



A comparison of methods for investigating the perceptual center of musical sounds

Justin London¹ · Kristian Nymo^{2,3,4} · Martin Torvik Langerød^{2,3} · Marc Richard Thompson⁵ · David Loberg Code⁶ · Anne Danielsen^{2,3}

Published online: 10 May 2019
© The Psychonomic Society, Inc. 2019

Abstract

In speech and music, the acoustic and perceptual onset(s) of a sound are usually not congruent with its perceived temporal location. Rather, these "P-centers" are heard some milliseconds after the acoustic onset, and a variety of techniques have been used in speech and music research to find them. Here we report on a comparative study that uses various forms of the method of adjustment (aligning a click or filtered noise in-phase or anti-phase to a repeated target sound), as well as tapping in synchrony with a repeated target sound. The advantages and disadvantages of each method and probe type are discussed, and then all methods are tested using a set of musical instrument sounds that systematically vary in terms of onset/rise time (fast vs. slow), duration (short vs. long), and center frequency (high vs. low). For each method, the dependent variables were (a) the mean P-center location found for each stimulus type, and (b) the variability of the mean P-center location found for each stimulus type. Interactions between methods and stimulus categories were also assessed. We show that (a) in-phase and anti-phase methods of adjustment produce nearly identical results, (b) tapping vs. click alignment can provide different yet useful information regarding P-center locations, (c) the method of adjustment is sensitive to different sounds in terms of variability while tapping is not, and (d) using filtered noise as an alignment probe yields consistently earlier probe-onset locations in comparison to using a click as a probe.

Keywords Rhythm · P-center · Tapping task · Alignment task · Negative mean asynchrony · Microtiming

APP Keywords Music cognition · Sound recognition · Psychoacoustics · Temporal processing

Introduction

When two sounds are synchronized, we do not precisely align their acoustic or perceptual onsets, but rather align their *perceptual centers* (P-centers). This is because there is a distinction between acoustic onset of a sound (which can be

subliminal), the perceptual onset (at which point a sound can be detected), and the P-center itself, which is the reference point for where a sound is placed relative to other sounds in a rhythmic sequence (Morton, Marcus, & Frankish, 1976). While originally conceived as a discrete location sometime after the acoustic onset of a sound, subsequent research has shown that P-centers may have some temporal spread and shape (Danielsen et al., 2019; Gordon, 1987; Wright, 2008). Likewise, when a sequence of events occurs, it is the timing between successive P-centers that determines whether the sequence is perceived as regular or irregular.¹

A variety of methods have been used to determine the P-center of a sound (Villing, 2010, provides an excellent overview of the history of P-center research, including detailed descriptions and analyses of the models developed by Gordon, 1987; Harsin, 1997; Howell, 1988; Pompino-

✉ Justin London
jlondon@carleton.edu

¹ Carleton College, Northfield, MN, USA

² RITMO Center for Interdisciplinary Studies of Rhythm, Time, and Motion, University of Oslo, Oslo, Norway

³ Department of Musicology, University of Oslo, Oslo, Norway

⁴ Department of Informatics, University of Oslo, Oslo, Norway

⁵ Department of Music, University of Jyväskylä, Jyväskylä, Finland

⁶ School of Music, Western Michigan University, Kalamazoo, MI, USA

¹ The term P-center has come to be associated with auditory and speech events, but in principle the P-center concept is applicable to events in any modality (Villing, 2010; see also Morton et al., 1976).

Marschall, 1989; Marcus, 1981; Scott, 1993; Vos & Rasch, 1981). First, one may use the psychophysical method of adjustment: a repetitive, isochronous series of target sounds is presented (i.e., a "loop"), along with either (a) another set of sounds, or (b) a series of clicks or very brief tones, the latter having the advantage of having a precise temporal location, given their extremely brief duration. 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. Second, one may have the participants produce a series of target sounds with systematic variations (e.g., "pa" vs. "la" syllables, which differ in initial consonant but not in vowel sound); these sounds are paced with a metronome, and then participants freely reproduce the sounds with or without the metronome while maintaining a steady, isochronous pace. Finally, participants can tap along with the sounds. One can monitor the alignment of the taps themselves, or one can perturb a target sound (presenting it a little early or late) and observe the phase-correction response, which has been well documented in tapping studies (see Repp, 2005; Repp & Su, 2013).

There are drawbacks, however, to any of these methods of 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 familiar task for musicians, since this is what they must do when playing together in an ensemble. While coarse-grained aspects of alignment in the in-phase task can be related to the respective onsets of the target and the click, fine-grained alignment may rely more on timbral cues – a change in the spectral quality of the articulation of the blended (click+target) sound – than on timing per se. Kochanski and Orphanidou (2008) had participants read a repetitive text with a pacing metronome, and found the loudest syllables were aligned with the metronome click – but this can also be regarded as a strategy for dealing with possible masking effects of the pacing metronome, rather than the loudness itself being the primary cue for the P-center's location. Bechtold and Senn (2018) presented click and target sounds dichotically, which manages the masking problem to some extent, but lessens the ecological validity of the task, since in most instances auditory cues for synchronization are heard non-dichotically.

While using an anti-phase click-alignment task addresses the masking/timbral blend problem, it raises other problems. The anti-phase alignment of the clicks with the target sounds creates a composite stream of sounds at twice the rate of the target sounds. Our perception of a rhythmic sequence differs, however, for inter-onset intervals (IOIs) within a range of 100–300ms versus those between 300 and 1,000 ms, with a preference for sequences in the 500–600 ms range (Fraisse, 1984; see London, 2012, for a review of recent literature). Thus, comparisons between in-phase and anti-phase

measurements may involve different timing mechanisms and/or strategies. Moreover, the anti-phase task presumes that participants will produce purely isochronous composite streams, from which the P-center of the target sounds can be inferred. However, in musical contexts isochronous "off-beat" locations are not always veridically perceived, as slight deviations from isochrony (which can be linked to the metrical position of a note/stimulus) are heard as normatively isochronous (Dixon, Goebel, & Cambouropoulos, 2006; Repp 1995, 1998). This is the case with stimuli, such as piano tones, whose articulation is relatively simple; more complex tones, and combinations of tones (i.e., as in the case of targets sampled from ensemble performance) may further influence the target location of the anti-phase clicks, as these sounds may influence the extent to which the sequence may be heard as "swung" versus "unswung," in particular if the target sounds are drawn from musical styles where rhythms are normally played in a manner producing a more or less non-isochronous pulse or subdivision (i.e., swing jazz, samba, funk). Note that the problem of off-beat timing (and the presumption of isochrony) also holds for the alternating-syllable-production tasks described above.

One may use the method of adjustment, but rather than aligning auditory clicks with the target sounds, participants align a visual signal (e.g., a flashing light) with the target signal. However, visual metronomes present other problems for P-center detection tasks, as it has been shown that our ability to synchronize with discontinuous visual cues such as flashing lights versus analogous auditory stimuli is slower/less accurate by an order of magnitude or more (Repp, 2003). While continuous visual stimuli afford much better synchronization (Hove et al., 2013; Iversen et al., 2015), the use of a visual metronome as a timing probe combined with an auditory target creates a cross-modal perception and integration problem, which is absent when the timing cues are all in the same sensory modality. Moreover, studies of coordination among ensemble musicians have shown that auditory cues alone are as good, if not better than, combined audio-visual cues for musical synchronization tasks (Thompson et al., 2015).

Most studies have used a metronome click as an alignment probe, either with in-phase or anti-phase alignment. In his seminal experiment, however, Gordon (1987) used a range of sounds as probes. His targets were a set of synthesized orchestral instrument sounds, and his probes were a subset of those sounds (E-flat clarinet, bassoon, and cello played *sul tasto*, as well as a conga drum sound). He presumed that the P-center measurement would be the same whether in-phase or anti-phase probe methods were used, as well as irrespective of the probe sound used; in his data analysis results were pooled (Gordon 1987, p. 90).

Another methodology involves synchronizing a repetitive action, such as tapping with the target rhythmic stimulus,

rather than an overt judgment of synchronicity, as tapping or drumming is a familiar and understandable response to a rhythmic stimulus. However, 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 brief tone in a simple in-phase synchronization task (see Aschersleben, 2002, and Repp, 2005, for recent reviews). NMAs can vary from 20–80 ms for untrained subjects, and while very small (10–30 ms), they may still persist for musicians (Repp & Doggett, 2007; see also Danielsen et al., 2019).

The current study reports on three experiments that investigated various methodological issues involved in studying the P-centers of musical sounds. The broader motivation for our study is to gain an understanding of the psychoacoustic landmarks that musicians use in ensemble performance. This involves the production of sounds in real time with others to create an aggregate sound that not only occurs at a given location in time, but also gives rise to a sense of rhythmic flow with a particular character. P-centers are properties of sounds that emerge in particular listening/experimental contexts, and, indeed, combinations of sounds may give rise to P-center percepts that are “more than the sum of their parts,” especially given (and as noted above) that P-centers are not simply points in time, but have temporal spreads and shapes (Danielsen et al. 2019; Gordon, 1987; Wright, 2008). Nonetheless, we wanted to assess the P-centers of a set of typical sounds used in musical contexts, and move toward an experimental task/context that is closer to what musicians do in actual performance. As a first step, we compare P-center results using the method of adjustment versus a coordinated rhythm production task. Thus, in the first experiment three different methods were tested using the same set of stimuli: in-phase alignment of a click probe, anti-phase alignment of the click probe, and in-phase tapping (see Table 1). The target sounds varied systematically in terms of three acoustic dimensions: attack/rise time, duration, and center frequency. Given the problems of both the temporal acuity of visual versus auditory modalities, and the added factor of cross-modal integration, we did not use a visual metronome or similar probe.² As musical performance typically involves the coordination of sounds other than clicks, the second experiment examined the characteristics of different probe sounds (2-ms click vs. 100-ms filtered noise burst) using clicks and various forms of noise as targets in an in-phase alignment task. The third experiment used the same target

stimuli as the first, but used the 100-ms noise burst as the probe in an in-phase alignment task. We focused on in-phase alignment in Experiments 2 and 3 both because it is analogous to the task involved in real-world music ensemble performance and because our first experiment showed little difference between anti-phase and in-phase alignment tasks.

By employing a range of P-center tasks and probes, the aim of these three experiments is to examine if and to what extent these different methods produce the same or different results in terms the location and variability of the P-center in general, and for each sound in particular.

Experiment 1

Participants

Twenty music students/semi-professional musicians (nine female) were recruited from the Oslo area. Musician participants were recruited because a pilot experiment showed that people without musical training often struggled to complete the experimental tasks. They received a gift card (value 400 NOK) for their participation in the experiments. Median age was 25.5 years (mean = 30.5, SD = 12.5; max = 60, min = 20). Two participants reported 1–4 years of music training, two participants had 5–10 years of training, and the remaining 16 participants had more than 10 years of training. As their main instrument, ten participants reported guitar/bass, two drums, three woodwind or brass, three vocals and two string instruments. All participants practiced on their instrument; ten participants practiced 1–6 h/week and ten more than 6 h/week. All participants reported an ability to read music.

Stimuli

The stimuli consisted of sounds of eight instruments that represent a balanced design of the three following acoustical factors, which we will refer to as Attack (shorter vs. longer rise time), Duration (of the stimulus sound, as opposed to the stimulus IOI), and Frequency (high vs. low spectral centroid).

Table 1 Overview of the three experiments

Experiment 1	Clicks aligned in-phase with target sounds, method of adjustment
	Clicks aligned anti-phase with target sounds, method of adjustment
	Tapping in-phase with target sounds
Experiment 2	Click-Noise benchmark tests
Experiment 3	Noise aligned in-phase with target sounds, method of adjustment

² Pilot studies with several visual metronomes were carried out, but then abandoned due to the inherent problems of a visual metronome design, as well as computer implementation and data synchronization issues in addition to the inherent cross-modal problem(s) of a visual metronome and an auditory target.

Manual measurements of the waveforms and results from the MIR toolbox for Matlab version 1.7 (Mathworks, Natick, MA, USA) are reported in Table 2.³ 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 one of the experimenters and controlled by a second.

Apparatus and method

In Experiment 1 we tested three separate tasks/methods:

- A) *Click alignment, in-phase condition (CA)*: During the CA trials, the participants’ task was to align a click track with the target stimulus; click and stimuli were both looped at a 600-ms interval (tempo=100 beats per minute (bpm)). Clicks were initially presented with a random offset, uniformly distributed between 100 and 200 ms before or after the target sound. In each trial, participants manipulated the offset of the two sounds by moving an on-screen cursor using the mouse and/or arrow keys; each individual press of the arrow key moved the click 1 ms. 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 quasi-random, constrained so that participants never heard the same stimulus on back-to-back trials.
- B) *Click alignment, anti-phase condition (AP)*: Stimuli, procedure, and number of trials were the same as in A, save for the task/instructions: rather than aligning the clicks on top of each stimulus sound, the task was to interleave the clicks and sounds to produce an even/isochronous sequence (i.e., with an effective IOI of 300 ms).
- C) *Tapping (TAP)*: In the tapping trials, participants used a pair of clave sticks to produce sounds in synchrony with the target stimulus (again looped at a 600-ms interval); claves were chosen as they produce a crisp percussive sound and are relatively easy to play. Each loop repeated for 20 s. 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 nine target stimuli was randomly ordered. Participants took from 5–10 min to finish the tapping trials.

³ The MIR toolbox used to analyze the stimuli systematically reports longer durations for short sounds than our manual measurements. This is due to the windowing technique used for calculating the amplitude envelope (window length 20 ms, with 98% overlap). Furthermore, durations of long sounds are under-reported 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 Nymoer et al. (2017).

Table 2 Sound stimuli and alignment probes used in Experiments 1 and 3

Stimulus parameters	Click			Noise*	Fast			Slow												
	Fast	Short	Low		Fast	Long	Low	Short	High	Long	High									
Instrument																				
Attack	Fast			Slow	Snare drum	Fast	Fast	Dark piano	Fast	Light piano	Fast	Arco bass	Slow	Cabasa	Slow	Synth bass	Slow	Fiddle	Slow	
Duration (waveform /MIR toolbox)	2 / 22 ms			100 ms	25 / 38 ms	487 / 371 ms	487 / 371 ms	487 / 371 ms	318 / 211 ms	318 / 211 ms	66 / 70 ms	66 / 70 ms	49 / 61 ms	49 / 61 ms	220 / 220 ms	220 / 220 ms	220 / 220 ms	105 / 89 ms	105 / 89 ms	
Frequency range (pitch in parenthesis where relevant)	3,000 Hz			Bandpass filter centered at 3,000 Hz	High	Low (65,4 Hz)	Low (65,4 Hz)	Low (65,4 Hz)	High (659,3 Hz)	High (659,3 Hz)	Low (65,4 Hz)	Low (65,4 Hz)	High	High	Low (32,7 Hz)	Low (32,7 Hz)	Low (32,7 Hz)	High (479 Hz)	High (479 Hz)	
Spectral Centroid (MIR toolbox)	3,720 Hz			4,809 Hz	2,831 Hz	623 Hz	623 Hz	893 Hz	893 Hz	538 Hz	538 Hz	8,199 Hz	8,199 Hz	781 Hz	781 Hz	781 Hz	781 Hz	1,581 Hz	1,581 Hz	

*Noise was not used in Experiment 1

Participants completed the CA and AP tasks 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. In the TAP task stimuli participants were listening through acoustically transparent headphones (Koss PortaPro), which allowed them to clearly hear their tapping during those trials. To eliminate timing latencies in the TAP setup, the stimulus was routed both to participants' headphones and to a mono recording channel on an audio interface (PreSonus Firebox); tapping sounds were recorded on a parallel mono channel using a Shure SM57 unidirectional microphone.

The order in which participants completed the tasks in Experiment 1 was counterbalanced. Between or after tasks, participants answered a series of background questions pertaining to their musical training and musical consumption, as well as age, gender, and nationality. For the CA and AP trials, between one and eight participants ran trials at individual workstations in the University of Oslo (UiO) computer music lab. The TAP trials participants were recorded 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. The experimenter waited nearby should any questions/problems arise.

In all three sets of trials probe locations are reported in milliseconds relative to the physical onset of the stimulus. A positive probe location means that the physical onset of the probe sound occurs after the physical onset of the stimulus sound. Participant responses for the CA and AP trials were averaged across four trials to produce a location for each participant per stimulus; standard deviations of each of the participants' responses were calculated to produce a measurement of participant variability per stimulus. Averages of probe location averages and averages of standard deviations for each participant per stimulus were then calculated across all participants to give the P-center location and P-center variability for each stimulus.

For the TAP trials, a MATLAB script was used to identify onsets of taps, as the time point where the value of the rectified tapping audio waveform first exceeded a predefined threshold close to the noise floor. An equal threshold was set across all recordings and verified by manually inspecting the audio waveforms and the detected taps of all recordings. 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 probe location for

each stimulus. One series by one participant had only 18 registered taps; here 14 consecutive taps from the fifth tap were used. 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.

Results

The location and variability for all stimuli in all tasks are provided in Table 3. No outliers were identified, indicating that all participants were capable of completing the tasks. For more details regarding the location and variability of the P-centers found in Experiment 1, see Danielsen et al. (2019).

Effect of method on probe location

A 3×9 ANOVA (Task \times Stimuli) was run for the CA, AP, and TAP data. There is a main effect of Task ($F(2, 38) = 12.225, p = .000; \eta_p^2 = .392$), a main effect of Stimuli ($F(8, 152) = 28.787, p = .000; \eta_p^2 = .602$), and a significant interaction between Task and Stimuli ($F(6.864, 130.416) = 4.337, p = .000; \eta_p^2 = .186$). One other concern, also evident in Fig. 1, is that significant effects due to Task and/or Stimulus might be strongly influenced by the click stimulus in the tapping task, due to the NMA produced when tapping to a metronome click. Thus, an additional 3×8 ANOVA was run without the click as a stimulus. Even without the click-as-target data, there was still a main effect of Task ($F(2, 38) = 7.399, p = .002; \eta_p^2 = .280$), a main effect of Stimuli ($F(7, 38) = 19.313, p = .000; \eta_p^2 = .504$), and a significant interaction between Task and Stimuli ($F(6.779, 128.804) = 3.078, p = .005; \eta_p^2 = .139$).

To further examine differences between tasks, *post hoc* tests were performed with Bonferroni corrections for multiple comparisons. The results show effect of task on location in the pairs involving TAP (see Table 4). There is no effect of task on variability.

To summarize:

- The CA and AP tasks did not produce significantly different probe locations.
- The TAP versus Alignment (CA or AP) tasks did produce significantly different locations.
- All methods were sensitive to stimulus differences.
- In the TAP trials the click-as-stimulus had a strong effect due to the NMA; this was not present in the alignment trials, where click-click alignment was nearly perfect.
- The CA and TAP tasks showed differential sensitivity to different categories of stimuli, most especially stimuli with short durations.

Table 3 Click/tap locations (average of average per sound per participant) relative to the physical onset of each stimulus for all four tasks, or hypothetical onset for anti-phase alignment task (N = 20)

Stimulus parameters	Fast		Fast		Fast		Slow		Slow		Slow		Fiddle		Grand mean
	Short Low	Short High	Long Low	Long High	Short Low	Short High	Long Low	Long High	Short Low	Short High	Long Low	Long High	Short Low	Short High	
Stimuli sound/ Task															
In-phase Click Alignment (CA)	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)						
Anti-phase Click Alignment (AP)	4.06 (11.92)	6.65 (12.39)	4.35 (13.11)	8.59 (17.33)	12.85 (15.40)	9.86 (16.03)	21.21 (19.21)	24.79 (17.49)	10.34 (14.63)						
Tapping (TAP)	-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)						
Click															
Click															
Click															

Standard deviations are included in parentheses. All data are presented in ms

Effect of method on probe variability

A 3 × 9 (Task × Stimuli) repeated-measures ANOVA, with the mean variability as the dependent variable, showed no main effect of Task, ($F(2, 38) = 1.472, p = .242; \eta_p^2 = .072$), but did find a main effect of Stimuli ($F(3.914, 74.373) = 9.720, p = .000; \eta_p^2 = .338$), and a significant interaction between Task and Stimuli ($F(6.897, 131.046) = 5.736, p = .000; \eta_p^2 = .232$). As again the click-click alignment task was fundamentally different from the other tasks, an additional 3 × 8 ANOVA was run (Task × Stimuli). There was again no main effect of Task ($F(2, 38) = .290, p = .750; \eta_p^2 = .015$), but again a main effect of Stimuli ($F(3.621, 68.799) = 6.986, p = .000; \eta_p^2 = .269$), but only a nearly significant interaction between Task and Stimuli ($F(6.658, 126.509) = 2.008, p = .062; \eta_p^2 = .096$); the interaction found previously thus seems driven by the click-click alignment trials.

In summary:

- The CA task was most sensitive to stimulus-driven differences in variability, ranging from near zero for the click-click alignment task to nearly 21 ms for the slow/long/low sound (stimulus #8).
- The click and the two percussive sounds (drum sounds) yield the least variability in the CA task.
- The TAP task was the least sensitive measure for stimulus-driven differences in variability, as the variability in the tapping task is driven by the timing and motor variance involved in producing a constantly repeated interval (Semjen, Schulze, & Vorberg, 2000; Repp, 2005; Vorberg & Wing, 1996).

Experiment 2: Probe comparison

The second experiment investigates the effect of the probe sound on P-center location in tasks where the probe and target sound are to be adjusted until they are perceived as simultaneous. As noted above, an inherent confound with the in-phase alignment task is that it involves adjusting two separate sounds (probe and target) until they form a fused, composite sound, one whose characteristics may be more than a simple sum of its parts. Here we use a click and a longer noise burst as probes, and we use a click and several different noise bursts as targets (stimulus details given below). The aim is to investigate the perceptual attributes of the probe used to determine a sound’s P-center, most importantly their own P-centers. A second aim was to investigate the effect of similarity/difference between the probe sound and the stimulus sound.

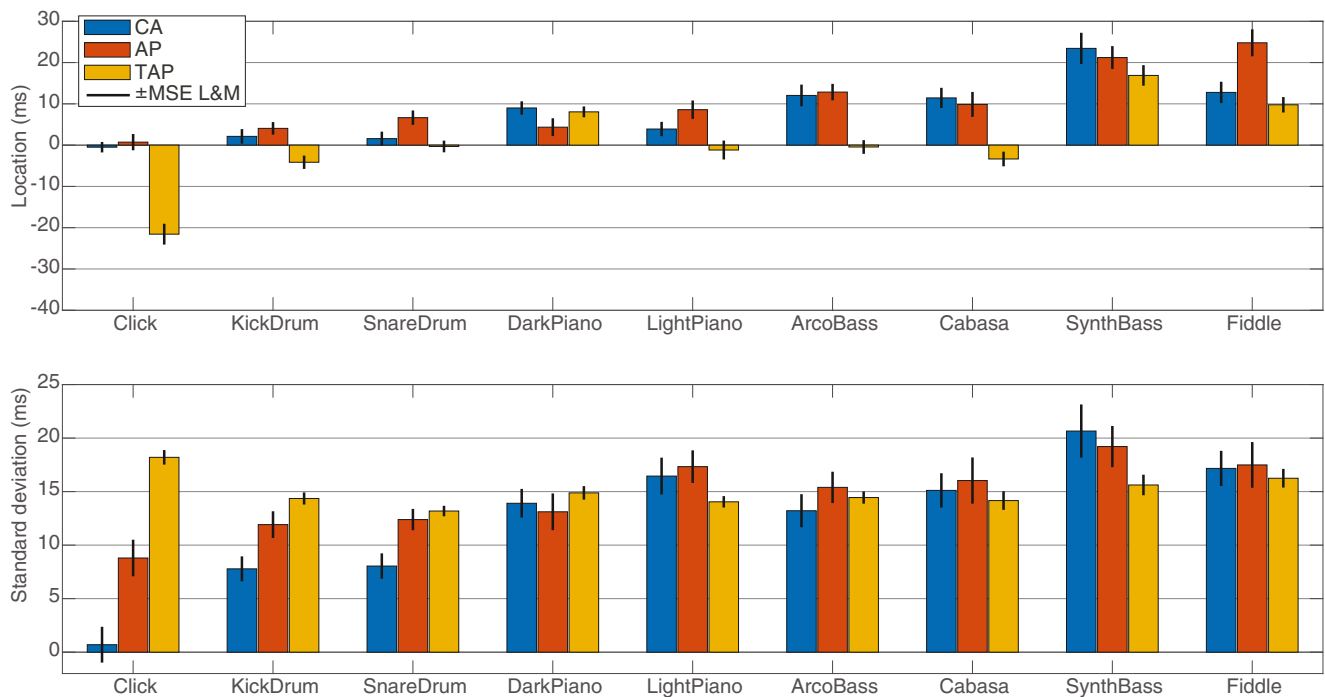


Fig. 1 Plots of probe location and standard deviation for all three tasks (N=20). Error bars calculated according to Loftus and Masson (1994)

Participants

Sixteen participants (seven female) were recruited from the Oslo area. One participant was not able to perform the task and was excluded. The median age of the remaining 15 participants was 30 years (mean = 31.8, SD = 7 years; max = 55, min = 24). Two participants reported 5–10 years of musical training; 13 participants had more than 10 years of training. As their main instrument, seven reported guitar/bass, one drums, three piano/keyboards, and four vocals. Thirteen out of the 15 participants practiced on their instrument: ten participants practiced 1–6 h/week and three more than 6 h/week. All participants reported an ability to read music.

Stimuli

The stimuli consisted of a click, a noise probe, and two variants of the noise probe with a different Attack and Center Frequency, respectively; see Table 5. The click sound was the same as used in Experiment 1. The noise probe was

generated via a narrow-band filter of random noise, with $Q = 10$ and a center frequency of 3,000 Hz. The noise probe had a 50-ms rise time with a linear slope, followed by a 50-ms decay (“Slow_High”). The two variants of the noise probe were altered in terms of center frequency (“Slow_Low,” shifted from 3,000 Hz to 100 Hz) or duration of rise-time (“Fast_High,” 3 ms rise time and 97 ms linear decay).

Apparatus and method

Following two practice trials, participants heard each target stimulus three times with each probe. The number of trials was reduced in comparison to Experiment 1 to save time and avoid possible effects of fatigue, one block of 12 trials with Click as probe (CA) and one block of 12 trials with Slow_High Noise as probe (NA). Note that blocking was essential for this study, as it makes clear which sound was the probe versus which sound was the target in trials that involved the Click and Slow_High Noise. The order of the blocks was randomized.

Table 4 Pairwise comparisons of tasks (N=20)

Pair	Probe location			Standard deviation of location		
	Mean difference (SD)	t	P	Mean difference (SD)	t	p
CA–AP	-1.93 (9.54)	-.905	1.000	-2.10 (7.60)	-1.236	.696
CA–TAP	8.01 (10.43)	3.434	.009	-2.50 (7.13)	-1.569	.399
AP–TAP	9.94 (8.54)	5.206	.000	.40 (5.92)	-.302	1.000

Bonferroni correction applied for multiple comparisons. Significant results are given in bold

Table 5 Sounds used as probes (Click or Noise_Slow_High) and stimuli in Experiment 2

Sound	Click	Noise Slow_High	Noise Slow_Low	Noise Fast_High
Rise time	0 ms	50 ms	50 ms	3 ms
Duration	2 ms	100 ms	100 ms	100 ms
Center frequency	3,000 Hz	3,000 Hz	100 Hz	3,000 Hz

Participants completed the tasks one at a time, using a Macbook Air computer (1.6 Ghz Intel core i5, OSX 10.10.5), listening via Marshall headphones (model Major II) at a comfortable intensity that could be further adjusted by the participant. All sessions were conducted in quiet rooms. Stimuli were presented using the same custom-made patch written in Max 7 (<http://www.cycling74.com>) as in Experiment 1, which also recorded participants’ responses. Participants were encouraged to proceed through the experiment at their own pace and to take breaks as needed. The experimenter waited nearby should any questions/problems arise.

Participants’ responses were averaged across the three trials to produce a mean probe location (reported in milliseconds relative to the physical onset of the stimulus) and a standard deviation for each participant per stimulus per task, using the same procedure as in Experiment 1.

Results

The P-center locations and variabilities for all stimuli in both tasks are provided in Table 6 and illustrated in Fig. 2. No outliers were identified, indicating that all participants were able to perform all tasks.

Effect of task and stimuli

Regarding P-center location, a 2 × 4 repeated measures ANOVA (Task = CA or NA, × Stimuli, four levels) found a main effect of Task ($F(1, 14) = 107.076, p = .000; \eta_p^2 = .884$), a main effect of Stimuli ($F(1.984, 27.782) = 41.601, p = .000; \eta_p^2 = .748$), and a significant interaction between Task and Stimuli, ($F(2.055, 28.770) = 2.601, p = .025; \eta_p^2 = .229$). *Post hoc* pairwise comparisons showed that the onset of the click probe (CA task) was on average located 28 ms later ($p = .000$) than

the onset of the noise probe (NA task). The probe location for click as stimulus was significantly earlier than all the three noise stimuli ($p = .000$). The fast-attack noise as stimulus was furthermore different from both the slow_high ($p = .003$) and the slow_low ($p = .000$) noise stimuli. The difference between slow_high and slow_low noise was not significant ($p = 1.000$).

We also ran an additional 2 × 3 RM ANOVA without the click as a stimulus. This showed a main effect of Task ($F(1, 14) = 83.403, p = .000; \eta_p^2 = .856$), a main effect of Stimuli ($F(2, 28) = 22.899, p = .000; \eta_p^2 = .621$), and significant interaction between Task and Stimuli ($F(2, 28) = 1.728, p = .050; \eta_p^2 = .193$), which means that click as stimulus only partly drives the interaction. Also, the fast-attack noise stimulus has an effect. When excluding the click as stimulus, a *post hoc* pairwise comparison showed that the onset of the click probe (CA task) was on average located 25 ms later ($p = .000$) than the onset of the noise probe (NA task).

Regarding variability, a 2 × 4 RM ANOVA (Task × Stimuli) shows a main effect of Task ($F(1, 14) = 12.385, p = .003; \eta_p^2 = .469$), and a significant effect of Stimuli ($F(3, 42) = 3.164, p = .034; \eta_p^2 = .184$). There was significant interaction between Task and Stimuli ($F(3, 42) = 6.848, p = .001; \eta_p^2 = .328$), such that Stimuli had greater effect on standard deviation when click was used as probe. *Post hoc* pairwise comparisons showed that the variability of the probe location was on average 6 ms higher ($p = .003$) in the NA task than in the CA task.

To investigate the effect of task further, we ran an additional 2 × 3 RM ANOVA without the click as a stimulus. This showed no effect of Task ($F(1, 14) = 1.548, p = .234; \eta_p^2 = .100$), no effect of Stimuli ($F(2, 28) = .614, p = .548; \eta_p^2 = .042$), and no significant interaction between Task and Stimuli ($F(2, 28) = 1.728, p = .196; \eta_p^2 = .110$), which means that click as stimulus drives the effect of task on variability.

To investigate further the effect of stimuli on probe variability, one-way RM ANOVAs were run for each task separately. The results show a main effect of Stimuli ($F(1.852, 25.935) = 13.838, p = .000; \eta_p^2 = .497$) on standard deviation in the CA task, but no effect in the NA task ($F(3, 42) = .449, p = .719; \eta_p^2 = .031$). *Post hoc* pairwise comparisons of stimuli in the CA task showed that the variability for click as stimulus was significantly different from all the three noise sounds ($p \leq .001$). No other pairwise comparisons were significant.

Table 6 Onset position of probe sound (average of all participant responses) relative to the physical onset of each stimulus for both tasks (N = 15)

Stimulus sound/ Probe sound	Click	Noise Slow_High	Noise Slow_Low	Noise Fast_High	Grand mean
CA Task (click probe)	0 (1)	14(14)	26 (14)	4 (9)	11 (16)
NA Task (noise probe)	-34 (18)	-4 (14)	-9 (14)	-20 (16)	-16 (21)

Standard deviations are included in parentheses. All data are presented in ms

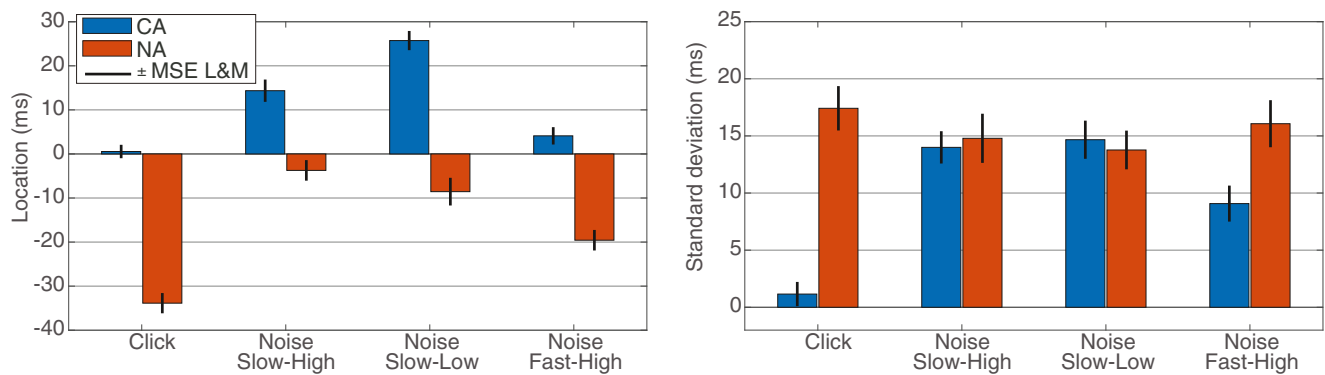


Fig. 2 Mean P-center location (**left panel**) and variability (**right panel**), Click Probe (CA) versus Noise Probe (NA). Error bars calculated according to Loftus and Masson (1994)

In summary:

- There was an effect of task and stimuli on location, but no interaction: NA locations are overall 28 ms earlier than CA locations (25 ms earlier if excluding the click).
- There was no effect of Task on variability when excluding the click.
- There is an effect of Stimuli on variability in the CA task, but no effect in NA task.
- The click differs from all three noise sounds regarding P-center locations.
- The two slow-attack noise sounds (high vs. low center frequency) produce very similar results for both location and variability in both tasks.
- Fast-attack noise differs from both slow-attack noise sounds regarding location.
- Click and fast-attack noise produce similar results for variability in the CA task.

Effect of similarity between probe and stimulus sound

To investigate further the effect of similarity between probe and stimulus, a 2×2 RM ANOVA (Task = CA or NA \times Probe-Stimulus Similarity = same (click-click or noise-noise) or different (click-noise or noise-click)) was conducted. Here we used the results for Slow_High Noise only, as this noise stimulus is identical to the noise probe sound. The analysis showed a main effect of Task ($F(1, 14) = 79.603, p = .000; \eta_p^2 = .850$), a main effect of Probe-Stimulus Similarity ($F(1, 14) = 30.382, p = .000; \eta_p^2 = .685$), and a significant interaction between Task and Stimuli ($F(1, 12) = 41.953, p = .000; \eta_p^2 = .750$).

Post hoc pairwise comparisons showed a significant difference ($p = .000$) in probe location between Click-Noise and Noise-Click: The click probe was on average located 14 ms after the onset of the noise stimulus whereas the Noise probe onset was on average located 34 ms before the click as stimulus (if mirroring the Click-Noise result, the expected value would have been -14 ms). This means that the order of

manipulation, that is, click-noise or noise-click, produces a difference in mean P-center value of 20 ms (see also Fig. 2). The difference in probe location between Click-Click and Noise-Noise (4 ms) was not significant ($p = .126$).

Regarding variability, a 2×2 ANOVA (Task \times Probe-Stimulus Similarity) showed a significant effect of Task ($F(1, 14) = 14.755, p = .002; \eta_p^2 = .513$), and Probe-Stimulus Similarity ($F(1, 14) = 12.673, p = .003; \eta_p^2 = .475$), and a significant interaction ($F(1, 14) = 6.102, p = .027; \eta_p^2 = .304$).

Post hoc tests of variability showed no significant difference ($p = .339$) between Click-Noise and Noise-Click, but a significant difference between Click-Click (average Standard Deviation = 1ms) and Noise-Noise (average Standard Deviation 15ms, $p = .000$).

In summary:

- Click-Click produces close to zero offset (i.e., perfect alignment).
- Noise-Noise produces an offset of 4 ms. The difference between click-click and noise-noise is not significant.
- The click-click task produces close to zero standard deviation while there are no significant differences in standard deviation between the three other targets, whether the click or noise is used as probe (all three in the 14- to 18-ms range).
- The order of manipulation, that is, Click-Noise or Noise-Click, produces a difference in mean probe location of 20 ms.

Experiment 3: In-phase alignment using noise as probe

This experiment is a variant/replication of the CA task in Experiment 1, with the click probe replaced by the Slow_High_Noise probe examined in Experiment 2. We recruited participants from Experiment 1, which allowed for a within-subjects comparison of CA and Noise Alignment (NA) data.

Participants

Fifteen of the original participants (eight female) from Experiment 1 were recruited for Experiment 3 to preserve a within-subjects design. Median age was 26.5 years (mean = 32.1, SD = 14.2; max = 60, min = 20). Two participants reported 1–4 years of musical training, one participant had 5–10 years of training, and the remaining 12 participants had more than 10 years of training. As their main instrument, eight reported guitar/bass, one drums, three woodwind or brass, and three vocals. All participants practiced on their instrument; nine participants practiced 1–6 h/week and six more than 6 h/week. All participants reported an ability to read music.

Stimuli

The target stimuli were the same as in Experiment 1, including the click as a target – thus the NA task used all ten stimuli listed in Table 2.

Apparatus and method

Following two practice trials, participants heard each target stimulus *three* times for a total of 30 trials.

The number of trials was reduced in comparison to Experiment 1 to save time and avoid possible effects of fatigue.

Participants completed the NA task one at a time, using a Macbook Pro computer (3.1 Ghz Intel core i7, OSX 10.13.2), listening via Beyerdynamic 770 headphones at a comfortable intensity that could be further adjusted by the participant. All sessions were conducted in quiet rooms. Stimuli were presented using the same custom-made patch written in Max 7 (<http://www.cycling74.com>) as in Experiment 1, which also recorded participants’ responses. Participants were encouraged to proceed through the experiment at their own pace and to take breaks as needed. The experimenter waited nearby should any questions/problems arise.

Participants’ responses were averaged across the three trials of the NA task and the first three trials of the CA task, respectively, to produce a probe location (reported in milliseconds relative to the physical onset of the stimulus) and a standard deviation for each participant per stimulus per task, using the same procedure as in Experiment 1.

Results

The P-center locations and variabilities for all stimuli in both tasks are provided in Table 7 and illustrated in Fig. 3. No outliers were identified, indicating that all participants were able to perform all tasks.

Table 7 Onset position of probe sound (average of all participant responses) relative to the physical onset of each stimulus for both tasks (N = 15)

Stimulus parameters	Fast			Slow			Grand Mean		
	Short	Long	Low	Short	Long	High	Short	Long	High
Stimuli sound/ Task	Click			Arco Bass			Fiddle		
CA Task (Exp. 1)	-0.47 (0.45)	11.84 (15.49)	5.07 (14.23)	12.93 (13.40)	8.18 (12.34)	20.20 (20.05)	10.73 (16.86)	8.13 (12.22)	
N=15, 3 trials									
NA Task (Exp. 2)	-24.22 (15.72)	-15.31 (13.78)	-9.53 (11.65)	-16.20 (13.63)	-17.11 (13.60)	0.02 (16.81)	-4.71 (15.31)	-11.67 (14.40)	
N=15, 3 trials									

Standard deviations are included in parentheses. All data are presented in ms

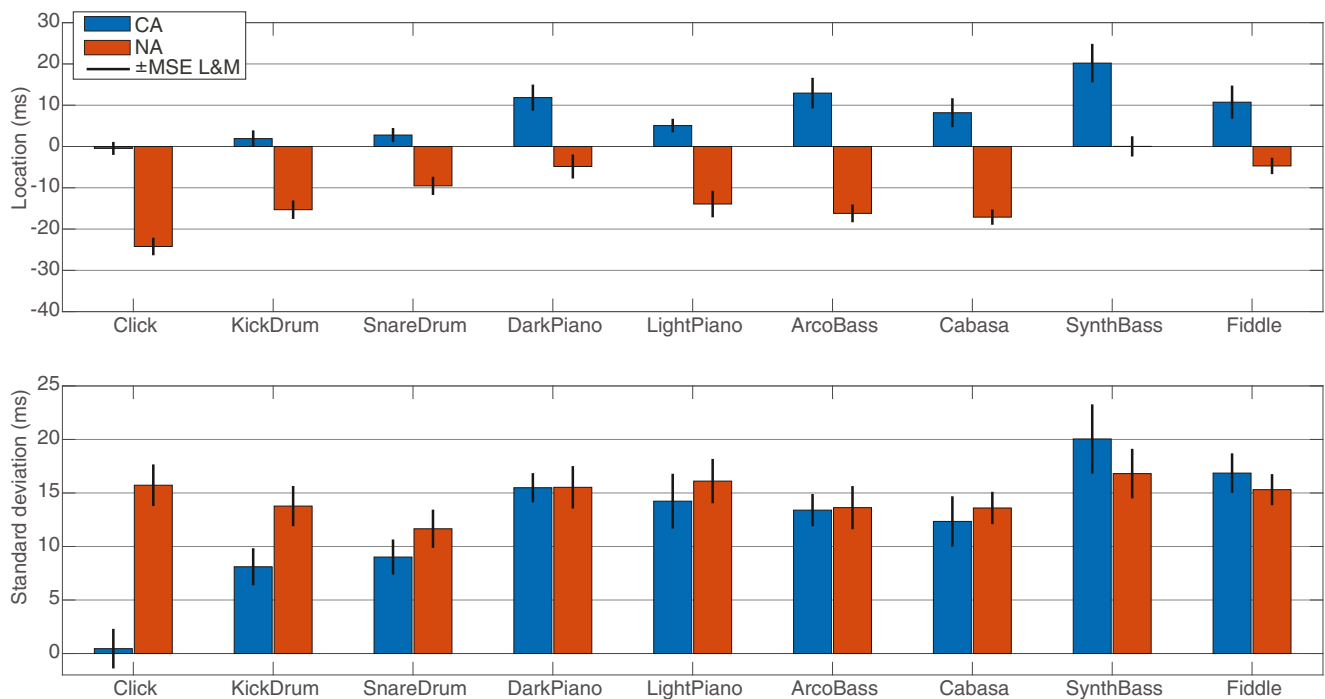


Fig. 3 Plots of probe location and standard deviation for CA (Exp. 1) and NA (Exp. 2) tasks ($N=15$), click plus eight core stimuli. Error bars calculated according to Loftus and Masson (1994)

Probe location: CA versus NA

A 2×9 repeated-measures ANOVA (Task, two levels, CA vs. NA, and Stimuli, click plus the eight original stimuli) was conducted, showing a main effect of Task, ($F(1,14) = 81.80$, $p = .000$; $\eta_p^2 = .854$), a main effect of Stimuli ($F(8,112) = 10.94$, $p = .000$; $\eta_p^2 = .439$), but no significant interaction between Task and Stimuli ($F(8,112) = 1.52$, $p = .159$; $\eta_p^2 = .098$).

As can be seen in Fig. 3, the NA task produces a pattern of results that are consistently earlier than the CA task (grand mean difference = 20 ms), though both tasks exhibit analogous effects of stimuli, as in both tasks stimuli with slower attacks and longer durations produced later P-center locations (Dark Piano, and especially Synth Bass and Fiddle). The difference between CA and NA is greatest for the two sounds that are most similar to the noise probe, that is, the Arco Bass and the Cabasa. Both of these musical sounds and the noise probe have slow attacks and short duration.

Variability

In terms of variability a 2×9 (Task \times Stimuli) repeated-measures ANOVA showed no effect of Task, ($F(1,14) = 2.72$, $p = .121$; $\eta_p^2 = .163$), a main effect of Stimuli ($F(8,112) = 4.65$, $p = .000$; $\eta_p^2 = .249$), and significant interaction between Task and Stimuli ($F(8,112) = 3.42$, $p = .001$; $\eta_p^2 = .196$), such that stimuli had a greater effect in the CA trials than in the NA trials; see Fig. 3. An additional 2×8 (Task \times Stimuli) RM ANOVA was run without the click as a stimulus. Again, there was no

effect of Task, ($F(1,14) = .13$, $p = .726$; $\eta_p^2 = .009$), a main effect of Stimuli ($F(7,98) = 3.33$, $p = .003$; $\eta_p^2 = .192$), but no significant interaction between Task and Stimuli ($F(7,98) = .90$, $p = .510$; $\eta_p^2 = .060$). Similar to in Experiment 1, the interaction found in the 2×9 RM ANOVA seems driven by the click-as-target in the CA trials.

Discussion

In three experiments, we explored various methods and materials that may be used to study the P-centers of musical sounds. In the first experiment, we used the method of adjustment, with a probe sound (a 2-ms 3,000-Hz click) either in-phase or anti-phase in relation to the target sound, as well as a synchronized tapping task. All three methods were used with a set of target stimuli that varied systematically in terms of attack (slow vs. fast), duration (short vs. long) and center frequency (high vs. low). In the second experiment, the characteristics of various probes (2-ms click vs. 100-ms 3,000-hz noise burst) were examined, with both the click and a variety of filtered noise sounds used as target sounds. In the third experiment, the 100-ms noise burst was used as the probe in an in-phase alignment task, using the same musical target stimuli as in Experiment 1.

The various methods and probes give different P-center locations and differing amounts of variability about the P-centers for each sound. These differences may be summarized as follows:

- In-Phase versus Anti-Phase alignment tasks (Experiment 1) produce very similar results in terms of P-center location and variability, save for the click-click alignment task, where in-phase variability is much lower (near zero) in comparison with anti-phase variability.
- Tapping versus Click Alignment results (Experiment 1) differ in terms of P-center location with some sounds but not others. For the Click-as-target, Light Piano, Arco Bass, and Cabasa sounds, mean tapping locations were consistently earlier than click alignment locations. Save for the fact that none of these sounds belong to the slow-long category, there is no consistent pattern of acoustic factors with these stimuli, as some have fast attacks (Click, Light Piano) while others have slow attacks (Arco Bass, Cabasa), and some are short (Click, Arco Bass, Cabasa) while others are long (Light Piano), and so forth. For the other sounds Tapping and Click Alignment produced similar results.
- Tapping versus Click Alignment results (Experiment 1) differ in terms of variability, and here a more consistent pattern emerges. In the tapping task the variability is more or less constant, which is to say, is insensitive to the target stimulus, while in the alignment task variability varies systematically with stimulus type: short sounds with fast onsets (Kick and Snare Drums) have the lowest variability, and long sounds with slow onsets (Synth Bass and Fiddle) have the highest variability.
- In comparing click versus noise probes (Experiment 2), click and noise probes produced parallel results, but with the noise probes marking P-centers an average of 28 ms earlier than click probes (25 ms earlier if excluding click as stimulus). Variability of click and noise probes were found to be the same, save for the click-click alignment task, where (as noted above) variability is near zero.
- When using Noise as a probe of musical stimuli (Experiment 3), results are analogous to Experiment 1, but with the noise probes marking P-centers an average of 20 ms earlier than click probes; likewise, the variability of P-center location does not significantly differ between the two probe methods used in Experiments 1 and 3. This can be interpreted as alignment between the P-center of the probe – which is essentially at 0 ms (sound onset) for the click, whereas it is in the 20- to 30-ms range for the noise probe – and the P-center of the target sound.

Regarding the effect of method – alignment via the method of adjustment versus a tapping task – one should ask whether or not these two methods are measuring the same percept. Alignment tasks, whether in-phase or anti-phase, and whether they use clicks or noise as probes, are overt judgment tasks. They are not time-pressured, in that participants may take as much time as they like and make as many adjustments as they wish until they obtain their

desired alignment. The goal of such tasks is either to produce perfect isochrony in the anti-phase task, or perfect alignment in the in-phase task. The latter task thus involves creating a blended sound in which any cues for the location of separate sounds are merged into the cue for a single sound. This task is most apparent in the click-click alignment task, for not only are the temporal thresholds for cue separation at their lowest, given the brevity of both probe and target (Hirsh, 1959), but also as there are clearly audible timbral/pitch differences amongst the different alignments within a 1-to 2-ms span around their absolute onset alignment. By contrast, tapping to sounds is a motor-synchronization task, which is time-pressured, has an implicit judgment task, also involves the production of a blended sound, the production of a repeated, stable inter-tap interval, and engages error correction mechanisms for period and phase correction. As tapping tasks are time-pressured in a way that alignment tasks are not, they are inherently more sensitive to event rate. The IOI interval of target sounds can thus affect alignment in so far as it affects sensory, perceptual, and motor production mechanisms at different absolute time scales (Bååth et al., 2016; London, 2012; Tierney & Kraus, 2016). The implicit judgment regarding the P-center of the target sound, which functions as a “pacing stimulus” for one’s taps, emerges through one’s physical/bodily interaction with the stimulus. As such, it is an example of embodied or extended cognition (Clark, 2008; Wilson, 2002). As in the in-phase alignment task, the goal of the tapping task is to produce a blended sound that signifies the desired synchronization. The tap itself makes a noise (i.e., the sound of the clave sticks), which fuses with the stimulus sounds to create a blended sound. This means that the judgment one makes regarding synchronization is as much about the resulting qualities of the sound as it is to the alignment between action and target.

Even in the absence of a pacing stimulus, tapping at a constant rate requires perception and maintenance of a stable inter-tap interval, which is then further complicated when error-correction mechanisms are engaged to maintain synchrony with an isochronous target sound (Repp, 2005; Repp & Su, 2013). Thus, while perception of the P-center of the target sound is involved in both alignment and tapping tasks, the different natures of the tasks interact with that perception in different ways, giving different measures of the location and variability of the P-center of the target sound. Nowhere is this more apparent than in tapping with the click as the target sound, which gives rise to the well-known negative mean asynchrony (NMA; see Repp, 2005, as well as Danielsen et al., 2019). Given all of these complications in a tapping task, alignment may be regarded as giving a “purer” sense of the P-center location. However, tapping has the twin advantages of (a) not involving ratiocination regarding one’s

judgment, and (b) for many participants is an easier and more natural task – almost everyone has tapped their toe or danced to music, while relatively few have performed what is essentially a digital music production task of loop or “track” alignment.

When using the method of alignment/adjustment, the choice of probe also affects the determination of P-centers. We found it does so in three ways. First, as different probes themselves have different P-centers, this difference must be taken into account when comparing results using such probes. Unsurprisingly, we found that, all other things being equal, the location of the noise probe was earlier than the location of the click probe (20 ms on average in Experiment 3), indicating that the P-center of the noise sound is much later than the click, but not as late as the energy peak of the noise sound (at 50 ms). The “all other things being equal” caveat was added above due to the second way the choice of probe may affect P-center determinations, and that is the degree of sonic similarity between probe and target sounds. In Experiment 1, all stimuli were equally mismatched to the sonic characteristics of the click used as the probe, save for the click-click alignment trials. In Experiment 3, the noise-probe differed from the target sounds to varying degrees. In terms of P-center location, we found results were comparable, save for those stimuli that were most *similar* to the sound of the probe, i.e., click-click alignment in Experiment 1 and noise-probe alignment with the Arco Bass and Cabasa sounds in Experiment 3. Likewise, in terms of P-center variability, results were similar, save for the stimuli that were most *different* from the sound of noise probe, namely the click, kick drum, and snare drum sounds (i.e., very fast onset).

Thirdly, whether one manipulates the probe versus the target may also affect the location of the P-center, as was found in Experiment 2: when the click is the probe and the noise is the target, mean alignment occurs 34 ms after the onset of the noise, whereas when the noise is the probe and the click is the target, mean alignment occurs 14 ms after the onset of the noise. This is in some ways our most puzzling result. In any given trial, “moving the target later” versus “moving the probe earlier” are epistemically equivalent, as these manipulations occur in the context of continually repeated sounds. Recall also that the offset of the probe was randomized in terms of temporal interval and position (before/after the target sound). However, as the use of different probes was blocked in our experiment – in each block the probe sound remained constant from trial to trial, while the target sound changed – this context framed participants’ sense that the probe was manipulated and the target was “stationary.” The fact that such a perceived “order of manipulation” might have produced the difference between click-noise and noise-click results may be related both to the details of the alignment task and to the noise probe sound having a larger “window” of possible synchronization

points. Regarding the former, as the goal of the task is to produce a blended sound that signifies that the probe and target occur simultaneously, when the click is the probe, the alignment has to be sufficient to indicate that the click occurs after the *acoustic* onset of the noise – but the click is not necessarily masked. In these cases we find the click placed at about 14 ms after the noise onset. When noise is the probe, we have the same problem, but to make sure alignment has occurred – and given the inherent fuzziness of both the perceptual onset and P-center of the noise probe itself – achieving the goal involves a more substantive masking of the target click, for this makes certain that the click has been aligned after the acoustic/perceptual onset of the noise probe (N.B.: the RMS volume of the noise probe is roughly double at 34 ms after onset vs. 14 ms). Regarding the latter, recall that the standard deviation of click-probe locations for the noise probe sound was 14 ms, which covers both the click-noise and the noise-click results. When used as a probe the noise sound thus establishes a context of looser synchronization, that is, a larger “beat bin” (Danielsen, 2010; Danielsen et al., 2019). An obvious line of future research would be to change the design by removing the blocked presentation of different probe sounds, which led to this framing of the order of manipulation.

It is of course axiomatic in psychological experiments, and perceptual experiments in particular, that one’s results are strongly dependent on the particular details of the stimuli, method, and task used. Indeed, the classical methods of psychophysics are a response to this basic problem (Boring, 1942). To that extent, what we have reported here is not surprising. The three experiments reported on here illustrate the usefulness of employing a varied set of tasks/responses for obtaining basic measurements of perceptual processes, as well as the importance of benchmarking sonic and perceptual aspects of materials used as both probes and target stimuli. More broadly, our study points to the difficulty involved in achieving any sort of ecological validity in even the simplest of perceptual tasks and judgments. In real-world musical contexts, musicians and their audiences integrate complex constellations of sonic onsets and their alignments into perceptions of temporal location and motion, hearing “fat” beats and “pushing” or “pulling” rhythms. The manifold ways listeners can interact with the very simple stimuli used in the experiments described here gives us a glimpse of the richness and complexity of musical experience.

Acknowledgements We want to thank Elise Måsvær for assistance with data collection to Experiment 2. 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.

Open practices statement None of the data or materials for the experiments reported here are available, and none of the experiments was preregistered.

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>
- Bååth, R., Tjøstheim, T. A., & Lingonblad, M. (2016). The role of executive control in rhythmic timing at different tempi. *Psychonomic Bulletin & Review* 23(6): 1954–1960.
- Bechtold, T. A., & Senn, O. (2018). Articulation and dynamics influence the perceptual attack time of saxophone sounds. *Frontiers in Psychology*, 9:1692. <https://doi.org/10.3389/fpsyg.2018.01692>
- Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology*. New York: D. Appleton-Century Co.
- Clark A. (2008). *Supersizing the mind*. New York: Oxford 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., Nymoen, K., Anderson, E., Câmara, G. S., Langerød, M. T., Thompson, M. R., & 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*, in press.
- Dixon, S., Goebel, W., & Cambouropoulos, E. (2006). Perceptual smoothness of tempo in expressively performed music. *Music Perception* 23(3): 195–214.
- Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology* 35: 1–36.
- 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>
- Harsin, C. A. (1997). Perceptual-centre modeling is affected by including acoustic rate-of-change modulations. *Perception & Psychophysics*, 59(2): 243–251. <https://doi.org/10.3758/BF03211892>
- Hirsh, I. J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America* 31(6): 759–767.
- Hove, M. J., Iversen, J. R., Zhang, A., & Repp, B. H. (2013). Synchronization with competing visual and auditory rhythms: Bouncing ball meets metronome. *Psychological Research* 77(4): 388–398.
- Howell, P. (1988). Prediction of P-center location from the distribution of energy in the amplitude envelope: I. *Perception & Psychophysics* 43(1): 90–93. <https://doi.org/10.3758/BF03208978>
- Iversen, J. R., Patel, A. D., Nicodemus, B., & Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals. *Cognition* 134: 232–244.
- Kochanski, G., & Orphanidou, C. (2008). What marks the beat of speech? *The Journal of the Acoustical Society of America* 123: 2780–2791. <https://doi.org/10.1121/1.2890742>
- Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476–490.
- London, J. (2012). *Hearing in time: Psychological aspects of musical meter*, 2nd Edition. New York, NY, Oxford University Press.
- Marcus, S. M. (1981). Acoustic determinants of Perceptual-center (P-Center) location. *Perception & Psychophysics* 30(3): 247–256. <https://doi.org/10.3758/BF03214280>
- Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, 83(5): 405–408. <https://doi.org/10.1037/0033-295X.83.5.405>
- Nymoen, 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.
- Pompino-Marschall, B. (1989). On the psychoacoustic nature of the P-center phenomenon. *Journal of Phonetics* 17(3): 175–192.
- Repp, B. H. (1995). Detectability of duration and intensity increments in melody tones: A partial connection between music perception and performance. *Perception & Psychophysics* 57(8): 1217–1232.
- Repp, B. H. (1998). The Detectability of Local Deviations from a Typical Expressive Timing Pattern. *Music Perception* 15(3): 265–289.
- Repp, B. H. (2003). Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits and costs of interval subdivision. *Journal of Motor Behavior* 35(4): 355–370.
- 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>
- Scott, S. K. (1993). *P-Centres in speech: an acoustic analysis*, PhD diss., University College London, London, UK.
- Semjen, A., Schulze, H.-H., & Vorberg, D. (2000). Timing precision in continuation and synchronization tapping. *Psychological Research* 63: 137–147.
- Thompson, M.R., Diapoulis, G., Johnson, S., Kwan, P.Y., & Himberg, T. (2015). Effect of tempo and vision on interpersonal coordination of timing in dyadic performance. In *Proceedings of the 11th International Symposium on Computer Music Multidisciplinary Research*, M. Aramaki, R. Kronland-Martinet, and S. Ystad (Eds.). University of Plymouth, Plymouth.
- Tierney, A., & Kraus, N. (2016). Getting back on the beat: Links between auditory–motor integration and precise auditory processing at fast time scales. *European Journal of Neuroscience* 43(6): 782–791.
- Villing, R. (2010). *Hearing the Moment: Measures and Models of the Perceptual Centre*. PhD diss., National University of Ireland, Maynooth, Ireland.
- Vorberg, D., & Wing, A. M. (1996). Modeling variability and dependence in timing. In *Handbook of perception and action, volume 2: Motor skills*, H. Heuer, & S. W. Keele, (eds.) New York, Academic Press: 181–262.
- Vos, J. & Rasch, R. A. (1981). The perceptual onset of musical tones. *Perception & Psychophysics* 29(4): 323–335. <https://doi.org/10.3758/BF03207341>
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review* 9(4): 625–636.
- Wright, M. J. (2008). *The shape of an instant: Measuring and modeling perceptual attack time with probability density functions*. PhD diss., Stanford University, Stanford, CA.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.