover, the specific combination of the stability provided by both tonality and that of meter may jointly define the overall stability of a musical event. Experiment 1 used pitch-time probe events following a musical context to investigate how pitch and time contribute to judgments of events in a musical context, and to determine the role of these dimensions' structural organization (i.e., tonality and meter).

Method

Participants

There were 19 adult musician participants in this experiment, each with at least eight years of formal training. Participants received either credit in an introductory psychology class at the University of Toronto at Mississauga or monetary compensation ($10). Recruitment occurred through an online experiment database and flyers posted on campus. The average years of training was 12.8 years ($SD = 4$) and the average age was 22.1 ($SD = 3.9$).

Stimuli

One of the authors (JBP) composed a tonal melody with harmonic accompaniment in the key of C major that followed the stylistic rules of Western tonal music (Aldwell & Schacter, 2002); Figure 2 shows this passage. For this musical context there were four beats per measure, with a tempo of 120 beats per minute, producing a single beat duration (i.e., a quarter note in musical terminology) of 500 ms. Created using a PC computer and Finale 2005 software, all stimuli used a harmonically complex piano timbre. All stimuli were exported to .wav files with a sampling frequency of 44.1 KHz. The loudness of the passage was set at a comfortable listening level, with all pitches of equal loudness. The entire length of the passage was 4 seconds. Simultaneous with the musical passage there was also a metronome click occurring on every beat (i.e., 500 ms intervals). The metronome click continued for one complete measure (2 sec) after the passage itself finished, in order to

maintain a constant metric framework for listeners while also providing a pause between the melody and probe event.

Subsequent to the measure of metronome clicks, listeners heard a 250 ms pitch-time probe event. These probe events consisted of one of the 12 pitch classes of the chromatic scale ranging from middle C (262 Hz) to the B above middle C (494 Hz), at one of 8 possible temporal positions in the following measure. The eight temporal positions corresponded to an eighth-note subdivision of the metric hierarchy established by the passage (and its corresponding metronome click), meaning that these probe events could occur at any 250 ms interval within the final measure. Therefore, the amount of time elapsed between the final metronome click and the probe event could be as short as 500 ms or as long as 2250 ms. Figure 2 also presents a sample probe event used in this study. Crossing the 12 different pitch classes with the 8 possible temporal positions produced a total of 96 possible probe events.

Apparatus

Participants did the experiment on a Macintosh G4 computer running OSX 10.3, with a Viewsonic VG175 monitor. The Experiment Creator package (http://www.ccit.utoronto.ca/billt/BillThompson_files/experiment.html) controlled presentation of the stimuli. Participants heard the stimuli through Sennheiser HD 280 Pro headphones and made their responses using the computer keyboard. The experiment took place in an IAC soundproof booth that provided a quiet listening environment.

Procedure

All participants gave informed consent and completed a questionnaire about musical experience prior to the experiment. On each trial, listeners heard a musical passage followed by the probe event. After the probe event, the computer prompted listeners to rate how well the probe fit with the preceding context on a Likert scale of 1 (fits poorly) to 7 (fits well). There were no explicit instructions to attend or ignore either the pitch or temporal aspects of the probe event. Participants completed between seven and ten practice trials to familiarize them with the nature of the stimuli and the task. Practice trials were randomly selected (no randomization constraints) from the 96 experimental stimuli and no feedback was provided. Altogether, listeners completed three blocks of 96 trials, for a total of 288 trials. The entire procedure took about one hour.

<div align="center">Results and Discussion</div>

To check the consistency of the data, we examined inter-block correlations ($M = .62$, $SD = .18$), inter-subject correlations ($M = .59$, $SD = .06$), and – for trials on which the probe event occurred on the downbeat of the test measure – correlations with the standard tonal hierarchy values reported by Krumhansl and Kessler (1982) (*Mean r* = .80, $SD = .13$)[1]. For each listener, if all three of these correlations were more than two standard deviations below the group mean, then further analyses did not use that participant's data. Two listeners met these exclusion criteria, but analyses that included their data produced the same pattern of results.

Ratings were analyzed using a three-way repeated measures Analysis of Variance (ANOVA), with the within-subjects factors of pitch class (C, C#,..., B), temporal position (position 1 through 8) and block (1, 2 and 3). Of the three main effects, both pitch class, $F(11,176) = 55.85$, $MSE = 13.4$, $p < .001$, $\eta_p^2 = .78$, and temporal position, $F(7,112) = 4.43$, $MSE = 3.32$, $p < .001$, $\eta_p^2 = .22$, were significant, although block failed to reach significance, $F(2,32) < 1$, $MSE = 1.2$. Importantly, the interaction between pitch class and temporal

position was not significant, $F(77,1232) = 1.17$, $MSE = 1.06$, $p = .15$, $\eta_p^2 = .07$. Neither the remaining two-way interactions, nor the three-way interaction, were significant.

The preceding analysis demonstrates that listeners' ratings varied systematically as a function of the pitch and temporal context of the probe event, and that these effects were independent. To explore the relative contributions of the structure of these dimensions (i.e., tonality and meter) on judgments, we regressed the average ratings on predictions based on the standard tonal hierarchy values reported by Krumhansl and Kessler (1982) and metric hierarchy values reported by Palmer and Krumhansl (1990). In the first step of this regression, the individual hierarchies predicted averaged probe event ratings, as separate variables. The second step in the analysis employed an interactive term, created by multiplying the tonal and metric hierarchies. For completeness, there were three other interactive terms as well (a division, maximum, and minimum model). Step one produced a multiple R = .95 ($r^2 = .89$), with both tonal and metric hierarchy factors contributing significantly ($\beta = .94$ and .09, respectively; $p < .001$ for both). Interestingly, of the 89% of the variance accounted for by these two factors, the tonal hierarchy uniquely explained all but 1%. In contrast, none of the interactive terms included in step two added any explanatory variance ($\Delta r^2 = 0$ for all). This analysis reveals that both the tonal and metric hierarchies influenced listeners' ratings of the pitch-time probe events, with the tonal hierarchy accounting for most of the variance in mean ratings.

Figure 3 visually demonstrates the strength of the pitch effect by displaying the ratings for the 12 probe pitch class of the chromatic scale, averaged across temporal position. The pattern of ratings associated with the standard tonal hierarchy reported by Krumhansl and Kessler (1982) is clearly discernible. In comparison, Figure 4 shows the ratings for the 8 temporal positions, averaged across pitch class. The metric hierarchy, as reported by Palmer and Krumhansl (1990), is also visible, although the pattern is more subtle. Furthermore, the longer wait between the final metronome click and the later temporal positions in the probe measure did not decrease the amplitude of the oscillation in ratings between on- and off-beat temporal positions. Overall, the analyses demonstrate that goodness of fit ratings differed as a function of both pitch and time, consistent with the dimensional structures of tonality and meter, and the two dimensions made independent contributions to judgments in this task and context.

Why did pitch exhibit such a strong effect in this experiment? One potential reason is that there were more possible pitch classes (12) than temporal positions (8). Unequal stimulus quantity could have led to better overall discriminability (Garner & Felfoldy, 1970; Melara & Mounts, 1994) for pitch, making it a more accessible psychological dimension and overshadowing differences in temporal position[2]. Another possibility is that without specific instructions to attend to either pitch or time, listeners may instinctively focus on the pitch dimension, magnifying the effect of pitch class relative to temporal position by default. The relative strength of pitch class variations raises questions as to whether this study provided the best possible context for uncovering an interaction between the two dimensions, and between the tonal and metric hierarchies. Given that temporal position was relatively weak in influencing ratings (compared to pitch class) the chance of finding an interaction between the two, should it exist, is small. One way to resolve this issue more fully would be to examine how pitch and time combine when temporal variation plays a stronger role in influencing listeners' ratings. Investigating this question was the goal of the next experiment.

Experiment 2: Pitch-time dimensional relations in selective attention to time

In Experiment 1, pitch structure (tonality) dominated judgments of musical probe events that varied in both pitch class and temporal position. In an attempt to increase the impact of time on ratings in Experiment 2, we asked listeners to attend selectively to temporal position and ignore pitch. Such a selective attention task enabled evaluation of the degree to which listeners are able to classify events in terms of temporal position, and it provided another test of how musical pitch and time combine, as well as the tonal and metric hierarchies.

Method

Participants

There were 22 adult musician participants in this experiment, recruited either from an introductory psychology class or through flyers posted at the University of Toronto at Mississauga; the average years of training was 13.1 years ($SD = 4.3$) and the average age was 21.9 ($SD = 4.2$). Participants received either course credit or $10 for their time.

Stimuli, Apparatus and Procedure

The stimuli, apparatus and procedure were the same as in Experiment 1, except for the instructions given to listeners. Specifically, the instructions given to participants directed them to consider only the temporal aspects of the melody and probe event, while ignoring its pitch when making goodness of fit ratings.

Results and Discussion

Data were again checked by examining inter-block correlations ($M = .49$, $SD = .25$), inter-subject correlations ($M = .63$, $SD = .08$), and correlations with the metric hierarchy of Palmer and Krumhansl (1990) for the trials whose probe event used the tonic pitch (the most psychologically stable pitch; $M = .72$, $SD = .16$). Overall, inter-block correlations were more variable in this experiment, resulting in a high standard deviation for the group. Although no one participant's inter-block correlation fell more than two standard deviations below the mean, the inter-subject and metric hierarchy correlations of some subjects were low enough to suggest that they were clearly not attending to the task. Therefore, the exclusion criteria were adjusted such that if both of these correlations of any given participant were more than two standard deviations below the mean, their data were excluded; six participants met these criteria and thus their data were removed from further analysis. However, analyses that included data from these excluded participants did not change the observed pattern of results, and furthermore, using inter-subject correlations as the only exclusion criterion resulted in removing the same six participants' data.

A three-way repeated-measures ANOVA tested for effects of pitch class (12 levels), temporal position (8 levels) and block (3 levels) on listeners' ratings. Pitch class and temporal position both exerted significant effects, $F(11,165) = 7.79$, $MSE = 3.09$, $p < .001$, $\eta_p^2 = .34$, $F(7,105) = 41.05$, $MSE = 20.34$, $p = .001$, $\eta_p^2 = .73$, respectively. There was no effect of block, $F(2,30) = 1.41$, $MSE = 1.74$, $p = .26$, $\eta_p^2 = .09$; neither were there any significant two-way or three-way interactions. Ratings clearly varied as a function of both pitch class and temporal position, even though listeners' instructions were to ignore pitch variation.

As in Experiment 1, we examined the relative contributions of tonality and meter on judgments using a stepwise linear regression. Step one introduced the tonal and metric hierarchies, and their (multiplicative) interaction entered in step two. Step one revealed that both tonal and metric hierarchies had significant predictive power in the model ($\beta = .191$ and .836, respectively; $p < .001$ for both), resulting in a multiple R of .86 ($r^2 = .74$). The interaction variable from step two added no additional predictive power ($\Delta r^2 = 0$). Thus the individual hierarchies explained 74% of the variance in ratings, with meter accounting for

70% and tonality the other 4%. Figures 3 and 4 show the variability in listeners' ratings as a function of pitch class and temporal position, respectively. Reversing the pattern observed in the previous study, variation in accordance with the tonal hierarchy in this experiment attenuated (although it is still present) compared to the now more pronounced metric hierarchy.

A comparison of Figures 3 and 4 clearly reveal an effect of instructions on ratings, with pitch dominating ratings in Experiment 1 and time prevailing in Experiment 2. Even though the degree of differentiation (or range) along the two dimensions varied across experiments, the profiles (or general shape) of these ratings were the same across experiments. That is, the ratings from both experiments reflected both the tonal and metric hierarchies, even though the relative range of their effect differed across experiments. As a further comparison across experiments, a four-way repeated measures ANOVA compared the ratings from Experiment 1 and 2, using within-subjects factors of pitch class, temporal position and block, and a between-subjects factor of instructions (experiment). Ratings were higher overall in Experiment 2, producing a main effect of instructions, $F(1,31) = 5.76$, $MSE = 72.38$, $p < .05$, $\eta_p^2 = .16$. Not surprisingly, both pitch class and temporal position showed significant main effects in this analysis, $F(11,341) = 59.27$, $MSE = 8.41$, $p < .001$, $\eta_p^2 = .66$; $F(7,217) = 47.17$, $MSE = 11.56$, $p < .001$, $\eta_p^2 = .6$, respectively. Block was not significant, $F(2,62) < 1$, $MSE = 1.46$, although it did interact with temporal position, $F(14,434) = 1.96$, $MSE = 1.41$, $p < .05$, $\eta_p^2 = .06$, reflecting the fact that participants differentiated more between temporal positions as the experiments progressed. Confirming that the instructions given to listeners did indeed modify how pitch class and temporal position influenced their ratings, experiment interacted with both pitch class and temporal position, $F(11,341) = 29.95$, $MSE = 8.41$, $p < .001$, $\eta_p^2 = .49$; $F(7,217) = 28.51$, $MSE = 11.56$, $p < .001$, $\eta_p^2 = .48$, respectively. The only significant three-way interaction was between pitch class, temporal position, and experiment, $F(77,2387) = 1.31$, $MSE = 1.21$, $p < .05$, $\eta_p^2 = .04$, although this interaction failed to produce any obviously interpretable pattern of results.

Therefore listeners could, to a large extent, voluntarily change what dimension to emphasize in forming a goodness of fit rating of probe events varying in pitch class and temporal position. The extent of listeners' success in this relative emphasis indicates the degree to which they can separate the dimensions of pitch and time in music in this task. Furthermore, linear and additive combinations of the hierarchical structures of the two dimensions (tonality and meter) predicted the ratings in both experiments. No interactive factors increased the predictive power of either model. Therefore, these findings fit with the extant body of results using a comparable approach to assessing how pitch and temporal structures combine (Palmer & Krumhansl, 1987a, 1987b) or melodic and harmonic information (Schmuckler, 1989).

On the other hand, the results of Experiment 2 revealed a failure to ignore completely the pitch information of the probes despite listeners' instructions. Previous work on dimensional processing proposed that an inability to attend selectively to one dimension while ignoring another dimension suggests they are not independent (Garner, 1974). Without making strong claims of integrality or separability, these results do suggest a more holistic perception of musical pitch-time probe events. However, it is important to keep in mind the much-attenuated size of the pitch effect in Experiment 2. Using a convergent method to explore this issue further, we used a speeded classification task in Experiments 3 and 4.

Experiment 3: Pitch-time dimensional relations in speeded temporal classification

We used a speeded-classification task in Experiment 3 to examine how musical pitch and time work together in perception of pitch-time probe events. Previous research in both auditory (Ben-Artzi & Marks, 1999; Dyson & Quinlan, 2002; Melara, Marks, & Lesko, 1992) and musical contexts (Krumhansl & Iverson, 1992; Tillmann & Lebrun-Guillaud, 2006) used speeded classification tasks. Specifically, Krumhansl and Iverson (1992) investigated the perceptual relations between pitch and timbre, both in isolation and embedded in a musical context. When using a memory paradigm in a musical context, these authors found independence between the dimensions. But in isolation, and using a speeded classification task, the dimensions interfered with each other, as they did in Melara and Marks (1990b). In this experiment, we adopted the orthogonal (filtering) task, in which all possible combinations of pitch class and temporal position formed the range of musical probe events.

Method

Participants

This experiment had 36 adult musician participants[3], again recruited via the introductory psychology class and through flyers posted at the University of Toronto at Mississauga. The average years of training was 9.67 years (*SD* = 3.19) and the average age was 19.4 (*SD* = 2.54).

Stimuli

There were three new melodies used as stimulus contexts in this experiment. This increase in the number of melodies was because participants in the previous experiments felt the repetition of only one melody made it challenging to remain engaged in the task. These melodies (notated in Figure 5) had the same tonality and harmonic rhythm as Experiments 1 and 2. Each block had a different melody, with the order of blocks counterbalanced across listeners.

Apparatus

The experimental apparatus were the same as in Experiments 1 and 2.

Procedure

All aspects of the experimental procedure were the same as in the previous studies, except for the instructions given to listeners. Specifically, the instructions provided to listeners were to indicate as quickly as possible whether the probe was on or off the beat (i.e., aligned in time with the metronome clicks, if they continued into the probe measure) after hearing the probe event. The "a" and ";" key were the two response keys, with the assignment of keys counterbalanced across listeners. The participants knew that the pitch of the probe would vary; the instructions were to disregard this variation, as it had no bearing on the temporal task they were to perform. The experiment took less than an hour.

Results and Discussion

Mean accuracy in temporal classification was 80% (*SD* = 14.7%); years of formal training did not correlate significantly with accuracy (*r* = .2, *p* = .25). Accuracy and RT correlated negatively (*r* = -.35), indicating that there was no speed-accuracy tradeoff.

Analysis of RT data customarily entails removing RTs for incorrect trials and replacing them with the mean value for that participant. Unfortunately, such a procedure would result in replacing an unacceptably high level of missing data (20%) with the mean, making statistics performed on the modified data set misleading. Accordingly, it was not possible to analyze RTs, and all analyses used only accuracy data.

A three-way repeated measures ANOVA on the accuracy data, using pitch class, temporal position and block as within-subjects variables revealed main effects for both pitch

class, $F(11,385) = 3.66$, $MSE = .21$, $p < .001$, $\eta_p^2 = .1$, and block, $F(2,70) = 14.89$, $MSE = .48$, $p < .001$, $\eta_p^2 = .3$, but no main effect of temporal position, $F(7,245) < 1$, $MSE = .2$. Averaging across temporal position, accuracy across pitch class did not show an interpretable pattern, and did not correlate with the tonal hierarchy ($r = -.31$, $p = .6$). The effect of block was due to the fact that accuracy decreased over block, perhaps due to fatigue. There was an interaction between temporal position and block, $F(14,490) = 5.56$, $MSE = .36$, $p < .001$, $\eta_p^2 = .14$. This interaction reflected the fact that listeners were disproportionately better at the earlier temporal positions in the first block, but not in the second or third block. Most importantly, there was an interaction between pitch class and temporal position, $F(77, 2695) = 1.48$, $MSE = .11$, $p < .01$, $\eta_p^2 = .04$, reflecting the fact that the accuracy at each temporal position systematically varied according to the pitch of the probe. No other interactions were significant.

It is difficult to understand the pitch-time interaction by inspecting all 96 conditions simultaneously (12 pitch classes by 8 temporal positions). However, correlating the accuracy across pitch classes with the tonal hierarchy at each temporal position reveals the nature of the pitch-time interaction, as well as correlating the accuracy across temporal positions with the metric hierarchy at each pitch class. Tables 1 and 2 display these correlations, respectively. Table 1 shows that the correlations between accuracy and the tonal hierarchy oscillate in accordance with the metric stability of each temporal position. Correlations with the tonal hierarchy are positive on temporal positions with high metric stability by virtue of being on-beat (positions 1, 3, 5 and 7; correlation $M = .47$), whereas they are negative for off-beat positions that are low in metric stability (positions 2, 4, 6 and 8; correlation $M = -.48$). Table 2, in comparison, shows that the correlation between accuracy and the metric hierarchy is positive for all pitch classes. However, even though the Table 2 correlations are all positive (for Experiment 3), they still show an interesting pattern across pitch class. Correlations with the metric hierarchy were highest for pitch classes with the greatest tonal stability (tonic triad members C, E and G; correlation $M = .87$), and less for the remaining pitch classes (correlation $M = .63$).

This pattern is interpretable as a congruity effect (Pomerantz & Garner, 1973), or a redundancy gain and loss (Garner, 1974) for the temporal classifications. Accuracy improved on metrically stable temporal positions when the probe was a stable pitch in the tonality of the passage (i.e., occupied a high position in the tonal hierarchy). Conversely, accuracy improved on metrically unstable temporal positions when the probe was tonally unstable. Thus congruity between the structures of the pitch and temporal dimensions improved accuracy, and incongruity was detrimental.

To determine if the effect of pitch, and its interaction with time could be due to a perceptual difficulty or a decision bias, signal detection theory was used to examine performance across pitch class (SDT, Green & Swets, 1966; Macmillan & Creelman, 1991). For each participant, a correct response of "on the beat" was a hit, whereas incorrectly responding "on the beat" was a false alarm. Obviously, by collating the data in this fashion, the data could not be analyzed across temporal position. To avoid infinite d' measures, any hit or false alarm proportions equally 0 or 1 were converted to $1/(2n)$ or $1-1/(2n)$ respectively, where n equals the number of trials included in the ratio (Macmillan & Creelman, 1991). We also use this correction for all subsequent experiments. A one-way repeated measures ANOVA on the discriminability (d') scores did not vary across the within-subjects factor of pitch class, $F(11,385) = 1$, $MSE = .27$, $p = .44$, $\eta_p^2 = .03$, denoting that probe events of one pitch were no more distinguishable than other pitch classes.

However, a separate one-way repeated measures ANOVA (again with pitch class as the within-subjects factor) testing response bias (c) did reveal a difference across pitch class, $F(11,385) = 2.67$, $MSE = .09$, $p < .01$, $\eta_p^2 = .07$. This finding indicates that the tonal stability of the probe event influenced the perceived metric stability of the probe event. In other words, listeners were more likely to report that the probe event was on-beat in accordance with greater tonal stability of the probe event's pitch. This pattern is visible in Figure 6, which shows that response bias followed the tonal hierarchy, $r = .87$.

There are two primary findings from this study. First, consistent with Experiment 2, pitch influenced listeners' speeded classifications of temporal position even when instructed to attend selectively to temporal information. Further, pitch variation was irrelevant to this classification task, which presumably would make it easier to ignore than in a goodness of fit rating task. Second, pitch affected listeners' responses in a manner consistent with the tonal hierarchy. Accordingly, changing the task (classification instead of rating goodness of fit) uncovered a different form of interference of pitch on temporal judgments of musical probe events. The fact that the tonal hierarchy affected response bias and not discriminability gives some insight as to the level at which the integration of pitch and temporal structure occurred. Because discriminability did not differ across pitch class, listeners were not more or less accurate at perceiving the probe event based on pitch class. Rather, the effect in response bias indicates that pitch affected listeners' judgments at a higher cognitive (decisional) level, such that the tonal stability of the pitch biased the assessment of metric stability in a congruent manner.

However, these data do not reveal whether this pattern is global or asymmetric. In other words, even though the pitch of the probe affected a temporal judgment, the corresponding task of varying the temporal position of the probe event while classifying its pitch may or may not show a similar pattern of interference. This distinction is critical for discerning the nature of dimensional processing. Garner and Felfoldy (1970) demonstrated that dimensions can appear to interact when in fact it is only differences in stimulus discriminability that are causing the effect. If the irrelevant dimension of a stimulus is substantially easier to perceive (and is thus more discriminable) than the attended one, then redundancy gains and losses will result whether or not the dimensions are integral. When systematically varying the relative discriminability of dimensions, Garner interference varies accordingly and practice does not affect this pattern (Melara & Mounts, 1993).

Therefore, the real test comes when listeners classify the probe based on its pitch while the temporal position varies. If stimulus discriminability is the cause of these results, then the congruity effect will disappear. Such a result would have occurred because the less discriminable, distracter dimension (temporal position) would not interfere with classification of the more discriminable, attended dimension (pitch). If, however, the pattern of joint contributions of pitch and time to the perception of musical probe events remains when switching the attended dimension, then there is evidence of a global failure of selective attention (Melara & Mounts, 1993). Such a finding would provide stronger evidence of joint contributions of pitch and time in a classification task. Therefore, modifying the task instructions to attend to pitch class instead of temporal position will shed light on the issue of asymmetric or global failure of selective attention in musical pitch and time. In turn, this modification will elucidate further how the dimensions of musical pitch and time, as well as their internal structures, combine. The goal of Experiment 4 was to investigate this possibility.

Experiment 4: Pitch-time dimensional relations in speeded pitch classification

Listeners were unable to attend selectively to temporal position and ignore pitch when rating the goodness of fit of probe events in Experiment 2. Experiment 3 provided a replication of this interference using a speeded-classification task of temporal position; furthermore the tonal hierarchy biased responses via a congruity effect. Experiment 4 further explored the combination of musical pitch and time by asking listeners to make speeded classifications based on pitch class while ignoring temporal position.

Method

Participants

The participants were 38 adult musicians recruited from an introductory psychology class, or through flyers posted at the University of Toronto at Mississauga, or in-class recruitment of music students at the University of Toronto at Scarborough; the average years of training was 10.7 years (SD = 2.35) and the average age was 19.8 (SD = 2.41).

Stimuli, Apparatus and Procedure

All stimuli and apparatus were the same as in the previous studies. The procedure was the same as in Experiment 3, with the exception of the instructions. In this experiment, participants indicated whether the pitch of the probe event was "in" or "out" of the key of the tonality of the passage by pressing either the "a" or ";" key as quickly as possible; again counterbalancing the assignment of keys to responses across listeners. Although the participants knew that the timing of the probe would vary, their instructions were to disregard the temporal dimension, both in the melody and the probe. Musical organization specifies that, of the 12 chromatic pitches, 7 of these notes are "in" a given key, or tonality; the remaining 5 pitches are "out" of the key. The pattern of "in" and "out" of key pitch classes follows closely, but not exactly, an oscillation as pitch height increases; accordingly judgments of in- versus out-key are not confounded with pitch height.

Results and Discussion

Preliminary inspection of the data indicated that accuracy in this task was lower than the previous experiment (M = 74%, SD =14%). Furthermore, average RTs were nearly twice as long as those in Experiment 3 (1322 ms versus 748 ms, respectively). Further exploration of these data revealed that the majority of errors occurred on the diatonic pitch classes that were not members of the tonic triad (the second, fourth, sixth, and seventh scale degrees). The mean accuracy on tonic triad pitch classes was 85%; for the remaining diatonic pitch classes mean accuracy was 59%; the non-diatonic pitch classes was 80%. These pitch classes (D, F, A and B in the key of C major) occupy an intermediate level in the tonal hierarchy, and hence, although technically part of the diatonic set (and classified accurately as in-key above chance, $t(37)$ = 2.1, p < .05), listeners might consider their key membership to be ambiguous. Given their intermediate status vis a vis their membership in the tonality of the context melody, and the fact that these pitch classes produced more errors and longer RTs, our analyses disregarded diatonic non-tonic triad pitch classes and focused on comparisons between the tonic triad and non-diatonic pitch classes (i.e., the two extreme levels of the tonal hierarchy).

Removing diatonic non-tonic triad pitch classes (D, F, A and B) resulted in increased overall accuracy (M = 82%, SD = 7%). As in Experiment 3, accuracy and RT correlated negatively (r = -.31), ruling out a speed-accuracy tradeoff, and years of formal training did not correlate significantly with accuracy (r = .2, p = .25).

We used a three-way repeated measures ANOVA with the within-subjects factors of pitch class, temporal position, and block to analyze the accuracy data. The only significant main effect was that of pitch class, $F(7,259)$ = 3.27, MSE = .53, p < .01, η_p^2 = .08. This

effect indicates that listeners were more accurate for pitch classes that had high tonal stability; accuracy across pitch class correlated significantly with the tonal hierarchy ($r = .72$, $p < .05$). There was no main effect of temporal position, $F(7,259) = 1.23$, $MSE = .11$, $p = .29$, $\eta_p^2 = .03$, nor was there a main effect of block, $F(2,74) < 1$, $MSE = .22$. There was an interaction between pitch class and temporal position, $F(49,1813) = 1.52$, $MSE = .1$, $p < .05$, $\eta_p^2 = .04$; no other interactions were significant.

As in Experiment 3, Tables 1 and 2 show the pattern of accuracy across temporal position and pitch class. Table 1 shows that the correlation with the tonal hierarchy did not change across temporal position as it did in Experiment 3. Instead, all of the correlations of accuracy with the tonal hierarchy were positive, showing that accuracy for the different pitch classes did not change as a function of temporal position. Table 2 shows that the correlations of accuracy with the metric hierarchy did change as a function of pitch class, but not in accordance with tonal stability (tonic triad members C, E and G; correlation $M = .18$), the correlation of the remaining pitch classes was -.06.

Unlike Experiment 3, neither discriminability (d') nor response bias measures (c) explain the nature of the interaction between pitch class and temporal position. For this analysis, hits were correct responses of "in the key" and incorrect responses of "in the key" were false alarms. Neither discriminability, $F(7,259) = 1.26$, $MSE = .21$, $p = .27$, $\eta_p^2 = .03$, nor response bias, $F(7,259) = 1.39$, $MSE = .05$, $p = .21$, $\eta_p^2 = .04$, differed across temporal position. Based on all of these analyses, the most intuitive explanation for the significant pitch class by temporal position interaction reported in the earlier omnibus ANOVA is to note that accuracy on the tonic triad pitch classes was better than the nondiatonic pitch classes, but only for the first four temporal positions; accuracy for the two types of pitch converged at the later temporal positions (see Figure 7).

Although there was an interaction between pitch and time in both Experiments 3 and 4, the nature of this interaction is different in important ways. First of all, the interference of the irrelevant dimension varied based on the task instructions. In Experiment 3, there was a strong main effect of the irrelevant dimension of pitch ($\eta_p^2 = .1$), whereas there was no main effect of the irrelevant dimension of time in Experiment 4 ($\eta_p^2 = .03$).

Second, and similarly, the structure of the irrelevant dimension governed the form of the pitch-time interaction in Experiment 3, but not in Experiment 4. That is, the tonal hierarchy dictated the pattern of interference of pitch on the temporal judgments (in the form of a response bias); yet the reverse did not occur. The metric hierarchy had nothing to do with performance in Experiment 4 (see Table 1), with neither discriminability nor response bias differing across temporal position.

Third, the interaction in Experiment 4 is understandable as perceptual decay over time. In Experiment 3, participants always performed better when the probe occupied a more stable position in the hierarchy of the relevant dimension (i.e., on-beat probes did better than off-beat probes). Homologously, accuracy was higher for in-key probes than out-key probes in Experiment 4. However, this advantage decreased with increasing temporal distance from the melody (i.e., further along in the probe measure). As with any stimulus, sensory and perceptual information fades after the event passes, thus listeners' perception of the difference between in-key and out-key probes decreased for probes that occurred later in the probe measure. Lastly, as will become clear in Experiment 6, we failed to replicate this interaction.

Accordingly, Experiment 4 points to an asymmetry in listeners' selective attention to pitch versus temporal dimensions in these classification tasks. That is, the nature of the interaction was asymmetric – musical pitch interfered with time, but not vice versa.

However, before considering the implications of this asymmetry, it is important to highlight the limitations of these data. One such limitation is that these experiments may have emphasized the pitch dimension more than the temporal dimension. If true, then the observed asymmetry of pitch affecting time and not vice versa is not a general process of musical probe event perception, but is instead idiosyncratic to the current experimental context. Along these lines, there are several factors that may have contributed to these findings, and thus suggest that a different experimental context could well produce a different pattern of findings.

The first such factor stems from the recognition that some of the pitch classes used as probe events were not in fact present in the initial musical context. In contrast, all of the temporal positions employed as probe events occurred within the context. Thus, it might simply be that the appearance of new pitch classes surprised the listener, resulting in accentuated attention to the pitch dimension relative to the temporal dimension.

A second and related factor derives from the fact that (as previously described) there were more pitch classes (12) than temporal positions (8) in these stimulus passages. This inequality in stimulus quantity could conceivably highlight for listeners the importance of the pitch dimension, relative to the temporal dimension. In fact, stimulus inequalities may also be linked to the most counterintuitive observed result in this work, namely that pitch had such a strong interfering effect on temporal judgments (Experiment 3), and yet participants performed relatively poorly when directly classifying pitch (Experiment 4). Recent work on the origin of Stroop effects (Melara & Algom, 2003) offers a potential explanation for such a finding. These authors propose that adding levels to one dimension (pitch in the present case) lowers the probability of hearing any particular level on a given trial, and leads to two related phenomena. First, when the dimension with more levels is irrelevant, it draws attention away from the relevant dimension with fewer levels; consequently the likelihood of interference from the irrelevant dimension increases, but accuracy is unaffected. Second, accuracy declines when classifying the dimension with more levels directly because the relevant dimension becomes less predictable, yet interference from the irrelevant dimension is less likely. The results of Experiments 3 and 4 are consistent with both of these phenomena – interference from the irrelevant dimension of pitch when classifying time (but unaffected accuracy), concomitantly decreased accuracy when classifying pitch (but no interference from time). As a counter example to these arguments, other research has demonstrated that in a non-musical context, increasing the quantity of pitches relative to the quantity of loudness levels actually decreased the degree to which pitch interfered with loudness judgments (Melara & Mounts, 1994). Nevertheless, it remains possible that the same pattern does not hold within musical contexts.

Differences in discriminability between dimensions comprise the most common explanation for asymmetric results such as observed in these studies. Even with dimensions previously demonstrated as independent, if discriminability varies across dimensions then the more discriminable of two dimensions tends to influence the less discriminable dimension, but not vice versa (Garner, 1974; Garner & Felfoldy, 1970; Melara & Algom, 2003; Melara & Mounts, 1993; Sabri, Melara, & Algom, 2001). Melara and Mounts (1993) eliminated and even reversed the classic Stroop asymmetry (Stroop, 1935) by manipulating the relative discriminability of color and word dimensions. In each case, these authors found that the more discriminable of two dimensions always interfered with judgments of the less

discriminable dimension. It is important to note, though, that such findings would actually predict the opposite of the results of Experiments 3 and 4. Given that performance was generally better when making temporal classifications relative to pitch classifications, the temporal dimension was presumably more discriminable and hence should have interfered with pitch classifications. Accordingly, of the two explanations – unequal stimulus quantity and discriminability – it seems that it would be the former factor, as opposed to the latter factor, that would account for these results. Regardless, it remains possible that both may have contributed to these findings.

The best way to address issues of unequal stimulus quantity and dimensional discriminability is to use an equal number of pitch classes and temporal positions in baseline tasks in addition to the filtering (orthogonal) task already employed. Baseline tasks determine if the discriminability is equal between the dimensions by varying only one dimension at a time while the other remains constant; baselines thus preclude the possibility of any interference from the irrelevant dimension. Using an equal number of levels for both pitch and temporal dimensions avoids any possible set size effects that might unbalance efforts to establish equal discriminability in the baseline tasks (Melara & Mounts, 1994). Therefore, if discriminability in the pitch and time baseline tasks is equal, while using equal numbers of pitch classes and temporal positions, then the validity and reliability of using the filtering task to investigate dimensional interference improves. The goal of the final two studies was to provide baseline measures (Experiment 5) for a subsequent filtering task (Experiment 6).

Experiment 5: Baseline pitch and time dimensional relations

In Experiment 3, listeners were instructed to ignore the pitch while judging the temporal position of probes, with probes events encompassing 12 possible pitch classes and 8 possible temporal positions. Experiment 4 reversed the relevant and irrelevant dimensions of this task by asking listeners to judge pitch while ignoring time, employing the same probe events. Experiment 5 examined the importance of this inequality in the number of pitch classes versus temporal positions by having listeners perform baseline classifications of pitch and time with the same stimuli consisting of only eight pitch classes and eight temporal positions. Furthermore, in an effort to control the first factor described earlier, all probe pitch classes and temporal positions were explicitly presented in the preceding context. For the two pitch and temporal baseline tests, only one dimension changed while the other remained constant, thus measuring the accuracy of dimensional classifications separately. If the accuracy of classification judgments for pitch classes and temporal positions is comparable, then subsequent tests of selective attention to pitch and time are not confounded by unequal discriminability between dimensions.

Experiment 5 also provided baseline goodness of fit ratings for the 12 pitch classes and 8 temporal positions used in the previous experiments. Baseline goodness of fit ratings offered the opportunity to ensure that the musical contexts successfully invoked the tonal and metric hierarchies when only one dimension was varied in the probe event. Comparison of such ratings provides an important baseline test ensuring that the hierarchical structure of each of the dimension (i.e., the tonal and metric hierarchies) are equally accessible within and induced by the preceding context.

Method

Participants

The participants in this experiment were 16 adult musicians, who received either credit in an introductory psychology class or monetary compensation. The average years of training for these participants was 10.5 years ($SD = 2.8$) and the average age was 20.1 years ($SD = 3.4$).

Stimuli

The classification baselines used a small subset of the probe stimuli from the previous experiments, with one additional pitch included as well. The pitch classification trials used the seven diatonic notes of C major and added an eighth position by using the C from the next upper octave, producing a probe set in increasing pitch height of C, D, E, F, G, A, B and C'. Using this probe set, half of the pitches belong to the C major tonic triad (C, E, G and C'), whereas the remaining half are diatonic pitch classes not found in the tonic triad. All pitch classes were played on the downbeat at the start of the probe measure, resulting in eight possible probe events. For the temporal classification trials, the probe events were always the same pitch (C), and used the same eight temporal positions as in the previous experiments (half on-beat, half off-beat), also producing eight unique probes. The goodness of fit rating baselines used the same stimuli as the classification baselines, but also added the nondiatonic pitch classes to produce 12 pitch classes total.

The trials were blocked by response type (classification or goodness of fit rating) and dimension (pitch or time). All blocks employed two repetitions of the melodies from Experiments 1 and 2 (Figure 2) and Experiments 3 and 4 (Figure 5), resulting in 64 trials in each of the classification blocks and 96 trials in each of the goodness of fit rating blocks. The entire experiment took approximately one hour to complete.

Apparatus and Procedure

The apparatus and procedure were the same as in the previous studies. Listeners indicated as quickly as possible whether the probe event was on/off the beat for the temporal classification trials, and in/out of the tonic triad for the pitch classification trials. The "a" and ";" key were the two response keys, with the assignment of these keys counterbalanced across listeners. For the goodness of fit ratings, listeners used a 1 to 7 rating scale, with 1 denoting the probe fit poorly with the context whereas 7 denoting a good fit.

<div align="center">Results</div>

Classification Tasks

Temporal classification accuracy was 80% (SD = 14%, RT = 1213 ms), and pitch classification accuracy was 87% (SD = 12%, RT = 1742 ms). The correlation between accuracy and RT was negative for temporal classifications (r = -.18) and pitch classifications (r = -.08), ruling out speed-accuracy tradeoffs in both tasks. Years of training and accuracy did not correlate significantly (r = .33, p = .1). Of most interest, a one-way repeated measures ANOVA on the accuracy scores using dimension (temporal or pitch) as the within-subjects factor showed that there was no difference in accuracy between the temporal and pitch dimensions, $F(1,15)$ = 2.8, MSE = .01, p = .12, η_p^2 = .16. A signal detection analysis converted the accuracy measures into d' values (discriminability); the average temporal classification d' was 2.13 (SD = 1.3), and 2.69 for pitch classification (SD = 1.2). Replicating the accuracy scores, d' did not differ between the two dimensions, $F(1,15)$ = 2.41, MSE = 1.03, p = .14, η_p^2 = .14.

Further analyses tested if accuracy differed across temporal position and pitch class, and the respective roles of the metric and tonal hierarchies. Temporal classifications were analyzed with a one-way repeated measures ANOVA, employing temporal position as a within-subjects variable. Accuracy differed significantly across temporal position, $F(7,105)$ = 5.67, MSE = .03, p <.001, η_p^2 = .27, with this profile reflecting the metric hierarchy, r = .78, p < .05; listeners were more accurate for on-beat than off-beat temporal positions. Looking at pitch classification, there was a significant difference in accuracy across pitch class, $F(7,105)$ = 6.78, MSE = .03, p < .001, η_p^2 = .31, but this profile did not reflect the tonal

hierarchy, $r = .02$, $p = .96$. Instead, listeners tended to perform slightly worse on the tonic triad members that were not the tonic (E and G); this finding may reflect the fact that as exemplars of the tonic triad, the third and fifth scale degrees are understandably not as prototypical as the tonic.

Goodness of Fit Ratings

To test if the musical contexts employed in these studies established the tonal and metric hierarchies, the goodness of fit ratings were analyzed with two separate one-way repeated measures ANOVAs, using either temporal position or pitch class as the within-subjects variable. For the temporal trials the goodness of fit ratings varied significantly across the eight temporal positions, $F(7,105) = 11.32$, $MSE = 1.09$, $p < .001$, $\eta_p^2 = .43$; this pattern emulated the metric hierarchy, $r = .80$, $p < .05$. Likewise for the pitch stimuli there was a significant difference across pitch class in the goodness of fit ratings, $F(11,165) = 70.76$, $MSE = .58$, $p < .001$, $\eta_p^2 = .83$, with the ratings consistent with the tonal hierarchy, $r = .99$, $p < .001$.

Overall, these baseline tests reveal that, in classification tests, these musical contexts contained pitch and temporal hierarchical information that was equally discriminable. Furthermore, the goodness of fit ratings indicate that the musical contexts were similarly successful in inducing perceived tonal and metric hierarchies in listeners. As such, it seems unlikely that the asymmetry in interference between pitch and time found in the filtering tasks of the earlier studies can be explained by simple differences in discriminability between these dimensions. However, it is important to test what pattern of interference (or lack thereof) results when the filtering task uses equal numbers of pitch classes and temporal positions. Therefore, armed with these baselines, Experiment 6 tested the filtering task to determine if the asymmetric pattern of interference obtained in Experiments 3 and 4 persisted.

Experiment 6: Pitch-time dimensional relations in equalized classifications

Experiments 3 and 4 demonstrated how the tonality of a probe event following a typical musical context can influence evaluations about the probe's timing, but not vice versa. However, differences in the set size of pitch classes and temporal positions may have artificially influenced this finding. Although Experiment 5 demonstrated equal discriminability between the pitch and temporal dimensions using a set of baseline tasks, this task intentionally employed stimuli containing equal levels of pitch and temporal information (at least in the classification task). As such, it nevertheless remains theoretically possible that the earlier asymmetry observed in Experiments 3 and 4 arose due to the unequal number of levels of these stimuli, possibly through varying the degree of discriminability between these dimensions. Accordingly, a final, and conclusive, test of this possibility involves replicating the earlier asymmetry in filtering, but this time employing stimuli that are known to be equally discriminable. Providing this test was the goal of the current experiment.

Specifically, this study employed the filtering task of Experiments 3 and 4, combined with the matched stimuli of Experiment 5. In this task half of the listeners received instructions to ignore pitch while judging the timing of the probe, and half received the opposite instructions. If the response bias of pitch influencing temporal classifications consistent with the tonal hierarchy was due to unequal stimulus quantity or discriminability, then the asymmetry between pitch and time should either disappear, or there should now be an effect of time on pitch judgments. In contrast, if the same asymmetry as in Experiments 3 and 4 emerges, then it is clear that these earlier results were not due to issues related to unequal stimulus quantity or differences in discriminability.

Method

Participants

Thirty-six adult musicians participated in this experiment, either for credit in an introductory psychology class or for monetary compensation. The average years of training for these participants was 10 years ($SD = 2.02$) and the average age was 18.8 ($SD = 1.41$). Participants were randomly assigned to either judge temporal position (N = 18) or pitch class (N = 18).

Stimuli

The stimuli in this experiment consisted of all possible combinations of the pitch classes and temporal positions used in the baseline classification tasks of Experiment 5. That is, there were eight pitch classes (half tonic triad, half non-tonic triad) and eight temporal positions (half on-beat, half off-beat), producing 64 unique probe events in all. There were four blocks of trials presented with a unique melody employed in each block; these four melodies were the same as those as used in Experiment 5.

Apparatus and Procedure

The apparatus and procedure were the same as in the previous experiments. Listeners indicated as quickly as possible whether the probe event was on- or off-beat, or in or out of the tonic triad, depending on their assigned condition. Four blocks of 64 trials produced an experimental session of just less than an hour.

Results and Discussion

Overall accuracy in temporal classification was 77% (RT $M = 1281$ ms) and in pitch classification it was 78% (RT $M = 1657$ ms). Accuracy and years of formal training did not correlate significantly ($r = .14$, $p = .41$). Accuracy and RT correlated negatively ($r = -.23$, $p = .17$), ruling out a speed-accuracy tradeoff.

A four-way repeated measures ANOVA on accuracy across the within-subjects variables of pitch class, temporal position and block, and instruction as a between-subjects variable revealed main effects of pitch class, $F(7, 238) = 4.31$, $MSE = .56$, $p < .001$, $\eta_p^2 = .11$, time, $F(7, 238) = 6.54$, $MSE = .39$, $p < .001$, $\eta_p^2 = .16$, and block, $F(3, 102) = 3.3$, $MSE = .37$, $p < .05$, $\eta_p^2 = .09$. These main effects show that accuracy differed depending on the pitch class and temporal position of the probe, and also that performance improved during the course of the experiment. There was no main effect of instruction, $F(1, 34) < 1$, $MSE = 4.03$. There were two-way interactions between pitch and instruction, $F(7, 238) = 4.32$, $MSE = .56$, $p < .001$, $\eta_p^2 = .11$, time and instruction, $F(7, 238) = 4.57$, $MSE = .39$, $p < .001$, $\eta_p^2 = .12$, and time and block, $F(21, 714) = 1.81$, $MSE = .13$, $p < .05$, $\eta_p^2 = .05$. These interactions indicate that, as expected, the effect of pitch and time were larger when they were the relevant dimensions, respectively. The interaction of time and block reflected the fact that the effect of time was not as differentiated in the first block compared to the remaining blocks. Of principal interest, however, was the significant pitch by time by instruction interaction, $F(49, 1666) = 1.76$, $MSE = .13$, $p = .001$, $\eta_p^2 = .05$. Inspection of this interaction suggests that pitch and time interacted in temporal classifications, but not pitch classifications. To verify this interpretation, two subsequent three-way repeated measures ANOVAs (one for each instruction type) tested the within-subjects variables of pitch, time and block. These analyses are described separately, in turn.

For temporal classification, there were significant main effects of time, $F(7, 119) = 6.12$, $MSE = .68$, $p < .001$, $\eta_p^2 = .27$, and block, $F(3, 51) = 3.44$, $MSE = .42$, $p < .05$, $\eta_p^2 = .17$. These effects indicated improved accuracy for on-beat probes compared to off-beat probes, and later blocks compared to earlier blocks. Accuracy across temporal position

therefore followed the metric hierarchy, $r = .75$, $p < .05$. There was no main effect of pitch class, $F(7, 119) = 1.74$, $MSE = .11$, $p = .11$, $\eta_p^2 = .09$. The interaction between time and block was significant, $F(21, 357) = 1.83$, $MSE = .15$, $p < .05$, $\eta_p^2 = .1$. Most importantly, and replicating Experiment 3, there was a significant interaction between pitch and time, $F(49, 833) = 1.97$, $MSE = .14$, $p < .001$, $\eta_p^2 = .1$. No other interactions were significant.

A signal detection analysis further explored this experimentally critical interaction between pitch and time. Just as in Experiment 3, discriminability (d') did not differ across pitch class, $F(7, 119) = 1.49$, $MSE = .17$, $p = .18$, $\eta_p^2 = .08$, but response bias (c) did, $F(7, 119) = 3.47$, $MSE = .11$, $p < .01$, $\eta_p^2 = .17$. Moreover, response bias again followed the tonal hierarchy, $r = .69$, $p = .06$ (see Figure 8). The marginal significance of this correlation coefficient is likely because of the removal of an entire level of the tonal hierarchy (nondiatonic pitch classes).

Table 1 displays the correlations between accuracy as a function of pitch class and the relevant tonal hierarchy values, for each temporal position. Again replicating Experiment 3, there was a regular oscillation in these correlations, with the correlations positive for on-beat temporal positions (positions 1, 3, 5 and 7; correlation $M = .38$), and negative for off-beat positions (positions 2, 4, 6 and 8; correlation $M = -.54$). Table 2 shows the correlations between the accuracy at each pitch class and the metric hierarchy. Also like Experiment 3, correlations with the metric hierarchy were higher for tonally stable pitches (correlation $M = .40$ and -.43 for the tonic triad and non-tonic triad tones, respectively). In short, the results of the temporal classifications replicate the congruity effect of Experiment 3.

For pitch classifications there was a main effect of pitch class, $F(7, 119) = 4.58$, $MSE = 1.02$, $p < .001$, $\eta_p^2 = .21$. However, this effect only weakly followed the tonal hierarchy ($r = .4$, $p = .16$); instead, this effect is explainable in terms of increased accuracy for probes with the tonic pitch (C), relative to the remaining pitch classes. No other effects or interactions were significant. Most importantly, there was no pitch by time interaction, $F(49, 833) < 1$, $MSE = .12$. Similar to Experiment 4, the signal detection analysis failed to reveal any differences in discriminability, $F(7, 119) = 1.69$, $MSE = .14$, $p = .12$, $\eta_p^2 = .09$, or response bias, $F(7, 119) < 1$, $MSE = .05$, as a function of temporal position. Table 1 shows that the correlations with the tonal hierarchy did not vary across temporal position, and instead were uniformly positive (correlation $M = .35$ and .41 for on- and off-beat temporal positions, respectively). Likewise, Table 2 shows that the correlations with the metric hierarchy across pitch class were also positive (correlation $M = .16$ and .29 for the tonic triad and non-tonic triad tones, respectively).

Overall, these results replicate Experiments 3 and 4, in that a response bias driven by the tonal hierarchy influenced listeners' temporal judgments, but there was no corresponding metric hierarchy bias affecting pitch judgments. Once again, temporal classification was characterized by a congruity effect in terms the impact of pitch on time. Finally, there is convergence between the current study and Experiments 3 and 4 in that discriminability itself did not differ across pitch class or temporal position, but rather, the difference across pitch class in temporal classification was specific to response bias. This finding reinforces the idea that a higher-level decisional bias rather than a perceptual difficulty underlies the observed effect of pitch on time.

These results also address the previously discussed limitations of Experiments 3 and 4. Specifically, this study employed equal numbers of pitch classes and temporal positions (all of which were in the context prior to the probe), and still produced the same asymmetries as observed in Experiments 3 and 4. Therefore, neither differences in stimulus

quantity nor the presence or absence of probe pitches in the context explain the continuing observed asymmetry of pitch affecting time and not vice versa. Further, accuracy between the temporal and pitch instructions, as well as the overall average d' values for the two instructions, did not differ. Accordingly, differences in stimulus discriminability cannot explain the persistent asymmetric effects of pitch on time in these musical contexts.

General Discussion

To summarize the main results, six experiments explored the relation between musical pitch and time in perceiving a musical passage typical of Western music. Experiments 1 and 2 collected ratings of goodness of fit for probe events that combined different pitch classes and temporal positions, whereas Experiments 3 and 4 used a speeded classification of either pitch class or temporal position, accompanied by variation along the other (irrelevant) dimension. Experiments 5 and 6 ruled out several alternative explanations of the results of Experiments 3 and 4, providing baseline measures and replicating the primary findings of these studies using stimuli that controlled for an array of extraneous factors. The effects of dimensional structure (tonality and meter) on performance were evident in all experiments, but there were important variations in how these two dimensional structures combined. Specifically, tonality and meter combined additively in goodness of fit ratings (Experiment 1), even when participants' instructions were to ignore pitch (Experiment 2). In a metric speeded classification task (Experiments 3 and 6), the tonal hierarchy biased responses, producing a congruity effect between the pitch and time dimensions, with this effect attributable to differential response bias but not discriminability. In contrast, in a pitch speeded classification task (Experiments 4 and 6), the metric hierarchy failed to influence responses, with no corresponding impact on either response bias or discriminability. Table 3 summarizes the classification judgment results to illustrate the asymmetrical effects of pitch on time.

Before discussing the implications of these findings, there are some alternative explanations for and limitations to these experiments that merit consideration. Perhaps the most obvious issue derives from the structure of the melodic context provided prior to the probe event. In Western music, the tonal and metric hierarchies show a strong positive correlation, with metrically stable temporal positions typically containing tonally stable pitches; the context melodies used here are no exception. This correlation makes it difficult to ascertain the degree of contribution of the tonal and metric hierarchies; however in tests of rhythm perception that put pitch and temporal factors in opposition, Dawe, Platt and Racine (1993, 1994, 1995) found independent contributions of both, and even a dominance of pitch. But to what extent do the current findings depend on the specific context used? Perhaps the rich pitch structure in these musical contexts magnified the relative importance of pitch compared to time. It is possible that other contexts with varied strength of pitch and temporal structure might elicit corresponding differences in the perception of subsequent pitch-time probe events. Consequently, these results cannot currently generalize to all forms of Western music.

While considering this limitation, it is important to remember that the statistical properties of a standard Western tonal context do not exclusively determine encultured listeners' subsequent perceptions. As already described, not any hierarchical ordering of pitch classes will result in a corresponding perceived tonal hierarchy, but only the one to which listeners have been exposed throughout life (see N. A. Smith & Schmuckler, 2004). This finding suggests that there may be a stored representation of the statistical properties of music that a suitable context activates in an all-or-none fashion, similar to a retrieval cue from memory. If so, any musical context that successfully activates tonal and metric

hierarchies should produce the pattern of findings reported here. Nevertheless, testing relative dimensional dominance while systematically varying the pitch and temporal structure in musical contexts is a fruitful avenue for further research.

Related to this concern is that perhaps as a dimension, time did not get a fair chance – that a larger effect of time requires presenting more levels of the temporal hierarchy. Such an argument implies that one needs to present a wider range of temporal events in the preceding musical contexts to activate the metric hierarchy in listeners' minds. However, this logic suggests that subdivisions of the metric hierarchy do not occur without explicit presentation of events at those temporal locations in preceding contexts. Moreover, and by extension, this argument also suggests a comparable effect for pitch, with the tonal hierarchy not instantiated unless all pitches are presented in a tonal melody. Ultimately, this argument is unsupportable based on the wealth of data throughout the music cognition literature that demonstrates relatively impoverished musical contexts, such as just a few notes, a simple chord, or short rhythmic patterns can activate both tonal (Cuddy & Badertscher, 1987; Krumhansl, 1990; Oram & Cuddy, 1995; N. A. Smith & Schmuckler, 2004) and metric hierarchies (Brochard, Abecasis, Potter, Ragot, & Drake, 2003; Desain & Honing, 2003; Large & Palmer, 2002; Palmer & Krumhansl, 1990; Povel & Okkerman, 1981; Stoffer, 1985; Tekman, 2001). For example, Palmer and Krumhansl (1990) provided only one level of the metric hierarchy (the tactus, or beat level) and yet still observed nuanced metric hierarchies of matching temporal locations. In addition, and more directly to the point, the baseline goodness of fit ratings of Experiment 5 demonstrated that these contexts were sufficient to induce strong tonal and metrical hierarchies in listeners. Thus, it strains credibility to assume that the perception of the metric hierarchy was somehow weaker in these studies than the perception of the tonal hierarchy.

Another qualification to these findings is the nature of the task. Presenting a tonal musical context followed by a pitched probe event means that these tasks were strongly pitch-based. Other tasks that used different methodologies may observe alternate patterns of dimensional salience[4]. For example, research paradigms in which participants tap in synchrony with a rhythmic sequence (and thus are inherently time-based) can show dominance of temporal factors over pitch (Pfordresher, 2003; Snyder & Krumhansl, 2001; Vos, Vandijk, & Schomaker, 1993). Perhaps, therefore, a given dimension is more likely to interfere with another dimension if the task focuses attention towards it. It is difficult to quantitatively asses the extent to which a task design favors one dimension over another, therefore this question is also an important area for future research on dimensional relations.

Despite these limitations, the current studies present intriguing results and implications. The congruity effect in Experiment 3 is one such finding, in which responses were more accurate to tonally stable events when they appeared in metrically stable positions. Thus, not only did pitch affect temporal judgments, but the hierarchical structure (tonality) of the irrelevant dimension governed the nature of this interference. This finding converges with a result reported by Lebrun-Guillaud and Tillman (2007), in which the tonal function of a note influenced the ability to detect a change in the temporal regularity of a sequence. In this study, listeners heard three chords followed by three notes, all in isochronous rhythm, and were asked to detect a temporal deviation on the fifth event. These researchers observed that performance improved with increasing tonal stability of the target note, an effect congruent with the current results. One notable distinction between these findings and Lebrun-Guillaud and Tillman (2007), however, is that the latter showed an effect in both discriminability (d') and response bias (c) measures, whereas the current findings only showed a difference in response bias. Accordingly, these experiments indicate a

later, decisional origin of pitch interference, conversely the results of Lebrun-Guillaud and Tillman (2007) suggest pitch interfered at both a perceptual and decisional level for their task.

The fact that congruity between the hierarchical organizations of stimulus dimensions can affect perception is a dramatic result, and has implications beyond the local domain of music. Take, for instance, the hierarchical organization in focal colors and color-neutral prototypical words. When combined in a single stimulus, one might reasonably expect that a more focal color would facilitate processing of a prototypical word. The current findings, however, suggest an additional counterintuitive result, which is that a less focal color would actually facilitate the processing of a less prototypical word. If true, such a result would indicate that the hierarchical structure of colors and words interact in a way that congruity between them influences the perception of a stimulus.

A main theme in these tests of how pitch and time combine is that when judging a pitch-time probe event following a typical tonal sequence, the pitch dimension dominated listeners' responses. The baseline tasks of Experiment 5, coupled with the filtering task of Experiment 6, ruled out both differences in stimulus quantity or discriminability as well as presence in the preceding context, as underlying these results. Indeed, when asking how pitch and time combine in the perception of these musical contexts, these data suggest a simple dominance of pitch rather than an increased discriminability.

Although the discussion has thus far focused on Experiments 3-6, considering the interference of pitch on time as greater in the classification judgments than the goodness of fit ratings is not the most accurate characterization of these data. In both classification judgments and goodness of fit ratings, the tonal hierarchy exerted a robust interfering effect on responses, but in different forms. In the rating data the tonal hierarchy affected judgments by influencing ratings of temporal goodness of fit in accordance with the probe's tonal stability. In the classification data the tonal hierarchy affected judgments of metric classification by biasing listeners to respond in accordance with the probe's tonal stability.

Why did the congruity effect occur only in classification judgments? In classification judgments, tonal-metric congruity (high tonal and metric stability, or low tonal and metric stability) increased accuracy. When classifying the metric stability of a probe that was off the beat, nondiatonic pitches aided accuracy likely because they reinforced and facilitated the judgment. However, for goodness of fit ratings, a congruity effect would be highly counterintuitive. Consider a probe that has both low tonal and metric stability (congruent). A congruity effect would require this probe to receive a higher rating than a probe with incongruent stability (high tonal and low metric, or low tonal and high metric), even though the incongruent probe was more stable.

One could similarly ask why tonally stable pitches were not always more accurate than unstable pitches in the temporal classification judgments; this question invokes the corollaries of the above explanations. When classifying metric stability, tonally stable pitches increase accuracy for on-beat probes, but tonally unstable pitches help for off-beat probes (the congruity effect). Thus averaging across temporal position eliminates this pattern such that accuracy across tonal stability remains fairly constant; at the least, any differences across pitch class are unlikely to correspond to tonal stability. However, when rating temporal goodness of fit, tonally stable pitches always received a higher rating regardless of the temporal position, whereas tonally unstable pitches had the opposite effect. This consistent effect of tonal stability caused the tonal hierarchy to be more apparent when averaging across temporal position for the goodness of fit ratings instead of classification accuracy.

Ultimately, therefore, the nature of the task instructions determined whether the interference of pitch on time (as the tonal hierarchy) appeared as a congruity effect for the classification judgments, or as an additive relationship for the goodness of fit ratings. Yet in both cases the tonal hierarchy strongly influenced responses, thus it is misleading to suggest that the interference of pitch on time was stronger in one task than another.

These findings require an account of why pitch predominated in these studies. One explanation for these results is that pitch information was simply more salient to listeners in these typical musical contexts, such that listeners could not ignore pitch even when it was irrelevant to the task. Although the stimulus inequality between pitch and time may have played a role in the results of Experiment 1, the findings of this study support the idea that without any explicit attentional instructions listeners' ratings demonstrated a clear dominance of pitch events. Accordingly, pitch was clearly the more salient dimension for listeners in these studies, a finding that aligns with previous research on long-term memory for music (Hébert & Peretz, 1997) and perception of metric structure (Dawe et al., 1994).

Of course, an explanation of these results based on pitch dominance begs the question of why pitch was more salient in these experiments. One possibility is that listeners' greater focus on pitch information might be inevitable given their general experience with and exposure to Western music. In most genres of Western tonal music, the complexity of the pitch structures dwarfs the complexity of the temporal structures. For example, Western music is overwhelmingly in binary or ternary meters, and tends to involve only a handful of different notated duration values; in contrast, the pitch structures in Western music use many different pitch sets in harmony and melody. Indeed, in Western music in general, the possible unique combinations of pitch are greater than those of duration. There tends to be comparatively little variability in duration, often using only two or three duration values (e.g., quarter and eighth notes). In contrast, the same music contains considerably more variability in pitch, with most of the seven pitches in the scale typically represented. But more importantly, the possible set of pitches that can define a tonal context is larger than the set of durations that (when organized properly) can create a metric framework. Therefore any set of pitches will tend to have a lower probability of occurrence than any set of durations. Events that have a low probability of occurrence should require greater cognitive resources to process than events that have a high probability of occurrence. Put differently, events with a high probability of occurrence should be more predictable and therefore easier to process. Over years of exposure, the habit of automatically allocating more mental effort to pitch than temporal structures means that pitch becomes more salient in typical Western music.

This idea has a precedent in research on picture-word interference. In this domain, words of low frequency are more distracting than higher frequency words when a participant is naming a picture (Miozzo & Caramazza, 2003). Because lower frequency words are more distinctive, the process of inhibiting their lexical entry takes longer, thus affecting picture naming more than higher frequency words. In a similar way, the fact that a large number of pitch sequences can successfully induce a tonal framework means that the pitch dimension of a melody attracts more attention than the temporal dimension. Therefore, it may be difficult to ignore variations along this dimension, even when actively trying to do so. If true, this argument predicts a reverse of this asymmetric interference for listeners principally familiar with musical systems that develop significantly more complex temporal than pitch patterns (e.g., Australian Aboriginal, African, South Asian or Eastern European cultures). Certainly, listeners from such musical cultures detect violations to musical passages using

complex temporal structures better than Western listeners (Hannon & Trehub, 2005), and as a result might be less able to ignore variation along the temporal dimension.

A second possible account for these findings arises from previous demonstrations of asymmetric interference in the absence of discriminability inequalities. Indeed, there are precedents for similar effects in both speech (Mullennix & Pisoni, 1990; Tong, Francis, & Gandour, 2008) and face perception (Atkinson, Tipples, Burt, & Young, 2005; Haxby, Hoffman, & Gobbini, 2000; Schweinberger, Burton, & Kelly, 1999). With respect to face perception, Haxby, Hoffman and Gobbini (2000) offer an explanation for asymmetric interactions with equal discriminability by delineating between invariant and changeable information. These authors propose that facial identity is invariant information, whereas the expression a faces uses is changeable. Furthermore, invariant characteristics will display asymmetric interference on changeable dimensions because when processing a facial stimulus, an invariant characteristic (facial identity) acts as a better referent than the changeable one. Atkinson et al. (2005) adapt this model by moving the source of interference from the perceptual stage to a later, decisional stage (in keeping with the current findings). These authors also allow for cultural learning, such as female faces associated with greater expressive range in Western society, an aspect similar to the culturally learned favoring of pitch in typical Western music. Interestingly, in their model of face perception, Haxby et al. (2000) also specify distinct neurobiological pathways for the processing of facial identity and expression; neuropsychological work on music perception offers a parallel proposal for pitch and time (Peretz & Coltheart, 2003; Peretz & Kolinsky, 1993; Peretz et al., 1994).

Although it is unlikely that pitch ubiquitously acts as an invariant characteristic, when the task is strongly pitch-based, pitch may in fact function as a better perceptual reference point than other dimensions (e.g., timing, timbre, etc.), thus acting as an invariant attribute. In these tasks, the pitch defined the presence of a musical event regardless of exactly when it occurred, whereas any given temporal position was not by itself an event unless a pitch occurred then. Therefore the pitch of a probe event may have provided a better reference point than its temporal position; as a result participants had difficulty ignoring pitch even when it was irrelevant to the task. Consequently, pitch would interfere in the temporal task while time would not interfere with the pitch task. Conversely, in a tapping task (e.g., Snyder & Krumhansl, 2001), the temporal position (the timing of the tap) is the main attribute of interest, whereas a pitch need not be present. In this case, the timing may function as the better referent and reverse the pattern of pitch-time dominance observed here. This concept resembles and extends the concept of physical primacy (Garner, 1974) discussed in the introduction.

Unfortunately, these data cannot unambiguously determine which of these explanations best accounts for the increased salience of pitch in the current experiments; moreover, these explanations are not necessarily mutually exclusive. Regardless of whether this increased salience results from a learned cultural pitch bias, or is in fact a more fundamental characteristic of musical events, these findings do reveal an important distinction between dimensional discriminability and salience. That is, the more salient dimension in a stimulus may or may not be more discriminable than a different dimension. As a result, asymmetric dimensional interactions may result from inequalities in discriminability (Garner, 1974; Garner & Felfoldy, 1970; Melara & Algom, 2003; Melara & Mounts, 1993; Sabri et al., 2001), or inequalities in salience. Therefore, investigations aimed at definitively establishing dimensional interaction versus independence would thus require ensuring not only that the dimensions are equally discriminable, but also that one of the two

dimensions does not dominate the other dimension in terms of its salience or importance to perceivers. Normally, one would expect a correlation between salience and discriminability, such that the more discriminable dimension is also more salient to perceivers. That is, our perceptual systems tend to maximize processing of the most biologically relevant stimuli (e.g., faces, motion, voices); this emphasis may serve to increase the salience of the dimensions that comprise these stimuli. Occasionally, however, the most salient dimension is not always the easiest to process because it is noisy, degraded or masked. In this case the more salient dimension is not the most discriminable. These findings demonstrate that under the proper conditions, dimensional discriminability and salience can be dissociable.

The current studies have provided insight into the nature of the relation between the processing of musical pitch and time, as well as how the hierarchical organization of these dimensions can affect their perceptual integration. They have also drawn a distinction between discriminability and salience. Hopefully, such findings will point the way towards fruitful investigations not only with the domain of musical perception and cognition, but also for the understanding of dimensional processing within other visual and auditory dimensions as well.

References

Abe, J. I., & Okada, A. (2004). Integration of metrical and tonal organization in melody perception. *Japanese Psychological Research, 46*(4), 298-307.

Aldwell, E., & Schacter, C. (2002). *Harmony and voice leading*: Thompson.

Amazeen, E. L., & Turvey, M. T. (1996). Weight perception and the haptic size weight illusion are functions of the inertia tensor. *Journal of Experimental Psychology: Human Perception and Performance, 22*(1), 213-232.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*(2), 154-179.

Atkinson, A. P., Tipples, J., Burt, D. M., & Young, A. W. (2005). Asymmetric interference between sex and emotion in face perception. *Perception and Psychophysics, 67*(7), 1199-1213.

Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception and Psychophysics, 57*(8), 1151-1162.

Ben-Artzi, E., & Marks, L. E. (1999). Processing linguistic and perceptual dimensions of speech: Interactions in speeded classification. *Journal of Experimental Psychology: Human Perception and Performance, 25*(3), 579-595.

Boltz, M. G. (1989a). Perceiving the end: Effects of tonal relationships on melodic completion. *Journal of Experimental Psychology: Human Perception and Performance, 15*(4), 749-761.

Boltz, M. G. (1989b). Rhythm and good endings - effects of temporal structure on tonality judgments. *Perception and Psychophysics, 46*(1), 9-17.

Boltz, M. G. (1989c). Time judgments of musical endings - effects of expectancies on the filled interval effect. *Perception and Psychophysics, 46*(5), 409-418.

Boltz, M. G. (1991). Recognition memory for auditory event duration. *Bulletin of the Psychonomic Society, 29*(6), 517.

Boltz, M. G. (1992). Time-estimation and expectancies. *Bulletin of the Psychonomic Society, 30*(6), 484.

Boltz, M. G. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception and Psychophysics, 53*(6), 585-600.

Boltz, M. G. (1995). Effects of event structure on retrospective duration judgments. *Perception and Psychophysics, 57*(7), 1080-1096.

Boltz, M. G. (1998a). The processing of temporal and nontemporal information in the remembering of event durations and musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 24*(4), 1087-1104.

Boltz, M. G. (1998b). Task predictability and remembered duration. *Perception and Psychophysics, 60*(5), 768-784.

Boltz, M. G. (1999). The processing of melodic and temporal information: Independent or unified dimensions? *Journal of New Music Research, 28*(1), 67-79.

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The "Ticktock" Of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science, 14*(4), 362-366.

Crowder, R. G., & Neath, I. (1995). The influence of pitch on time perception in short melodies. *Music Perception, 12*(4), 379-386.

Cuddy, L. L., & Badertscher, B. (1987). Recovery of the tonal hierarchy - some comparisons across age and levels of musical experience. *Perception and Psychophysics, 41*(6), 609-620.

Dawe, L. A., Platt, J. R., & Racine, R. J. (1993). Harmonic accents in inference of metrical structure and perception of rhythm patterns. *Perception and Psychophysics, 54*(6), 794-807.

Dawe, L. A., Platt, J. R., & Racine, R. J. (1994). Inference of metrical structure from perception of iterative pulses within time spans defined by chord changes. *Music Perception, 12*(1), 57-76.

Dawe, L. A., Platt, J. R., & Racine, R. J. (1995). Rhythm perception and differences in accent weights for musicians and nonmusicians. *Perception and Psychophysics, 57*(6), 905-914.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception, 32*(3), 341-365.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception and Psychophysics, 28*(5), 381-389.

Dyson, B. J., & Quinlan, P. T. (2002). Within- and between-dimensional processing in the auditory modality. *Journal of Experimental Psychology: Human Perception and Performance, 28*(6), 1483-1498.

Fries, W., & Swihart, A. A. (1990). Disturbance of rhythm sense following right-hemisphere damage. *Neuropsychologia, 28*(12), 1317-1323.

Garner, W. R. (1974). *The processing of information and structure* (Vol. 203). Oxford, England: Lawrence Erlbaum.

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology, 8*(1), 98-123.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*(3), 225-241.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Griffiths, T. D., Johnsrude, I., Dean, J. L., & Green, G. G. R. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neuroreport, 10*(18), 3825-3830.

Handel, S., & Imai, S. (1972). Free classification of analyzable and unanalyzable stimuli. *Perception and Psychophysics, 12*(1B), 108-116.

Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science, 16*(1), 48-55.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223-233.

Hébert, S., & Peretz, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory and Cognition, 25*(4), 518-533.

Jones, M. R. (1976). Time, our lost dimension - toward a new theory of perception, attention, and memory. *Psychological Review, 83*(5), 323-355.

Jones, M. R. (1987). Dynamic pattern structure in music - recent theory and research. *Perception and Psychophysics, 41*(6), 621-634.

Jones, M. R., & Boltz, M. G. (1989). Dynamic attending and responses to time. *Psychological Review, 96*(3), 459-491.

Jones, M. R., Boltz, M. G., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Perception and Psychophysics, 32*(3), 211-218.

Jones, M. R., Johnston, H. M., & Puente, J. (2006). Effects of auditory pattern structure on anticipatory and reactive attending. *Cognitive Psychology, 53*(1), 59-96.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13*(4), 313-319.

Jones, M. R., & Ralston, J. T. (1991). Some influences of accent structure on melody recognition. *Memory and Cognition, 19*(1), 8-20.

Jones, M. R., Summerell, L., & Marshburn, E. (1987). Recognizing melodies - a dynamic interpretation. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 39*(1), 89-121.

Kelley, Z. A., & Brandt, J. F. (1984). Pitch change recognition as a function of duration in successive dichotic stimuli. *Psychology of Music, 12*(1), 43-59.

Kidd, G., Boltz, M. G., & Jones, M. R. (1984). Some effects of rhythmic context on melody recognition. *American Journal of Psychology, 97*(2), 153-173.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* Oxford University Press, New York, NY: US.

Krumhansl, C. L. (1991). Memory for musical surface. *Memory and Cognition, 19*(4), 401-411.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 739-751.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*(4), 334-368.

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science, 26*(1), 1-37.

Lebrun-Guillaud, G., & Tillmann, B. (2007). Influence of a tone's tonal function on temporal change detection. *Perception and Psychophysics, 69*(8), 1450-1459.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, Massachusetts: MIT Press.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* New York: Cambridge University Press.

Maddox, W. T. (2001). Separating perceptual processes from decisional processes in identification and categorization. *Perception and Psychophysics, 63*(7), 1183-1200.

Maddox, W. T., & Dodd, J. L. (2003). Separating perceptual and decisional attention processes in the identification and categorization of integral-dimension stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(3), 467-480.

Makris, I., & Mullet, E. (2003). Judging the pleasantness of contour-rhythm-pitch-timbre musical combinations. *American Journal of Psychology, 116*(4), 581-611.

Marks, L. E. (1987). On cross-modal similarity - auditory visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance, 13*(3), 384-394.

Mavlov, L. (1980). Amusia due to rhythm agnosia in a musician with left-hemisphere damage - a non-auditory supramodal defect. *Cortex, 16*(2), 331-338.

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance, 15*(1), 69-79.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110*(3), 422-471.

Melara, R. D., & Marks, L. E. (1990a). Dimensional interactions in language processing - investigating directions and levels of crosstalk. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 539-554.

Melara, R. D., & Marks, L. E. (1990b). Perceptual primacy of dimensions - support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance, 16*(2), 398-414.

Melara, R. D., & Marks, L. E. (1990c). Processes underlying dimensional interactions - correspondences between linguistic and nonlinguistic dimensions. *Memory and Cognition, 18*(5), 477-495.

Melara, R. D., Marks, L. E., & Lesko, K. E. (1992). Optional processes in similarity judgments. *Perception and Psychophysics, 51*(2), 123-133.

Melara, R. D., & Mounts, J. R. W. (1993). Selective attention to Stroop dimensions - effects of base-line discriminability, response-mode, and practice. *Memory and Cognition, 21*(5), 627-645.

Melara, R. D., & Mounts, J. R. W. (1994). Contextual influences on interactive processing - effects of discriminability, quantity, and uncertainty. *Perception and Psychophysics, 56*(1), 73-90.

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General, 116*(4), 323-336.

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General, 132*(2), 228-252.

Monahan, C. B., & Carterette, E. C. (1985). Pitch and duration as determinants of musical space. *Music Perception, 3*(1), 1-32.

Monahan, C. B., Kendall, R. A., & Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception and Psychophysics, 41*(6), 576-600.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech-perception. *Perception and Psychophysics, 47*(4), 379-390.

Nittono, H., Bito, T., Hayashi, M., Sakata, S., & Hori, T. (2000). Event-related potentials elicited by wrong terminal notes: Effects of temporal disruption. *Biological Psychology, 52*(1), 1-16.

Oram, N., & Cuddy, L. L. (1995). Responsiveness of western adults to pitch-distributional information in melodic sequences. *Psychological Research/Psychologische Forschung, 57*(2), 103.

Palmer, C., & Krumhansl, C. L. (1987a). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 116-126.

Palmer, C., & Krumhansl, C. L. (1987b). Pitch and temporal contributions to musical phrase perception - effects of harmony, performance timing, and familiarity. *Perception and Psychophysics, 41*(6), 505-518.

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 728-741.

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

Peretz, I. (1996). Can we lose memory for music? A case of music agnosia in a nonmusician. *Journal of Cognitive Neuroscience, 8*(6), 481-496.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*(7), 688-691.

Peretz, I., & Kolinsky, R. (1993). Boundaries of separability between melody and rhythm in music discrimination - a neuropsychological perspective. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology, 46*(2), 301-325.

Peretz, I., Kolinsky, R., Tramo, M. J., Labrecque, R., Hublet, C., Demeurisse, G., et al. (1994). Functional dissociations following bilateral lesions of auditory-cortex. *Brain, 117*, 1283-1301.

Pfordresher, P. Q. (2003). The role of melodic and rhythmic accents in musical structure. *Music Perception, 20*(4), 431-464.

Pitt, M. A., & Monahan, C. B. (1987). The perceived similarity of auditory polyrhythms. *Perception and Psychophysics, 41*(6), 534-546.

Pomerantz, J. R. (1983). Global and local precedence - selective attention in form and motion perception. *Journal of Experimental Psychology: General, 112*(4), 516-540.

Pomerantz, J. R., & Garner, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception and Psychophysics, 14*(3), 565-569.

Potts, B. C., Melara, R. D., & Marks, L. E. (1998). Circle size and diameter tilt: A new look at integrality and separability. *Perception and Psychophysics, 60*(1), 101-112.

Povel, D. J., & Okkerman, H. (1981). Accents in equitone sequences. *Perception and Psychophysics, 30*(6), 565-572.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology, 4*(3), 328-350.

Rosch, E. H. (1975). Cognitive reference points. *Cognitive Psychology, 7*(4), 532-547.

Rosch, E. H., & Mervis, C. B. (1975). Family resemblances - studies in internal structure of categories. *Cognitive Psychology, 7*(4), 573-605.

Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyesbraem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382-439.

Sabri, M., Melara, R. D., & Algom, D. (2001). A confluence of contexts: Asymmetric versus global failures of selective attention to Stroop dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 27*(3), 515-537.

Schellenberg, E. G., Krysciak, A. M., & Campbell, R. J. (2000). Perceiving emotion in melody: Interactive effects of pitch and rhythm. *Music Perception, 18*(2), 155-171.

Schmuckler, M. A. (1989). Expectation in music - investigation of melodic and harmonic processes. *Music Perception, 7*(2), 109-150.

Schmuckler, M. A., & Boltz, M. G. (1994). Harmonic and rhythmic influences on musical expectancy. *Perception and Psychophysics, 56*(3), 313-325.

Schön, D., & Besson, M. (2002). Processing pitch and duration in music reading: A RT-ERP study. *Neuropsychologia, 40*(7), 868-878.

Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. *Perception and Psychophysics, 61*(6), 1102-1115.

Smith, J. D., & Baron, J. (1981). Individual differences in the classification of stimuli by dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 7*(5), 1132-1145.

Smith, K. C., & Cuddy, L. L. (1989). Effects of metric and harmonic rhythm on the detection of pitch alterations in melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 457-471.

Smith, L. B., & Kemler, D. G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology, 24*(2), 279-298.

Smith, L. B., & Kemler, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cognitive Psychology, 10*(4), 502-532.

Smith, N. A., & Schmuckler, M. A. (2004). The perception of tonal structure through the differentiation and organization of pitches. *Journal of Experimental Psychology: Human Perception and Performance, 30*(2), 268-286.

Snyder, J. S., & Krumhansl, C. L. (2001). Tapping to ragtime: Cues to pulse finding. *Music Perception, 18*(4), 455-489.

Stoffer, T. H. (1985). Representation of phrase structure in the perception of music. *Music Perception, 3*(2), 191-220.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662.

Tekman, H. G. (2001). Accenting and detection of timing variations in tone sequences: Different kinds of accents have different effects. *Perception and Psychophysics, 63*(3), 514-523.

Tekman, H. G. (2002). Perceptual integration of timing and intensity variations in the perception of musical accents. *Journal of General Psychology, 129*(2), 181-191.

Thomas, R. D. (2001). Perceptual interactions of facial dimensions in speeded classification and identification. *Perception and Psychophysics, 63*(4), 625-650.

Thompson, W. F. (1993). Modeling perceived relationships between melody, harmony, and key. *Perception and Psychophysics, 53*(1), 13-24.

Thompson, W. F. (1994). Sensitivity to combinations of musical parameters: Pitch with duration, and pitch pattern with durational pattern. *Perception and Psychophysics, 56*(3), 363-374.

Thompson, W. F., Hall, M. D., & Pressing, J. (2001). Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 128-140.

Tillmann, B., & Lebrun-Guillaud, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgments of chord sequences. *Psychological Research, 70*(5), 345-358.

Tong, Y. X., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes, 23*(5), 689-708.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327-352.

Vos, P. G., Vandijk, A., & Schomaker, L. (1993). Melodic cues for meter. *Perception, 23*, 965-976.

Wood, C. C. (1974). Parallel processing of auditory and phonetic information in speech discrimination. *Perception and Psychophysics, 15*(3), 501-508.

Author Notes

Jon B. Prince, Department of Psychology, University of Toronto Mississauga; William F. Thompson, Department of Psychology, Macquarie University; Mark A. Schmuckler, University of Toronto Scarborough.

Notes

[1] Correlations with the metric hierarchy were much more variable (M = .17, SD = .5), thus an exclusion criterion based on them was superfluous.

[2] Experiments 5 and 6 addressed this issue more explicitly and in greater detail.

[3] In this and the next experiment there initially appeared to be two groups of participants based on accuracy. Therefore, there were twice as many participants to obtain sufficient power in each group. However, this dichotomy failed to reach significance as a bimodal distribution (Hartigan dip test p > .05), and the results were qualitatively the same between groups. Therefore all participants were analyzed as one group.

[4] We thank an anonymous reviewer for suggesting this idea.

Table 1

Correlations of accuracy with the tonal hierarchy at each temporal position

Temporal Position	Metric Stability	Experiment 3	Experiment 4	Experiment 6	
				Temporal Condition	Pitch Condition
1	High	0.68	0.77	0.75	0.27
2	Low	-0.52	0.87	-0.79	0.42
3	Med	0.76	0.57	0.46	0.43
4	Low	-0.68	0.58	-0.34	0.45
5	High	0.28	0.45	0.48	0.26
6	Low	-0.27	0.44	-0.59	0.35
7	Med	0.18	0.46	-0.12	0.45
8	Low	-0.45	0.47	-0.45	0.42

Note: Positive correlations indicate agreement with the tonal hierarchy

Table 2
Correlations of accuracy with the metric hierarchy at each pitch class

Pitch Class	Tonal Stability	Experiment 3	Experiment 4	Experiment 6	
				Temporal Condition	Pitch Condition
C	High	0.80	0.36	0.56	0.32
C#	Low	0.62	0.37	-	-
D	Med	0.43	-0.59	0.10	0.22
D#	Low	0.68	0.36	-	-
E	High	0.90	0.10	0.07	0.07
F	Med	0.68	-0.58	-0.52	0.35
F#	Low	0.73	0.16	-	-
G	High	0.92	0.08	0.36	0.28
G#	Low	0.79	0.06	-	-
A	Med	0.71	0.32	-0.48	0.08
A#	Low	0.44	-0.34	-	-
B	Med	0.55	-0.28	-0.82	0.51
C'	High	-	-	0.59	-0.04

Note: Positive correlations indicate agreement with the metric hierarchy

Table 3
Summary of experimental findings on classification judgments (Experiments 3-6)

Experiment	Main Effect of Pitch Class on Accuracy	Main Effect of Temporal Position on Accuracy	Response Bias From Irrelevant Dimension
3 Classify Metric Stability	Yes	No	Yes, Follows Tonal Hierarchy
4 Classify Tonal Stability	Yes, Follows Tonal Hierarchy	No	No
6 Classify Metric Stability	No	Yes, Follows Metric Hierarchy	Yes, Follows Tonal Hierarchy
6 Classify Tonal Stability	Yes, Weakly Follows Tonal Hierarchy	No	No

Figure 1

Tonal and metric hierarchies (Krumhansl & Kessler, 1982; Palmer & Krumhansl, 1990)
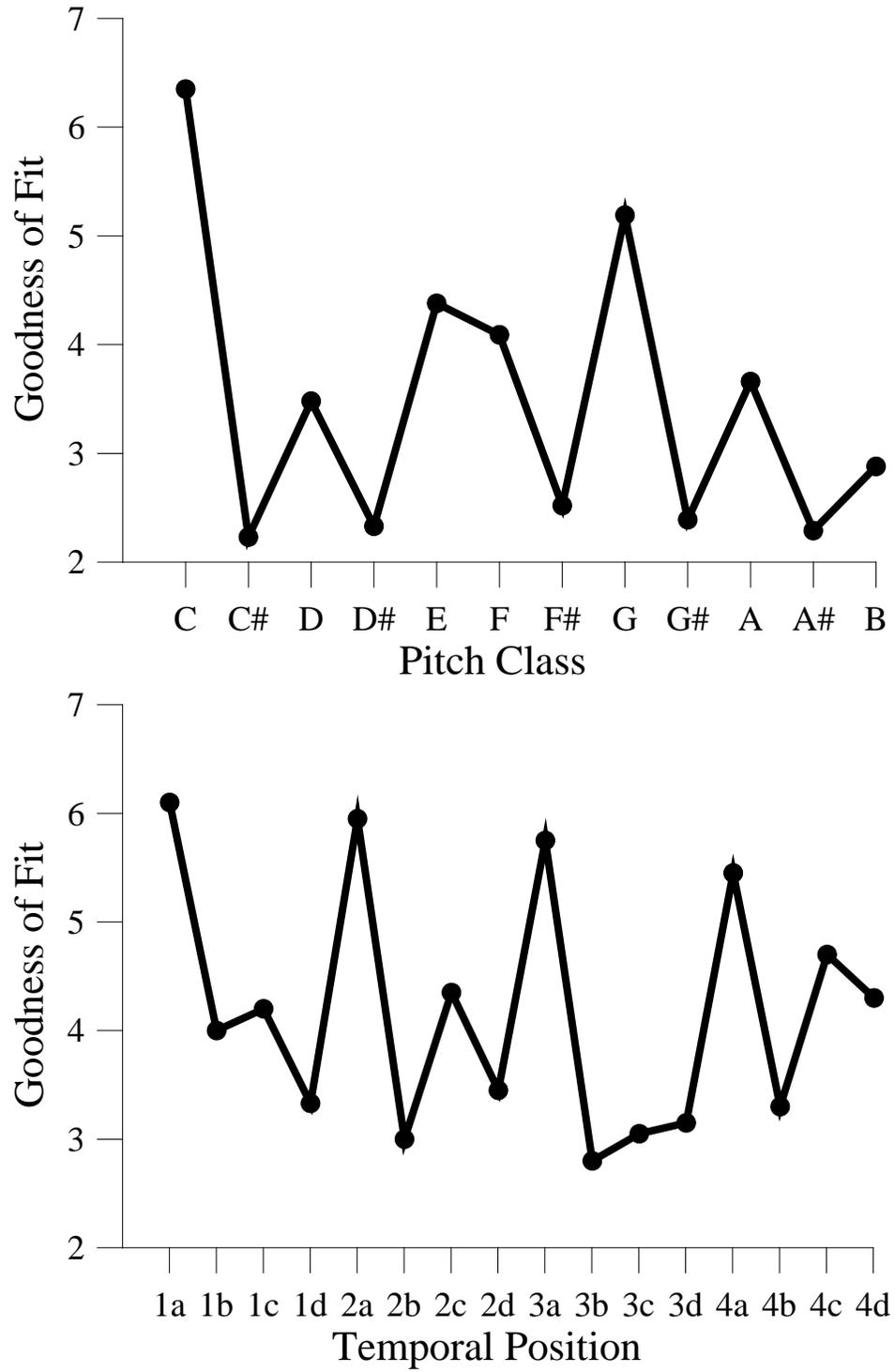
Figure 2
Musical context and example probe event rated for goodness of fit in Experiments 1 and 2

Figure 3
Interaction of the size of the pitch effect and experiment
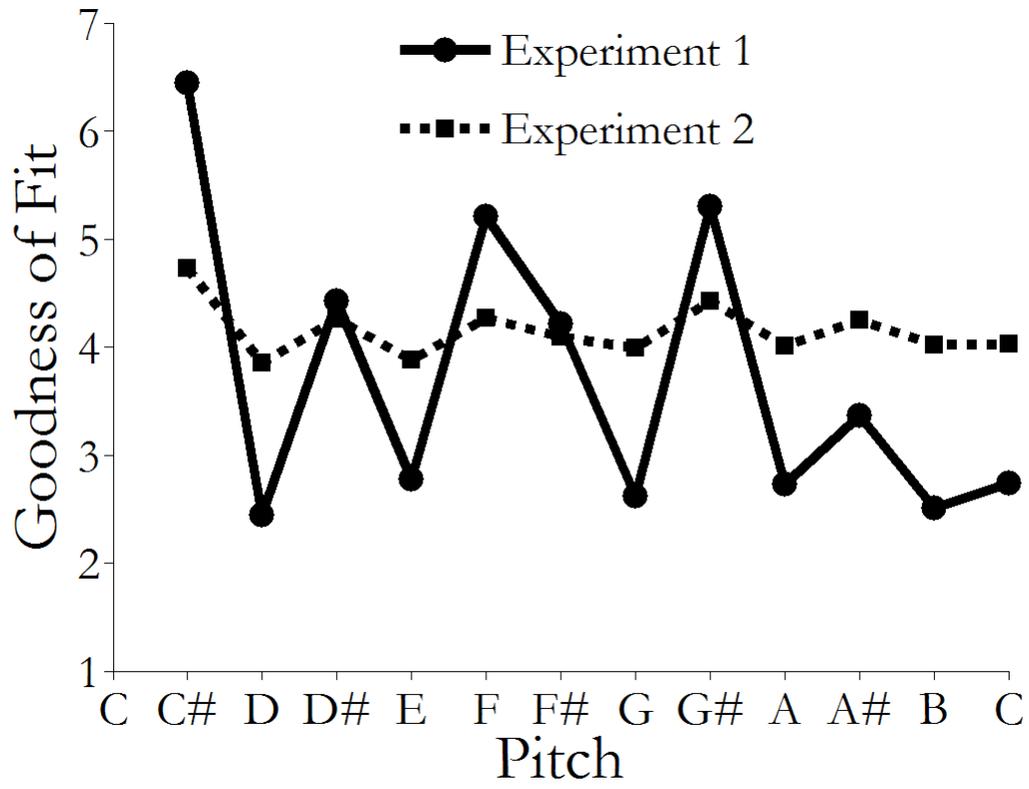
Figure 4

Interaction of the size of the temporal position effect and experiment
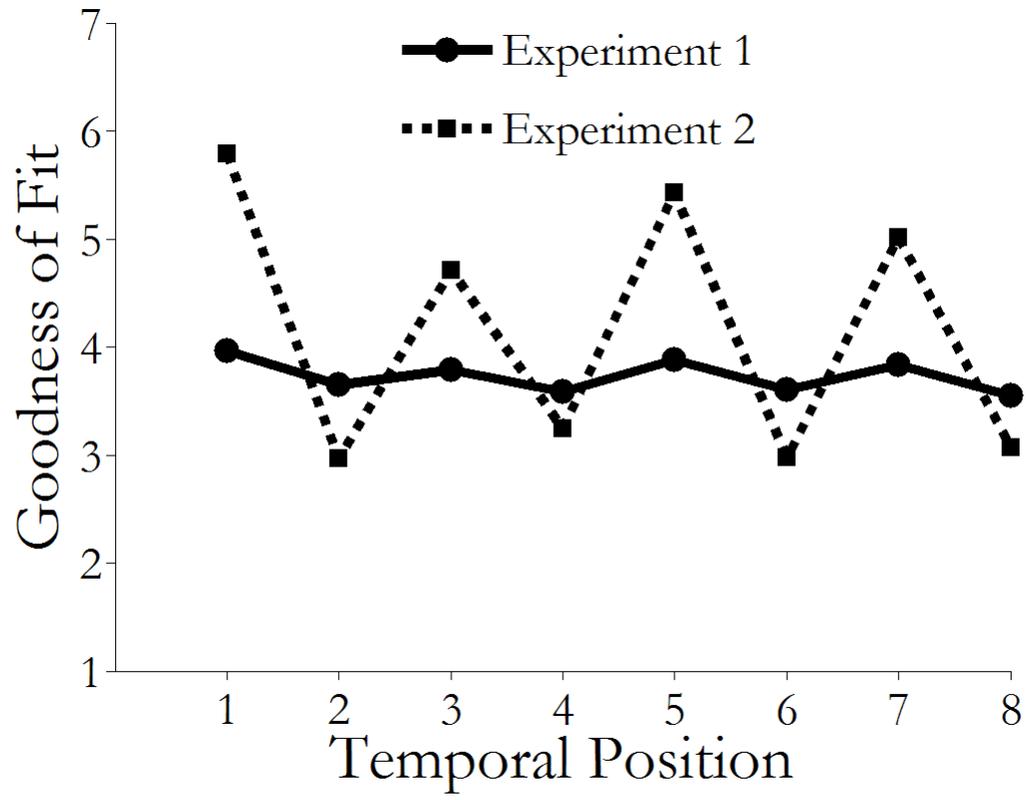
Figure 5
Musical contexts and example probe event used in Experiments 3 and 4

Figure 6
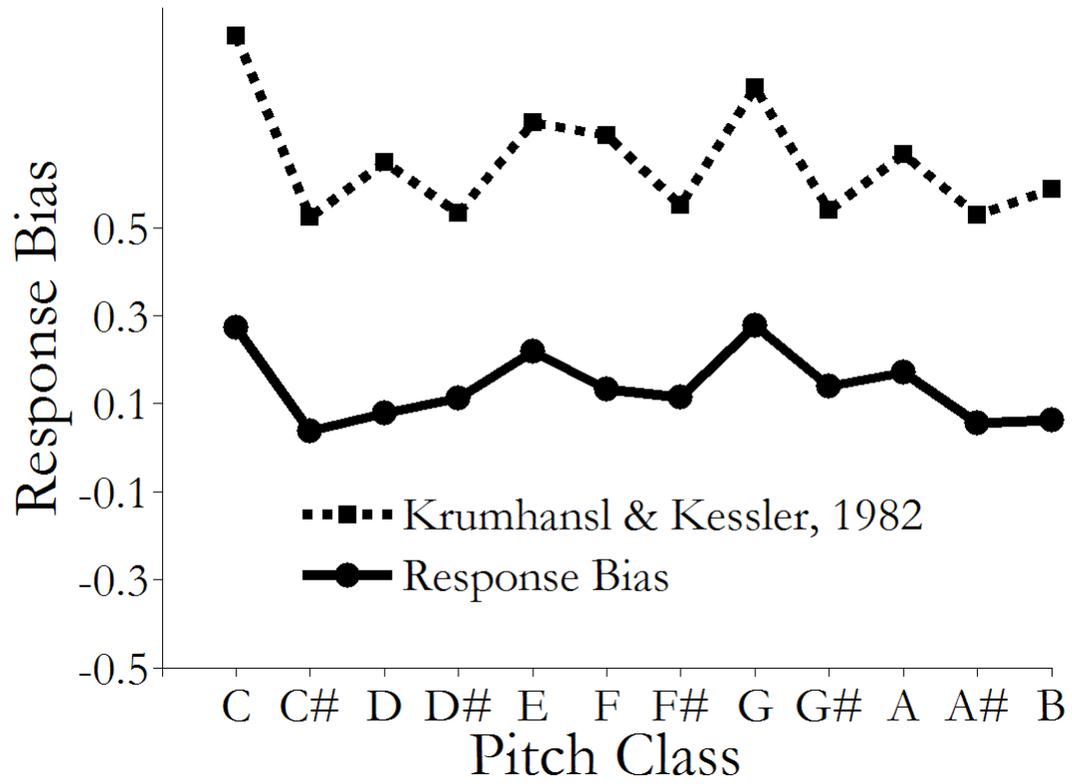*Response bias to respond "on the beat" across pitch for Experiment 3.*

Figure 7

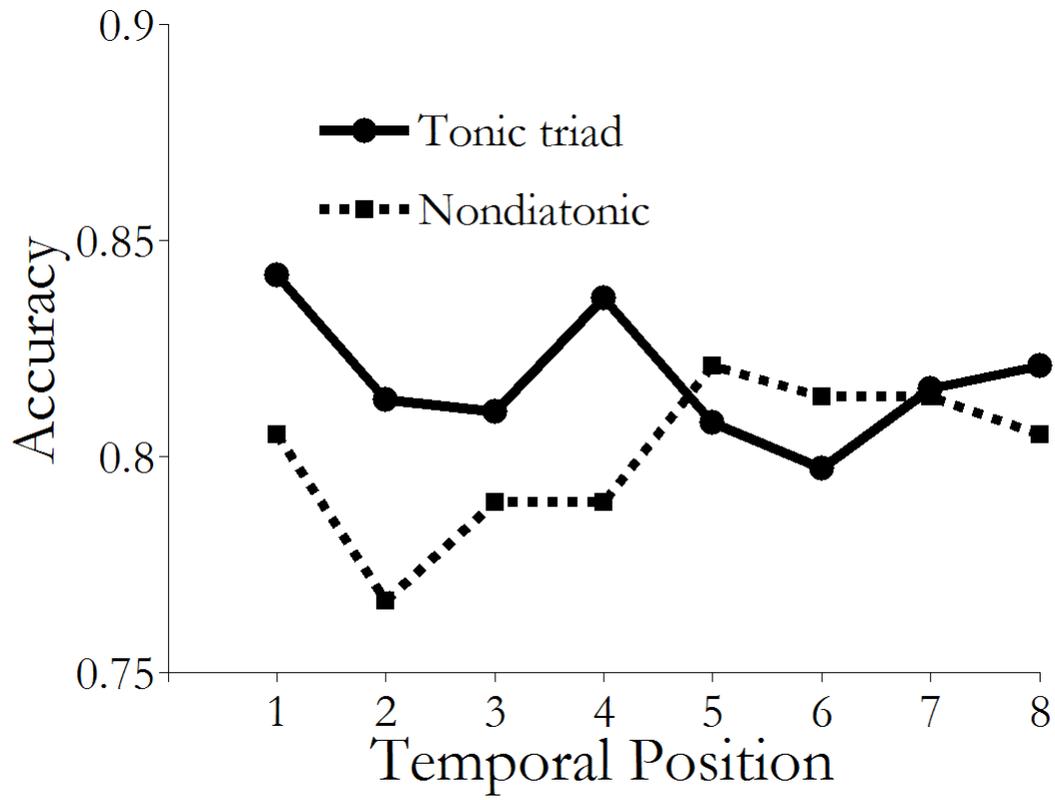*Accuracy for the tonic triad members and nondiatonic pitches across temporal position for Experiment 4*

Figure 8
Response bias to respond "on the beat" across pitch for Experiment 6, temporal condition.