

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Danielsen, Anne; Nymoén, Kristian; Anderson, Evan; Câmara, Guilherme Schmidt; Langerød, Martin Torvik; Thompson, Marc; London, Justin

Title: Where is the beat in that note? Effects of attack, duration, and frequency on the perceived timing of musical and quasi-musical sounds

Year: 2019

Version: Accepted version (Final draft)

Copyright: © 2018, American Psychological Association.

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Danielsen, A., Nymoén, K., Anderson, E., Câmara, G. S., Langerød, M. T., Thompson, M., & London, J. (2019). Where is the beat in that note? Effects of attack, duration, and frequency on the perceived timing of musical and quasi-musical sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 45(3), 402-418. <https://doi.org/10.1037/xhp0000611>

Where is the beat in that note?

Effects of attack, duration and frequency on the perceived timing of musical and quasi-musical sounds

Anne Danielsen^{ab}, Kristian Nymo^{abc}, Evan Anderson^d, Guilherme Schmidt Câmara^{ab}, Martin Torvik Langerød^b, Marc R. Thompson^e, Justin London^d

^aRITMO Center for Interdisciplinary Studies of Rhythm, Time, and Motion, University of Oslo, Norway

^bDepartment of Musicology, University of Oslo, Norway

^cDepartment of Informatics, University of Oslo, Norway

^dCarleton College, USA

^eDepartment of Music, Art & Culture (MACS), University of Jyväskylä, Finland

© 2018, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/xhp0000611

Abstract

When coordinating physical actions with sounds, we synchronise our actions with the *perceptual center* (P-center) of the sound, understood as the specific moment at which the sound is perceived to occur. Using matched sets of real and artificial musical sounds as stimuli, we probed the influence of Attack (rise time), Duration, and Frequency (center frequency) on perceived P-center location and P-center variability. Two different methods to determine the P-centers were used: Clicks aligned in-phase with the target sounds via the method of adjustment, and tapping in synchrony with the target sounds. We found that attack and duration are primary cues for P-center location and P-center variability, and that the latter is a useful measure of P-center shape. Probability density distributions for each stimulus display a systematic pattern of P-center shapes ranging from narrow peaks close to the onset of sounds with a fast attack and short duration, to wider and flatter shapes indicating a range synchronization points for sounds with a slow attack and long duration. The results support the conception of P-centers as not simple time points, but "beat bins" with characteristic shapes, and the shapes and locations of these beat bins are dependent upon both the stimulus and the synchronization task.

Public significance statement

In music and dance, as well as many other contexts, we coordinate our physical actions with sounds. Our research shows how the fine-grained details of a sound interact in our temporal perception of it. This has implications for a wide range of applications that involve timing, from rehearsing musical ensembles to the sonification of complex patterns of information.

Keywords: P-center; synchronization; timing; tapping; click alignment; beat perception

1. Introduction

How do we know if two sounds (or more precisely, two events which generate those sounds) occur "at the same time"? A simple answer might be "if their onsets appear to occur at the same time, they are simultaneous events." But the onset of a sound is a complex event. It is well known, for example, that sounds played by different musical instruments have different onset and attack phase characteristics related to the manner of their sound production (blowing a reed, bowing a string, plucking a string, striking a membrane, etc.; see Rossing, Moore & Wheeler, 2002, pp. 190-334) and that musicians have to take these differences into account to achieve ensemble synchrony (Rasch 1979).

A presumption of previous research has been that for each kind of vowel/phoneme or instrumental sound, there is a specific location that is heard as the point in time where that sound is located. This gave rise to the notion of the perceptual center or "P-center," that is, the specific moment at which a sound is perceived to occur (Morton, Marcus & Frankish, 1976.) Sound synchronization, then, becomes a matter of aligning P-centers, which may be achieved with greater or lesser precision. Musicians know that temporal synchronization (a) admits degrees, such that one can speak of "tight" versus "loose" synchronization, (b) that there can be a character to this synchronization, such that players can "push" or "pull" the sense of beat, and (c) that some sounds are more forgiving/elastic than others in terms of achieving synchrony (e.g., aligning two bowed string instruments versus two drum hits). In some funk and funk-derived musical genres, for example, we find considerable and varying onset discrepancies between rhythmic events articulating the same beat (Danielsen, 2006; Bjerke, 2010; Carlsen & Witek, 2010; Danielsen, 2010, Brøvig-Hanssen & Danielsen, 2016). P-centers, then, seem to be more than a particular moment within the microstructure of a musical or speech sound. Rather, they have a temporal extent, and a temporal shape; in an iterated context they function as "beat bins" (Danielsen, 2010).

1.1 Previous research into musical P-centers

P-centers are not the same as the acoustic or psycho-acoustic onset of a sound, the latter based upon some absolute or relative onset threshold (Gordon, 1987). Rather, the P-center seems to be

located somewhere in between the perceptual onset and the energy peak of a sound (see Figure 1).

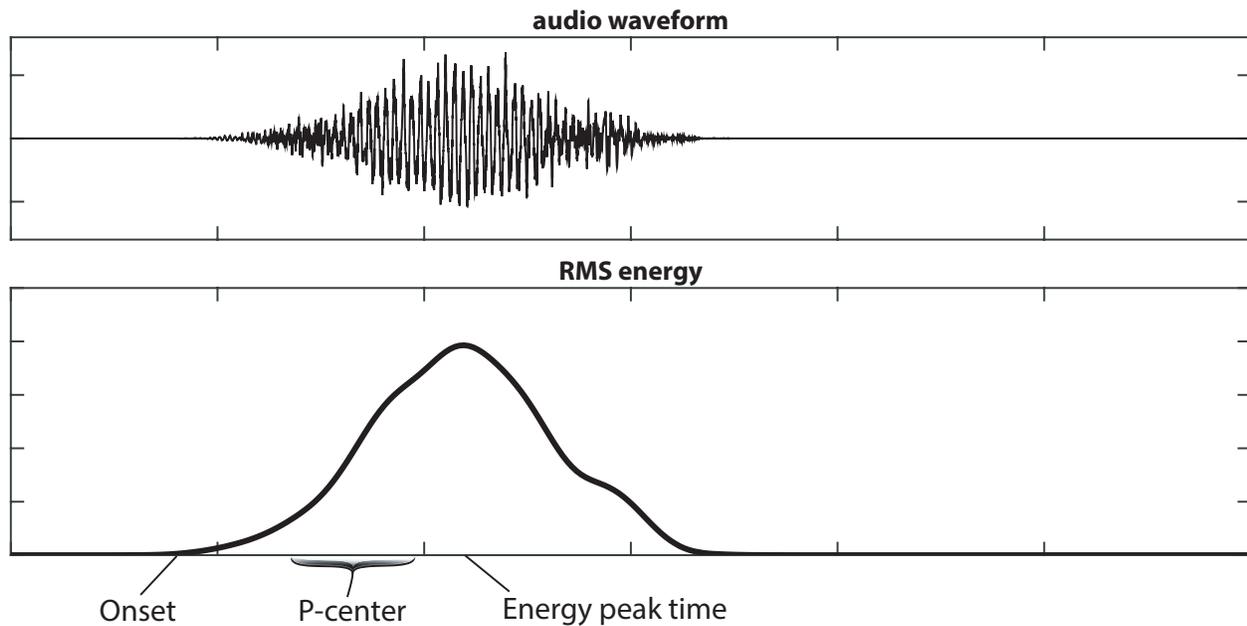


Figure 1. Schematic illustration of sound shape (amplitude/time), audio waveform (top) and amplitude envelope (bottom).

Perceptually isochronous natural speech, for example, is objectively non-isochronous (Morton et al., 1976; Fowler, 1979). The initial consonant duration, that is, the attack phase, has been proposed as an important cue (Marcus, 1981; Howell, 1984; Scott, 1998). However, other features that have been shown to be salient in some studies were not in others (e.g., local or global intensity peaks, vowel onset, vowel quality, or consonant structure; for a review see Villing, 2010, p. 17 ff.). Thus cues for P-center location in speech appear to be complex and context sensitive.

Previous research on the P-centers of musical sounds found that shorter rise times (duration from onset to energy peak) lead to earlier P-centers, and conversely, longer rise times lead to later P-centers. Vos and Rasch (1981) investigated the perceptual onset¹ of sawtooth tones (400 Hz), asking participants to adjust the timing of the test sound by altering its onset time while keeping

¹ Vos and Rasch (1981) define perceptual onset as “the moment at which the temporal envelope during the rise portion passes a certain relative threshold amplitude” (p. 325). However, their method of adjustment was to produce perceptually isochronous sequences of sound, which implies that they measured a percept very similar to the P-center of the sounds.

its offset time fixed. Their results show that lengthening the rise time shifted the P-center later. A later study by Vos, Mates and Kruysbergen (1995) confirmed this finding. Furthermore, in his seminal study of the perceptual attack times (PAT)² of 16 re-synthesised orchestral instrument tones representing varying timbres, rise times and envelope shapes, Gordon (1987) found that the distance from onset to the PAT generally increases with longer rise times. He also found that for sounds with short rise times the PAT was primarily determined by amplitude cues, whereas when a tone's rise time is long, its PAT was also influenced by spectral cues.

Even though the differences in rise times amongst different acoustic instruments have long been known (Rasch, 1979; Vos & Rasch, 1981), the perceptual attributes that can affect the timing of most musical sounds have received little scrutiny. In a survey of 118 empirical papers, Schutz and Vaisberg (2014) found that a limited range of sounds tended to be used in studies of rhythmic and tonal perception: pure sine tones (mostly sharply ramped with a constant amplitude and then a fairly sharp release), synthesized piano tones, and various percussion sounds. Many aspects of those sounds (duration, amplitude envelope, etc.) were unspecified.

As regards other acoustical factors, the picture is even less clear. Previous research suggests a weak effect of *duration* on P-center, that is, longer durations tend to produce later P-centers (Vos et al., 1995; Scott, 1998; Seton, 1989). Very few studies have been conducted on the effect of *frequency* on P-center location. Seton (1989), investigating the effect of auditory streaming on P-center perception, found that all other factors kept constant, high frequency tones (4000 Hz) perceptually occurred 9 ms later on average than middle frequency tones (1000 Hz). This is an unexpected result given the higher perceptual threshold, that is, the delayed perceptual onset, of the 1000 Hz sound compared to the 4000 Hz sound (as specified by the perceptual equal loudness curve [ISO/TC43 2003]). On the contrary, in a study of P-centers of two-tone piano chords with an onset asynchrony between the high and the low tone in the chord, Hove, Keller and Krumhansl (2007) found that P-centers were later when the low-tone onset followed the high-tone onset rather than the reverse in both tapping and an anti-phase click alignment task. Moreover, taps preceded click positions by 42.8 milliseconds on average, which probably reflects

² Perceptual attack time or PAT was defined as the time a tone's moment of attack or most salient metrical feature is perceived relative to its physical onset and resembles closely the definition of the perceptual center in speech research.

the common tendency of a negative mean asynchrony (NMA) in tapping tasks (see discussion below).

1.2 P-center variability and probability distributions

According to Gordon (1987), PAT or P-center is close to a single value for more impulsive sounds but tends to be a range of values for tones with gradual rises in amplitude. The idea of representing the P-center not as a single point in time but as a probability distribution has been further investigated by Wright (2008). As Wright points out, there is often a range of values that sound equally correct when aligning sounds rhythmically, and this range depends on perceptual characteristics of the specific sounds, such as the sharpness of their attacks. Using the clarinet, trumpet and violin tones of Gordon's study (1987), as well as clicks and snare drum, Wright conducted an online listening test where the participants adjusted the relative timing of two repeating sounds (tempo 100 BPM) until they sounded synchronous. Similar to Gordon's study, Wright's results show that both the location of the P-center and the shape of the probability distribution vary with the sharpness of the attack: both found narrow distributions for sharp attacks/fast rise times and wider distributions for slower rise times.

1.3. The beat bin hypothesis

Conceptualizing the location of a rhythmic event as a probability distribution rather than a point in time departs from more conventional means of representing temporal location in music, such as the metric grid inherent in standard musical notation, and resembles the beat bin hypothesis put forward by Danielsen (Danielsen, 2010; Danielsen et al., 2015). The “beat bin” is defined as the perceived temporal width of a beat according to the musical context. Multiple onsets falling within the boundaries of the perceived beat bin will be heard as merging into one beat, whereas onsets falling outside these boundaries will be heard as belonging to another category—namely, that of “not part of the beat” (Danielsen, 2010, p. 29–32).

The beat bin hypothesis grew out of analyses of beat-inducing rhythmic events in repeated musical patterns used in musical genres such as dance and hip-hop that have a temporal shape that makes their location in time unclear or at least difficult to locate relative to a single point in time. Often this comes as a consequence of digital sound processing or relocation of beat-related rhythmic events along the time axis such that they, instead of being completely in synchrony,

jointly form a beat (Bjerke, 2010; Carlsen & Witek, 2010; Danielsen, 2010, Brøvig-Hanssen & Danielsen, 2016). However, several studies argue that the tolerance for what might be heard as on-the-beat, that is, fall inside the beat bin, vary considerably across different musical genres (Haugen, 2016; Johansson, 2010; Stover, 2009).

1.4 Methods for probing P-centers and Negative Mean Asynchrony

A variety of methods have been used to determine the P-center of a sound. The method of adjustment used by Gordon and Wright consists of a repetitive, isochronous series of target sounds that are repeated (i.e., a "loop"), along with either (a) another set of sounds, or (b) a series of clicks or tones. The participant's task is to adjust the timing of the second set of sounds so that they are either (a) perfectly aligned with the target sounds, or (b) in perfect anti-phase alignment with the target sounds, bisecting the temporal interval between the target sounds. P-center location may also be probed by having participants tap along with the sounds, again, either with in-phase or anti-phase alignment.

It is important to recognize that the different methods and experimental configurations may have had an effect on the results obtained. As to the methods used in the present study, the alignment of a click that is in phase with the target stimulus (i.e., on top of the P-center) creates a problem of masking and sonic blend, though this represents a task that is of high ecological validity for musicians, since this is precisely their task in playing together in an ensemble. Tapping studies create a different problem, namely that of the negative mean asynchrony (NMA), the well-established tendency for musically untrained participants to tap slightly ahead of a metronome click or tone in a simple in-phase synchronization task (see Aschersleben, 2002 and Repp, 2005 for reviews). NMAs can vary from 20-80ms for untrained subjects to 10-30ms for musicians (Repp & Dogget, 2007). With real music, the NMA has been found to diminish or disappear (Repp 2005, Repp & Su, 2013), and it systematically varies according to acoustic factors. Indeed, the observed variation of NMA due to rise time and tone duration led Vos et al. (1995) to claim that we use the P-center rather than the acoustic onset of a sound as the cue/target for synchronization. Thus, the NMA relative to the P-center may be present for all sounds, but may be masked as the P-center shifts to a later position relative to acoustic onset (see further discussion below).

Here we provide evidence for P-centers as "bins" that vary in both location and shape according to selected sound factors. Rather than identifying the P-center as a specific point of synchronization and regarding its variability as normally distributed noise around this mean, we claim that P-center variability and probability density distributions are ways to understand crucial features of the P-center of an auditory event, that is, its temporal extent and shape. A wide beat bin affords a broader range of alignments. This will show up as uncertainty in an early/on time/late judgment task, but alternatively (as in an alignment task) can be regarded as a sign of increased 'rhythmic tolerance' (Johansson 2010), that is, synchronization to the event becomes more flexible.

The current study consists of two experiments which probe the influence of various acoustic factors on the location, temporal extent (width), and shape of auditory P-centers in a systematic fashion, using matched sets of musical and artificial sounds as our stimuli. Both experiments used two different methods to determine the location and variability of P-centers of a set of repeated (looped) sounds:³

- Clicks aligned in-phase with target sounds via the method of adjustment
- Tapping in phase with target sounds

Three acoustical factors were investigated: Rise time (which we will refer to as "Attack" in the discussion below), Duration, and Frequency (center frequency). We were also looking for systematic relations between Click Alignment and Tapping, which would shed light on the NMA in relation to the P-center of repeated musical sounds.

We hypothesized the following effects of acoustical factors:

- a) Longer *rise time* leads to later P-center and higher standard deviation, that is, to a later and wider beat bin.

³ In addition to the CA and TAP trials we also tested two other methods, a click alignment anti-phase task and a visual metronome. Except for the click, the anti-phase trials yielded very similar results to those produced by the CA trials. This concurs with previous research, which found that these two methods most likely measure the same percept (Villing, 2012, p. 107). The fourth method was a silent visual metronome. This last method turned out to have implementation problems (video frame rate), and so its data are not analysed here.

- b) Longer *duration* leads to later P-center and higher standard deviation, that is, to a later and wider beat bin.
- c) We expect wider and flatter probability density distributions for sounds with slow attack and/or long duration.
- d) We expect an effect of frequency but make no specific hypotheses regarding its effect.
- e) We expect some interaction between these acoustical factors but make no specific hypotheses regarding their interaction.

2. Experiment 1 – Musical Sounds

2.1 Participants

Twenty music students / semi-professional musicians (9 female) were recruited from the Oslo area. They received a gift card (value 200 NOK) for their participation in the experiment. A power calculation conducted in G*Power indicates that for the planned repeated measures ANOVAs, 16 participants are needed to have 80% power for detecting a medium sized effect when employing the traditional .05 criterion of statistical significance. (Expected effect sizes were obtained through a pilot experiment conducted in the Dept. of Musicology at the University of Oslo.) No participants were excluded. Median age was 25.5 years (Mean = 30.5, SD = 12.5 years; max = 60, min = 20). 2 participants reported 1-4 years of training, 2 participants had 5-10 years of training, and the remaining 16 participants had more than 10 years of training. As their main instrument 10 reported guitar/bass, 2 drums, 3 woodwind or brass, 3 vocals and 2 string instruments. All participants regularly practiced on their instrument, 10 participants practicing 1-6 hours/week and 10 more than 6 hours/week. All participants reported an ability to read music.

2.2 Stimuli

The stimuli consisted of eight musical sounds that represent a balanced design of three acoustical factors: Attack (shorter, impulsive vs. longer, gradual rise time), Duration (of the stimulus sound, as opposed to the stimulus IOI) and Frequency (high vs. low center frequency). We started by qualitatively assessing a range of acoustical/waveform features, using our own knowledge of musical instrument timbres, to find representative sounds in each category, and then verified those assessments with subsequent acoustical analysis. We sought psycho-acoustically salient differences between categories (Schutz and Vaisberg, 2014). In our search for sounds with a slow

attack, for example, we looked for sounds with a gradual rather than impulsive attack, as we regard this feature more important than the duration of the attack phase (which in slow attack sounds is rather difficult to estimate in a precise way). When estimating the duration of the percussive sounds, we looked for sounds with a fast decay after the energy peak, and regarding frequency, we were concerned with identifying sound pairs with a qualitative difference between high and low frequency. These choices were most sensible both from a psycho-acoustic perspective and based on our (considerable) musical expertise.

Manual measurements of the waveforms and results from the MIR toolbox for Matlab, version 1.7 are reported in Table 1. Some perceptually salient microtemporal aspects of sounds are particularly difficult to capture using signal processing techniques like those in the MIR toolbox, and manual judgments are thus required to balance out errors in the MIR toolbox measurements.⁴ A click sound (i.e., the same as the click probe in the CA task) was also included amongst the stimuli; click-click and tap-click data are analyzed separately below. Because there is no way of arriving at an objectively equal level of loudness for sounds with these different sonic characteristics, the relative loudness level of the different sounds was adjusted by ear by two of the experimenters.

Table 1. Sound Stimuli Experiment 1—Musical Sounds

Sound Factor	Click	Fast Short	Fast Short	Fast Long	Fast Long	Slow Short	Slow Short	Slow Long	Slow Long
		<u>Low</u> Kick Drum	<u>High</u> Snare Drum	<u>Low</u> Dark Piano	<u>High</u> Light Piano	<u>Low</u> Arco Bass	<u>High</u> Cabasa	<u>Low</u> Synth Bass	<u>High</u> Fiddle
Rise Time (from onset to energy peak)		5 ms	2 ms	18 ms	19 ms	31 ms	22 ms	33 ms	43 ms
Duration (waveform / MIR toolbox)	1 ms	130 / 123 ms	25 / 38 ms	487 / 371 ms	318 / 211 ms	66 / 70 ms	49 / 61 ms	220 / 220 ms	105 / 89 ms
Frequency Range or Pitch	3000 Hz	Low	High	65,4 Hz	659,3 Hz	65,4 Hz	High	32,7 Hz	479 Hz
Spectral Centroid (MIR toolbox)		780 Hz	2831 Hz	623 Hz	893 Hz	538 Hz	8199 Hz	781 Hz	1581 Hz

Note. Measures of the waveform were obtained using Amadeus Pro, Version 2.4.5 (HairerSoft, London, UK). Fast = fast attack; Slow = slow attack; Short = short duration; Long = long duration; Low = low-frequency range; High = high-frequency range.

⁴The version of the MIR toolbox that we used is not suited for capturing rise times with the precision that the present study requires. It also systematically reports longer durations for short sounds than our manual measurements. Both are due to the windowing technique used for calculating the amplitude envelope (window length 20 ms, with 98 % overlap). Furthermore, durations of long sounds are underreported by the MIR toolbox, because of the way it estimates the start and end points of sound events (applying a thresholding technique to the amplitude envelope). See also (Nymoen et al., 2017).

2.3 Apparatus and Method

During the CA trials the participants' task was to align a click track with the target stimulus; click and stimuli were both looped at 600ms interval (tempo=100 bpm) with a random offset (+/- 100-200 ms). In each trial, participants manipulated the offset of the two sounds by moving an on-screen cursor using the mouse and/or arrow keys. Participants were also able to adjust the volume of the click track. When satisfied that the target stimulus was synchronized with the click track, participants moved to the next trial. Following two practice trials, participants heard each target stimulus four times for a total of 36 trials. The order of stimulus presentation was randomized, but constrained so that participants never heard the same stimulus on back-to-back trials. The time for each participant to perform all CA trials varied from 30 to 60 minutes.

Participants completed the CA task using iMac computers (3.1 Ghz Intel core i7, OSX 10.11.16), listening via AKG K171 MkII headphones at a comfortable intensity that could be further adjusted by the participant. Stimuli were presented using a custom-made patch written in Max 7 (<http://www.cycling74.com>), which also recorded participants' responses. All participants' responses were averaged across the four trials to produce a P-center location for each stimulus; P-centers are reported in milliseconds relative to the physical onset of the stimulus. Average standard deviations were calculated for each stimulus by participant, and then the grand average of participant standard deviations was used as a measure of the P-center variability for each stimulus.

In the Tapping trials, the task was to tap along using a pair of clave sticks in synchrony with the target stimulus (again looped at a 600ms interval). Each loop repeated for 20 seconds.

Participants were given two practice trials to gain familiarity with the clave sticks as well as with the task at hand. The presentation of the 9 target stimuli was randomly ordered. Participants took from 5 to 10 minutes to finish the Tapping trials.

In the TAP task participants used acoustically transparent headphones (Koss PortaPro) which allowed them to clearly hear their tapping during those trials. To eliminate timing latencies during the Tapping task, the stimulus was split and routed both to participants' headphones and to a mono recording channel on an audio interface (PreSonus Firebox); tapping data were recorded

on another mono channel using a Shure SM57 unidirectional microphone. A MATLAB script was used to identify tap onsets, as the time point where the rectified tapping audio waveform first exceeded a predefined threshold close to the noise floor. For each registered tap, the time difference between its detected onset and the first zero crossing of the closest stimulus sound was calculated. The locations of 24 consecutive taps from the fifth tap of each trial were averaged to give a P-center location for each stimulus. One series by one participant had only 18 registered taps; in that case 14 consecutive taps from the fifth tap were used. For each participant, the standard deviation of the tap locations was calculated within each trial, and then (as in the CA task), the grand average of participant standard deviations was calculated for each stimulus.

The order in which participants completed the two tasks was counterbalanced. Between or after experimental tasks, participants answered a series of background questions pertaining to their musical training and musical listening preferences, as well as age, gender, and nationality). For the CA trials, between one and eight participants ran trials at individual workstations in the University of Oslo (UiO) computer music lab. The TAP trials were conducted as individual sessions in UiO's motion capture lab. Participants were encouraged to proceed through the experiment at their own pace and to take breaks as needed. One of experimenters waited nearby should any questions/problems arise.

2.4 Data analysis

In order to test the effect of the acoustical factors on P-center location and P-center variability, repeated measures ANOVAs with Attack, Duration and Frequency as independent variables and (a) mean P-center location or (b) standard deviation of mean location as dependent variables were conducted for each task (click alignment vs. tapping) separately and for both tasks combined. Note that here and in Experiment 2, reported mean P-center "locations" should be understood as the peaks of the beat bins, which are described in the general discussion below. Click-Click and Tap-Click alignments were not included in the ANOVAs, but analyzed separately. In addition, paired samples t-tests (Tapping vs. CA condition) of P-center mean and standard deviation were conducted for each of the nine sounds to examine possible NMA. All statistical tests were performed in SPSS (ver. 24) (IBM).

2.5 Results

The mean location and variability (per stimulus) for both click alignment and tapping trials are given in Figures 2 and 3, as well as Table A1 in the Appendix. No outliers were identified, indicating that all participants were capable of completing the task.

2.5.1 Click Alignment Task

Regarding P-center location, a 2x2x2 (Attack x Duration x Frequency) repeated-measures ANOVA showed a main effect of Attack, ($F(1,19) = 42.428, p = .000; \eta_p^2 = .691$), a main effect of Duration ($F(1,19) = 9.702, p = 0.006; \eta_p^2 = .338$), and a main effect of Frequency, ($F(1,19) = 12.018, p = 0.003; \eta_p^2 = .387$). Slow Attack, long Duration and low Frequency all led to later P-center location. There was also a significant interaction between Duration and Frequency, such that frequency has a larger effect when the duration is long ($F(1,19) = 4.739, p = .042; \eta_p^2 = .200$).

In terms of P-center variability a 2x2x2 (Attack x Duration x Frequency) repeated-measures ANOVA showed a main effect of Attack, ($F(1,19) = 37.026, p = .000; \eta_p^2 = .661$), a main effect of Duration ($F(1,19) = 10.733, p = .004; \eta_p^2 = .361$), but no effect of Frequency. Slow Attack and long Duration led to higher variability.

2.5.2 Tapping Task

In the Tapping task, regarding P-center location a 2x2x2 (Attack x Duration x Frequency) repeated-measures ANOVA showed a main effect of Attack, ($F(1,19) = 9.255, p = .007; \eta_p^2 = .328$), a main effect of Duration, ($F(1,19) = 65.529, p = .000, \eta_p^2 = .775$), and a main effect of Frequency, ($F(1,19) = 5.553, p = .029; \eta_p^2 = .226$). Again, slow Attack, long Duration and low Frequency all led to later P-center location. Moreover, similar to in the click alignment task, there was significant interaction between Duration and Frequency such that frequency has a larger effect when the duration is long, ($F(1,19) = 16.361, p = .001; \eta_p^2 = .463$). In addition there was significant interaction between Attack and Duration such that there was a larger effect of Duration when the attack was slow than when it was fast ($F(1,19) = 15.389, p = .001; \eta_p^2 = .447$).

In terms of P-center variability, a 2x2x2 (Attack x Duration x Frequency) repeated-measures ANOVA showed a main effect of Attack, ($F(1, 19) = 5.478, p = .030; \eta_p^2 = .224$), a trend

towards a main effect of Duration ($F(1, 19) = 3.193, p = .090; \eta_p^2 = .144$), but no effect of Frequency. Similar to in the CA task, slow Attack and long Duration yielded higher variability.

2.5.3 Task Comparison

Regarding P-center location, a $2 \times 2 \times 2 \times 2$ (Task x Attack x Duration x Frequency) repeated-measures ANOVA showed a main effect of Task ($F(1, 19) = 7.139, p = .015, \eta_p^2 = .273$). As expected there were also main effects of Attack ($F(1, 19) = 42.641, p = .000, \eta_p^2 = .692$), Duration ($F(1, 19) = 59.595, p = .000, \eta_p^2 = .758$), and Frequency ($F(1, 19) = 16.402, p = .001, \eta_p^2 = .463$), with slow Attack, long Duration and low Frequency all leading to later P-center.

The task comparison also revealed significant interaction effects. Attack had a greater effect in the CA trials than in Tapping, ($F(1, 19) = 6.136, p = .023, \eta_p^2 = .245$) and duration had a greater effect in Tapping than in CA trials, ($F(1, 19) = 4.683, p = .043, \eta_p^2 = .198$). As expected there was also stronger effect of duration when the attack was slow than when it was fast ($F(1, 19) = 5.514, p = .030, \eta_p^2 = .225$) and frequency range had a greater effect when the duration was long, ($F(1, 19) = 12.985, p = .002, \eta_p^2 = .406$).

In terms of P-center Variability a $2 \times 2 \times 2 \times 2$ (Task x Attack x Duration x Frequency) repeated-measures ANOVA revealed no significant effect of Task, but again main effects of Attack ($F(1, 19) = 65.428, p = .000, \eta_p^2 = .775$) and Duration ($F(1, 19) = 11.664, p = .003, \eta_p^2 = .380$).

Unsurprisingly, slow Attack and long Duration both produced higher variability. There was no effect of Frequency. Again there were significant interaction effects involving Task: Attack had a greater effect on variability in the CA trials than in Tapping, ($F(1, 19) = 13.684, p = .002, \eta_p^2 = .419$) and duration had a greater effect in Tapping than in the CA trials, ($F(1, 19) = 7.422, p = .013, \eta_p^2 = .281$).

Paired samples t-tests (Tapping vs. CA condition) of P-center location for each sound show significant differences for four out of the nine stimuli. For kick drum there is a significant difference of 6 ms, and for the two slow-short sounds (cabasa and arco bass) the mean difference range between 12.5-15 ms. For click stimulus the mean difference is 21 ms. Regarding variability, paired samples t-tests (Tapping vs. CA condition) show significant differences for

click-click versus tap-click, and the two fast-short sounds. In all cases the standard deviation is significantly higher in the CA condition; see Table 2.

Table 2. Paired Samples t Tests (Tapping vs. CA Condition) for Each Musical Sound

	P-center Location		Paired Differences (ms)		t	df	Sig. (2-tailed)
	Stimulus	Test condition	Mean	Std. Dev.			
Pair 0	Click	CA – TAP	21.07	15.52	6.070	19	.000
Pair 1	Fast_Short_Low	CA – TAP	6.27	9.43	2.973	19	.008
Pair 2	Fast_Short_High	CA – TAP	1.90	12.38	.686	19	.501
Pair 3	Fast_Long_Low	CA – TAP	.94	13.92	.302	19	.766
Pair 4	Fast_Long_High	CA – TAP	5.08	15.88	1.432	19	.168
Pair 5	Slow_Short_Low	CA – TAP	12.49	18.00	3.104	19	.006
Pair 6	Slow_Short_High	CA – TAP	14.79	18.54	3.569	19	.002
Pair 7	Slow_Long_Low	CA – TAP	6.55	21.36	1.372	19	.186
Pair 8	Slow_Long_High	CA – TAP	3.00	18.50	.725	19	.477

	P-center Standard deviation		Paired Differences (ms)		t	df	Sig. (2-tailed)
	Stimulus	Test condition	Mean	Std. Dev.			
Pair 0	Click	CA – TAP	-17.51	5.15	-15.189	19	.000
Pair 1	Fast_Short_Low	CA – TAP	-6.58	8.07	-3.644	19	.002
Pair 2	Fast_Short_High	CA – TAP	-5.14	8.93	-2.575	19	.019
Pair 3	Fast_Long_Low	CA – TAP	-.971	10.79	-.402	19	.692
Pair 4	Fast_Long_High	CA – TAP	2.40	11.33	.949	19	.354
Pair 5	Slow_Short_Low	CA – TAP	-1.23	9.15	-.603	19	.554
Pair 6	Slow_Short_High	CA – TAP	.95	10.53	.404	19	.691
Pair 7	Slow_Long_Low	CA – TAP	5.039	16.98412	1.327	19	.200
Pair 8	Slow_Long_High	CA – TAP	.91800	11.35785	.361	19	.722

Note. Significant results in boldface. Fast = fast attack, Slow = slow attack, Short = short duration, Long = long duration, Low = low center frequency, and High = high center frequency. CA = click alignment and TAP = tapping.

Summing up, the results show that slow attack, long duration and low frequency all lead to later P-centers. There is significant interaction between Duration and Frequency such that frequency has a larger effect when the duration is long. Attack and Duration also lead to higher variability of P-center locations. Attack had a greater effect in the CA trials than in Tapping. Conversely, duration had a greater effect in Tapping than in CA trials and the effect is larger when the attack

is slow than when it is fast. Tapping trials generally locate the P-center earlier than CA trials, and these differences were statistically significant for click plus three out of four of the short sounds. CA trials yield higher variability than tapping trials, in particular for the click and for fast-short sounds.

3. Experiment 2 – Quasi-musical Sounds

3.1 Participants

Thirty participants (11 female) were recruited from the Northfield, Minnesota community. Participants were unpaid but were entered into a drawing for a gift card from a local coffee shop. The required sample size was calculated in a manner similar to in experiment 1; because the pilot experiment demonstrated considerable differences in performance between musicians and non-musicians, and the participants in this second experiment were non-musicians, we increased the number of participants. One participant was rejected due to their inability to perform either experimental task; five other participants were unable to perform the click alignment task, and their data (including tapping trials) were excluded from all analyses. Median age of the 24 remaining participants was 21 years (Mean = 30.2, SD = 14.7 years; max = 63, min = 18). Two participants had no musical training, 6 participants had 1-4 years of training, 9 participants had 5-10 years of training, and the remaining 7 participants had more than 10 years of training. Twelve participants reported that they play an instrument at least once a week, 5 of whom play daily. Twenty-one participants reported an ability to read music. One participant identified themselves as a professional musician, and another had at least 2 years of experience as a sound engineer.

3.2 Stimuli

The sound stimuli used in Experiment 2 were patterned after the musical sounds used in the first experiment, with an aim at having more precise control of Attack (rise time), Duration, and Frequency (center frequency). A click sound (i.e., the same as the click probe in the CA task) was also included amongst the stimuli; click-click and tap-click data are analysed separately below. The sound files were generated in Max 7, using white noise and bandpass-filters with a Q-factor of 10. The amplitude of the sound files was scaled linearly from 0 (beginning of file) to 1 (at the indicated rise time in Table 3), immediately followed by a linear decay to silence at the end of each sound file.

Table 3. Synthesis parameters for sound stimuli in Experiment 2: Quasi-musical sounds.

	Click	Fast Short Low	Fast Short High	Fast Long Low	Fast Long High	Slow Short Low	Slow Short High	Slow Long Low	Slow Long High
Attack/Rise Time	0 ms	3 ms	3 ms	3 ms	3 ms	50 ms	50 ms	50 ms	50 ms
Total Duration	1 ms	100 ms	100 ms	400 ms	400 ms	100 ms	100 ms	400 ms	400 ms
Center Frequency	3000 Hz	100 Hz	700 Hz	100 Hz	700 Hz	100 Hz	700 Hz	100 Hz	700 Hz

Note. Fast = fast attack, Slow = slow attack, Short = short duration, Long = long duration, Low = low center frequency, and High = high center frequency.

3.3 Apparatus and Method

The tasks were identical to experiment 1, that is, click alignment trials and tapping trials in counterbalanced order with a background questionnaire administered between trial blocks. All experimental sessions took place in a recording studio with a high level of sound attenuation. Participants were encouraged to move through the experiment at their own pace, while one of the experimenters waited nearby to deal with any questions/problems that might arise. On average, experimental sessions lasted approximately 30 minutes.

The procedure of the Click Alignment trials was the same as for experiment 1, except for the number of trials: Following two practice trials, participants heard each target stimulus twice for a total of 18 trials. The time it took to perform each trial was also recorded. On average, participants finished the Click Alignment trials in approximately 15 minutes, and took 5 minutes, on average, to finish the Tapping trials. The procedure of the Tapping trials was also identical to in Experiment 1. Mean P-center locations and standard deviations for both Click Alignment and Tapping conditions were calculated in the same manner as in Experiment 1.

Participants completed both experimental tasks using a MacBook Pro Laptop (15-inch screen, 2.3 GHz Intel Core i7, running macOS Sierra 10.12.5) via a Max 7 patch, which also recorded participants' responses. All auditory stimuli were presented to participants via Beyerdynamic 990 Headphones, an over-the-ear but acoustically transparent model which allowed them to clearly hear their tapping during tapping trials.

Again, timing latencies during the Tapping task were eliminated by splitting the stimulus and routing it both to participants' headphones and to a mono recording channel on an audio interface

(Zoom UAC-2); tapping data were recorded on another mono channel. The same MATLAB script was used to identify onsets of taps. The locations of 23 consecutive taps from the fifth tap of each trial were averaged to give a P-center location for each stimulus relative to the first zero crossing of the closest stimulus sound. 23 taps instead of 24 (as in Experiment 1) were selected because the 24th was the last tap in many series and we wanted to exclude this last tap from the data.

3.4 Data analysis

Similar to in experiment 1, repeated measures ANOVAs with the acoustical factors Attack, Duration and Frequency as independent variables and (a) mean P-center location or (b) standard deviation of mean location as dependent variables were conducted for each task separately and for both tasks combined. Click-Click and Tap-Click alignments were not included in the ANOVAs but analyzed separately. Paired samples t-tests (Tapping vs. CA condition) of P-center mean and standard deviation were conducted for each of the nine sounds. All statistical tests were performed in SPSS (ver. 22) (IBM).

3.5 Results

The P-center mean location and variability for both click alignment and tapping trials are provided in Figures 2 and 3, as well as in Table A2 in the Appendix.

3.5.1 Click Alignment Task

Regarding P-center location, a 2x2x2 (Attack x Duration x Frequency) repeated-measures ANOVA showed a main effect of Attack ($F(1, 23) = 19.384, p = .000, \eta_p^2 = .457$) and a main effect of Duration ($F(1, 23) = 10.340, p = .004, \eta_p^2 = .310$). Slow Attack and long Duration both led to later P-center location. There was no main effect of Frequency, but there was a significant interaction between Attack and Duration such that there was a larger effect of Duration when the attack was slow than when it was fast ($F(1, 23) = 4.856, p = .038, \eta_p^2 = .174$).

In terms of P-center variability a 2x2x2 (Attack x Duration x Frequency) repeated-measures ANOVA showed a main effect of Duration ($F(1, 23) = 9.486, p = .0053, \eta_p^2 = .292$) and a nearly significant effect of Attack ($F(1, 23) = 4.042, p = .056, \eta_p^2 = .149$). Both slow Attack and long

Duration produced higher variability. Again, there was no effect of Frequency ($F(1, 23) = .010, p = .920, \eta_p^2 = .000$).

3.5.2 Tapping Task

In the Tapping task a 2x2x2 (Attack x Duration x Frequency) repeated-measures ANOVA showed a main effect of Attack ($F(1, 23) = 56.268, p = .000, \eta_p^2 = .710$) and a main effect of Duration ($F(1, 23) = 37.895, p = .000, \eta_p^2 = .622$) on P-center location. Again, slow Attack and long Duration led to later location. There was no effect of Frequency. There were no significant effects on P-center variability in the tapping task.

3.5.3 Task Comparison

Regarding P-center location, a 2x2x2x2 (Task x Attack x Duration x Frequency) repeated-measures ANOVA showed there was no effect of Task ($F(1, 23) = .634, p = .434, \eta_p^2 = .027$). In this omnibus ANOVA there were, as one would expect, main effects of Attack ($F(1, 23) = 56.584, p = .000, \eta_p^2 = .711$) and Duration ($F(1, 23) = 44.258, p = .000, \eta_p^2 = .658$), with slow Attack and long Duration leading to later P-center location. There was no significant effect of Frequency. There was a significant interaction of Attack and Duration, such that there was a larger effect of Duration when the attack was slow than when it was fast ($F(1, 23) = 4.899, p = .037, \eta_p^2 = .176$).

In terms of P-center variability a 2x2x2x2 (Task x Attack x Duration x Frequency) repeated-measures ANOVA revealed no significant effect of Task, but again main effects of Attack ($F(1, 23) = 6.072, p = .022, \eta_p^2 = .209$) and Duration ($F(1, 23) = 5.076, p = .034, \eta_p^2 = .181$), with slow Attack and long Duration producing higher variability. There was no effect of Frequency, and there was a significant interaction between Task and Duration such that duration had a greater effect in Tapping than in CA trials, $F(1, 23) = 11.839, p = .002, \eta_p^2 = .340$.

While the RM ANOVA found no significant effect of task, a separate analysis of the click-click and tap-click alignment trials regarding P-center location was statistically significant (mean paired difference = 27 ms; $t(23) = 3.983; p = .001$), as the NMA was clearly apparent in the tapping trials when the click was the target stimulus. There was also a significant difference in

variability between click-click and tap-click (mean paired difference = -17 ms; $t(21) = -8.449$; $p < .001$; note two additional participants were excluded due to extreme outliers).

The results of Experiment 2 show that slow attack and long duration lead to later P-center also when using the more controlled quasi-musical sounds as stimuli (see Figure 2). There was a significant interaction between Attack and Duration, such that duration had stronger effect with slow attacks. In contrast to the results for musical sounds, however, we found no effect of Frequency on P-center location. Experiment 2 also confirmed that slow Attack and long Duration lead to higher P-center variability (see Figure 3). Again, Duration had a greater effect in Tapping than in Click Alignment trials. Moreover, a majority of the average P-center locations were earlier in the tapping tasks than in the click alignment task. However, apart from trials where a click was the target stimulus, none of these differences reached statistical significance.

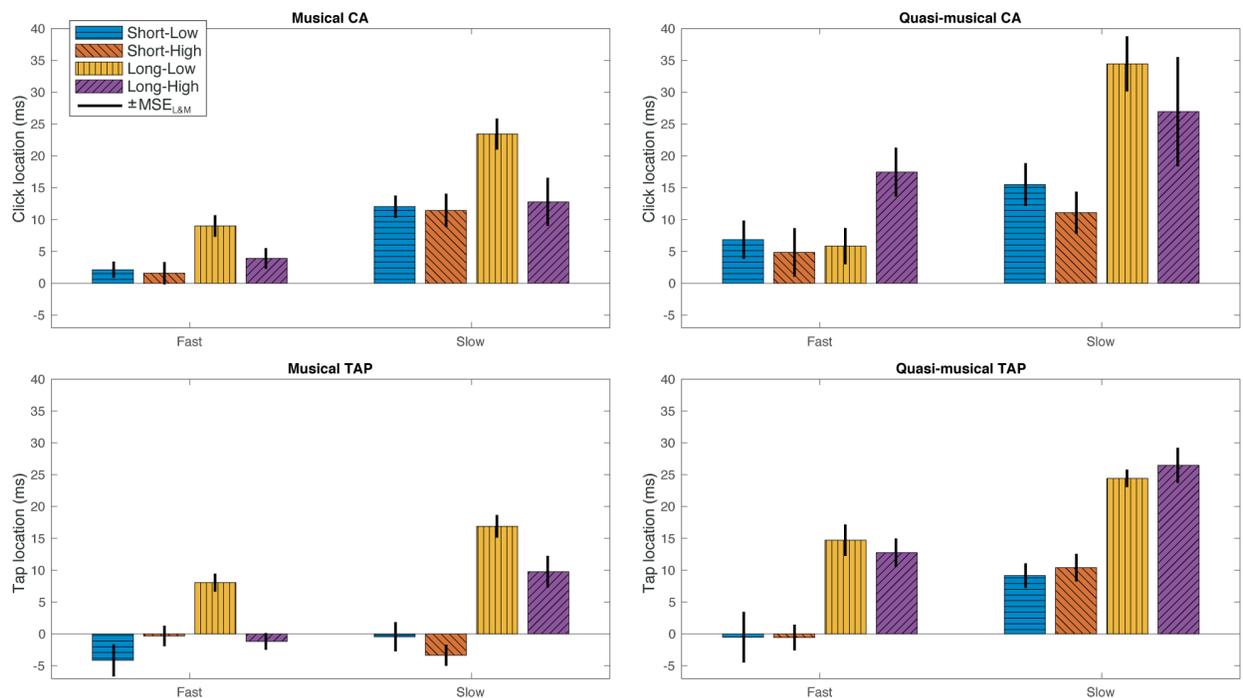


Figure 2. A summary comparison of P-center locations for all conditions in both experiments (click-as-target stimuli excluded). P-center locations are given relative to the physical onset of stimuli. All data are presented in milliseconds. Error bars calculated according to Loftus and Masson (1994). Fast = fast attack, Slow = slow attack, Short = short duration, Long = long duration, Low = low center frequency, and High = high center frequency. CA = click alignment, and TAP = tapping.

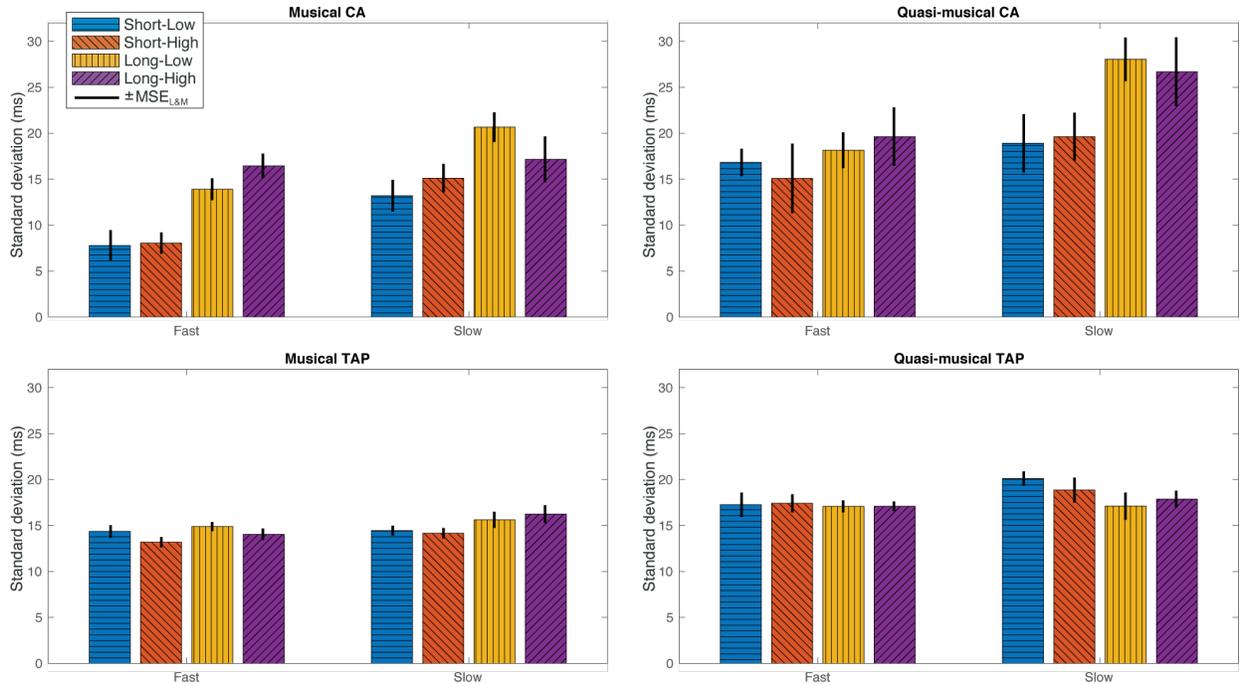


Figure 3. A summary comparison of standard deviations for all conditions in both experiments (click-as-target stimuli excluded). All data are presented in milliseconds. Error bars calculated according to Loftus and Masson (1994). Fast = fast attack, Slow = slow attack, Short = short duration, Long = long duration, Low = low center frequency, and High = high center frequency. CA = click alignment, and TAP = tapping.

4. Comparison of Experiments 1 and 2

4.1 Data Analysis Overview

P-center mean locations and average variability are compared across stimulus types (musical sounds of Experiment 1 vs. quasi-musical sounds of Experiment 2) with experiment as the between-groups variable, and Attack, Duration, and Frequency as within-group variables. Then, to include the effect of task, a more fine-grained ANOVA was run, with Experiment as the between-groups variable, and Task, Attack, Duration, and Frequency as within-group variables.

In order to inspect the shapes of the beat bins for the different sounds, we also produced probability density graphs (Gordon, 1987) of all the stimuli sounds based on the click alignment results. If the probability distributions functioned merely as a measure of the participants' accuracy, we would expect to see a symmetrical distribution around a centered peak, that is, the same Gaussian shape, for each sound. The graphs were produced using the *fitdist*-function in MATLAB 2017b. The curves were fitted to the data using kernels based on normal distributions and a bandwidth of 6 milliseconds.

After having conducted the above *a priori* planned analyses, we decided to also conduct non-parametric Friedman tests on differences between distributions of click locations from the CA tasks produced by the acoustical factors Attack and Duration (results for the two frequency levels were collapsed due to the absence of significant results for this factor in the across experiment analyses). Subsequently, we conducted Friedman's tests on paired differences between distributions. The aim was to test the differences between beat bin shapes without assuming a normal distribution of the data for each sound. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons.

4.2 CA trials Comparison

In the CA trials, the P-center locations of the nine pairs of corresponding musical and quasi-musical stimuli were highly correlated ($r = .847, p = .004$). The $2 \times (2 \times 2 \times 2)$ (Experiment x (Attack x Duration x Frequency)) mixed-design ANOVA showed that there was no main effect of Experiment ($F(1, 42) = 2.772, p = .103, \eta_p^2 = .062$), while there were main effects of Attack ($F(1, 42) = 43.774, p = .000, \eta_p^2 = .510$) and Duration ($F(1, 42) = 15.989, p = .000, \eta_p^2 = .276$). There was no effect of Frequency, and there was a nearly significant interaction of Attack and Duration, $F(1, 42) = 3.968, p = .053, \eta_p^2 = .086$ (i.e., for fast attacks, grand mean of P-centers for short durations vs. long durations was 3.85 ms vs. 9.05 ms, while for slow attacks the means were 12.52 ms vs. 24.51 ms; longer duration had a proportionally greater effect on P-center location in the context of slow attacks). Likewise, the average variabilities of pairs of corresponding musical and quasi-musical stimuli were highly correlated for click alignment trials ($r = .942, p < .001$). The analogous $2 \times (2 \times 2 \times 2)$ mixed design ANOVA for variability (Experiment x (Attack x Duration x Frequency)) showed a main effect of Experiment ($F(1, 42) = 5.134, p = .029, \eta_p^2 = .109$), a main effect of Attack ($F(1, 42) = 10.849, p = .002, \eta_p^2 = .205$), and a main effect of Duration ($F(1, 42) = 19.957, p = .000, \eta_p^2 = .322$). The grand mean of P-center variability was 20.37 ms for the artificial stimuli, versus 14.04 ms for the musical stimuli. Again, there was no effect of Frequency, and there were no significant interactions.

4.3 Tapping trials Comparison

In the tapping trials the P-center locations of pairs of corresponding musical and quasi-musical stimuli were again highly correlated ($r = .929, p < .001$). The $2 \times (2 \times 2 \times 2)$ (Experiment x (Attack x Duration x Frequency)) mixed-design ANOVA showed a main effect of Experiment ($F(1, 42) =$

4.864, $p = .033$, $\eta_p^2 = .104$), a main effect of Attack ($F(1, 42) = 52.665$, $p = .000$, $\eta_p^2 = .556$), and a main effect of Duration ($F(1, 42) = 75.963$, $p = .000$, $\eta_p^2 = .644$) on P-center location. The grand mean for P-center location with artificial stimuli was +12.11 ms, versus +3.16 ms for real stimuli. There was no effect of Frequency. Thus, while there was no significant difference between the grand means for real and artificial stimuli in the CA task, there was such a difference in the tapping task.

In the tapping task there were also a number of interactions between musical vs. quasi-musical sounds, though effect sizes are uniformly modest. There was an interaction of Attack and Experiment such that the effect of Attack was larger for quasi-musical sounds than it was for musical sounds ($F(1, 42) = 7.065$, $p = .011$, $\eta_p^2 = .144$). There was an interaction of Frequency and Experiment such that the effect of Frequency only occurred for real sounds ($F(1, 42) = 4.373$, $p = .043$, $\eta_p^2 = .094$). The mixed-design ANOVA also confirmed the following previously reported interaction effects: the effect of Duration was larger when the attack was slow than when the attack was fast ($F(1, 42) = 14.035$, $p = .0005$, $\eta_p^2 = .183$), and the effect of frequency only occurred when the duration was long ($F(1, 42) = 4.495$, $p = .040$, $\eta_p^2 = .097$). Finally, there was a three-way interaction of Attack, Duration, and Experiment ($F(1, 42) = 7.783$, $p = .008$, $\eta_p^2 = .156$).

In the tapping trials the correlation between the average variabilities of pairs of corresponding musical and quasi-musical stimuli did not reach statistical significance ($r = .615$, $p = .078$). The $2 \times (2 \times 2 \times 2)$ (Experiment \times (Attack \times Duration \times Frequency)) mixed-design ANOVA showed a main effect of Experiment ($F(1, 42) = 9.784$, $p = .003$, $\eta_p^2 = .189$) and a main effect of Attack ($F(1, 42) = 6.218$, $p = .017$, $\eta_p^2 = .129$).⁵ There were no effects of Duration or Frequency, but there was an interaction of Duration and Experiment such that variability shrank as quasi-musical sounds got longer, but variability grew as musical sounds got longer ($F(1, 42) = 5.090$, $p = .029$, $\eta_p^2 = .108$).

⁵ To further assess the tapping performance, we analyzed the lag1 autocorrelations (Lag1AC) from the tapping data in both experiments. Overall, the results show that the tapping performance was quite stable in both experiments (StdDev of Lag1AC in Exp. 1 = .204; Exp 2. = .183), which confirms that the variability in P-center location reflects the size of the beat bins and was not caused by erroneous tapping. However, as in Hove et al. (2007), the more musically experienced participants in Exp. 1 had significantly lower lag1 autocorrelation than the less musically experienced participants in Exp. 2 (mean Lag1AC in Exp. 1 = -.228; Exp. 2 = -.316; Mann-Whitney U test $Z = -4.365$, $p < .001$).

Figure 4 summarizes the correlations between the two experiments for P-center location and variability in both click-alignment and tapping task conditions; it makes clear that the correlations are not due to a clustering of main effects (Fast vs. Slow onset, and/or Short vs. Long duration). The uniform standard deviation in the tapping condition (save for the tap-click trials) is also clear from the scatterplot.

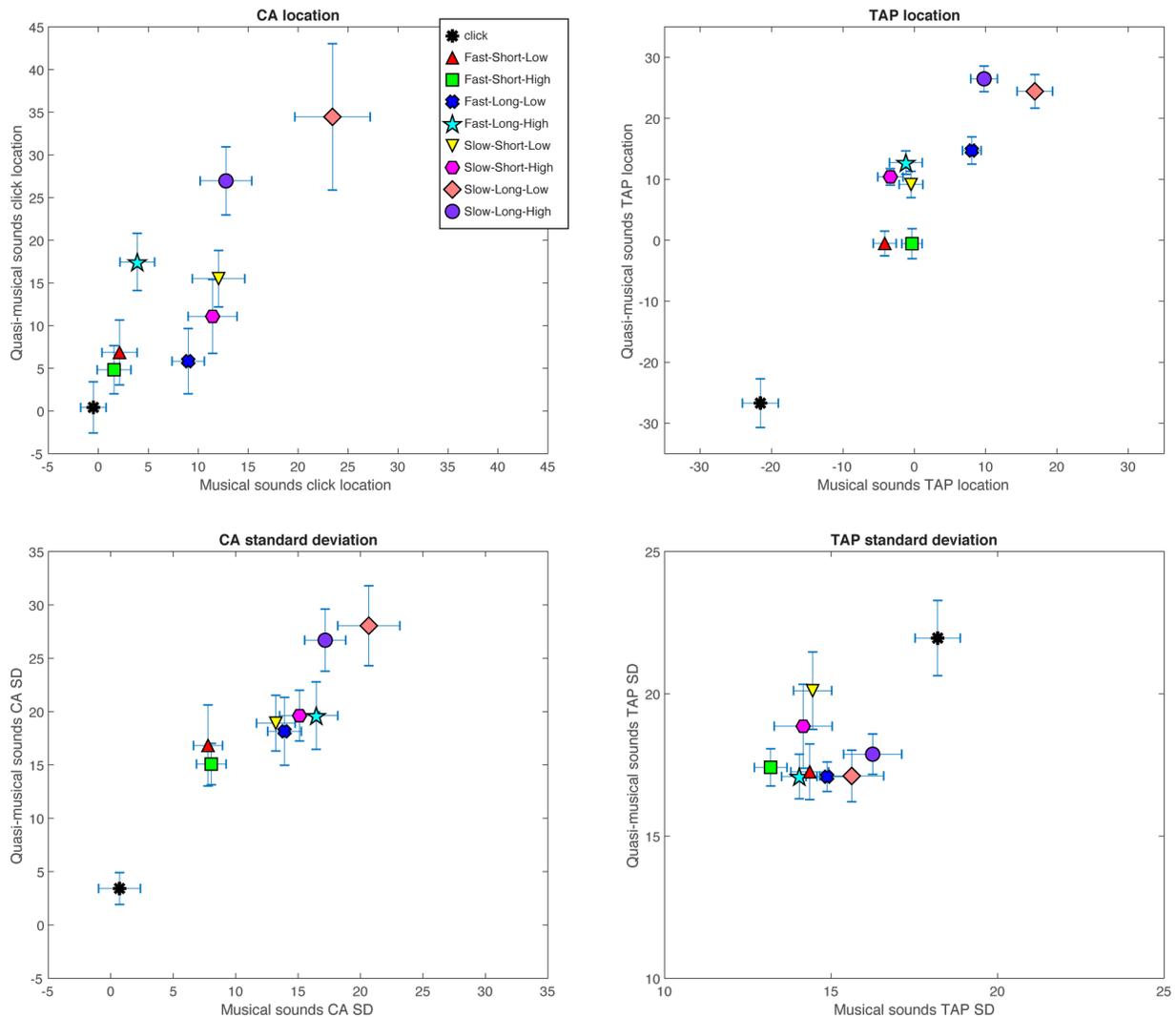


Figure 4. Scatterplots comparing Experiment 1 (x-axis) and Experiment 2 (y-axis) in terms of P-center location and standard deviation in each experimental task. Plots include the click-click and tap-click tasks (black dot). Note that while x- and y-axis scales are always equivalent, they differ from panel to panel. All data are presented in milliseconds.

4.4 Task Comparison

Regarding P-center mean, a $2 \times (2 \times 2 \times 2 \times 2)$ (Experiment x (Task x Attack x Duration x Frequency)) mixed-design ANOVA showed a main effect of Experiment ($F(1, 42) = 6.719, p = .013, \eta_p^2 = .138$), a main effect of Attack ($F(1, 42) = 92.400, p = .000, \eta_p^2 = .688$), and a main effect of Duration ($F(1, 42) = 79.400, p = .000, \eta_p^2 = .654$). Task was nearly significant ($F(1, 42) = 3.724, p = .060, \eta_p^2 = .081$), as was Frequency ($F(1, 42) = 3.769, p = .059, \eta_p^2 = .08$). There was an interaction between Duration and Experiment such that the effect of Duration was larger for quasi-musical sounds than for musical sounds ($F(1, 42) = 4.976, p = .031, \eta_p^2 = .106$), an interaction of Attack and Duration such that the effect of Duration was larger when the attack was slow than when it was fast ($F(1, 42) = 9.722, p = .003, \eta_p^2 = .188$), and there was a nearly significant interaction of Attack and Experiment, $F(1, 42) = 4.035, p = .051, \eta_p^2 = .088$.

Regarding P-center variability, a $2 \times (2 \times 2 \times 2 \times 2)$ (Experiment x (Task x Attack x Duration x Frequency)) mixed-design ANOVA showed main effects of Experiment ($F(1, 42) = 8.645, p = .0053, \eta_p^2 = .171$), Attack ($F(1, 42) = 16.248, p = .000, \eta_p^2 = .279$), and Duration ($F(1, 42) = 16.226, p = .000, \eta_p^2 = .279$). There were no effects of Frequency or Task. There was an interaction between Task and Attack such that the effect of Attack on variability occurred only during click alignment ($F(1, 42) = 5.767, p = .021, \eta_p^2 = .121$), and there was an interaction between Task and Duration such that the effect of Duration occurred only during click alignment, $F(1, 42) = 18.554, p = .000, \eta_p^2 = .306$.

4.5 Probability Density Graphs and Non-Parametric Tests of CA-distributions

As Figures 5 and 6 show, the nine stimulus sounds in each experiment yielded a wide variety of shapes, from narrow peaks (fast-short sounds) on the one extreme, to wide, flat shapes (slow-long sounds) on the other. A Friedman's test showed that there was a significant difference among the distributions of musical sounds ($\chi^2(3) = 28.254, p = .000$), and quasi-musical sounds ($\chi^2(3) = 27.038, p = .000$). Post hoc pairwise Friedman's tests demonstrate a systematic pattern produced by the two factors Attack and Duration (see Table 4). For both musical and quasi-musical stimuli, slow-long sounds are significantly different from all other categories. As to the musical sounds, fast-short is also significantly different from slow-short. The artificial sounds also exhibit more complex beat bin shapes, with more distinct secondary peaks (compare, for example, the slow-

short high and slow-long-low real (Figure 5) versus artificial (Figure 6) stimuli). For descriptive statistics, see Table A3 in the Appendix.

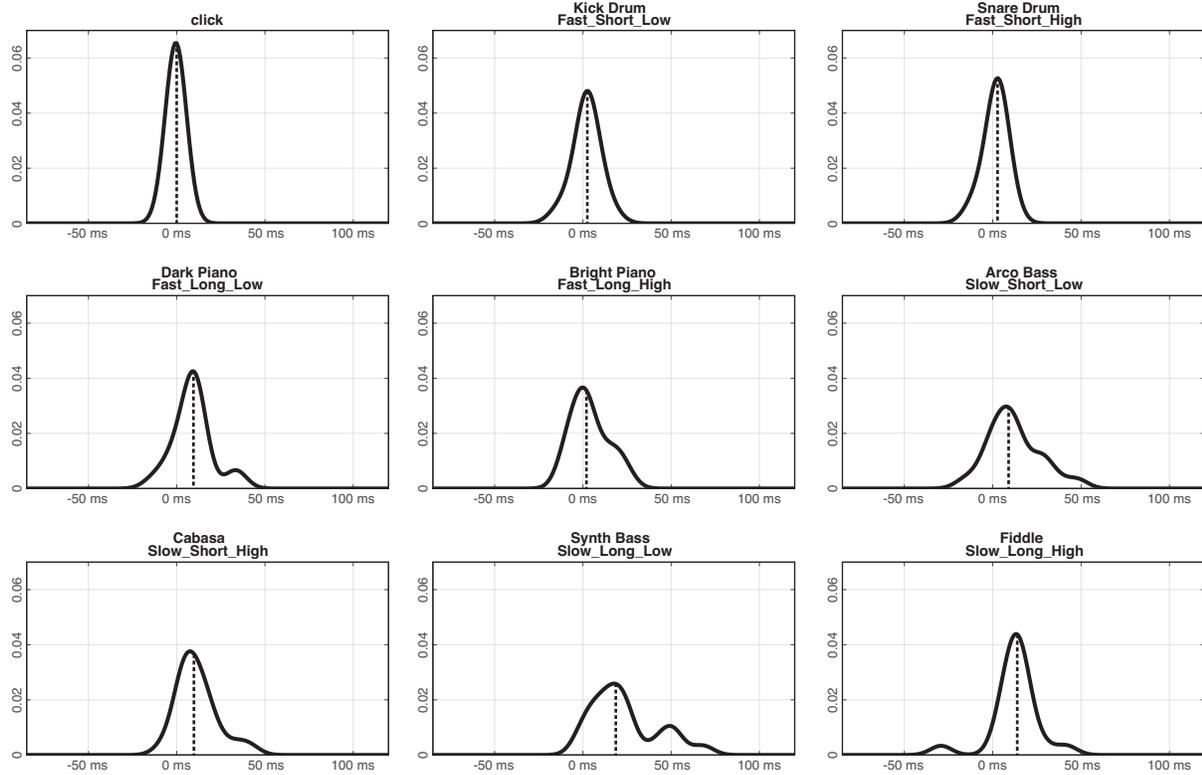


Figure 5. Probability density distributions (probability/time) for musical sounds (experiment 1) in the CA task. Median indicated by vertical stippled line. Fast = fast attack, Slow = slow attack, Short = short duration, Long = long duration, Low = low center frequency, and High = high center frequency.

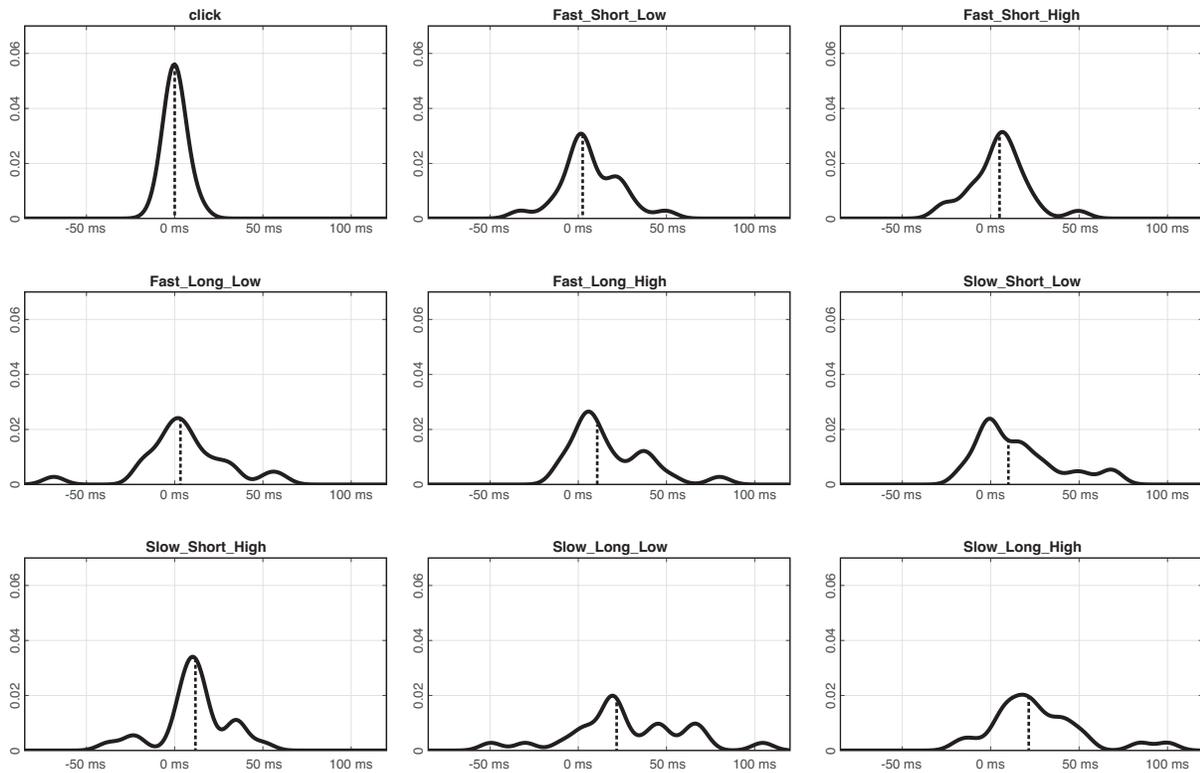


Figure 6. Probability density distributions (probability/time) for quasi-musical sounds (experiment 2) in the CA task. Median indicated by vertical stippled line. Fast = fast attack, Slow = slow attack, Short = short duration, Long = long duration, Low = low center frequency, and High = high center frequency.

Table 4. Results (*p* Values) from Nonparametric Friedman Tests on Paired Differences Between Distributions of Click Alignment Click Locations

Stimuli	Musical Sounds		
	Fast_Short	Fast_Long	Slow_Short
Fast_Long	.354		
Slow_Short	.030	.942	
Slow_Long	.000	.006	.042
Quasi-musical Sounds			
	Fast_Short	Fast_Long	Slow_Short
Fast_Long	.366		
Slow_Short	.132	1.000	
Slow_Long	.000	.006	.006

Note. Significant results in boldface type (Bonferroni corrected for multiple comparisons). Fast = fast attack; Slow = slow attack; Short = short duration; Long = long duration; Low = low center frequency; High = high center frequency.

5. General Discussion

The main findings across both experiments were:

- Slow attack and long duration both lead to a later P-center location, but duration has less effect when the attack is fast;
- Low center frequency leads to later P-center location only for musical sounds, and primarily for longer sounds with slow attack;
- Slow attack and long duration also lead to greater variability in the location of the P-center; that is, to wider beat bins
- The probability density distributions display a systematic pattern of different beat bin shapes with the combination of slow attack and long duration leading to the flattest shape, which indicates a wider tolerance/broader "beat bin". Non-parametric statistical tests confirmed this pattern.
- Slow attack and long duration also produced distributions with complex shapes that suggest these sounds afford multiple locations for beat placement.
- Apart from the click, there is no NMA relative to onset (<5ms), but there is a significant NMA relative to P-center for three out of four of the short musical sounds.

In the following we will look closer into the findings for negative mean asynchrony, before proceeding to P-center location, P-center width (variability), and P-center shape (probability density distributions), respectively.

5.1 Negative mean asynchrony (NMA)

Relevant to our investigation of P-center location is the well-known phenomenon of "negative mean asynchrony" (NMA), the tendency to tap slightly ahead of a series of metronome clicks or tone bursts. The NMA is typically 10-30ms, though musicians (especially percussionists) exhibit reduced or even no NMA (Repp, 2005; Repp & Su, 2013). As with previous studies, all of our participants exhibited an NMA when tapping with the click stimuli (see Table 5), and as in previous studies, there was a great degree of individual variation, with NMAs ranging from -85 to 0 ms (three participants had positive mean asynchronies, ranging from 5-48 ms). When tapping to musical or quasi-musical sounds, however, all NMAs relative to onset are small (less than 5 milliseconds for musical sounds), or non-existing (quasi-musical sounds).

Table 5. Descriptive statistics for click trials in both experiments. P-center locations relative to the physical onset of stimuli. Standard deviations are included in parentheses. All data are in milliseconds.

	Experiment 1 (N=20)	Experiment 2 (N=24)
Click Alignment to Click	-0.50 (0.70)	0.42 (3.73)
Tapping to Click	-21.57 (18.20)	-26.71 (21.96)

Little has been done to study how the microstructure of the target sound affects NMA; our research methodology, which combines both a tapping and an alignment tasks with systematic variations in the target stimulus provides a framework for further investigation of the NMA. Vos et al. (1995) hypothesized that in synchronization tasks participants use the P-center, rather than the physical onset of a tone, as the target for the synchronization task. Thus, when the P-center occurs later relative to the sound onset of a target sound/tone, the NMA (relative to the onset) is correspondingly shifted. As can be seen in Figures 2 and 3 above, for most stimuli the Tap-based P-center occurs earlier than the CA-based center, save for one stimulus (fast attack, long duration, low frequency). Tests of the difference sound by sound show that the average tap location is significantly early compared to the parallel click location for three out of the four short musical sounds. This is in accordance with the results of Vos et al. (1995), who found reduced negative mean asynchrony when increasing stimulus duration. Vos et al. also found effect of rise time, but no such pattern was found in the present study. Though only three of these individual paired differences were significant (in part due to the small magnitude of the differences between Tap- vs. CA-based P-centers), this pattern is suggestive; future studies with a set of more expert tappers may yield more significant results.

5.2 Effects of acoustical factors on P-center location

Both experiments show that sounds with a fast attack lead to P-centers that are very close to the attack peak of the sounds (see Figure 2). Duration also has a strong effect, and as hypothesized longer duration generally leads to later P-centers. However, this effect is significantly reduced in the presence of a fast attack. These results confirm previous studies on P-center perception and synchronization for musical and synthesized sounds (Vos & Rasch, 1981; Gordon, 1987; Vos et al., 1995; Scott, 1998; Seton, 1989).

The effect of Frequency on P-center location was evident only in the musical stimuli, and only with longer durations, where low frequency led to later mean P-center locations. Previous studies on the effect of frequency are very few and have different designs, but point in the same direction as our result (Seton 1989; Hove et al. 2007). Slower response to low frequencies in the cochlea (Wojtczak 2012), as well as the physical fact that a low frequency sinusoid takes longer time to complete a wave cycle, can explain the later P-centers of low-frequency musical sounds found in the present study. Relatedly, Wojtczak et al. (2017) found a robust asymmetry in the perception and neural coding of synchrony that reflects greater tolerance for delays of low- relative to high-frequency sounds than vice versa. They suggest that the auditory pathways may have developed a higher tolerance to the *de facto* low-frequency delays that happen in the auditory periphery, thereby providing veridical perceptual experiences of simultaneity.

The interaction effect found in the present study, that is, the effect of Frequency is stronger when Duration is long, could be seen to confirm Gordon's (1987) finding that only sounds with longer rise times were influenced by spectral cues because sounds with longer rise times also tend to be long.

The effect of frequency was not found in our second experiment using quasi-musical sounds. Comparing the frequency registers used in the two experiments, the low center frequencies used in the musical stimuli (Exp. 1) were slightly more extreme than the low center frequency of the quasi-musical stimuli (Exp. 2; see Tables 1 & 2). Perhaps more important, however, is the richness of information in the frequency domain of musical sounds compared to filtered noise, as well as the conventional musical roles of the different musical sounds used (bass sounds and kick drum versus snare drum, fiddle and percussion). This may have produced a musically meaningful distinction between sounds with low and high center frequency that was not present in Experiment 2. Interestingly, Wojtczak et al. (2017) found that the effect of low leading tones was only observed in the conditions in which the low and high tones did not substantially overlap in spectrum. Overlap between the spectra of the two filtered-noise tones might thus explain the missing effect of frequency in this case. However, further research is needed to understand the role of Frequency in different contexts and different critical ranges.

5.3 Effects of acoustical factors on P-center Variability and Beat Bin Width

Approaching the perceptual center as a bin of possible locations rather than a single point in time, the standard deviation of P-center location, rather than a source of "noise" (produced by participants with greater or lesser consistency in task performance), becomes a useful dependent variable both within and across participant responses. As expected, we found that slow attack and long duration both lead to greater variability in the location of the P-center (see Figure 3; note relative lengths of the error bars), that is, to wider beat bins in both experiments. This has also been found in previous studies (Gordon, 1987; Wright, 2008). The click alignment task was most sensitive to stimulus-driven effects on P-center variability, whereas Tapping shows a constant level of variability across all stimuli.

5.4 Effects of Acoustical Factors on Beat Bin Shape

Non-parametric Friedman's tests of differences between click alignment distributions showed a systematic pattern produced by the sound factors Attack and Duration. Inspecting the probability density graphs, we see that responses for fast-attack musical sounds (Figure 5) generally cluster around a close-to-zero positioned attack peak. This is most salient for the fast-short sounds, which were significantly different from both slow-long and slow-short sounds. Regarding the musical sounds with fast attack and *long* duration (that is, the two piano sounds), the point preferred by most participants is still this peak, but the peaks are less pronounced and the shape of the beat bin is slightly right-tailed.

As regards slow-attack musical sounds, their peaks occur later (relative to acoustic onset), and the beat bins are generally skewed and left-leaning with right-ended tails. We also see a clear effect of frequency: the slow, *low*-frequency sounds have the widest and most left-leaning beat bins (positive skewness) and a longer probability "tail" (right-tailed kurtosis). Accordingly, slow-long-low displays the lowest peak and the flattest probability density distribution of all sounds; here all the sound factors seem to work in the direction of widening the beat bin.

Looking into the probability density graphs for the quasi-musical sounds (Figure 6), we generally find lower peaks and wider bins, which reflects the higher grand mean standard deviation (20 ms compared to 14 ms in Experiment 1). This could be partly explained by the participants in experiment 1 having more musical training and thus being better at the tasks. Another possible

explanation is that synchronizing to more traditional musical sounds is easier due to their familiarity—a practice effect, if you will. Yet another possibility is that the notched noise stimuli in Experiment 2, while acoustically and psycho-acoustically simpler, give rise to a more complex perceptual and sensori-motor response because these sounds are not the natural result of a typical sound-producing action (musical instruments do not, by and large, produce narrow-band notched noise). Thus, while there are normative perception-action coordination aspects to the musical sounds, as the sonic results of human agents, this does not hold for the stimuli used in Experiment 2.

Nonetheless, overall the probability density distributions of the two experiments display similar patterns, which were confirmed in the statistical analysis (see Table 4): Slow attacks and longer durations lead to later, wider and flatter beat bins. Interestingly, in both experiments the pairs for fast-long and slow-short sounds resemble each other more closely than fast-short and slow-short (the latter two were significantly different in the case of musical sounds). In the former pair, both factors are different, whereas in the latter the Attack factor is changed whereas Duration is held constant. This indicates that there is sensory-perceptual interference between the two factors, that is, the effect of one factor tends to be canceled out when the other factor works in the opposite direction. Using the terminology of Melara and Marks (1990), this means that positively correlated combinations of Attack and Duration, that is, fast-short and slow-long, might cause a redundancy gain whereas negatively related combinations, such as fast-long and slow-short, cause a redundancy loss. Similar effects have been found for several paired dimensions in research into isolated sounds (see, for example, Grau & Kemler-Nelson [1988] on pitch and loudness, Melara & Marks [1990] on pitch and timbre, and timbre and loudness, and Tekman [2002] on timing and intensity). The extent to which this is a systematic pattern, as well as whether it also extends to Frequency, are topics for future research.

6. Conclusion and Future Research

P-centers, rather than being durationless moments within the microstructure of a musical or speech sound, have a temporal extent or width, and a temporal shape. Depending on a range of acoustical factors, a musical P-center may vary from a narrow “metronomic” point in time to a wide “beat bin” (Danielsen, 2010). It should be noted here that the source of this variation is perceptual: while a P-center is produced by a sonic event, the event is not a point or a bin in

itself. P-centers are psycho-acoustic phenomena, with an emphasis on the "psycho" side of the equation. The two experiments reported on here confirm previous studies which have shown that attack and duration are key cues for the location of P-centers: slower attacks and longer duration lead to later P-center location. However, we also show that Attack and Duration together produce a systematic pattern of P-center *shapes*: from narrow peaks close to the onset of the sounds (fast attack, short duration), to onset peaks combined with clearly left-leaning bins of moderate width (fast attack, long duration), to wide bins displaying different numbers and positions of peaks, as well as varying skewness and kurtosis (slow attack, long duration). The results indicate that positively correlated combinations of Attack and Duration, that is, fast-short and slow-long sounds, cause a redundancy gain whereas negatively related combinations, such as fast-long and slow-short sounds, cause a redundancy loss.

P-centers/beat bins are affordances for action, especially in musical contexts. We care about the P-center of a musical sound not only so we can know what kind of sound it is, but also (and perhaps primarily) so we can hear and move in synchrony with it. The two experiments reported on here illustrate the usefulness of a varied set of tasks/responses for obtaining data on sensorimotor perception and action. Tapping tells one story: the movement dynamics of the tapping task are likely to be the cause of the reduced variability in P-center location in the tapping condition, as the necessity of maintaining a stable tapping rate while executing a repetitive motion is a "ballistic constraint" on the variability of P-center location as well as the inter-tap interval. The CA data tell another story—one with different P-center locations, as well as varying beat bin widths and shapes. Moreover, tapping trials tended to locate the P-center earlier (relative to stimulus onset) than CA trials, indicative of a persistence of the negative mean asynchrony across a range of stimuli, and not just metronome clicks.

Our current results suggest several avenues for future research. The alignment task should investigate the use different "probes" beyond a click, such as short tones whose center frequency matches (or does not match) that of the target tone, as alignment tasks inherently involve the production of a fused sound. Also, additional evidence is needed to build the case for beat bins. In addition to the alignment task, an experimental task where participants have to judge whether a probe click appears early, on time, or late relative to a target tone could yield more information on beat-bin structure, as well as avoid some of the difficulties involved in using the method of

adjustment. A wider range of frequencies for the target sounds, both real and artificial, should be explored. The extent to which the sonic dimensions produced by the various acoustical factors are separable or integral is also a topic for further research. In the future, we also wish to further examine the effects of training and musical enculturation on P-center perception; participants from a broader range of musical backgrounds and cultures will give greater insight on the effect of familiarity on P-center perception, and hence on rhythmic perception more generally.

Acknowledgments

This work was partially supported by the Research Council of Norway through its Centres of Excellence scheme, project number 262762, and the TIME project, grant number 249817.

References

- Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition*, 48(1), 66–79. <https://doi.org/10.1006/brcg.2001.1304>
- Bjerke, K. (2010) Timbral relationships and microrhythmic tension: shaping the groove experience through sound. In A. Danielsen (Ed.), *Musical Rhythm in the Age of Digital Reproduction* (pp. 85-101). Farnham: Ashgate/Routledge, UK.
- Brøvig-Hanssen, R., & Danielsen, A. (2016). *Digital signatures: The Impact of Digitization on Popular Music Sound*. Cambridge, Mass.: MIT Press.
- Carlsen, K., & Witek, M. A.G. (2010). Simultaneous rhythmic events with different schematic affiliations: Microtiming and dynamic attending in two contemporary R&B grooves. In A. Danielsen (Ed.), *Musical Rhythm in the Age of Digital Reproduction* (pp. 51–68). Farnham: Ashgate/Routledge, UK.
- Danielsen, A. (2006). *Presence and Pleasure: The Funk grooves of James Brown and Parliament*. Middletown, Connecticut: Wesleyan University Press.
- Danielsen, A. (2010). Here, there, and everywhere. Three accounts of pulse in D'Angelo's 'Left and Right'. In A. Danielsen (Ed.), *Musical Rhythm in the Age of Digital Reproduction* (pp. 19-36). Farnham: Ashgate/Routledge, UK.
- Danielsen, A., Haugen, M. R., & Jensenius, A. R. (2015). Moving to the beat: Studying entrainment to micro-rhythmic changes in pulse by motion capture. *Timing & Time Perception*, 3(1–2), 133–154. <http://dx.doi.org/10.1163/22134468-00002043>
- Fowler, C. A. (1979). “Perceptual centers” in speech production and perception. *Attention, Perception, & Psychophysics*, 25(5), 375–388. <https://doi.org/10.3758/BF03199846>
- Gordon, J. W. (1987). The perceptual attack time of musical tones. *The Journal of the Acoustical Society of America*, 82(1), 88–105. <https://doi.org/10.1121/1.395441>
- Grau, J. W., & Nelson, D. K. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General*, 117(4), 347-370. doi:10.1037/0096-3445.117.4.347
- Haugen, M. R. (2016). *Music–Dance: Investigating Rhythm Structures in Brazilian Samba and Norwegian Telespringar Performance* (Doctoral dissertation), University of Oslo, Oslo, Norway.

- Hove, M. J., Marie, C., Bruce, I. C., & Trainor, L. J. (2014). Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. In *Proceedings of the National Academy of Sciences*, *111*(28), 10383–10388. <https://doi.org/10.1073/pnas.1402039111>
- Howell, P. (1984). An acoustic determinant of perceived and produced anisochrony. In *Proceedings of the 10th International Congress of Phonetic Sciences* (pp. 429-433). Foris Publications: Amsterdam, Netherlands.
- Johansson, M. (2010). The concept of rhythmic tolerance. Examining flexible grooves in Scandinavian folk fiddling. In A. Danielsen (Ed.), *Musical Rhythm in the Age of Digital Reproduction* (pp. 19-36). Farnham: Ashgate/Routledge, UK.
- Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323–355. <http://dx.doi.org/10.1037/0033-295X.83.5.323>
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review* *106*(1), 119–159. <http://dx.doi.org/10.1037/0033-295X.106.1.119>
- Large, E. W., & Snyder, J. S. (2009). Pulse and meter as neural resonance. *Annals of the New York Academy of Sciences*, *1169*(1), 46–57. <https://doi.org/10.1111/j.1749-6632.2009.04550.x>
- Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic bulletin & review*, *1*(4), 476-490.
- Marcus, S. M. (1981). Acoustic determinants of perceptual center (P-center) location. *Attention, Perception, & Psychophysics*, *30*(3), 247–256, <http://dx.doi.org/10.3758/BF03214280>
- Melara, R., & Marks, L. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Attention, Perception & Psychophysics*, *48*(2), 169–178.
- Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, *83*(5), 405-408. <http://dx.doi.org/10.1037/0033-295X.83.5.405>
- Nymoén, K., Danielsen, A., & London, J. (2017). Validating attack phase descriptors obtained by the Timbre Toolbox and MIRtoolbox. In *Proceedings of the SMC Conferences* (pp. 214–219). Aalto University, Finland.

- Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., & Peretz, I. (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia*, *49*(5), 961–969. <https://doi.org/10.1016/j.neuropsychologia.2011.02.002>
- Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acoustica*, *43*(2), 121–131.
- Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*(6), 969–992. <https://doi.org/10.3758/BF03206433>
- Repp, B. H., & Doggett, R. (2007). Tapping to a very slow beat: A comparison of musicians and non-musicians. *Music Perception*, *24*(4), 367–376. <https://doi.org/10.1525/MP.2007.24.4.367>
- Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, *20*(3), 403–452. <https://doi.org/10.3758/s13423-012-0371-2>
- Rossing, T., Moore, F., & Wheeler, P. (2002). *The Science of Sound. Third Edition*. San Francisco: Addison Wesley
- Schutz, M., & Vaisberg, J. M. (2014). Surveying the temporal structure of sounds used in *Music Perception*. *Music Perception: An Interdisciplinary Journal*, *31*(3), 288–296. <https://doi.org/10.1525/mp.2014.31.3.288>
- Scott, S. K. (1998). The point of p-centres. *Psychological Research*, *61*(1), 4–11. <http://dx.doi.org/10.1007/PL00008162>
- Seton, J. C. (1989). *A Psychophysical Investigation of Auditory Rhythmic Beat Perception* (Doctoral dissertation), University of York., York, UK.
- Stover, C. (2009). *A Theory of Flexible Rhythm Spaces for Diasporic African Music* (Doctoral dissertation), University of Washington, Washington, DC.
- Tekman, H. G. (2002). Perceptual integration of timing and intensity variations in the perception of musical accents. *The Journal of General Psychology*, *129*(2), 181–191. <http://doi.org/10.1080/00221300209603137>
- Villing, R. (2010). *Hearing the Moment: Measures and Models of the Perceptual Centre* (Doctoral dissertation). National University of Ireland, Maynooth, Ireland.
- Vos, J., & Rasch, R. (1981). The perceptual onset of musical tones. *Perception & Psychophysics*, *29*(4), 323–335. <http://dx.doi.org/10.3758/BF03207341>
- Vos, P. G., Mates, J., & van Kruysbergen, N. W. (1995). The perceptual centre of a stimulus as the cue for synchronization to a metronome: Evidence from asynchronies. *The Quarterly*

Journal of Experimental Psychology Section A, 48(4), 1024–1040.

<https://doi.org/10.1080/14640749508401427>

Wojtczak, M., Beim, J. A., Michey, C. & Oxenham, A. J. (2012) Perception of across-frequency asynchrony and the role of cochlear delays. *Journal of the Acoustical Society of America*, 131(1), 363–377.

Wojtczak, M., Mehta, A. H., & Oxenham, A. J. (2017). Rhythm judgments reveal a frequency asymmetry in the perception and neural coding of sound synchrony. *Proceedings of the National Academy of Sciences*, 114(5), 1201–1206. <http://doi.org/10.1073/pnas.1615669114>

Wright, M. J. (2008). *The Shape of an Instant: Measuring and Modeling Perceptual Attack Time with Probability Density Functions* (Doctoral dissertation), Stanford University, Stanford, CA.

APPENDIX

	Click	Fast Short Low	Fast Short High	Fast Long Low	Fast Long High	Slow Short Low	Slow Short High	Slow Long Low	Slow Long High	Average
Instrument		Kick Drum	Snare Drum	Dark Piano	Light Piano	Arco Bass	Cabasa	Synth Bass	Fiddle	
Click Alignment	-0.50 (0.70)	2.13 (7.78)	1.58 (8.04)	9.00 (13.91)	3.90 (16.45)	12.04 (13.21)	11.44 (15.11)	23.44 (20.66)	12.78 (17.17)	8.42 (12.46)
Tapping	-21.57 (18.20)	-4.15 (14.36)	-0.32 (13.19)	8.06 (14.88)	-1.18 (14.05)	-0.46 (14.45)	-3.36 (14.16)	16.89 (15.62)	9.77 (16.25)	0.41 (15.02)
Average	-11.04 (9.45)	1.01 (11.07)	0.63 (10.62)	8.53 (14.40)	1.36 (15.25)	5.79 (13.83)	4.04 (14.64)	20.17 (18.14)	11.28 (16.71)	

Table A1. Descriptive statistics of results for musical sounds. P-center average locations relative to the physical onset of each stimulus for both tasks (N = 20). Standard deviations are included in parentheses. All data are presented in milliseconds. Fast = fast attack, Slow = slow attack, Short = short duration, Long = long duration, Low = low frequency range, and High = high frequency range.

	Click	Fast Short Low	Fast Short High	Fast Long Low	Fast Long High	Slow Short Low	Slow Short High	Slow Long Low	Slow Long High	Average
Click Alignment	0.42 (3.73)	6.85 (16.82)	4.83 (15.08)	5.83 (18.15)	17.46 (19.62)	15.50 (18.92)	11.08 (19.62)	34.46 (28.05)	29.96 (26.69)	13.71 (18.52)
Tapping	-26.71 (21.96)	-0.52 (17.26)	-0.56 (17.41)	14.72 (17.09)	12.74 (17.09)	9.15 (20.11)	10.41 (18.86)	24.43 (17.11)	26.48 (17.88)	7.79 (18.31)
Average	-13.15 (12.84)	3.17 (17.04)	2.14 (16.25)	10.28 (17.62)	15.10 (18.36)	12.33 (19.51)	10.74 (19.24)	29.44 (22.58)	26.72 (22.29)	

Table A2. Descriptive statistics of results for quasi-musical sounds. P-center locations relative to the physical onset of each stimulus for both tasks (N = 24). Standard deviations are included in parentheses. All data are presented in milliseconds. Fast = fast attack, Slow = slow attack, Short = short duration, Long = long duration, Low = low center frequency, and High = high center frequency.

	N	25th percentile	Median	75th percentile
CA_Fast_Short_musical	20	-2,25	3,00	4,75
CA_Fast_Long_musical	20	-1,50	6,00	10,25
CA_Slow_Short_musical	20	4,25	10,50	23,50
CA_Slow_Long_musical	20	10,25	15,50	20,50
CA_Fast_Short_quasi-musical	24	-2,25	2,50	13,50
CA_Fast_Long_quasi-musical	24	1,25	9,00	19,50
CA_Slow_Short_quasi-musical	24	2,50	13,50	26,50
CA_Slow_Long_quasi-musical	24	14,25	30,50	46,00

Table A3. Non-parametric descriptive statistics of distributions of CA click locations.