MURDOCH RESEARCH REPOSITORY

http://researchrepository.murdoch.edu.au

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.

Pitch structure, but not selective attention, affects accent weightings in metrical grouping

Jon B. Prince

Murdoch University, Perth, Australia

Corresponding Author: Jon Prince
Department of Psychology and Exercise Science
Murdoch University
90 South Street
Murdoch, WA 6150
Australia
tel (+61) 8 9360 6670
fax (+61) 8 9360 6878
j.prince@murdoch.edu.au

Abstract

Among other cues, pitch and temporal accents contribute to grouping in musical sequences. However, exactly how they combine remains unclear, possibly because of the role of structural organisation. In three experiments, participants rated the perceived metrical grouping of sequences that either adhered to the rules of tonal Western musical pitch structure (musical key), or did not (atonal). The tonal status of sequences did not provide any grouping cues and was irrelevant to the task. Experiment 1 established equally strong levels of pitch leap accents and duration accents in baseline conditions, which were then recombined in subsequent experiments. Neither accent type was stronger or weaker for tonal and atonal contexts. In Experiment 2, pitch leap accents dominated over duration accents, but the extent of this advantage was greater when sequences were tonal. Experiment 3 ruled out an attentional origin of this effect by replicating this finding while explicitly manipulating attention to pitch or duration accents between participant groups. Overall, the presence of tonal pitch structure made the dimension of pitch more salient at the expense of time. These findings support a dimensional salience framework in which the presence of organisational structure prioritises the processing of the more structured dimension regardless of task relevance, independent from psychophysical difficulty, and impervious to attentional allocation.

Introduction

The seminal Gestalt approach to the study of perception provided principles that describe how external stimuli are organised into coherent percepts. Although the primary application of these rules are in visual perception {Wagemans, 2012 #3313}, they have also been applied to auditory processing, particularly in the context of auditory stream segregation {Bregman, 1990 #59;Alain, 2000 #2951}. Gestalt principles are also relevant to the larger concept of auditory objects {Snyder, 2007 #1572;Kubovy, 2001 #229}, based on extracting regularities from the environment {Winkler, 2009 #2848}. Music provides an ideal domain for investigating grouping in auditory sequences, as it is central to the perceptual organisation of individual musical notes {Jones, 1976 #1822}, musical phrases {Deutsch, 1981 #106;Sloboda, 1980 #2856}, separating instruments {Handel, 1995 #1226}, and large-scale structure {Lerdahl, 1983 #240;Clarke, 1990 #1668}.

Musical accents, or points of stress, make individual events stand out from the surrounding notes. They are the basis of grouping in music {van Noorden, 1975 #2702;Deutsch, 1999 #561;Thomassen, 1982 #406}, and occur in a number of dimensions, or information sources. A pitch leap (a large frequency change between adjacent notes), or a change in pitch contour (switching between frequency increases/decreases in a melody) highlights the note following the change, and is fundamental to melodic perception {Schmuckler, 2009 #3266;Dowling, 1978 #112}. There are numerous other accent types, such as variations in duration {Vos, 1977 #1767;Cooper, 1960 #1654}, a silent gap between notes {Jones, 1997 #1797;Preusser, 1970 #2456}, a change in loudness {Repp, 1995 #2203;Tekman, 2002 #1160}, or a timbral change {Cusack, 2000 #3324}.

An emergent property of accents in musical sequences is metre, the periodic oscillation between strong and weak points in time more commonly known as the beat or pulse of a given musical passage {Lerdahl, 1983 #240}. The placement of accents determines

the metric framework {Parncutt, 1994 #2542;Povel, 1985 #1756;Hannon, 2004 #178}, but metre is an abstract and hierarchical structure that is extracted from the accent patterns comprising the musical surface features {London, 2004 #3098}. Many different accent sequences can nonetheless invoke the same metric framework {cf. \Povel, 1985 #1756}. Similarly, the exact same accent pattern can with conscious effort be interpreted in different metric frameworks {Repp, 2007 #2369}. In Western music, musical metre largely uses either binary (strong-weak) or ternary (strong-weak-weak) groupings, although both types can be combined when nested hierarchically {London, 1995 #2555}. Despite its abstract nature, metre is a fundamental component of music {Palmer, 1990 #293}, which is observable from infancy {Hannon, 2005 #1542;Zentner, 2010 #2277}, arises spontaneously {Ladinig, 2009 #2032} – even from isochronous sequences of identical notes {Brochard, 2003 #1755;Potter, 2009 #1815} – is neurally observable {Snyder, 2005 #1997}, and dissociable from the processing of rhythm {Geiser, 2009 #1807}.

Given its centrality to musical behaviour, it is hardly surprising that metre influences the perception of musical sequences. Detecting changes to a temporal pattern is easier if it conforms to a regular metrical framework {Grube, 2009 #1808;Hébert, 2002 #1747}. Indeed, metric sequences are overall more accurately encoded, processed and remembered than nonmetric sequences {Povel, 1985 #1756;Keller, 2005 #2942}. The benefits of metrical structure extend to production paradigms – strong beats provide stable reference points for tapping {Repp, 2008 #2054}, conversely, synchronisation is more difficult when sequences are only weakly metrical {Patel, 2005 #2357}. Metrical patterns even aid tapping on weak metric locations {Keller, 2005 #3223} and in unfamiliar metrical frameworks {Tillmann, 2011 #2586}. The presence of a metrical framework can also create distortions, such that the detection of timing perturbations is superior for strong metric locations, but inferior at weak positions {Repp, 1992 #1556}. Estimations of event duration are similarly affected – listeners

perceive sequences whose metric structures suggest an earlier ending point as longer than identical sequences that suggest a later ending point, giving systematic over/under-estimations consistent with the metric framework {Jones, 1989 #205;Jones, 1993 #1559}. One possible mechanism for these effects is that metric frameworks direct attention towards metrically strong points in time rather than systematic perceptual distortions {Repp, 2010 #2367;Ellis, 2010 #2420}. However, there is no benefit for implicit learning of metric patterns in a serial reaction time task (involving online attending), suggesting that the source is not attentional but relevant only for encoding and retrieval {Schultz, 2013 #3199}.

Metre is not the only abstract and hierarchical structure in music. Tonality (or musical key) provides a hierarchical arrangement of the 12 pitch classes per octave, and is central to the perception of music. For one of many examples, tonality affects the detection of mistuned notes in musical sequences {Warrier, 2002 #2872;Cuddy, 1979 #91;Cohen, 1989 #2920;Cohen, 1990 #2921;Marmel, 2008 #1672;Trainor, 1993 #421}. Also, recognition memory is superior for tonal melodies than atonal (no musical key) ones {Schulze, 2012 #2882;Dowling, 1991 #111;Bharucha, 1983 #32;Freedman, 1999 #147}. The role of tonality in music perception is so strong that it creates a processing cost for less tonally-related events {Tillmann, 2003 #416;Tillmann, 2008 #1673} and can mask otherwise noticeable changes in melodies {Trainor, 1992 #3326;Trainor, 1994 #420}. Neural evidence shows specific neural loci and activation patterns for tonality {Fedorenko, 2012 #3136;Janata, 2002 #201;Koelsch, 2005 #215;Marmel, 2011 #2987}, and that tonal relations are processed preattentively {Brattico, 2001 #3215;Brattico, 2006 #1685;Koelsch, 2002 #1686}. A substantive difference between tonality and metre is that although some form of tonality is universal across musical cultures {Krumhansl, 2010 #2873}, its specific form is not innate but must be acquired through exposure {Trainor, 2010 #3177}.

The effects of pitch and temporal structure are not limited to their respective dimensions, but can influence their integration. In general, having a coherent pitch-time structure (i.e., accent patterns in both dimensions suggesting the same grouping) enhances time judgements {Boltz, 1992 #1692;Boltz, 1995 #1092;Boltz, 1998 #1087;Boltz, 1999 #1084}, as well as detection of pitch and temporal changes {Brown, 2002 #2991;Jones, 1982 #1179;Monahan, 1987 #271}, melody recognition {Jones, 1991 #206}, and production {Jones, 1997 #1797;Pfordresher, 2003 #1588}. However, more recent research reports that the presence of hierarchical structure in one dimension can make it more salient at the expense of another dimension. {Prince, 2009 #1774@@author-year} found that confirming temporal expectancies (playing a note exactly when the listener expects to hear it) only improved accuracy on a pitch comparison task when the sequences were atonal (unstructured pitch). When sequences were tonal, these authors reported that confirming or denying temporal expectancies had no effect on performance. Similarly, the influence of harmonic accents (chord changes) on perceived rhythm {Dawe, 1993 #1606;Dawe, 1994 #1744} decreases for atonal patterns {Dawe, 1995 #1604}. These authors instructed listeners to indicate the starting point of repeating rhythm patterns, and manipulated the irrelevant temporal position of alternations between two chords. Responses were most consistent when chord changes were congruent with duration accents (the emphasis created by making one note longer than its neighbours), when both cues accented the same event in the sequence. When the sequences were atonal, chord changes no longer influenced judgements.

The more general question of how information from pitch and time combines in music perception has been an area of continual interest {for reviews`, see \Krumhansl, 2000 #218;Prince, 2011 #2468}, with unfortunately little consensus on the issue of if the integration of these dimensions is linear/additive {Palmer, 1987 #1123;Palmer, 1987 #1129} or interactive {Jones, 1987 #1126;Jones, 1989 #205}. One possible source of contradictory

findings is the relative discriminability of pitch and time. A critical step in assessing how

stimulus dimensions combine in perception is to assess their relative discriminability

{Garner, 1974 #1147}, because unequal discriminability across dimensions can result in

demonstrably independent dimensions functioning in an interactive manner {Garner, 1970

#1144;Garner, 1976 #1547}. There are only two studies in the literature on pitch-time

integration in music that satisfy this prerequisite {Ellis, 2009 #1881;Prince, 2009 #1609}.

Prince, Thompson, et al. {, 2009 #1609} showed that tonal structure influences the

perception of the temporal properties of a probe note following a melody, but not vice versa;

this relationship persisted after equalising the dimensions in terms of discriminability. These

authors proposed the concept of dimensional salience to explain these results. A more salient

dimension is prioritised in perceptual processing, and contributes more strongly to forming

the mental representation of the external stimulus. As a result, the more salient dimension is

more likely to interfere with the processing of another dimension, and also less likely to

experience interference from a less salient dimension. Importantly, this concept is separate

from the issue of psychophysical difficulty, which can lead to demonstrably independent

dimensions appearing interactive {Garner, 1970 #1144;Melara, 1993 #1170;Melara, 1994

#1164}.

The only other study to assess baseline discriminability in pitch-time integration

comes from Ellis and Jones {, 2009 #1881}. These authors equalised the strength of accents

induced by pitch leaps with those from duration accents to create either a duple (groups of

two) or triple (groups of three) metric structure. After this equalisation, they combined these

accent types in both congruent and incongruent manners and found that pitch leap accents

accounted for three times the variance of duration accents. These results contrast with earlier

research discounting the role of pitch accents compared to temporal accents in perception

{Huron, 1996 #1837;Monahan, 1993 #2635;Monahan, 1985 #1178}, as well as production

{Drake, 1993 #1760;Drake, 1991 #2194;Snyder, 2001 #1739;Pfordresher, 2003 #1588}.

Importantly, none of this earlier research addressed the issue of discriminability.

Taken together, these studies motivate an investigation of the role of tonal structure on the relative strength of pitch and temporal grouping cues of equal discriminability. In the context of music, tonality can attenuate the effects of temporal expectancies {Prince, 2009 #1774}, but how might it affect the combination of the pitch and temporal dimensions in metrical grouping? The results of Dawe et al. {, 1995 #1604} suggest that removing tonality could weaken the effect of pitch grouping cues, however their participants were instructed to indicate rhythm patterns only, and there was no baseline equalisation of the strength of accents in both dimensions. In the present study, three experiments comprised an exploration of this question. In the first experiment, different sizes of pitch leap and duration accents were tested, in order to establish equally strong accent levels, both in a tonal and atonal context. In subsequent experiments, participants heard stimuli consisting of sequences of equal-strength accents, combined in both a congruent and incongruent fashion, and also varying in the presence of tonal structure. Experiment 2 had no specific instructions to attend selectively to one dimension, whereas Experiment 3 manipulated this instruction between subjects.

In all experiments, participants rated the perceived metrical grouping of the stimuli on a five-point rating scale – a subjective measure instead of an objective test. This choice follows from an established precedent in research on metrical grouping that specifically investigates the respective roles of pitch and time using a rating scale {Hannon, 2004 #178;Nittono, 2000 #1082;Ellis, 2009 #1881}, a homologous 2AFC task {Dawe, 1994 #1744;Thomassen, 1982 #406}, or similar subjective reports of metrical structure {Dawe, 1993 #1606;Dawe, 1995 #1604}. The main theoretical test of dimensional salience in this research derives from a synthesis of the two previous papers that equalised the strength of

pitch and time prior to combining them {Ellis, 2009 #1881;Prince, 2009 #1609}. Continuity with these papers was desirable, both of which employed a rating task. Additionally, Ellis and Jones explained that this methodology extends the magnitude matching paradigm of Stevens and Marks {Marks, 2002 #3567;Stevens, 1980 #3566}, in which participants used the same ordinal scale to rate the dimensions of brightness and loudness (independently). The present research furthers this extension.

A different methodological approach to metrical grouping is to use a tapping task {e.g.`, \Povel, 1985 #1756;Pfordresher, 2003 #1588;Snyder, 2001 #1739}, which ventures into the sensorimotor literature {for reviews`, see \Repp, 2005 #2575;Repp, 2013 #3557} beyond pure perception and thus the purview of the current manuscript. Additionally, the advantage of using a more objective task (in perception or production) comes with the limitation of being an indirect measure of metrical interpretation. Ultimately the issue comes down to how best to measure participants' perception of metre. The most direct approach is to ask the participants exactly that – what metre do they perceive? Establishing continuity with earlier research and the specific focus on perception made a rating task the most appropriate choice, particularly as a first step towards investigating dimensional salience in the context of metrical grouping.

Experiment 1: Baseline tests

The goal of Experiment 1 was to determine levels of pitch and duration accents that resulted in equally strong percepts of groupings. In other words, for inducing duple/triple groupings, how big does a pitch leap need to be in order to be equally strong as a duration accent? Three different levels of pitch leap and duration accent were used in order to select the closest-matching values across dimension. Previous research has derived such values {Ellis, 2009 #1881}, but all durations were below the temporal integration threshold of 200ms. Below this threshold, changes in duration and intensity (loudness) mutually interfere;

accordingly a duration accent in Ellis and Jones may have been perceived as an intensity accent. In order to remove any potential contribution of changes in perceived loudness to the duration accent, all events had durations at or above the temporal integration threshold.

Method

Participants

There were 12 participants in Experiment 1, with an average age of 29 ($SD = 11.5$) and an average of 2 years of musical training ($SD = 3.1$). Participants consisted of students at Murdoch University and members of the nearby community. Students enrolled in an introductory psychology class received course credit for their participation; others received $10 compensation.

Stimuli

All stimuli began with 3 chords of unequal duration (1750ms, 250ms, and 750ms, respectively) in order to prevent a bias towards any particular metric grouping. After the chords there was a silent gap of 1250ms, followed by an isochronous sequence of 24 notes, with an inter-onset interval of 500ms (120bpm). The total duration of each stimulus (chords and notes) was therefore 16s. Each sequence contained either a pitch leap accent or a duration accent, but not both. Accents occurred either every 2 (duple grouping) or 3 (triple grouping) notes. Each grouping type of both dimensions had three levels of strength: weak, medium, and strong, described below. Stimuli were generated as .wav files in Matlab 7.0 {Mathworks, 2004 #2275} and were harmonically rich, consisting of the fundamental frequency and its first 3 partials (relative energy of 50%, 25%, 12.5%, and 12.5%, respectively). All frequencies were restricted to the 5-octave range of 77.8 Hz to 2349 Hz, following Krumhansl, Bharucha, and Kessler {, 1982 #221}.

For the trials using pitch leaps as accents, the three-chord prefix was followed by an ascending pitch sequence of 12 notes, which then repeated once (24 notes total). Pitch leap accents occurred every 2 or 3 notes, in order to induce a duple or triple grouping. Within a group the notes ascended by a smaller interval. Sequence starting notes were either C3 (130.8Hz), C#3 (138.6Hz), D3 (146.8Hz), or D#3 (155.6Hz). Each note lasted for 500ms, such that there was no silent gap between notes (i.e., the inter-stimulus interval was 0ms).

In the tonal trials, all notes were selected from the major diatonic scale of the first note (e.g., C major). The three-chord prefix served to orient the listener to the musical key of the subsequent 24-note sequence (in musical terms, chords with the harmonic function of I, IV, and V). The pitch accent consisted of a leap between adjacent notes of 3, 4, or 5 scale steps; larger scale steps correspond to stronger pitch leap accents. Due to the structure of the major diatonic scale, the absolute size of a given leap was either 3 or 4 semitones (3 scale step interval, $M = 3.5$), 6 or 7 semitones (4 scale steps, $M = 6.1$), or 7 or 8 semitones (5 scale steps, $M = 7.9$)[1]. Within a group, adjacent notes ascended by either 1 or 2 semitones (base interval, $M = 1.7$), also varied as necessary to use only diatonic notes. The sequences were clearly tonal – the average maximum key correlation (MKC) using the Krumhansl-Schmuckler keyfinding algorithm {Krumhansl, 1990 #219;Krumhansl, 1986 #2256} was .87 ($SD = .03$); without the preceding chords the mean MKC was .75 ($SD = .08$). Figure 1a depicts an example sequence suggesting a duple grouping, in the key of C major. Figure 1b shows a tonal example with a triple grouping, also in a C major tonality. Example stimuli from all experiments are available as supplemental material.

The atonal trials were designed to prevent the establishment of any musical key (thus atonal). The 3 starting chords were the same as those in the tonal trials, but 4 critical notes were shifted up or down by one semitone (smallest possible interval in Western music) so as to destroy its tonality. Within a given scale step type (3, 4, or 5), the atonal trials had exactly

the same number and size of intervals as the tonal trials, such that the theoretical pitch accent strength would match across tonality. The difference was in the placement of these intervals, which was manipulated such that the notes did not fall into any one musical key (MKC $M =$ .46, $SD = .06$; without the preceding chords $M = .50$, $SD = .17$). Figure 1c shows an example sequence suggesting a duple grouping, starting on the note C. Figure 1d shows a triple grouping example, also starting on C.

Insert Figure 1 about here

Sequences with duration accents also had the same chords as the pitch accent trials, with half being tonal and the other half atonal. After the chords, the sequences were monotonic (same pitch throughout). In place of the pitch accent, there was a duration accent on one out of every two (duple grouping) or three (triple grouping) notes. Within a group, notes lasted for 200ms (base duration), whereas the accented note lasted for either 250ms, 333ms, or 450ms. Sequences remained isochronous (equal inter-onset interval), but the gap between notes varied as a function of their duration.

In sum, the design was a fully crossed 2 (Dimension: pitch/time) by 2 (Grouping: duple/triple) by 2 (Tonality: tonal/atonal) by 3 (Strength: weak/medium/strong) by 4 (Starting Pitch: C3, C#3, D3, D#3) design, giving 96 unique conditions. Participants did each condition twice (in random order), yielding a total of 192 trials. There were two expected findings – first, a main effect of Grouping, as triple groupings should be rated differently (higher) than duple groupings. Second, participants should rate sequences as higher (more triple) with increasing strength of triple grouping accents, but lower (more duple) with increasing strength of duple grouping accents. Thus there should be an interaction between Grouping and Strength.

Apparatus

The experiment ran on a PC with the Windows XP operating system. The experimental interface was programmed in Matlab 7.0, using the Psychophysics toolbox {Brainard, 1997 #917}. Participants heard the sequences over Sennheiser HD280 Pro headphones, alone in a quiet room.

Procedure

Participants gave informed consent and completed a background questionnaire on their musical experience. The experimenter guided each participant through 4 practice trials, which consisted of hearing an example sequence and entering a rating on a scale of 1 (strong duple) to 5 (strong triple) indicating how the sequence sounded to them. The experimenter answered any remaining questions and commenced the full version of the experiment (192 trials). After completing the experiment they received a debriefing on its purpose and methodology. The entire procedure took approximately one hour.

<div style="text-align:center">Results</div>

There was a high level of agreement across participants in their perceived grouping ratings, with an average intersubject correlation of .70 ($SD = .10$). Participants' ratings were then averaged across the four starting pitches and two repetitions, yielding 24 data points per participant (from the original 192 trials). These ratings were analysed with a repeated measures ANOVA with the within-subjects factors of Dimension (2) by Grouping (2) by Tonality (2) by Strength (3), following the design described in the Method section.

The ANOVA revealed slight sphericity violations for Tonality by Strength (Mauchly's $p = .042$), Grouping by Strength ($p = .030$), and Tonality by Grouping by Strength ($p = .037$). These violations disappear when excluding the participant with the lowest intersubject correlation ($r = .50$, two standard deviations below the mean). The pattern

of findings remains the same with or without these data, thus the participant's data were retained in all analyses. Further, using the Greenhouse-Geiser correction (with all participants) does not change any of the results, for any experiment. Accordingly, all analyses use the more conventional reporting of degrees of freedom and original p-values. There was no main effect of Dimension, $F(1,11) = 1.18$, $p = .301$, $\eta^2 < .01$, indicating that participants were equally likely to rate sequences as duple or triple regardless of if the accent type was a pitch leap or duration[2]. There was an expected main effect of Grouping, $F(1,11) = 256.88$, $p < .001$, $\eta^2 = .85$, because participants rated duple sequences lower (more duple) than triple sequences. There was no main effect of Tonality, $F(1,11) < 1$, *ns*, meaning that overall, the presence or absence of tonality did not bias participants to rate sequences as more duple or triple. There was a weak but unanticipated main effect of Strength, $F(2,22) = 4.34$, $p = .026$, $\eta^2 < .01$, because there was a slight trend towards more triple ratings with higher levels of grouping strength. There is no obvious explanation for this result other than a potential ceiling effect with the triple groupings. As this effect has no direct bearing on the theoretical question under investigation, it will not be discussed further. The only other significant effect was the expected interaction between Grouping and Strength, $F(2,22) = 18.01$, $p < .001$, $\eta^2 = .02$, reflecting the important fact that the effect of Strength went in opposite directions for different groupings. Consistent with the rating scale, stronger duple groupings gave lower numerical ratings (more duple), whereas stronger triple groupings gave higher numerical ratings (more triple). Figure 2 displays this pattern, for pitch and time separately. No other interactions were significant.

Insert Figure 2 about here

Because the main goal of Experiment 1 was to establish equal levels of accent strength across dimension, there were additional analyses of Dimension, Grouping, and Tonality, despite their lack of interaction. Confidence interval comparisons (see Table 1) revealed that the best match of accent strength across dimension was the 5 scale step level (pitch leap accent) and 333ms (duration accent). For duple groupings, 5 scale steps gave a mean rating of 1.30, and 333ms (duration) gave 1.28. For triple groupings, 5 scale steps (pitch) yielded 4.57; 333ms (time) was 4.66. These levels were significantly different from lower grouping levels (3 scale steps and 250ms).

Insert Table 1 about here

Discussion

Experiment 1 provided levels of equal strength for pitch leap accents and duration accents in both tonal and atonal sequences, and gave two main findings. First, varying the structural characteristic of tonality does not influence perceived grouping based on pitch or temporal surface cues in baseline sequences. In sequences with no temporal variation, the strength of pitch leap accents was the same regardless of their level of tonality. Similarly, the tonality of a chord prefix to a monotonic sequence did not affect listeners' ease of extracting duration grouping accents. Stated differently, the presence or absence of pitch structure did not change the effectiveness of pitch leap and duration accents when presented in isolation (baseline conditions). This finding is consistent with Prince, Thompson, et al. (2009), who found that the tonal stability of a probe tone biases judgements of whether it is on or off the beat, but without any effect on discriminability (instead, purely response bias). It also concurs with Prince, Schmuckler, et al. (2009), who found that a pitch height comparison task was equally difficult for both tonal and atonal sequences.

Second, using only durations above the temporal integration threshold of 200ms gave equivalent results to research using a shorter base duration of 60ms {Ellis, 2009 #1881}.

These authors found that duration accents of 100ms gave equivalent grouping strength to a 5 semitone pitch leap (base interval of 1 semitone). Converting these values to Weber fractions (WF) gives .67 (duration) and 4 (pitch leap). In the current study, the 333ms duration accent (200ms base) also corresponds to a WF of .67, and the equally strong 5 scale step (1 step base) pitch accent gives a WF of 4. These equalised values of accent strength in the current experiment have exactly the same WFs as Ellis and Jones, suggesting that the relative strength of duration and pitch leap accents may remain constant across multiple base durations. This accordance has the notable limitation that the scale step size in the current study did not equate to a constant semitone size (as used in Ellis and Jones). Converting the scale steps to their average absolute size (7.9 and 1.7 semitones) gives a slightly different WF of 3.41. Additionally, the semitone size of pitch intervals was variable in the present study, in order to accommodate the construction of diatonic sequences (cf. Method), necessitating further caution in interpreting this finding. Regardless, it seems unlikely that any perceived changes in loudness from duration accents below the temporal integration threshold had any sizable influence on earlier findings, at least when tested in baseline conditions. Following this related thread of determining the exact nature of the relationship between pitch leap, intensity, and duration accents at various base durations and WFs would be a useful extension of this literature. The main goal of this experiment, however, was to determine the relative sizes of pitch leap and duration accents needed to result in equally strong percepts of metrical grouping while ruling out the potential contribution of the temporal integration threshold to duration accents. This baseline experiment sets the stage for the primary question at hand – how does tonality affect the combination of both accent types in metrical grouping?

Experiment 2: Combining pitch and duration accents

Having determined what sizes of pitch leap accent and duration accent result in equally strong perceived metrical groupings (Experiment 1), the sequences of Experiment 2

combined these accent types, such that participants heard sequences with both pitch leap accents and duration accents. For each dimension, there were three different levels of accent – duple, neutral (accent every six notes), and triple. A 3 by 3 factorial combination of the accent levels enables testing how congruent (e.g., duple-duple) and incongruent (e.g., triple-duple) accent patterns affect perceived metrical grouping. In particular, it is possible to test if – once combined – one dimension is more effective than the other dimension, despite being equally strong when presented in isolation (Experiment 1). For example, when the pattern has a duple pattern of duration accents and a triple pattern of pitch leap accents, which one prevails? Previous research which used this design {Ellis, 2009 #1881} suggests that pitch accents should be more powerful.

Of course, the manipulation of greater theoretical interest is that of tonality. There are several possible outcomes of this manipulation; the first is that pitch leap and duration accents function no differently for tonal and atonal sequences (null hypothesis). That is, one dimension may be more powerful than the other, but unaffected by tonality. The second potential result is that tonality provides a framework that enhances the ability to extract grouping accents in general, regardless of the dimension in which the accent is presented, such that tonal sequences aid the processing of accents in both pitch and time (a "more structure is better" hypothesis). The third possibility is that tonal sequences enhance the strength of pitch leap accents, whereas atonal sequences benefit duration accents (dimensional salience hypothesis). Of these hypotheses, the second is the least likely – tonality did not affect the strength of accents in the baseline conditions of Experiment 1, so there is little reason to expect it to influence the perception of accents when the dimensions are combined.

Method

Participants

A new set of 12 listeners participated in Experiment 2. These participants were similar in age (M = 28.4 years, SD = 12.2), and musical training (M = 2.3, SD = 3.3).

Stimuli

New sequences were created in Experiment 2, based on the size of pitch leap accent (5 scale steps) and duration accent (333ms) selected from Experiment 1. In addition to the duple and triple groupings, there was a neutral grouping which had an accent on every sixth note. This grouping was labelled neutral because it could be heard as either two triple groups or three duple groups, and follows Ellis and Jones (2009). Sequences were created in the same manner as Experiment 1, except that different groupings of the selected pitch and duration accent were recombined to create congruent, unbiased (a neutral grouping in one or both dimensions), and incongruent groupings. Accordingly, the stimuli varied the following variables in a factorial manner: Pitch Grouping (duple, neutral, triple), Time Grouping (duple, neutral, triple), and Tonality (tonal, atonal). Having equal observations per condition while maintaining a manageable number of trials for each participant required using only three starting pitches for each condition: C3, C#3, and D3 (discarding the D#3 from Experiment 1). Crossing 3 pitch groupings with 3 time groupings, 2 levels of tonality, and 3 starting pitches resulted in 54 stimuli. Each participant did three blocks, giving 162 trials total. All other aspects of the stimuli were the same as Experiment 1.

Apparatus

The apparatus were the same as Experiment 1.

Procedure

The procedure was the same as Experiment 1. Practice trials included only congruent trials (i.e., both pitch and duration accents suggested either a duple or a triple grouping). These trials crossed tonal/atonal sequences with duple/triple groupings.

Results

Ratings were consistent across participants, with an average intersubject correlation of .53 ($SD = .09$). As in Experiment 1, participants' ratings were averaged across starting pitch (C, C#, D) and block (1, 2, and 3) prior to subjecting the data to a 3 (Pitch Grouping) by 3 (Time Grouping) by 2 (Tonality) repeated measures ANOVA. A preliminary ANOVA added Block as a within-subjects variable. It exerted neither a main effect nor any interactions, and was therefore omitted from the main analysis presented below. As expected, there were robust effects of Pitch Grouping, $F(2,22) = 84.93$, $p < .001$, $\eta^2 = .56$, and Time Grouping, $F(2,22) = 31.83$, $p < .001$, $\eta^2 = .19$, but not Tonality, $F(1,11) < 1$, ns. There was an interaction between Pitch Grouping and Time Grouping, $F(4,44) = 8.77$, $p < .001$, $\eta^2 = .03$. Figure 3a displays this interaction, showing that participants differentiated more between Time Grouping levels at the neutral Pitch Grouping level than at other levels of pitch grouping (i.e., the slope is steeper for Pitch Neutral).

Insert Figure 3 here

The only other significant interactions were between Pitch Grouping and Tonality, $F(2,22) = 7.88$, $p = .003$, $\eta^2 = .01$, as well as between Time Grouping and Tonality, $F(2,22) = 7.82$, $p = .003$, $\eta^2 = .01$. There were two additional repeated measures ANOVAs (one for tonal trials, one for atonal) to explore these interactions, both of which had the within-subjects variables of Pitch Grouping and Time Grouping. The interaction occurred because the effect size of Pitch Grouping was .64 in the tonal trials but only .49 in the atonal trials. Conversely, Time Grouping had a larger effect size in atonal trials (.24) than in tonal trials (.14). That is, tonality magnified the contribution of pitch accents and attenuated that of duration accents.

Because effect sizes describe the percent variance accounted for by each variable, they do not indicate the absolute size of the effects. It is possible that both pitch accents and duration accents were stronger in the tonal trials (the "more structure is better" hypothesis), but that the pitch dimension benefited more, yielding the observed interaction. Effect size analyses therefore cannot speak to this issue. However, raw difference scores between duple and triple grouping conditions enable examination of this question, as they represent absolute values rather than the proportion of total variance. Averaging across Time Grouping levels, subtracting the mean rating for triple Pitch Grouping levels from the mean rating of duple Pitch Grouping levels gives a raw difference score representing the strength of Pitch Grouping cues. The complementary calculation provides the strength of Time Grouping cues (time triple minus time duple, after averaging across pitch levels). Figure 4a shows these difference scores, separately for tonal and atonal trials. It depicts how the influence of pitch leap accents declines when sequences are atonal (smaller difference score), whereas the effect of duration accents increases for these sequences (larger difference score). A 2 (Dimension) by 2 (Tonality) repeated measures ANOVA of these difference scores thus recovered not only a main effect of Dimension, $F(1,11) = 5.73$, $p = .036$, $\eta^2 = .32$ (pitch was stronger overall), but also an interaction, $F(1,11) = 91.05$, $p < .001$, $\eta^2 = .04$. Tonality was not significant as a main effect in this analysis, $F(1,11) = 2.03$, $p = .181$, $\eta^2 < .01$. An additional ANOVA included Block as a factor, and again found neither a main effect nor an interaction with any other variable. Further exploring the interaction between Dimension and Tonality with two-tailed paired t-tests[3] revealed that atonal and tonal trials differed significantly for both pitch leap accents, $t(11) = 5.37$, $p < .001$, and duration accents, $t(11) = 4.30$, $p = .003$. Tonality did not benefit both accent types, but selectively boosted the effectiveness of pitch leap accents at the expense of duration accents.

Insert Figure 4 about here

Discussion

In Experiment 2, pitch leap accents and duration accents were combined in individual sequences, forming sequences with congruent, unbiased, and incongruent grouping cues across dimension. Despite equating the strength of accents across dimension (Experiment 1), pitch accents nonetheless accounted for 3 times the variance that duration accents did. The interaction between pitch and time occurred because the effect of one dimension was larger when the other dimension was neutral. The more intriguing results come from the interaction of tonality with both pitch and duration accents: tonality enhanced the effect of pitch accents and reduced the influence of duration accents. These data therefore rule out the null hypothesis of no effect of tonality, and detract further from the notion of tonality providing an overall benefit to extracting any type of accent (second hypothesis). Instead, and in accordance with the last (dimensional salience) hypothesis, tonality benefited selectively the dimension for which it provides an organisational framework, namely pitch.

Previous work on metrical grouping also found that recombining pitch leap and duration accents equalised for strength in a baseline experiment resulted in a dominance of pitch {Ellis, 2009 #1881}. Specifically, pitch accounted for three times as much variance that time did (η^2 = .51 and .17, respectively), using only atonal sequences. In the atonal trials of the current study – arguably the most comparable condition between studies – the advantage of pitch was somewhat smaller in that it accounted for (only) twice the variance of time (η^2 = .49 and .24, respectively). If using durations below the temporal integration threshold inflated the contribution of duration accents to grouping (although Experiment 1 downplays this possibility), then one would predict that using durations above this threshold would eliminate this benefit. That is, pitch should account for more than three times the variance of time when

using longer base durations. However, the current study's finding of a smaller advantage of pitch in atonal trials suggests exactly the opposite, instead that the strength of pitch accents was actually hampered by exceeding the temporal integration threshold. This interpretation is speculative, particularly because of the remaining methodological differences between Experiment 2 and Ellis and Jones.

The most obvious of these differences is the variable size of the pitch interval between adjacent notes in the present study. Instead of 1 semitone intervals within a group (or 5 semitones between groups), the current experiments use a wider variety of step sizes: 1 or 2 semitones within a group, and 7 or 8 between a group (when using the 5 scale step accent size). A by-product of this variability is that a wider diversity of pitch classes occur in the present study ($M = 7.5$ unique pitch classes in any given sequence, compared to $M = 5$ unique pitch classes in Ellis and Jones). Another difference is the chord prefix before the sequences, although there is no intuitive mechanism by which it could affect the relative strength of these accent types in the subsequent sequence. Nevertheless, the potential effect of these methodological differences remains a question for future research.

The more interesting finding of Experiment 2 is how tonality influenced the relative strength of pitch leap and duration accents. For tonal trials, pitch accents were nearly four times stronger than time ($\eta^2 = .64$ and $.14$, respectively), but only twice as strong in the atonal trials ($\eta^2 = .49$ and $.24$). These sequences were carefully constructed so as to avoid any difference between tonal and atonal sequences (other than their tonality, of course). In particular, the size and distribution of intervals was equal for tonal and atonal sequences, the chord prefixes had only minute (but tonally critical) changes, the presence/absence of tonality was not predictive of the pitch groupings, and no group boundaries coincided with a tonal accent. In short, tonality had nothing to do with grouping in these sequences, and moreover, as an abstract organisational principle of musical syntax, its presence or absence is irrelevant

to the surface features of music such as pitch leaps and duration accents. Nonetheless, tonality increased the strength of pitch accents and decreased the contribution of duration accents. This finding complements Prince, Schmuckler, et al.'s {, 2009 #1774} report that violating temporal expectancies affected pitch height comparisons only when the (irrelevant) intervening sequences were atonal, but not when the sequences were tonal. It also aligns with Dawe, Platt, and Racine (1995), who found that chord changes were ineffective as rhythmic cues for atonal sequences.

Respecting a learned organisational framework increases the informative value of a dimension. For example, tonality provides a structure that listeners use to encode a musical sequence {Cuddy, 1981 #90;Krumhansl, 1990 #219;Bharucha, 1983 #32;Boltz, 1989 #1113}. By providing this structure, the consequently more useful dimension of pitch becomes more salient and is prioritised in perceptual processing {Prince, 2009 #1774;Prince, 2012 #3090}. This prioritisation means that *all* cues of this dimension (in this case, pitch leap accents) are more influential in perceptual processing, at the expense of cues in other dimensions (e.g., time). In the current experiment, pitch became more salient in tonal sequences, increasing the contribution of pitch accents to listeners' perceived grouping ratings, as well as decreasing the weighting of duration accents. Tonality is specific to music, but the same principle of passive and culture-specific learning affecting dimensional processing also arises in language processing – cultural background influences the relative strength of pitch and time accents in language {Cumming, 2010 #2684;Iversen, 2008 #1764;Yoshida, 2010 #2560`, but see Hay & Diehl`, 2007;, #2562}.

Another possible explanation of the effects of tonality in Experiment 2 is that listeners simply paid more attention to pitch when the sequences were tonal. Listeners did not receive any instructions to privilege one dimension over another, but it is possible that they chose to attend more to pitch when the sequences were tonal. The findings of Experiment 1 argue

against this interpretation, as tonality affected neither the strength of pitch nor duration accents (when tested separately in baseline conditions). Further, one could in theory also argue the opposite – that when sequences were atonal, they would violate listeners' expectations of a normal tonal sequence, in turn drawing more attention to pitch. Obviously the data do not support this latter explanation, but it bears mentioning as it reveals that both explanations are equally arbitrary a priori. In any case, the role of attention is uncertain in this experiment. Clarifying this issue was the goal of Experiment 3.

Experiment 3: Selective attention in grouping

Experiment 2 found that the presence of tonality exaggerated the influence of pitch accents at the expense of duration accents, but the exact cause of this effect was unclear. Specifically, effects of tonality may have been caused by listeners attending more to pitch when sequences were tonal, or by increased salience of this dimension. In Experiment 3 participants received selective attention instructions in order to manipulate explicitly their attention toward one dimension and away from the other. If the effects of tonality are artefacts of attentional focus, then they should disappear when listeners consciously direct their attention to one dimension. There were two groups of participants in Experiment 3 – one set instructed to attend selectively to pitch accents, the other to duration accents (i.e., manipulated between-subjects). In every other way the design of the current experiment was the same as Experiment 2.

Selective attention is an effective technique to modify auditory processing {for a review`, see \Snyder, 2012 #2867}. With specific regard to tonality, brain responses to violations of harmonic expectancies in chord sequences are larger and faster when the listener is actively attending the stimuli {Loui, 2005 #2135}. Although harmonic expectancies are not the same as tonality (chord sequences could be tonal without having expectancies), establishing a key area (i.e., tonality) is a prerequisite for the formation of harmonic

expectancies. Lastly, selective attention instructions are effective at changing the relative contribution of manipulations of tonality and metre to melodic goodness ratings {Prince, 2011 #2468}. In particular, interactions between pitch and time were stronger in that study when the dimensions were more equally matched in their main effect size (eta-squared).

A dimensional salience hypothesis predicts that pitch accents should be more influential for tonal sequences than atonal sequences regardless of attentional instruction, so the current experiment should replicate the findings of Experiment 2 for both experimental groups. Neither attentional nor psychophysical factors (cf. Experiment 1) could explain such results. The only remaining explanation would be that the effects of tonality on metrical grouping represent changes in dimensional salience – the priority of pitch or time in perceptual process of forming the mental representation of a given sequence.

Method

Participants

There were 24 participants in Experiment 3. The average age was 26.7 years (SD = 11.5) and they had on average 2.9 years of musical training (SD = 3.9).

Stimuli

The stimuli of Experiment 3 were the same as those used in Experiment 2.

Apparatus

Experiment 3 apparatus were the same as the previous experiments.

Procedure

In Experiment 3, half of the participants received instructions to base their ratings on pitch accents (ignore duration accents), whereas the other half were told the opposite. A given participant experienced only one instruction type for the entire experimental session. To ensure that participants understood what exactly they were to attend and ignore, the experimenter explained the concept of pitch accents and duration accents prior to beginning

any trials, and played examples of each (using baseline sequences from Experiment 1). As in Experiment 2, the subsequent 4 practice trials consisted only of congruent pairings of pitch and duration accents. Thus the practice and experimental trials were exactly the same as Experiment 2, except for the addition of the selective attention instructions.

Results

The average intersubject correlation was .58 (SD = .08) for the Attend Pitch instruction, and .53 (SD = .07) for the Attend Time instruction, again showing consistency across participants. Ratings were again averaged across starting pitch (C, C#, D) and block (1, 2, 3), and then entered into a mixed-design ANOVA with the between-subjects variable of Instruction (2 levels: Attend Pitch, or Attend Time). The within-subjects variables were the same as Experiment 2: Pitch Grouping (3 levels: Duple, Neutral, or Triple), Time Grouping (3 levels: Duple, Neutral, or Triple), and Tonality (2 levels: Tonal, Atonal). A preliminary ANOVA included Block as a within-subjects variable but it did not exert a main effect, nor did it interact with another variable.

There was no main effect of Instruction, $F(1,22) < 1$, ns, nor was there a main effect of Tonality, $F(1,22) < 1$, ns. Conversely, there were strong main effects of both Pitch Grouping, $F(2,44) = 143.87$, $p < .001$, $\eta^2 = .51$, and Time Grouping, $F(2,44) = 68.69$, $p < .001$, $\eta^2 = .22$. These effects represent the expected higher ratings (more triple) for Triple groupings than Duple.

Interestingly, Instruction did not interact with Pitch Grouping, $F(2,44) = 1.31$, $p = .280$, $\eta^2 < .01$, nor was there an interaction between Instruction and Time Grouping, $F(2,44) < 1$, ns. These null results suggest that the perceptual effects of Pitch Grouping and Time Grouping were the same for both instruction groups (i.e., the instruction to attend one dimension was ineffective). However, the selective attention instructions may have been more successful at the beginning of the experiment than at the end. When analysing only the

first block of data, there was a marginal interaction between Instruction and Pitch Grouping, $F(2,44) = 2.68$, $p = .080$, $\eta^2 = .01$ and a significant interaction between Instruction and Time Grouping, $F(2,44) = 3.50$, $p = .039$, $\eta^2 = .01$, both in the expected direction. These effects shrank in analyses of the third (last) block of participants' data, $F(2,44) = 1.13$, $p = .332$, $\eta^2 = .01$ and $F(2,44) < 1$, ns, respectively. The most likely explanation of the failure of the selective attention instructions is the unfortunately commonplace phenomenon of participant fatigue. This explanation is tentative given that there were no significant effects of Block in the preliminary analysis.

The interaction between Pitch Grouping and Time Grouping recurred, $F(4,88) = 21.06$, $p < .001$, $\eta^2 = .04$, depicted in Figure 3b and 3c, which replicates Experiment 2. Also as in Experiment 2, Tonality interacted with both Pitch Grouping, $F(2,44) = 30.75$, $p < .001$, $\eta^2 = .01$ and with Time Grouping, $F(2,44) = 14.48$, $p < .001$, $\eta^2 < .01$. These interactions replicated the pattern from Experiment 2, where the effect of Pitch Grouping was larger in the tonal trials ($\eta^2 = .58$) than the atonal trials ($\eta^2 = .44$), and the reverse for the effect of Time Grouping ($\eta^2 = .16$ and .29, respectively). There was no 3-way interaction between Pitch Grouping, Time Grouping, and Tonality, $F(4,88) = 1.02$, $p = .402$, $\eta^2 < .01$, nor was there a 3-way interaction of Pitch Grouping and Time Grouping with Instruction, $F(4,88) < 1$, ns.

Although Instruction did not interact with Pitch Grouping or Time Grouping in 2-way interactions, there was a significant 3-way interaction between Instruction, Pitch Grouping, and Tonality, $F(2,44) = 6.56$, $p < .001$, $\eta^2 < .01$, as well as a marginal interaction between Instruction, Time Grouping, and Tonality, $F(2,44) = 2.91$, $p = .065$, $\eta^2 < .01$. The initial mixed ANOVA does not specify the nature of these three-way interactions, nor enable inspection of how effect sizes changed across instruction. Therefore, separate analyses for each instruction explored these interactions and provided a comparison of findings across instruction, using the same within-subjects analysis design of Experiment 2. These analyses

are detailed below, but can be summarised as the interactive effect of Tonality (as in Experiment 2) being stronger in the Attend Time instruction than Attend Pitch. The 4-way interaction (between all variables in the analysis) was thankfully not significant, $F(4,88) < 1$, *ns*.

Attend Pitch Instruction

The same 3 (Pitch Grouping) by 3 (Time Grouping) by 2 (Tonality) ANOVA design as in Experiment 2 was used to examine the data from the 12 participants who received the instruction to attend selectively to pitch accents in forming their grouping ratings. Not surprisingly, this analysis revealed significant effects of Pitch Grouping, $F(2,22) = 92.64$, $p < .001$, $\eta^2 = .59$ and Time Grouping, $F(2,22) = 25.71$, $p < .001$, $\eta^2 = .17$, and an interaction between them, $F(4,44) = 14.60$, $p < .001$, $\eta^2 = .04$, following the same pattern as the overall analysis immediately above. Of greater interest, the Pitch Grouping by Tonality interaction was significant, $F(2,22) = 17.04$, $p < .001$, $\eta^2 < .01$, as was Time Grouping by Tonality, $F(2,22) = 5.33$, $p = .013$, $\eta^2 < .01$. The 3-way Pitch Grouping by Time Grouping by Tonality interaction was not significant, $F(4,44) < 1$, *ns*. The 2-way interactions followed from the overall analysis – Pitch Grouping had a larger effect in the tonal trials ($\eta^2 = .62$) than the atonal trials ($\eta^2 = .56$), whereas Time Grouping had a smaller effect in the tonal trials ($\eta^2 = .15$) than the atonal trials ($\eta^2 = .21$). The difference scores as calculated in Experiment 2 revealed a similar, but smaller, pattern (Figures 4a and 4b). The 2 (Dimension) by 2 (Tonality) repeated measures ANOVA of the difference scores yielded a main effect of Dimension, $F(1,11) = 5.32$, $p = .042$, $\eta^2 = .28$ (again pitch was stronger overall), and an interaction, $F(1,11) = 23.59$, $p = .001$, $\eta^2 = .01$. Tonality was again not significant as a main effect, $F(1,11) = 1.57$, $p = .236$, $\eta^2 < .01$. As before, Block was not significant as a main effect in an additional ANOVA, nor did it interact with another variable. Finally, the paired two-tailed t-test analysis of the difference scores verified that pitch leap accents were stronger

for tonal sequences, $t(11) = 3.56$, $p = .009$, and duration accents were stronger for atonal sequences $t(11) = 3.88$, $p = .005$.

Attend Time Instruction

The data from the 12 participants who received the instruction to attend selectively to time accents in forming their grouping ratings was analysed in the same manner as the Attend Pitch Instruction group. Of course, the main effects of Pitch Grouping, Time Grouping, and their interaction were again significant, $F(2,22) = 55.24$, $p < .001$, $\eta^2 = .43$, $F(2,22) = 45.34$, $p < .001$, $\eta^2 = .27$, $F(4,44) = 7.91$, $p < .001$, $\eta^2 = .04$, following the same pattern as the previous analyses. There were 2-way interactions of Tonality with Pitch Grouping and Tonality with Time Grouping, which were stronger than in the Attend Pitch instruction, $F(2,22) = 19.20$, $p < .001$, $\eta^2 = .01$, $F(2,22) = 9.62$, $p < .001$, $\eta^2 = .01$, respectively. There was no 3-way interaction of Pitch Grouping by Time Grouping by Tonality, $F(4,44) < 1$, *ns*. As before, the effect size of Pitch Grouping was larger in the tonal trials ($\eta^2 = .52$) than the atonal trials ($\eta^2 = .31$), and Time Grouping showed the opposite pattern ($\eta^2 = .17$ and .37, respectively). Figure 4c depicts these difference scores, which when subjected to the 2 (Dimension) by 2 (Tonality) repeated measures ANOVA, revealed no main effect of Dimension (in contrast to the Attend Pitch instruction), $F(1,11) < 1$, *ns*, nor a main effect of Tonality, $F(1,11) < 1$, *ns*. The critical interaction remained, $F(1,11) = 46.96$, $p < .001$, $\eta^2 = .08$, and when analysed with two-tailed paired t-tests, again showed that tonality magnified the effect of pitch accents, $t(11) = 5.58$, $p < .001$, and attenuated that of duration accents, $t(11) = 4.26$, $p = .003$.

Block Effects

For the Attend Pitch instruction, there were no interactions with Block and any other variable, including Tonality and Dimension, $F(2,22) = 2.43$, $p = .111$, $\eta^2 < .01$. Thus the effects of tonality did not differ across block, despite a nominal trend towards tonality having the strongest effect in Block 2, as shown in Figure 5a. The only significant interaction was

Block with Tonality and Dimension for the Attend Time instruction, $F(2,22) = 5.86$, $p = .009$, $\eta^2 = .01$. Figure 5b depicts this interaction, showing that the effects of tonality were strongest in the first block (i.e., the slopes are steepest in Block 1). This figure also offers support to the participant fatigue explanation of the null Time Grouping by Instruction interaction reported earlier (third paragraph of the Results section). Specifically, duration accents appear equally effective as pitch accents in Block 1, but not for the remaining blocks. More simply, only the solid lines cross. Nevertheless, Block did not interact with Dimension, $F(2,22) = 2.34$, $p = .119$, $\eta^2 < .01$.

Insert Figure 5 about here

Discussion

One proposed explanation of the pitch-tonality and time-tonality interactions in Experiment 2 was that listeners may have chosen to attend more to pitch when sequences were tonal (see Experiment 2 Discussion). The results of Experiment 3 refute this explanation, because when participants received explicit instructions to attend selectively to one dimension, they had difficulty doing so. The instructions were marginally more successful at the beginning of the experiment, and the Dimension by Tonality analysis of the raw difference scores showed that although pitch accents were stronger overall in the Attend Pitch condition, they were equal for the Attend Time condition (compare Figures 4b and 4c). Therefore, instructions did have some measurable effect on ratings, but the effects do not suggest a robust ability to ignore either dimension. This limited effectiveness of *intentional* direction of attention to either dimension in Experiment 3 undermines the possibility that *unintentional* allocation of attention explains the results of Experiment 2.

Yet stronger evidence of this point emerges from Figure 5b, which demonstrates that earlier in the experiment (when the possibility of participant fatigue is lowest and attentional instructions were most effective), the tonality manipulation actually had its strongest effects. That is, variations in dimensional salience invoked by the presence/absence of tonality were actually most apparent when participants were better at ignoring pitch. If the effects of tonality were due to attention, then they should decrease with selective attention success – not increase. To elaborate, any shift of attention induced by tonality should be weakest when participants' selective attention is best, and stronger as fatigue reduces their resistance to its effects. However, the pattern is exactly the opposite of the predictions of an attentional account, dissociating the two processes. This finding strengthens the case for dimensional salience as independent from attention.

Further discrediting an attentional origin of Experiment 2's effects is the replication and extension of the interactions of interest in Experiment 3. Tonal sequences exaggerated the influence of pitch leap accents at the expense of duration accents, whereas atonal sequences decreased the effects of pitch leap accents in favour of duration accents. To be clear, pitch leap accents were always more effective, but were less or more so based on the tonality of the sequences, as described above.

The strength of the interactions between dimensions and tonality also varied across instruction, as evidenced by two separate 3-way interactions (Instruction by Pitch Grouping by Tonality; Instruction by Time Grouping by Tonality). These interactions showed that the influence of tonality on the relative effect size of pitch and duration accents was larger for the participants instructed to attend solely to duration accents. Consult Figure 4b and 4c to visualise this effect – in the Attend Time instructions, the gap between pitch leap and duration accent difference scores changed more from atonal to tonal trials than in the Attend Pitch instructions. In other words, the slopes are steeper in Figure 4c than 4b.

Overall, Experiment 3 replicated the finding of Experiment 2 that the presence of tonality heightens the salience of the pitch dimension as a whole, such that pitch leap accents become more effective despite being exactly the same size in both tonal and atonal sequences. Concomitantly, duration accents become less effective when sequences are tonal, even though they were exactly the same physical size across tonality. These findings were not abolished by instructions to shift attention toward one dimension or another. Instead, they are consistent with a dimensional salience hypothesis {Prince, 2009 #1774;Prince, 2009 #1609} that prioritises a more informative dimension in constructing the mental representation of an external stimulus.

General Discussion

These three experiments reveal how tonality influences the relative strength of pitch and duration accents on ratings of perceived metrical grouping. Finding pitch leap accents and duration accents of equivalent strength was the goal of Experiment 1. Experiments 2 and 3 recombined these equalised accent types in sequences such that pitch and duration accents were either congruent, unbiased, or incongruent. When recombined, pitch leap accents always predominated over duration accents in their influence on participants' ratings. But more importantly, the advantage of pitch accents was larger when the sequences were tonal and smaller for atonal sequences, whereas duration accents followed the opposite pattern. These results could not be accounted for by psychophysical differences, because they used accents that were equally strong across dimensions (cf. Experiment 1). Experiment 3 ruled out an attentional origin of these effects, showing that the effects of tonality remained when participants had instructions to ignore one dimension.

Failures of selective attention can provide evidence of integral dimensions {Garner, 1974 #1147}, but definitive conclusions about the separability of dimensions require convergent evidence from different tasks and processes {Garner, 1976 #1547}. The results of

Experiment 3 do not permit strong interpretations of the independent versus interactive nature of pitch and time in music cognition, partly because the instruction manipulation did have some observable effects (in early blocks), and also because participant fatigue may have contributed to the small effect sizes. Indeed, auditory selective attention generally becomes more effective with practice, not less {Best, 2008 #3207}. Regardless, a critical prerequisite to reaching conclusions about how dimensions combine is ensuring that the dimensions are equally effective when tested separately, in baseline conditions. Other than the current research, the only study to do these baselines in tests of metrical grouping is Ellis and Jones (2009). They also found a dominance of pitch over time, and their methodology corresponds most closely with the atonal sequences of Experiments 1 and 2 in the current findings (they used neither tonal sequences nor selective attention instructions). Interestingly, their analyses of metrical clarity ratings suggested an interactive pattern of pitch-time integration, yet their perceived grouping ratings (as used in the current research) supported an independent relationship.

Combined with the other literature on pitch-time integration, it is clear that attempting to determine if pitch and time are independent or interactive is not the right question. Instead, the more interesting and generally applicable psychological issue is *why* the dimensions combine in different fashions. Discarding the view of a dichotic classification of these possibilities, several authors have explored how pitch-time integration can vary {Tillmann, 2006 #1072;Prince, 2011 #2468;Prince, 2009 #1774;Schellenberg, 2000 #1081}. For example, Schellenberg et al. mooted the possibility that subjective tasks (goodness of fit ratings, liking, etc.) give rise to interactive patterns, whereas objective tasks (classification, recall, etc.) foster independence. Yet both patterns can occur in both subjective {Palmer, 1987 #1129;Boltz, 1989 #1117} and objective {Jones, 1982 #1179;Smith, 1989 #1116} tasks. Tillmann and Lebrun-Guillaud (2006) propose that global tasks lead to interactive

relations, and local tasks favour independence {see also \Jones, 1989 #205}. Some arguably

global tasks can result in independence, and local tasks in interactions {Prince, 2011

#2468;Prince, 2009 #1609}, but this interpretation requires further exploration.

　　　　Another possible explanation of the variations in pitch-time integration comes from

the concept of dimensional salience, developed in the context of music perception {Prince,

2009 #1609}, but with more general applicability to other auditory dimensions, task domains,

and sensory modalities. What makes a dimension more salient? It is likely that numerous

variables influence dimensional salience, but the mechanism with the most support thus far is

based on the extent to which it adheres to a learned schema. This concept converges with the

Fuzzy Logical Model of Perception developed in the context of speech perception, in which

stimulus dimensions are weighted in accordance with the degree to which they match a

learned prototypical representation of a speech sound {Oden, 1978 #2899;Massaro, 1990

#2909}. There is also some overlap with the TRACE model {McClelland, 1986 #3590} in

that one source of information (e.g., pitch) can influence the sensitivity to another source

(e.g., time). However, the interactive-activation nature of the TRACE model (activating

higher-order units influences lower-order, and vice versa) predicts effects of context on

perceptual sensitivity {Massaro, 1989 #2904}, whereas demonstrations of dimensional

salience occur only when differences in discriminability are eliminated (via baseline testing).

For instance, Prince, Thompson, et al. (2009) found that the effects of tonality on temporal

classification were in response bias, not discriminability, aligning more with the principles of

the FLMP {Massaro, 1989 #2904}. Cutting's directed perception model {Cutting, 1986

#3589} also has relevance as it allows adjustable weightings of information sources, but it

also assumes a Boolean additive integration, whereas the current data show that neither

dimension could be ignored entirely. In the current study, the manipulation of adhering to a

prototypical structure was accomplished by varying the presence of tonality, the central

organisational principle of musical pitch {Krumhansl, 2010 #2873}. Adhering to this hierarchical structure makes the dimension of pitch more informative, and more useful as a framework for encoding additional information, such as the duration accents of the present study.

An inherent advantage of pitch over time in Western music has both theoretical {Schenker, 1935/1979 #2374;Everett, 2000 #2372} and experimental {Krumhansl, 2000 #218;Hébert, 1997 #640;Prince, 2009 #1609} support. A learned preferential weighting of pitch may occur as the stimuli increasingly resemble Western music – such as using tonal sequences. In the context of visual object identification and categorisation, selective attention can increase the discriminability of items along a given stimulus dimension at the expense of another dimension {Nosofsky, 1986 #3188}. According to Goldstone {, 1998 #2969}, this "attentional weighting" of dimensions may occur via perceptual or strategic processes. In the current research tonality provided neither a perceptual advantage nor a strategic benefit for pitch – it was truly irrelevant to the task. Thus dimensional salience is not simply an outgrowth of attentional processes highlighting pitch in tonal sequences.

Previous research has also shown the influence of tonality in pitch-time integration {Prince, 2009 #1774;Dawe, 1995 #1604}, and has manipulated metre, the homologous structure in the time dimension {Tillmann, 2006 #1072;Prince, 2011 #2468}. Yet other work explores the role of diversity, or the number of unique categories defined within a dimension {Prince, 2012 #3090;Sarrazin, 2007 #3155}; research in visual perception has also tested how diversity (quantity) affects dimensional processing {Melara, 1994 #1164}, but not dimensional salience *per se*. Tonality and metre are examples of schemas that provide organisational frameworks for use in perceiving musical stimuli specifically, but the applicability of dimensional salience is not limited to this domain - work in visual perception and attention provides convergent evidence of perceptual organisation influencing visual

processing automatically {Kimchi, 2009 #2447;Kimchi, 2007 #2177;Strother, 2012 #3274;Russell, 2005 #3306;Hermens, 2010 #3304;Parton, 2001 #2500;De Freitas, 2013 #3375}. Moreover, the concept of dimensional salience fits well with work on asymmetric dependencies in other musical dimensions {Cousineau, 2009 #2164;Tekman, 2002 #1160;Krumhansl, 1992 #1101;Warrier, 2002 #2872}, speech perception {Tong, 2008 #1763;Mullennix, 1990 #1708;Green, 1997 #3291}, and face perception {Graham, 2007 #3289;Zhao, 2013 #3294;Karnadewi, 2011 #3292}.

There are some limitations of these findings that deserve consideration. First, the current study (and previous literature) used only one measure of how strong each dimension was – the perceived grouping rating. Despite the careful procedure of equalising the dimensions in baseline conditions, a stronger test of relative dimension strength might include some external measure of the effects of pitch and time. For instance, a matching and/or adjustment technique could provide a more sensitive and hopefully convergent observation of these effects. Objective measures of processing facilitation (e.g., tapping) in accordance with metrical structure could also be useful.

Second, it is possible that ceiling/floor effects restricted the range of observable effects, particularly when pitch and duration accents were congruent. Perhaps the dimensions were not as equally matched as hoped, qualifying the conclusions regarding which dimension was stronger. The fact that these results converge so well with previous observations of metrical grouping (Ellis & Jones, 2009) provides some counter-evidence to this criticism. More importantly, however, the most interesting result is not about which dimension is stronger as a main effect but about how the presence of structure manipulates their *interaction*. In fact, the presence of floor/ceiling effects would only make for a more conservative test of this phenomenon, as the resultant range restriction would compress the

raw difference scores for both the tonal and atonal trials, underestimating the observed effects across tonality.

Third, in the baseline tests of Experiment 1, sequences had either pitch leap accents or duration accents – not both (by definition). Whereas pitch leaps create a clear grouping such that the first note after a leap begins the group, the interpretation of duration accents is more flexible – it is possible to perceive the longer note as the first *or* last note of a group {Iversen, 2009 #2748}. This flexibility may undermine the strength of duration accents when combined with the unambiguous pitch leap accents. Regardless, this interpretation remains consistent with the concept of dimensional salience – a dimension with more reliable cognitive reference points (in this case, pitch) provides a more robust framework for encoding stimulus information. In the context of face perception, facial identity is an invariant attribute, functioning as a more stable reference point than facial emotion {Haxby, 2000 #1704}. Accordingly, aspects of facial identity such as sex asymmetrically interfere with judgements of facial emotion {Atkinson, 2005 #1698}.

Fourth, only one tempo (120bpm, or 500ms between notes) was used in the current experiments. Listeners synchronise most easily to (and produce spontaneously) tempos with intervals of 600ms, but can entrain to a wide range of tempos {McAuley, 1999 #2540}, as long as they remain between the range of 300-1800ms apart {Fraisse, 1982 #2224}. Changing tempo can alter the metrical interpretation of rhythmic sequences {London, 2009 #2165;Handel, 1983 #1660}. For instance, Hannon et al. {, 2004 #178} found that listeners favoured interpretations of 3/4 over 6/8 for slower tempos. In the current experiments, the triple grouping would complete a full cycle after 1500ms, and the duple grouping would be every 1000ms. The duple is closer to the ideal 600ms interval, but both are well within the range of perceivable tempos. There was no strong evidence of an overall preference for duple versus triple groupings, as the grand mean average rating for each experiment was

indistinguishable from the midpoint of 3 on the 1-5 scale (Experiment 1: 2.97, Experiment 2: 2.81, Experiment 3: 2.96 and 2.97). It is unlikely, but as yet untested, that a different tempo would change the effects of tonality on perceived grouping.

Future research might examine the effects of different types of structure, and in other dimensions. This study varied the presence of tonality, the central organisational principle of musical pitch. Varying the presence of metre (whether sequences conform to a regular pulse or beat) may have similar effects {but cf. \Tillmann, 2006 #1072}. The dimensional salience framework would predict an affirmative answer, as it does not differentiate between which dimension exhibits greater structure, consistent with Pressing's notion of cognitive isomorphisms {, 1983 #2574}. Both tonality and metre are abstract organisational forms of structure, whereas more surface-based forms of structure may also play a role in dimensional processing {Melara, 1994 #1164;Prince, 2012 #3090;Schellenberg, 2013 #3268}. It may also be useful to explore the role of training on pitch-time interactions in metrical grouping. Although perceptual tests tend to reveal only quantitative (not qualitative) differences across expertise {Bigand, 2006 #1615}, it would be reassuring to verify that metrical grouping is not an exception. Another line of enquiry could explore how pitch-time integration changes in contexts where the dimensions are purposely unbalanced. Garner-type interference may occur {Garner, 1974 #1147}, or perhaps a different pattern could emerge when dealing with musical sequences, as they provide their own self-contained context. Whatever the application, investigations of dimensional salience must first equalise the dimensions in terms of their baseline discriminability (cf. Experiment 1, Ellis & Jones, 2009, and Prince, Thompson, et al., 2009) in order to distinguish the otherwise potentially conflated factors of discriminability and salience.

The present research expands on the concept of dimensional salience in pitch-time integration by manipulating the presence of irrelevant pitch structure in the global temporal

task of metrical grouping. Consistent with earlier findings, tonal sequences have the effect of prioritising the dimension of pitch, even when all other aspects of the stimuli are carefully controlled. Furthermore, this prioritisation cannot be attributed to unequal discriminability, or allocation of attention (neither conscious nor inadvertent) to pitch in the context of tonal sequences. These findings, and their implications for the dimensional salience framework, have applications beyond the domain of pitch-time integration in music because all stimuli in our natural environments are multidimensional. The process of integrating these dimensions is common to all perceptual modalities and is thus a crucial component of cognition.

References

Author Notes

Address correspondence concerning this article to Jon B. Prince, School of Psychology and

Exercise Science, Murdoch University, 90 South Street, Murdoch, WA 6150 (email:

j.prince@murdoch.edu.au).

Footnotes

[1] Note that stimuli used only whole number intervals; the mean is provided for convenience of comparison.

[2] All eta-squared values are full eta-squared (i.e., SS_{effect}/SS_{total}), not partial eta-squared, as the latter measure inflates the reported value.

[3] Bonferroni corrections were applied to all t-tests.

Table 1

Confidence intervals of accent levels tested in Experiment 1. Underlined values are the levels chosen for subsequent experiments.

Dimension	Grouping	Level	Mean	Lower Bound	Upper Bound
Pitch	Duple	1 (3 steps)	1.63	1.27	1.98
		2 (4 steps)	1.48	1.16	1.81
		3 (5 steps)	1.30	1.07	1.53
	Triple	1 (3 steps)	3.99	3.66	4.32
		2 (4 steps)	4.55	4.26	4.85
		3 (5 steps)	4.57	4.30	4.84
Time	Duple	1 (250ms)	1.78	1.33	2.23
		2 (333ms)	1.28	1.06	1.49
		3 (450ms)	1.43	0.93	1.93
	Triple	1 (250ms)	4.15	3.83	4.46
		2 (333ms)	4.66	4.46	4.85
		3 (450ms)	4.81	4.63	5.00

Figure captions

Figure 1. Example sequences with pitch leap accents of five scale steps, including irrelevant three-chord prefix. (A) Tonal sequence, duple grouping (accent every two notes); (B) tonal sequence, triple grouping (accent every three notes); (C) atonal sequence, duple grouping; (D) atonal sequence, triple grouping.

Figure 2. Perceived grouping of triple and duple accents utilising pitch leap and duration accents. The x-axis indicates the accent size (duration of accented note in group or scale step size between groups). Duration values are the length of the accented note; all other notes in a group were 200ms long. The y-axis indicates the perceived grouping, where lower numbers correspond to duple groupings and higher numbers to triple groupings. Error bars are standard error of the mean.

Figure 3. Interaction of Pitch Grouping and Time Grouping across experiment. The x-axis represents the duration accent pattern; series are patterns of pitch accent. (A) Experiment 2, with no instructions to attend either dimension; (B) Experiment 3, Attend Pitch participant group; (C) Experiment 3, Attend Time participant group. Error bars are standard error of the mean.

Figure 4. Difference score of perceived grouping (mean triple grouping rating minus mean duple grouping rating) as a function of tonality and dimension. The x-axis shows whether sequences were tonal or atonal, and series denote the accent type (pitch leap or duration). Higher scores indicate a more effective accent cue (greater rating difference between triple and duple groupings). (A) Experiment 2, with no instructions to attend either dimension; (B) Experiment 3, Attend Pitch participant group; (C) Experiment 3, Attend Time participant group. Error bars are standard error of the mean.

Figure 5. Effect of block on how tonality influenced the strength of pitch and duration accents in Experiment 3. (A) For the Attend Pitch instruction, this interaction was not significant; the overall effect of tonality was as shown in Figure 4b. (B) For the Attend Time instruction, pitch and time accents were equally strong only in the first block (i.e., only the solid lines cross). The effect of tonality on the pitch/duration accent balance remained the same across block, but was nominally largest in Block 1, when selective attention instructions were most successful. Error bars are standard error of the mean.

a)

b)

c)

d)

a) Experiment 2
(no selective attention instructions)

Pitch Triple
Pitch Neutral
Pitch Duple

b) Experiment 3
Attend Pitch Instruction

Pitch Triple
Pitch Neutral
Pitch Duple

a) Experiment 3 Attend Pitch
Block*Tonality*Dimension

b) Experiment 3 Attend Time
Block*Tonality*Dimension