Investigating the Perception of Pitch and Volume Changes in Rhythm

Steven D. Shaw

A Biology 4999E Thesis Submitted to

Western University, London, Ontario in Partial Fulfillment of the Requirements for the

Degree of

Bachelor of Science

Supervisor: Dr. Jessica Grahn

Signature of Approval: _____________________________

Advisors: Dr. Scott MacDougall-Shackleton and Dr. Bryan Neff

April 30th, 2013
Abstract

Dynamic Attending Theory (DAT) holds that attention is highest at musical time-points coinciding with the beat. According to DAT, on-beat attentional maxima aid in the detection of stimuli characteristics (e.g. volume changes). However, DAT may overlook emphasis due to expectancy for on-beat salience. Two change detection tasks were used to determine participants’ ability to detect on- and off-beat pitch and volume changes in rhythms. The hypothesis was that pitch and volume change detection would differ based on whether changes occurred on- or off-beat in a rhythm. Masking, due to expectancy of on-beat external salience, was predicted to hinder on-beat volume change detection, but not pitch change detection. No significant differences between rhythm type/ change location for pitch or volume thresholds were observed. These results suggest that any attentional maxima in auditory cognition, as proposed by DAT, or masking, due to expectancy effects, are not behaviourally relevant when listening to music.
Investigating the perception of pitch and volume changes in rhythm

Auditory and musical perception are subjective experiences in humans. Individuals pay attention (i.e. attend) to their current auditory landscape differentially. As a result, musical experience and perception can vary greatly between individuals (Matsubayashi et al., 2008). Survival-based explanations of auditory attentional mechanisms are quite clear for auditory stimuli; threatening auditory stimuli are prioritized to maximize information regarding a threat and increase the potential of avoidance (Numan Ermutlu et al., 2005; Boyera and Bergstrom, 2011). For musical perception however, threatening stimuli are not present, and the mechanisms of attentional differences are not clearly understood. This experiment investigates how pitch and volume changes in music are related to how humans perceive and attend to musical rhythms.

Auditory patterns unfold over time and because of this, individuals are able to establish a sense of beat and use this beat perception to anticipate future auditory events on the basis of previous auditory patterns. In rhythms, constant time intervals between accented notes maintain a beat and thus listeners can anticipate the beat. This anticipation of the beat has consequences for musical perception (Jones et al., 2002).

Dynamic attending theory (DAT) is a framework of these consequences; it attempts to explain how individuals perceive and analyze complex auditory sequences. Musical sequences are assembled in a hierarchical structure, and DAT holds that individuals have a tendency to pick up on regular time relationships within this musical hierarchy. The hierarchical structure of music, illustrated in Figure 1, is a way of understanding note patterns by looking at how the beat can be broken down into smaller
and smaller subdivisions. Individuals develop attentional reference points in musical sequences based on the auditory reference level (equivalent to the musical term ‘tactus’). Listeners attend to the hierarchical level at which they are currently referenced, however this level does not necessarily remain static. Attending only occurs on a single hierarchical level but listeners can transition between hierarchical levels from the current reference level of attending (Drake et al., 2005).

The regularity of reference points plays a key role in DAT. It is proposed that listeners focus their attention on notes of increased emphasis or salience (reference points) at a given hierarchical level throughout a musical sequence. In addition, the same time interval is expected between reference points, establishing a beat. Attention and expectancy are thought to increase nearing the reference point, creating a maximum attentional pulse that is centered on the reference point (see Figure 2; Jones et al., 2002). With regular rhythms, these expectancies can affect performance in auditory judgment tasks. People can more accurately judge the time intervals of auditory sequences when the ending occurs at the expected time compared to those that end earlier or later than expected (Barnes and Jones, 2000).

Some support for DAT has been shown through previous studies of beat perception. Reference point accentuations are commonly achieved in music by using notes that are louder than surrounding notes on reference points. The louder note is an example of a note that is externally salient. That is, the note is quantifiably louder than the notes surrounding it; it has a larger amplitude. Previous studies have shown that changes in external salience are not the only way to convey a sense of beat. If all aspects of an auditory sequence are held constant, for example pitch and volume, humans can
perceive beats in temporal patterns (Parncutt, 1994). Duration differences between the notes lead to some being perceived as more salient, without any increases in external salience. The timing differences create accented notes that lead to a sense of underlying beat.

One potential flaw in DAT, however, is that it does not account for the expectancy of salience on the beat. Expectancy may in fact mask the perception of true increases in external salience (e.g. a volume change). If this were the case, a note with true external salience, in the form of a volume change on-beat, would be masked (hindering detection) when it occurred at an expected beat location. Furthermore, detection of other aspects that are not expected on-beat, such as pitch change, would be facilitated. That is, since there is no expectancy for other types of stimuli changes on-beat (e.g. a pitch change), no masking would occur and the attentional mechanism proposed by DAT would remain the most important aspect of beat perception.

Expectancy in musical hierarchy has previously been shown to relate to masking effects in speech recognition tasks (Shi and Law, 2010). When asked to recognize words or sentences in the presence of music intended to mask the test stimuli, participants performed better when the masking music had a metric structure (a greater degree of expectancy present). Shi and Law (2010) propose that individuals can take advantage of the regularity in the music (cognitively) when this masking effect occurs. More cognitive resources can be diverted to distinguishing the words or sentences because less attention is needed for the highly predictable musical sequence. The key finding is that expectancy effects can potentially mask aspects of auditory perception and both expectancy and masking are related to reference point attending.
Similarly, speech audition has been shown to be subject to the “picket-fence effect”. When an individual is listening to soft speech, a separate interrupting loud tone or short interrupting quiet period does not disrupt the continuity of the spoken sentence. Rather, the spoken sentence is perceived as continuous and the interrupting tone is perceived as a separate event (Miller and Licklider, 1950). A similar effect when listening to music could occur; a continuous musical sequence with an interrupting volume change could result in the volume change being perceived as a separate event. Since the musical sequence is the primary focus of attention, detection of the volume change is masked due to its secondary importance to the individual.

Naturally, in order to perceive a beat, the music or rhythm being listened to must actually contain a musical structure that has a beat. Regular, or metric rhythms, use time intervals related by ratios of 1:2:3:4. Within this category of metric rhythms, simple and complex rhythms can be identified by the regularity or irregularity of accented intervals. Metric simple rhythms inherently induce a sense of beat, establishing clear reference points (on-beat notes) and non-reference points (off-beat notes). Metric complex rhythms do not induce a sense of beat, and thus reference points are not clearly defined. Metric simple rhythms use integer-ratio relationships between time intervals and have accented notes occurring at regular intervals. Metric complex rhythms also have integer-ratio relationships, but have irregular intervals between accented notes (see Figure 3; Grahn and Brett, 2005). The main difference between a metric simple and a metric complex rhythm is that metric simple rhythms induce a sense of beat due to regular intervals of accented notes, whereas metric complex rhythms do not.
A great deal of previous research has been conducted on how humans pay attention to different aspects of music. Dynamic Attending Theory remains a central dogma of auditory beat processing, however not all aspects of this theory have been fully tested. The purpose of my experiment was to determine whether internal expectancy of salience on the beat can mask the presence of external salience.

For the purpose of this experiment, the term on-beat refers to the times that coincide with notes in the implied beat of an auditory sequence (also called reference notes). The term off-beat refers to the times that do not coincide with notes in the underlying imposed beat. To determine if expectancy of salience on-beat can mask true external salience, two tasks were used, pitch and volume change detection, for three rhythm type/change locations: metric simple on-beat, metric simple off-beat, and metric complex off-beat. The dependent variable was the threshold of detection, or the smallest change in frequency (Hz) and amplitude (dB SPL) in the pitch and volume change detection tasks, respectively.

I hypothesized that in the pitch detection task, participants would have the smallest thresholds (most accurate detection) in the metric simple on-beat condition, followed by the metric simple off-beat and metric complex off-beat conditions respectively. In the volume change detection task, the reverse was predicted; participants should have the most accurate thresholds in the metric complex off-beat condition followed by metric simple off-beat, and metric simple on-beat. These predictions arose from a potential oversight in DAT; the expectancy of external salience on-beat may mask true increases in external salience. Thus, a note with true external salience in the form of a volume change on-beat would be masked (hindering detection), whereas detection of
other aspects that are not internally expected on-beat, such as a pitch change, would be facilitated.

Methods

Participants

Thirty-nine undergraduate students (26 males and 13 females) ranging in age from 18 to 26 years of age ($M = 19.12$, $SD = 1.67$) from the University of Western Ontario subject pool participated as part of the requirements for undergraduate psychology courses. One research study credit was given towards their psychology course as compensation. Participants were tested individually in an isolated room, had no prior experience with the experiment, and completed all conditions ($n = 39$). Participant testing took approximately one hour and all procedures were carried out with approval from the Psychology Research Ethics Board (PREB) at Western University.

Stimuli

Fifteen sequences for each rhythm type/ change location condition were generated using MATLAB (MathWorks). Each sequence was composed of 10, 12, or 14 intervals with 50 ms tones, separated by intervals that varied from 250, 500, 750, or 1000 ms. A single pitch or volume change was introduced into each sequence, on-beat or off-beat for metric simple rhythms, and off-beat for metric complex rhythms. All changes occurred in the second half of the rhythm to ensure participants were accustomed to the sequence and had ample time to recognize the beat (if applicable). Metric simple rhythms easily established a beat in listeners based on accentuated note placement, while metric complex rhythms did not establish a beat (Grahn & Brett, 2007).
Volume change stimuli were created at a constant pitch, 440 Hz. The baseline volume was 0.2 MATLAB volume units. Seventy equally sized volume changes of 0.1 MATLAB volume units were generated for each of the 15 rhythmic sequences per rhythm type/change location. Similarly, pitch change stimuli were created at constant volume, 0.2 MATLAB volume units. The baseline pitch was 440 Hz. Seventy 1 Hz pitch changes were generated for each of the 15 rhythmic sequences per rhythm type/change location (He et al., 2012).

**Testing Procedure**

Stimuli were presented on a computer over professional-grade headphones using E-Prime 2.0 software suite (Psychology Software Tools, Inc.) in a two-down one-up adaptive staircase design. The amount of frequency change (pitch) and amplitude change (volume) was adjusted between trials based on participants’ responses to the previous trial in that particular staircase. For added control, catch trials were inserted randomly every one in seven stimuli in the staircase. Catch trial stimuli were exactly the same rhythmically as rhythm type/change location pitch and volume change stimuli, but did not contain a pitch or volume change, and were selected from randomly. Pitch and volume thresholds were obtained separately, in a counterbalanced fashion between participants, and rhythm type/change location stimuli were intermixed within each task type.

In a given rhythm type/change location staircase, a reversal was recorded each time the participant recorded an incorrect answer following two previous correct answers, or, when the previous answer was incorrect, upon two subsequent correct answers. The initial step size used was 25 stimuli for both the pitch and volume staircases. After two
reversals, the step size changed to five stimuli, and after five reversals the step size changed to one stimuli. Step size refers to the stimuli change size (and indirectly Hz or dB SPL change size) after two correct answers in the downward direction, or one incorrect answer in the upward direction. Each rhythm type/ change location staircase concluded after 14 reversals and the threshold was determined as the average across the last six reversals in a given condition.

Upon the completion of both pitch and volume threshold detection tasks participants completed a standardized musical questionnaire. The questionnaire contained general questions about musical experience and experimental strategies/ problems encountered.

Data Analysis

A repeated measures analysis of variance (ANOVA; with Greenhouse-Geisser correction) was conducted in SPSS (IBM) for pitch and volume detection tasks, each with three levels of rhythm type/ change location: metric simple on-beat, metric simple off-beat, and metric complex off-beat. The dependent variable was the threshold of detection, or the smallest detectable change in frequency (Hz) or amplitude (dB SPL).

Musical training of participants was divided into three categories: less than one year of musical training, between one and five years of musical training, and greater than five years of musical training. A repeated measures ANOVA, using musical training as a between-subjects factor was performed to assess any effect of musical training.

Means and standard errors were calculated for each experimental group and all statistics were assessed using a significance criteria of alpha = 0.05.
Results

Pitch Detection Task

A Mauchly’s test of circularity of the covariance matrix was conducted and was not significant, therefore circularity of the covariance matrix was assumed, Mauchly’s $W = 0.93$, $\chi^2(2) = 2.58$, $p = 0.28$. Pitch thresholds did not significantly differ between metric simple on-beat ($M = 11.97$ Hz, $SE = 2.04$ Hz), metric simple off-beat ($M = 11.44$ Hz, $SE = 2.06$ Hz), and metric complex ($M = 12.18$ Hz, $SE = 2.06$ Hz) rhythms, $F(2, 72) = 0.18$, $p = 0.83$, $\eta^2_p = 0.005$, observed power = 0.08. That is, no main effect of rhythm type/change location was observed for the pitch detection task. Mean pitch thresholds are depicted in Figure 4.

Volume Detection Task

A Mauchly’s test of circularity of the covariance matrix was conducted and was significant, therefore circularity of the covariance matrix was not assumed, Mauchly’s $W = 0.73$, $\chi^2(2) = 11.91$, $p = 0.003$. Volume thresholds did not significantly differ between metric simple on-beat ($M = 3.50$ dB SPL, $SE = 0.25$ dB SPL), metric simple off-beat ($M = 3.55$ dB SPL, $SE = 0.31$ dB SPL), and metric complex ($M = 3.48$ dB SPL, $SE = 0.32$ dB SPL) rhythms, $F(2, 60) = 0.12$, $p = 0.84$, $\eta^2_p = 0.003$, observed power = 0.07. That is, no main effect of rhythm type/change location was observed for the volume detection task. Mean volume thresholds are depicted in Figure 5.

Catch Trials

A median split of pitch detection task catch trial accuracy ($M = 87.12\%$, $SE = 2.11\%$) showed that rhythm type/change location had no effect on pitch thresholds for participants above the median catch trial accuracy, $F(2, 24) = 1.31$, $p = 0.32$, $\eta^2_p = 0.06$. 
observed power = 0.19, or below the median catch trial accuracy, $F(2, 36) = 0.72, p = 0.50, \eta^2_p = 0.04$, observed power = 0.16. Participants pitch change thresholds for metric simple on-beat, metric simple off-beat, and metric complex conditions did not differ regardless of whether they were better or worse than the median performance in catch trial accuracy.

Similarly, a median split of volume detection task catch trial accuracy ($M = 94.28\%, SE = 1.01\%$) showed that rhythm type/ change location had no effect on volume thresholds for participants above the median catch trial accuracy, $F(2, 24) = 0.38, p = 0.60, \eta^2_p = 0.02$, observed power = 0.10, or below the median catch trial accuracy, $F(2, 35) = 0.36, p = 0.69, \eta^2_p = 0.20$, observed power = 0.10. Participants volume change thresholds for metric simple on-beat, metric simple off-beat, and metric complex conditions did not differ regardless of whether they were better or worse than the median performance in catch trial accuracy.

**Musical Training**

Musical training had no significant effect on pitch thresholds, $F(2, 36) = 2.28, p = 0.12, \eta^2_p = 0.11$, observed power = 0.43. Similarly, musical training had no significant effect on volume thresholds, $F(2, 36) = 0.45, p = 0.64, \eta^2_p = 0.024$, observed power = 0.12. That is, pitch and volume thresholds did not differ between individuals who had less than one year, between one and five years, or greater than five years of musical training.

**Discussion**

My hypothesis was that the detection of on-beat volume changes would be masked by individuals’ expectancy for external on-beat salience. In general, the results of
this study were not consistent with my hypothesis. Pitch and volume thresholds for metric simple on-beat, metric simple off-beat, and metric complex off-beat conditions did not significantly differ. That is, participants did not differ in their ability to detect pitch or volume changes regardless of the rhythm type/change location.

Interestingly, although these results do not support the hypothesis of this study, which proposed a flaw in Dynamic Attending Theory (DAT), the results do not support the outcome that would be predicted by DAT either. According to DAT, the detection of all types of stimuli changes (including pitch and volume changes) would be facilitated when these changes occur on the beat due to attentional maxima. Thus, the metric simple on-beat condition would have lower thresholds (better change detection) for both pitch and volume (Jones et al., 2002), but this result was not observed.

Previous work by Jones et al. (2002) had participants attempt to detect pitch changes between an initial note and a final note that were separated by a series of isochronous tones to create a rhythm. The final note, which was the test tone that participants needed to judge in its relative pitch, was manipulated to occur at expected or unexpected timing. Participants were better at detecting pitch changes when the test tone occurred at an expected timing following the isochronous rhythm when compared to an unexpected timing. This result would equate to better pitch detection in the metric simple on-beat condition in the present experiment, a result that was not observed. The results of previous work by Jones et al. (2002) are in contrast with the pitch change detection task results obtained in the present study.

A recent study by Repp et al. (2012) suggests that metrically created accents in music, similar to the ones used in the present study, may not induce the perception of
physically prominent events (e.g. a louder note). The researchers induced a beat using a series of isochronous tones and then presented two experimental probe tones. One or none of the probe tones fell on the induced beat, and the probe tones were manipulated to have different possible volume intensity relationships (equal volume or one note louder than the other by various magnitudes). Participants were asked to identity which probe tone was louder. The results of the study did not consistently find that on-beat probe tones were judged more accurately than off-beat tones. This result is comparable to the volume detection task results found in the present study, as no difference in volume detection between on and off the beat conditions was observed.

One of the major differences between my study and both the Jones et al. (2002) and the Repp et al. (2012) studies, is that the previous work used isochronous rhythms, whereas the present study used metric rhythms. This could explain the discrepancy between previous work and the results of this study (some results being comparable while others are in contrast). Isochronous rhythms have exactly equal spacing between notes and a note occurs at every beat location. For this reason, they are highly standardized, but much less similar to popular music. Metric rhythms, as described previously, use metric timing and are similar to the structure of popular music. The fact that previous research using similar tasks with isochronous rhythms has produced varying results when compared to this experiment, using metric rhythms, could suggest that the cognitive mechanisms by which we perceive changes in pitch and volume in isochronous rhythms are different than those used in metric rhythms, and subsequently popular music.

The addition of catch trials (control stimuli that do not have pitch or volume changes) into the threshold staircase could provide an alternative explanation for the lack
of observed differences between rhythm type/ change location conditions. Prior to this experiment, I conducted a pilot study. In the pilot study, all aspects of the experiment were the same, except that no catch trials were present in the staircase procedure. Catch trials were subsequently added to the experimental run to control for fatigue or lack of attention in participants during the procedure. The variability in scores drastically decreased when comparing the pilot experimental scores to the full experimental scores. Interestingly, overall condition thresholds also decreased (improved) quite drastically; in the experimental procedure with catch trials, participants were able to identify smaller changes in pitch and volume for all conditions. Two proposed mechanisms responsible for the decrease in thresholds with the addition of catch trials include: a comparison effect or a reverse-transparency effect. During the staircase with catch trials, participants were more frequently reminded of what rhythms without pitch or volume changes sounded like. This would allow them to frequently compare catch trials to experimental trials, resulting in overall better detection (comparison effect). Similarly, the addition of catch trials into the staircase could have made it more difficult for participants to determine the procedure of the experiment. This reverse-transparency in the mechanism of the procedure would better ensure that participants remained naïve to the nature of the program; it became less obvious that the procedure was honing in on the smallest amount of pitch or volume change the participant could detect. Regardless of which mechanism may have led to lower overall detection thresholds, the addition of catch trials effectively increased control in the experiment and may have a variety of implications for future research.
One limitation of my study was the statistical power. Observed power, for both pitch and volume change detection statistical tests, was very low. As a result, the estimated effect size of the results was also very low. Low observed power allows for the potential of Type II error. It is possible that there may have been a small effect size present in the data, but at the sample size used it would be difficult to detect. However, the number of participants needed to obtain a high observed power, given the estimated effect size, would be large and unrealistic. Furthermore, if an extensive sample size were to be used in order to detect a potential small effect of rhythm type/change location for each of the pitch and volume change detection tasks, the behavioural relevance of such an effect should be questioned. That is, if the effect size of rhythm type/change location on pitch or volume change detection is so small that a massive number of participants are needed to observe the effect, the effect itself is likely irrelevant when individuals are listening to popular music.

**Conclusion**

The results of this experiment indicated no differences between pitch or volume change detection for changes on or off the beat in metric simple or metric complex rhythms. These results did not support the experimental hypothesis, however the predictions of Dynamic Attending Theory (DAT) were not supported either. The lack of support for both the predictions of DAT, and the hypothesis of the present study, could suggest that the attentional maxima proposed by DAT, as well as the masking effects due to expectancy of on-beat salience, do not exist. An alternative explanation is that these mechanisms do exist, however they occur at such a small effect size that their behavioural relevance when listening to music is in question. Further research is
necessary in order to fully understand the cognitive mechanisms underlying human beat perception.
Acknowledgments

This research was made possible by the Western University Department of Biology thesis course (Biology 4999E), and the Natural Science and Engineering Research Council of Canada (NSERC). I would like to thank my supervisor Dr. Jessica Grahn, my advisors Dr. Scott MacDougall-Shackleton and Dr. Bryan Neff, and my lab mates for their continual constructive feedback and support.
References


**Figure 1.** The Hierarchy of Musical Structure. Modified from Drake et al. (2005).

Illustrates the hierarchical nature of musical structure. Any given level could be accentuated as the reference level and attending can occur to any hierarchical level. Individuals only attend to a single hierarchical level at a time when listening to music, however the level at which they attend is not necessarily static.

Modified from Jones et al. (2002). Illustrates a proposed profile of attentional pulses taking into account the expectancy of reference points in an isochronous rhythm. Attentional pulses build with the expectancy of a reference note and a climax in attention occurs on the reference note once the reference levels temporal arrangement has been established.
Figure 3. Metric Simple and Complex Rhythm Type Schematic. Modified from Grahn and Brett (2005). Stimuli schematic for metric simple and metric complex rhythm types. Auditory perceptual accents are indicated with “>”. Metric simple and metric complex rhythms use integer intervals of 1:2:3:4, with regularly or irregularly spaced accents respectively.
Figure 4. Mean Pitch Thresholds for Each Rhythm Type/Change Location. Depicts the mean ± SEM frequency thresholds (Hz) for each rhythm type/change location (metric simple on-beat, metric simple off-beat, and metric complex) determined as the average of the last six reversals out of 14 total reversals in a two-down one-up adaptive staircase design. No statistical differences in frequency threshold were observed between rhythm type/change location groups.
Figure 5. Mean Volume Thresholds for Each Rhythm Type/Change Location. Depicts the mean ± SEM volume thresholds (dB SPL) for each rhythm type/change location (metric simple on-beat, metric simple off-beat, and metric complex) determined as the average of the last six reversals out of 14 total reversals in a two-down one-up adaptive staircase design. No statistical differences in volume threshold were observed between rhythm type/change location groups.