Aalborg Universitet

The Effect of Auditory Pulse Clarity on Sensorimotor Synchronization

Kantan, Prithvi Ravi; Stefan Alecu, Rares; Dahl, Sofia

Published in:
Perception, Representations, Image, Sound, Music - 14th International Symposium, CMMR 2019, Revised Selected Papers

DOI (link to publication from Publisher):
10.1007/978-3-030-70210-6_25

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Kantan, P. R., Stefan Alecu, R., & Dahl, S. (2021). The Effect of Auditory Pulse Clarity on Sensorimotor Synchronization. In R. Kronland-Martinet, S. Ystad, & M. Aramaki (Eds.), *Perception, Representations, Image, Sound, Music - 14th International Symposium, CMMR 2019, Revised Selected Papers* (Vol. 12631, pp. 379-395). Springer. https://doi.org/10.1007/978-3-030-70210-6_25

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: March 27, 2025

The Effect of Auditory Pulse Clarity on Sensorimotor Synchronization

Prithvi Kantan[1][0000−0001−5618−4265], Rareș Ștefan Alecu[1][0000−0003−4169−5200], and Sofia Dahl[2][0000−0002−0687−7810]

[1] Sound and Music Computing, Aalborg University, Copenhagen, Denmark
{pkanta18,ralecu18}@student.aau.dk
[2] Department of Architecture, Design and Media Technology, Aalborg University, Copenhagen, Denmark sof@create.aau.dk
https://vbn.aau.dk/da/persons/118552

Abstract. This study investigates the relationship between auditory pulse clarity and sensorimotor synchronization performance, along with the influence of musical training. 29 participants walked in place to looped drum samples with varying degrees of pulse clarity, which were generated by adding artificial reverberation and measured through fluctuation spectrum peakiness. Experimental results showed that reducing auditory pulse clarity affected phase matching through significantly higher means and standard deviations in asynchrony across musical sophistication groups. Referent period matching ability was also degraded, and non-musicians were impacted more than musicians. Subjective ratings of required active concentration also increased with decreasing pulse clarity. These findings point to the importance of clear and distinct pulses to timing performance in synchronization tasks such as music and dance.

Keywords: pulse clarity, sensorimotor synchronization, rhythm, movement, perception, musical training, timing

1 Introduction

Sensorimotor synchronization (SMS) [1] is a form of referential behavior in which an action is coordinated with a predictable external event, the referent, where both are usually periodic. Examples of SMS are dance (where movements are synchronized with both music and the movements of other dancers), music performance and marching. The vast majority of studies primarily use finger tapping [1,2]. While, for instance, foot tapping, could be assumed to be just as common as tapping in every-day SMS, the prevalence of finger tapping studies most likely has its explanation in the fact that finger tapping is the most practical to measure. There are, however, an increasing number of studies on SMS involving lower-limb activities such as stepping in place [3], locomotion [4], and dance [5,6].

Another noteworthy observation concerns the choice of stimuli in SMS studies, where music is comparatively rare [1]. Instead, studies tend to use auditory stimuli consisting of brief tones or clicks. Such stimuli generally have sharp temporal profiles, which along with low noise in testing environments are likely to exhibit prominent and effortlessly perceptible periodicities or *pulses*. Real-life SMS referents, however, such as music performed in a reverberant hall, often have less pulse salience due to time-smearing and masking effects. Dynamics processing and speaker distortions can further undermine the strength of rhythmic pulsations. In extreme situations, the pulse may no longer even be readily perceptible, subject to individual perceptual ability and musical training or experience.

In this study, we used computational methods for *pulse clarity (PC)* estimation [7], to design ecologically valid stimuli for the investigation of systematic relations between pulse clarity and SMS task performance. In-phase walking in place was chosen for greater ecological validity with regard to general bodily movement.

2 Related Work

The extensive study of SMS is comprehensively reviewed in [1] and [2]. With respect to tapping to simple isochronous stimuli, Madison et al. [8] showed that the response mode, feedback received and training have an effect on SMS, with feedback and training resulting in lower local variability. In general, taps tend to precede sequence tones by a few tens of milliseconds, rather than being symmetrically distributed around tone onsets. Possible causes of this *Negative Mean Asynchrony (NMA)* have generated a considerable amount of research (see [2] for a summary). Research has shown that the magnitude of NMA is influenced by a number of factors such as effector used, feedback, event density (subdivision) and expertise. Trained musicians exhibit both smaller NMA [9,10] and a smaller standard deviation in asynchrony than non-musicians [11].

Chen et al. [12] studied the behavioral and neural effect of increased metrical complexity on SMS in musicians and non-musicians. Measuring SMS performance in terms of intertap interval and asynchrony, they found that musicians were significantly more accurate than non-musicians in terms of period and phase synchronization. Although these abilities were degraded in both groups with increasing rhythmic complexity, there was a significant interaction between asynchrony and music training across rhythm types. In other words, the performance of the musician group was not as severely degraded as the nonmusician group due to greater stimulus complexity (c.f. Fig. 3 in [12].) The authors attributed this superior performance to a more efficient recruitment of motor neural regions, as well as superior abilities in timing, error correction and general auditory discriminatory processes.

In another study, Chen et al. [3] investigated lower-limb SMS to a variable metronome for different types of response (uni- or bilateral heel tapping when sitting, bilateral heel tapping when standing, or the toe or heel strike when stepping on the spot). For the steady-state baseline part of the stimuli, they reported more negative asynchronies for stepping on the spot than bilateral heel tapping in standing. The details of the automatic detection of heel and toe strike during stepping were not reported. While the constraints of keeping balance in bilateral stepping and standing did show an effect on error correction, results also showed reduced variability for stepping compared to bilateral heel tapping in standing. The authors suggested that the reduced variability could be an effect of the increased sensory feedback to participants from the load of the body weight, which would explain the difference between stepping compared to heel tapping. Similarly, Palmer et al. [13] reported that tactile feedback appears to be important in reducing timing errors for clarinetists.

Lower-limb SMS such as stepping in place is particularly interesting because it relates to every-day activities such as walking, dancing and also to the use of SMS in rehabilitation [14,15]. While these activities can be synchronized with metronomes it is also common with live or recorded music, often with a clear pulse. With varying stimuli in less controllable surroundings, it seems reasonable to assume that most real-life SMS occurs in situations where the referent characteristics is distinct from the majority of laboratory settings.

Perceptual Centers. Ecological stimuli may exhibit varied temporal envelopes with distinct sub-band spectral evolution. For instance, the sound of a regular snare drum hit has a very well-defined onset with a sharp attack and fast decay, whereas a wind or string instrument can produce onset envelopes of varying length and shape. The *Perceptual Center* (P-center) of a sound is understood as the specific moment of perceived occurrence [16]. Synchronization then involves aligning P-centers, which studies [16,17] have shown to depend upon envelope characteristics. The P-center seems to be located between the perceptual onset and the energy peak of a sound. For impulsive sounds this is close to a single location, while tones with gradual onsets tend to show a range of equally 'correct' sounding locations [17]. Danielson et al. [16] found that in general, sounds with slower attacks and longer durations had later P-centers with greater variability in their exact location. In a synchronization task with musical and quasi-musical stimuli, NMA with respect to the physical onsets was small to non-existent, but significant with respect to the P-center. This aligned with the hypothesis of Vos et al. [18] that participants use P-centers, rather than the physical onset of the tone as the target for SMS tasks.

Pulse Clarity. The ease of perception of the underlying pulsation in temporally organized stimuli such as music may determine the ease with which the body can be synchronized to them. Lartillot et al. [7] define *Pulse Clarity (PC)* as a high-level musical measure conveying how easily listeners can perceive the underlying rhythmic or metric pulsation. There is evidence that pulse clarity affects corporeal movement characteristics during SMS tasks [5,6]. Burger et al. [6] found that music with greater pulse clarity elicited greater temporal and spatial regularity in dance. Van Dyck et al. [19] found that when the bass drum was made louder in the music mix, dancers increased their motor activity and entrained better to the beat. In addition to exploring internal SMS mechanisms

[20,21], neuroscience studies show that the extent to which cortical or subcortical motor activations are coupled with the auditory cortex depends on beat salience and music training [22]. As stable pulse perception underlies all SMS tasks, lowering the perceived pulse salience of a referent is likely to have a detrimental impact on SMS performance.

We now examine some ways in which Pulse Clarity has been modelled in the past. Lartillot et al. [7] quantified the temporal evolution of music in terms of its *Onset Detection Curve (ODC)*, where peaks indicate pulses. They then characterized pulse clarity by describing the ODC in terms of local configurations or the presence of periodicities. In the former, PC characterizations do not relate to periodicity, and focus on articulation and attack characterization. Specifically, the attack slope [23] and local maxima of the time derivative of the amplitude envelope [24] can be considered as possible factors for PC prediction. Periodic characterizations are made in terms of the autocorrelation function of the ODC, which can be either full-band or multi-band. The principle is that peaks in the autocorrelation function indicate the most probable periodicities, and resonance functions can be used to model these in terms of perceptual salience [24]. Lartillot et al. [7] described pulse clarity in terms of the global maximum, global minimum, kurtosis and entropy of the ODC autocorrelation curve. In perceptual tests, they found that the best predictor of perceptual pulse salience was the global minimum.

Pampalk et al. [25] extracted a rhythm pattern representation of music pieces based on the *amplitude modulation(AM)* of the loudness sensation per critical band [26], weighting its coefficients based on the psychoacoustic model of fluctuation strength [27]. The effect on hearing depends on the modulation frequency, and is most intense around 4 Hz [25]. Within each frequency band, gradient and Gaussian filters were used to emphasize distinctive beats, characterized through high fluctuation strength at these modulation frequencies relative to neighboring frequencies. They found that pieces of music dominated by strong beats had high fluctuation strength values, which were also correlated to bass. PC estimation metrics based on both ODC methods [7] and fluctuation strength [25] can be easily evaluated for audio recordings using the MIRtoolbox [28]. In the stimulus design process of the present study, both types of methods are assessed in terms of their correspondence to subjective beat perception.

3 Stimulus Design

The creation of suitable auditory referents necessitated 1) The design of rhythmic stimuli spanning the entire range of perceptual pulse salience by systematic manipulation of a single base stimulus. 2) The objective assessment of these stimuli by PC measures.

We made drastic changes to perceived pulse salience by simply altering the decay time of a digital reverberation (reverb) effect with a flat frequency decay, applied to a looped snare drum sample (EZDrummer 2 VST instrument). Increasing the decay time reduced perceptual pulse salience, due to the increasing masking effect of previous decay tails on subsequent onsets, along with the reduction in overall dynamic range. The reverb plugin used was the 64 bit version of WAVES TrueVerb, with early reflections and high frequency roll-off disabled. Finally, the audible pulsation of the primary resonance of the drum sample was suppressed using a narrow notch filter.

To determine the range of reverb decay times that fit the required perceptual range, we blindly adjusted decay time to yield pulses that were subjectively *'Very Clear', 'Moderately Clear', 'Moderately Unclear' and 'Very Unclear (but perceptible)'*. We then analyzed these four preliminary sample points of the perceptual range using the MIRToolbox [28] and 1) Entropy of *Onset Detection Function (ODF)* Autocorrelation [7]; 2) Max ODF Autocorrelation; 3) Min ODF Autocorrelation; and 4) Peakiness of Fluctuation Spectrum [5,25]. The fourth approach estimates PC by the relative Shannon entropy of the fluctuation spectrum [25], in terms of peak magnitude, regularity of spacing and noise between peaks. The calculated values can be seen in Table 1.

Table 1. Comparison between different methods for PC computation [7] [28]

Sr. No.	Salience	Fluct. Spectrum Peak	EntropyAutocor	MinAutocor	MaxAutocor
1	Very Clear	293961.77	0.5066	0.398	0.9899
2	Moderately Clear	146677.66	0.6314	0.335	0.9749
3	Moderately Unclear	24224.66	0.721	0.3891	0.7672
4	Very Unclear (but Perceptible)	17082.21	0.7439	0.3342	0.5134

Similar to what was reported in [5], the fluctuation spectrum of the perceptually clearer stimuli exhibited peaks with markedly higher magnitude at the beat frequency, and less inter-peak noise than the unclear stimuli (see Figure 1). Perceptually, increasing the reverb decay time was always found to reduce pulse salience, and this behavior was reproduced in preliminary tests with the fluctuation spectrum peakiness but not always with ODF-based measures. Hence, we used the fluctuation spectrum peakiness to model the perceptual range. Note that the 4 Hz peaks in the fluctuation spectrum plots are higher than the 2 Hz peaks (true frequency), and this could be attributed to the weighting of the AM coefficients in the algorithm [25].

From the total range of reverb times, we empirically found that *nine* total stimuli would sample the perceptual range with enough inter-stimulus difference to minimize redundancy. To determine the necessary fluctuation peak magnitudes, we fitted a 3rd order polynomial curve to the four previously determined values and designed stimuli S1 - S9 to match nine equi-spaced curve values in the same range, in decreasing order of PC. Stimulus tempo was centered around 120 BPM, close to the preferred human movement tempo [29], but varied by ± 1 BPM between successive stimuli to prevent short-term training effects [8]. The onset peak amplitude was kept constant across stimuli. The loudness of the base stimulus (dry snare drum sample) was kept constant between stimuli, so the perceived overall loudness increased with reverb decay time. This is an ecologically

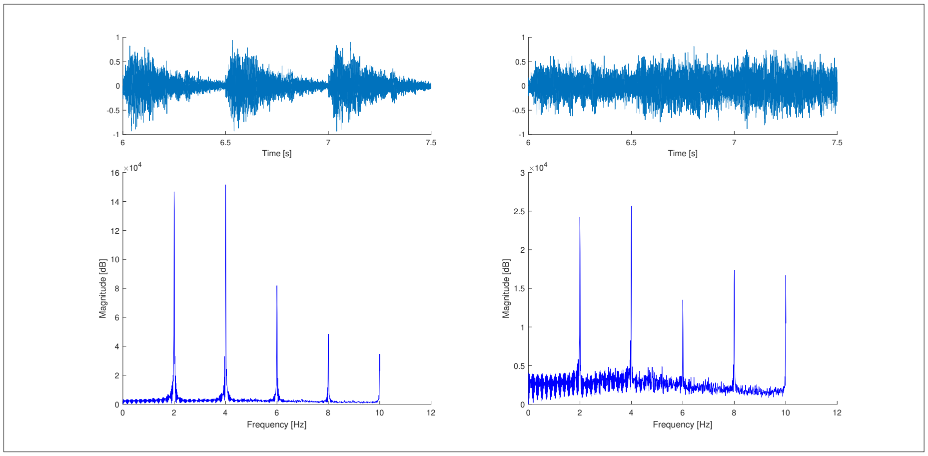

Fig. 1. Pulse clarity measurement using fluctuation spectra. For the signal to the right, the spectrum peaks have lower overall magnitude, and greater noise in between peaks, implying lower pulse clarity. This is also evident from the time domain waveform.

valid effect, as sounds are generally perceived to be louder in acoustically 'live' environments than in 'dead' (reverberation-free) ones.

4 Experiment

In order to test whether our stimuli with degrading pulse clarity would affect SMS performance we designed a within participant experiment. We hypothesized the following effects of pulse clarity manipulations of the referent stimulus:

- Decreased PC leads to smaller NMA across participants and greater asynchrony variance (greatest for nonmusicians).
- Decreased PC leads to greater tempo deviation from the referent tempo, and greater variance in the reproduced interval duration. Nonmusicians are affected to a greater extent than musicians.
- Decreased PC leads to higher ratings of required concentration across participants, independent of music training.

4.1 Participants

A convenience sample of 29 participants (6 women, 21-35 years, $MeanAge = 26$), mainly students at Aalborg University, volunteered in return for a film voucher. Participants were briefed on the length of the experiment (9 stimuli x 50 seconds) and that they could withdraw at any time without losing their remuneration.

4.2 Experimental Setup and Procedure

Participants were tested individually in a quiet, medium-sized room on campus. The stimuli was played via a set of Focusrite Studio Headphones, while recordings of the activity were captured with a Focusrite CM25 large-diaphragm cardioid condenser microphone. The audio was digitized to a 44.1 KHz/24-bit WAV format using a Focusrite Scarlett Solo Studio audio interface.

After obtaining the participants' informed consent, we asked them to complete an online musical background questionnaire to determine their Ollen Musical Sophistication Index (OMSI) [30]. The OMSI reflects the probability that a music expert would categorize a respondent as *"more musically sophisticated"*, with regard to musical knowledge, skill, and composition ability.

Subsequently, participants were instructed to assume a standing position in front of the microphone such that their feed were approximately 30 cm from the diaphragm of the microphone. They were then asked to walk in place, stepping in exact synchronization with each of the stimuli, which were presented in random order with brief pauses in between. After each trial, participants were asked to rate on a scale of 1-10 the amount of active concentration required to maintain synchronization with the stimulus. This procedure was repeated for all 9 stimuli.

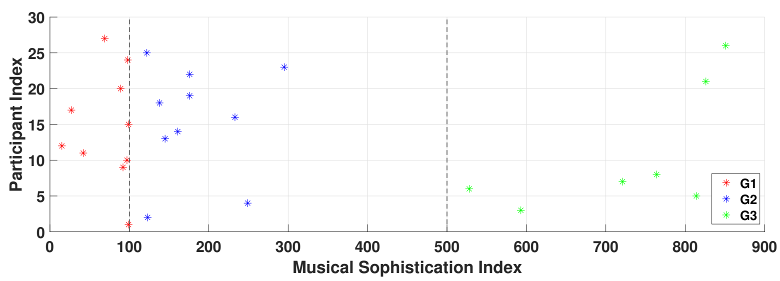

Fig. 2. Distribution of participants based on Musical Sophistication Index and the design of clusters for statistical analysis. The dotted lines indicate the segregation thresholds between groups G1, G2 and G3. Participants with OMSI scores greater than 500 should be classified as "more musically sophisticated" [30].

4.3 Data Analysis

Recordings of two participants were discarded in entirety due to poor signal quality, yielding 27×9 trials = 243 recordings for analysis.

The OMSI scores of the participants covered a considerable range (see Figure 2). In order to study the effect of musical training on task performance, we therefore further grouped participants on the basis of their OMSIs into 3 *MSoph* groups G1 (OMSI $<$ 100, **10** participants), G2 (OMSI 100 - 500, **10** participants), and G3 (OMSI 500 - 1000, **7** participants).

Onset Detection Algorithm From each recording, the extraction of footstep timestamps was carried out in MATLAB using an onset detection algorithm, based on [31]. The first 10 seconds of each recording (containing the initial rhythm acquisition phase) were discarded. The remaining audio was processed using a sliding window approach, with a frame size of 512 samples and a hop ratio of 0.5. From the obtained audio frames, the algorithm computes a signal reflecting the temporal evolution of spectral magnitude difference between short-term Fourier spectra of successive frames across bands. This is used as the *onset detection function* [31]. Spectral difference is a useful indicator in this case, since footstep onsets are accompanied by transient increases in spectral magnitude, resulting in detection function peaks. Finally, an adaptive threshold in the form of a moving median filter is used to pick peaks and record their associated timestamps.

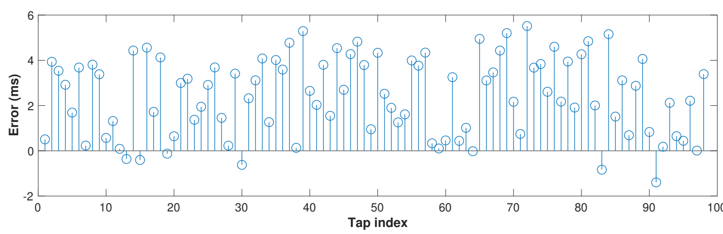

Fig. 3. Stem plot of error deviation between the ground truth timestamps and detected timestamps, computed for each foot.

Detection accuracy of the algorithm was gauged through a test conducted on one trial recording containing 100 valid steps. One author manually annotated the footstep timestamps using the REAPER tab-to-transient functionality to obtain "ground truth" timestamps for comparison with algorithm-detected timestamps. No false positives or false negatives were observed. As shown in Fig. 3 the algorithm detections showed a high degree of agreement with the annotated ground truths over 100 steps, with a mean error of 2.41±1.72 ms. The positive mean error can be attributed to the sliding window (11.6 ms) and hop size (50%) of the spectral energy difference computation. For the purposes of data analysis, we deemed the onset detection error and sound propagation delay low enough to be neglected.

Statistical Analysis From the timestamps obtained using the onset detection algorithm, SMS performance was analyzed in terms of the participants' ability to match stimulus period and phase. For phase matching, *mean asynchrony (MA)* and *standard deviation - asynchrony (STD-A)* were compared across stimuli (S1 - S9) using mixed-design repeated-measures ANOVAs, with 'Stimulus' as the within-subject factor and musical sophistication 'MSoph Group' as the between-subjects factor. For period matching, *inter-tap interval coefficient of variation (ITI-CoV)* and mean tempo deviation (MTD) in *beats per minute* were similarly compared. Additionally, we calculated the groupwise percentages of participants deviating from the correct stimulus tempo by over *0.2 BPM* for all stimuli. The threshold of 0.2 BPM corresponds to a timing mismatch of 100 ms after one minute at 120 BPM assuming the first tap was in phase, and this mismatch exceeds most echo perception thresholds [32]. Finally, A Friedman Test was conducted on the participants' subjective concentration ratings. Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons. All statistical analysis was done in SPSS 25.0 (IBM Corp).

5 Results

5.1 Phase Matching

Figures 4 and 5 show the average MA across participants and STD-A for each of Stimulus 1 to 9 (S1 to S9, decreasing PC) The effects of the independent variables on each dependent variable are considered in turn.

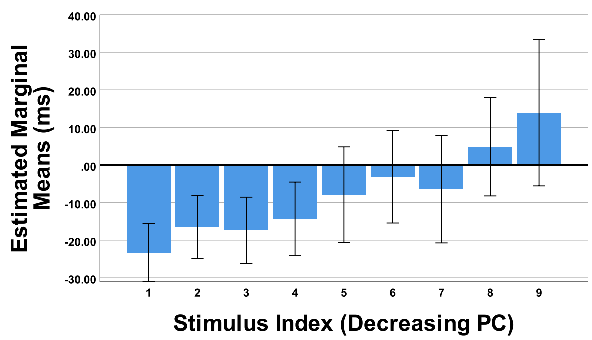

Fig. 4. Average MA across all participants for Stimuli 1 (maximum PC) through 9 (minimum PC). Vertical axis ticks are in ms and 95% confidence intervals are shown in error bars. The negative asynchrony gradually beocomes more positive with decreasing pulse clarity.

We tested the hypothesis that lower PC would lead to smaller NMA in a 9 Stimuli × 3 MSoph Groups mixed-design ANOVA. Results showed a significant main effect of stimulus (F(2,24) = 7.351, p = <.0001, η_p^2 = 0.776) and no significant interaction between stimulus and MSoph Group (F(2,24) = 1.379, n.s., η_p^2 = 0.393). A Tukey post-hoc test revealed that MA was statistically significantly more negative for stimulus S1(-23.3 ± 19.6 ms) as compared to S7(-6.4 ± 36.13 ms, p = 0.028), S8(4.8 ± 33.1 ms, p <.0001), and S9(13.9 ± 49.1 ms, p< .0001, see Figure 4). There were no significant differences in pairwise comparisons between MSoph groups.

Another mixed-design ANOVA tested the hypothesis that STD-A would increase with decreasing PC. We found a main effect of stimulus (F(2,24) = 5.628,

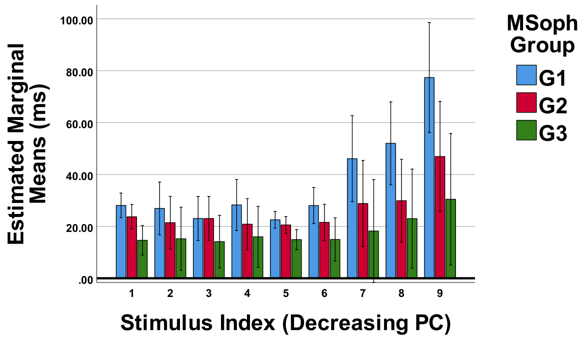

Fig. 5. Group wise *mean* STD-A for Stimuli 1 (maximum PC) through 9 (minimum PC). Vertical axis ticks are in ms and 95% confidence intervals are shown in error bars. Overall standard deviations increase with decreasing pulse clarity.

p =.001, $\eta_p^2 = 0.726$) with no significant interaction between stimulus and MSoph group. Post-hoc pairwise comparisons showed that STD-A was significantly less for S1(23.0 ± 8.8 ms) than for S9(53.9 ± 36.8 ms, p =.002), with a clear positive trend from S7 onward (see Figure 5). On the basis of MSoph, significant and nearly-significant differences exist between G1 (lowest OMSI group) and G3 (highest OMSI group) (p = 0.02), and G1 and G2 (p =.069) respectively.

5.2 Period Matching

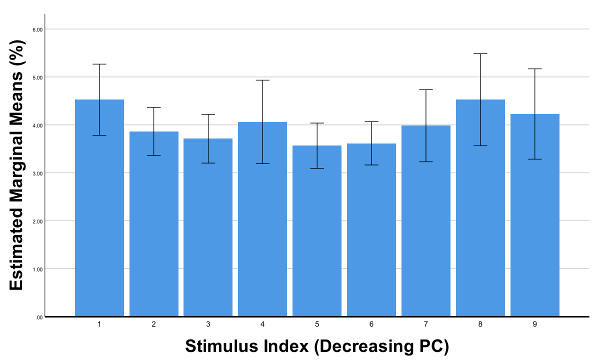

Fig. 6. Average ITI CoV across participants for Stimuli 1 (maximum PC) through 9 (minimum PC). Vertical axis ticks are in percentage and 95% confidence intervals are shown in error bars.

We tested the next hypothesis that decreasing PC would lead to increased ITI-CoV and Mean Tempo Deviation. The mixed ANOVA for ITI-CoV revealed

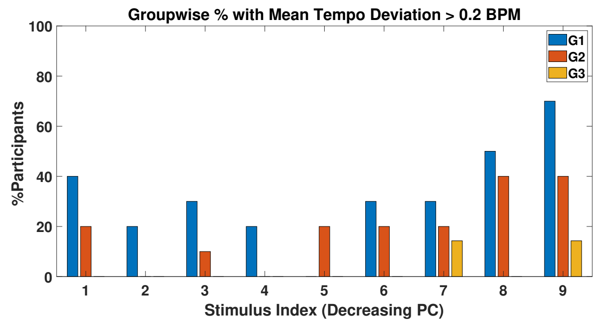

Fig. 7. Group wise % of participants with mean tempo deviation over 0.2 BPM for Stimuli 1 (maximum PC) through 9 (minimum PC). Vertical axis ticks are in percentage.

a significant main effect of stimulus ($F(2,24) = 2.399$, $p = .017$, $\eta_p^2 = 0.091$), although pairwise post-hoc tests showed non-significant differences between stimuli and MSoph groups. Despite this, it is noteworthy that S5 had the lowest mean ITI-CoV (3.6%) across participants (see Fig. 6) while S1 (max PC) had a mean value (4.59%) similar to S8 (4.64%) and S9 (4.32%) (min PC). The mixed ANOVA for MTD did not show a significant main effect of stimulus ($F(2,24) = 1.022$, $p = .421$, $\eta_p^2 = 0.041$). Fig. 7 shows the groupwise percentages of participants with MTD $>$ 0.2 BPM. For the majority of stimuli, this percentage appears inversely related to musical sophistication; G1 has the highest incidence and G3 the minimum. Contrary to our hypothesis, however, the percentages were minimum *not* for S1 but for S4 and S5, before increasing as expected towards S9.

5.3 Concentration Ratings

For the stimulus-wise subjective concentration ratings, a Friedman Test found significant differences among stimuli ($\chi^2 = 144.12$, p <.001), and Dunn-Bonferroni-based post-hoc comparisons showed significant differences between multiple pairs of stimuli and non-significant differences between MSoph groups, with a general increasing trend from S1 to S9 as shown in Fig. 8.

6 Discussion

The purpose of this study was to investigate the relationship between auditory pulse clarity and SMS performance (as measured by phase and period matching measures), as well as the impact of music training on this relationship. The clear trend of higher subjective concentration ratings with decreasing pulse clarity indicates that participants attended more closely to less clear stimuli to deduce their underlying pulsations, and maintain their level of synchronization performance. The ratings corroborate the good correspondence we found between

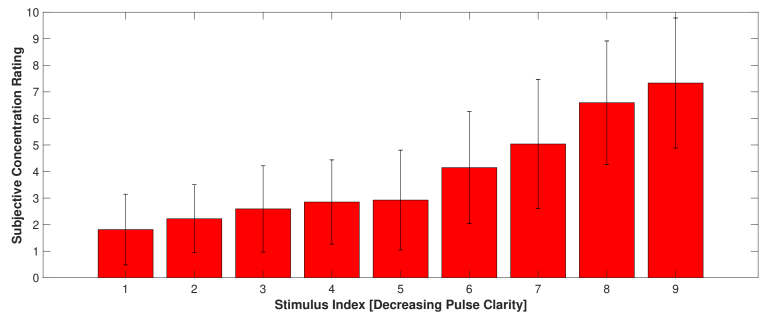

Fig. 8. Subjective concentration ratings for each of the nine stimuli across participants and groups. Vertical axis ticks are in the subjective scale unit, and 95% confidence intervals are shown in error bars.

fluctuation spectrum peakiness and perceived pulse salience, and were aligned with our hypothesis regarding them.

The phase-related results from the mixed ANOVAs showed good agreement with finger-tapping literature [2], firstly in that the mean asynchrony across participants was found to be negative for most high PC stimuli (see Figure 4). Secondly, the hypothesis of lower PC leading to smaller NMA was supported. Measured MA for all MSoph groups showed an increasing trend, with G2 and G3 having more positive mean values. This could be attributed to the masking of *true* perceptual onsets by previous reverb tails, and increased stimulus duration, ultimately leading to later P-centers and correspondingly later taps [16,18]. However, we did not find significant mean asychrony differences across MSoph groups resembling the findings of Chen et. al. [12]. This could be attributed to the use of OMSI as a general measure of musical training as opposed to the more detailed and stringent group segregation done by Chen et. al. [12] on the basis of several skill criteria.

G3 uniformly exhibited lower average STD-A than G1 and G2, in line with past literature [11]. We hypothesized that reducing PC would increase STD-A, but the results indicate that this relationship may not be linear, as the STD-A values across groups only begin to increase markedly beyond S6. The result seems reasonable, considering that most people would have experienced synchronizing to stimuli with some degradation in pulse clarity, e.g. due to natural reverberation. This points to there being a tolerated level of pulse degradation before regular phase synchronization begins to break down, which may affect individuals differently depending on their level of musical training. The increase in STD-A beyond S6 occurred across MSoph groups but was lower for trained musicians, in agreement with our hypothesis. An explanation is that the diminished extent of amplitude fluctuation of these stimuli during transients implied a smaller attack slope within the auditory temporal integration window, leading to a similar temporal P-center spread to those observed for slow-attack sounds by Danielsen et. al [16].

Next, we hypothesized that reduced PC would lead to group-specific degradation in period matching ability, measured in terms of Mean Tempo Deviation and ITI-CoV. We did not find a significant effect of PC on this outcome, nor any significant group differences. We did, however, find that was generally only participants from the lower MSoph groups (G1 and G2) who exceeded the 0.2 BPM deviation threshold, while G3 seemed largely unaffected in this regard. However, the percentage of these incidences did not steadily increase from S1-S9 as we had hypothesized. Interestingly, the percentages were lower for S4 and S5 than S1-3 as shown in Figure 7. A possible explanation could stem from the differences between the stimuli. S1-3 had greater PC with shorter reverb tails, meaning that they were sparser in the time domain. S6-9 had the lowest PC but were also the least sparse signals. We deduce that similar to the subdivision advantage [2], the presence of non-zero auditory information (reverb tails) in between two pulses may have an assistive effect on synchronization ability. S4 and S5 may thus have been 'sweet-spot' referents balancing the tradeoff between providing non-sparse inter-pulse assistive information and compromising the salience of the underlying pulsation itself. A possible support for this would be that listeners have been reported to use different strategies for judging duration of tones with flat or decaying envelope shape [33]. Another possibility is that higher PC gave participants greater awareness of their own synchronization errors, affecting confidence and worsening performance. Simultaneous masking of the higher PC transients due to bone conduction while stepping may also have imparted advantages to synchronizing with stimuli having longer decay times.

The tactile feedback from whole-body weight shifting when stepping is different from that of finger tapping but, as noted by Chen et. al. [3], also involves the constraint of maintaining balance. Chen et. al. suggested that the lower variability in stepping compared to heel tapping found in their study might be explained by the tactile feedback. Unlike their study, we recorded the sounds of the steps rather than the movement of markers. Comparing their reported mean asynchronies for toe (-61.45 ms) and heel (33.88 ms) [3], our mean values fall in between (see Figure 4). While we cannot be exactly sure how participants purposefully aligned their steps and the perceived tactile feedback to the stimuli sounds, we would argue that our our approach of sound detection makes sense from an ecological perspective.

Our hypothesis that ITI CoV would increase with decreasing PC was not supported by the data. In particular for G1, ITI CoV did not increase proportionally with STD-A as would be expected. This conflict is explained by these participants stepping in a regular fashion but drifting from the stimulus tempo. The lack of significant differences between groups resembling the results of [12] could similarly be attributed to our choice of musical sophistication index. Another measure more sensitive to differences in general rhythmic ability, along with a stringent group segregation based on concrete skill criteria might have given another result. Nevertheless, it is noteworthy that mean ITI-CoV was lowest for S5, which when combined with the superior tempo performance and low STD-A indicates that participants performed best here in terms of both period and phase matching. A more thorough investigation of this performance improvement observed in the intermediate stimuli would make an interesting topic for a follow-up study on a larger population sample.

Overall, these findings indicate that the clarity of the periodic referent has a considerable influence on SMS performance, which would have direct implications for music and dance performance. The type of degradation present in our stimuli bears resemblance to what might appear in real environments where music or dance activities are performed. For lower PC, beat entrainment not only consumes more cognitive resources, but is also less accurate and stable to a perceptible extent (mean STD-A of 54 ms for S9 v/s 23 ms for S1), highlighting the importance of clear pulse audibility for timing during performance. Interestingly, performance appeared to remain fairly consistent until a certain 'threshold' was crossed, around S7 ('Moderately Unclear') (see Figure 5), implying a certain sensory 'robustness' to referent degradation.

Limitations of the study include the static modality of PC manipulation, short length of the trials and relatively small number of participants, particularly with extensive musical training. Another shortcoming was the lack of detailed background information on the level of training the participants had; the use of OMSI may have blurred inherent inter-group differences. Pulse degradations in real-life situations may be time-varying due to the changing spectral content of referents, and unpredictably varying masking effects. Interpersonal entrainment and visual cues during group performance are also important factors. Future studies can address whether the different types of pulse degradation similarly impact SMS performance, and whether these can be accurately modeled by fluctuation spectrum measurements.

7 Conclusion

The present study concluded that reducing auditory pulse clarity influences sensorimotor synchronization performance in terms of both phase and period matching abilities, in addition to subjective ratings of required concentration. We found evidence that a certain degree of temporal pulse degradation is not only tolerated, but may also lead to performance improvements upto a point, beyond which pulse degradation detrimentally affects the performance of musically untrained participants more than trained ones. These results have direct relevance to timing performance in dance, music and timing practice, although further studies must be conducted on a larger sample, exploring other ecological pulse degradation methods to explore their true implications for real SMS contexts.

Acknowledgements

We thank the participants in our experiment. Authors PK and RSA were mainly responsible for the experiment, data analysis and writing of the manuscript. Author SD supervised the project, and assisted in writing. SD's contribution is partially funded by NordForsk's Nordic University Hub Nordic Sound and Music Computing Network NordicSMC, project number 86892.

References

1. Repp, B. H.: Sensorimotor Synchronization: A Review of the Tapping Literature. Psychonomic Bulletin & Review **12**(6), 969–992 (2005). https://doi.org/10.3758/BF03206433
2. Repp, B. H., Su, Y. H.: Sensorimotor Synchronization: A Review of Recent Research (2006–2012). Psychonomic Bulletin & Review **20**(3), 403–452 (2013). https://doi.org/10.3758/s13423-012-0371-2
3. Chen, H. Y., Wing, A. M., Pratt, D: The Synchronisation of Lower Limb Responses with a Variable Metronome: the Effect of Biomechanical Constraints on Timing. Gait & Posture **23**(3), 307–314 (2006). https://doi.org/10.1016/j.gaitpost.2005.04.001
4. Styns, F., van Noorden, L., Moelants, D., Leman, M: Walking on Music. Human Movement Science **26**, 769–785 (2007). https://doi.org/10.1016/j.humov.2007.07.007
5. Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., Toiviainen, P.: Influences of Rhythm- and Timbre-related Musical Features on Characteristics of Music-induced Movement. Frontiers in Psychology **4**, 183 (2013). https://doi.org/10.3389/fpsyg.2013.00183
6. Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., Toiviainen, P.: Music Moves Us: Beat-related Musical Features Influence Regularity of Music-induced Movement. In: Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, pp. 183–187 (2012)
7. Lartillot, O., Eerola, T., Toiviainen, P., Fornari, J.: Multi-Feature Modeling of Pulse Clarity: Design, Validation and Optimization. In: ISMIR, pp. 521–526 (2008)
8. Madison, G., Karampela, O., Ullén, F., Holm, L.: Effects of Practice on Variability in an Isochronous Serial Interval Production Task: Asymptotical Levels of Tapping Variability After Training Are Similar to Those of Musicians, Acta Psychologica **143**(1), 119–128 (2013). https://doi.org/10.1016/j.actpsy.2013.02.010
9. Fujii, S., Hirashima, M., Kudo, K., Ohtsuki, T., Nakamura, Y., Oda, S.: Synchronization Error of Drum Kit Playing with a Metronome at Different Tempi by Professional Drummers. Music Perception: An Interdisciplinary Journal, **28**(5), 491–503 (2011). https://doi.org/10.1525/mp.2011.28.5.491
10. Krause, V., Pollok, B., Schnitzler, A.: Perception in Action: The Impact of Sensory Information on Sensorimotor Synchronization in Musicians and Non-Musicians. Acta Psychologica **133**(1), 28-–37 (2010). https://doi.org/10.1016/j.actpsy.2009.08.003
11. Repp, B. H.: Sensorimotor Synchronization and Perception of Timing: Effects of Music Training and Task Experience, Human Movement Science **29**(2), 200-–213 (2010). https://doi.org/10.1016/j.humov.2009.08.002
12. Chen, J., Penhune, V., Zatorre, R.: Moving on Time: Brain Network for Auditory–Motor Synchronization is Modulated by Rhythm Complexity and Musical Training, Journal of Cognitive Neuroscience **20**(2), 226–239 (2008). https://doi.org/10.1162/jocn.2008.20018
13. Palmer, C., Koopmans, E., Loehr, J. D., & Carter, C., Movement-related feedback and temporal accuracy in clarinet performance. Music Perception: An Interdisciplinary Journal **26**(5), 439—449 (2009). https://doi.org/10.1525/mp.2009.26.5.439

14. Mainka, S., Wissel, J., Völler, H., & Evers, S., The Use of Rhythmic Auditory Stimulation to Optimize Treadmill Training for Stroke Patients: A Randomized Controlled Trial. Frontiers in Neurolog **9**, 755 (2018). https://doi.org/10.3389/fneur.2018.00755
15. Schaffert, N., Janzen, T. B., Mattes, K., & Thaut, M. H., A Review on the Relationship Between Sound and Movement in Sports and Rehabilitation. Frontiers in psychology **10**, 244 (2019). https://doi.org/10.3389/fpsyg.2019.00244
16. Danielsen, A., Nymoen, K., Anderson, E., Câmara, G. S., Langerød, M. T., Thompson, M. R., London, J.: 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 (2019). https://doi.org/10.1037/xhp0000611
17. Gordon, J. W.: The Perceptual Attack Time of Musical Tones. The Journal of the Acoustical Society of America **82**(1), 88—105 (1987). https://doi.org/10.1121/1.395441
18. Vos, P. G., Mates, J., van Kruysbergen, N. W.: 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 (1995). https://doi.org/10.1080/14640749508401427
19. Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., Coussement, P., Leman, M.: The Impact of the Bass Drum on Human Dance Movement, Music Perception: An Interdisciplinary Journal **30**(4), 349–359 (2012). https://doi.org/10.1525/mp.2013.30.4.349
20. Large, E. W.: On Synchronizing Movements to Music, Human Movement Science **19**(4), 527–566 (2000). https://doi.org/10.1016/S0167-9457(00)00026-9
21. Fujioka, T., Trainor, L. J., Large, E. W., Ross, B.: Internalized Timing of Isochronous Sounds Is Represented in Neuromagnetic Beta Oscillations, Journal of Neuroscience **32**(5), 1791–1802 (2012). https://doi.org/10.1523/JNEUROSCI.4107-11.2012
22. Chen, J. L., Penhune, V. B., Zatorre, R. J.: The Role of Auditory and Premotor Cortex in Sensorimotor Transformations. Annals of the New York Academy of Sciences **1169**(1), 15–34 (2009). https://doi.org/10.1111/j.1749-6632.2009.04556.x
23. Peeters, G.: A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO project (version 1.0). In: Report, Ircam (2004)
24. Klapuri, A.,Eronen, A., Astola, J., Analysis of the Meter of Acoustic Musical Signals. IEEE Transactions on Audio, Speech and Langage Processing **14**(1), 342–355 (2006), https://doi.org/10.1109/TSA.2005.854090
25. Pampalk, E., Rauber, A., Merkl, D.: Content-based Organization and Visualization of Music Archives. In: Proceedings of the 10th ACM International Conference on Multimedia, pp. 570–579 (2002)
26. Feiten, B. and Günzel, S., Automatic Indexing of a Sound Database Using Self-organizing Neural Nets. Computer Music Journal **18**(3), 53–65 (1994), https://doi.org/10.2307/3681185
27. Fastl, H., Fluctuation Strength and Temporal Masking Patterns of Amplitude-Modulated Broad-band Noise. Hearing Research **8**(1), 59–69 (1982). https://doi.org/10.1016/0378-5955(82)90034-X
28. Lartillot, O., Toiviainen, P., Eerola, T., A Matlab Toolbox for Music Information Retrieval. In: C. Preisach, H. Burkhardt, L. Schmidt-Thieme, R. Decker (Eds.), Data Analysis, Machine Learning and Applications, Studies in Classification, Data Analysis, and Knowledge Organization, Springer, Berlin, Heidelberg (2008)
29. Fraisse, P.: Rhythm and Tempo. The Psychology of Music **1**, 149–180 (1982)

30. Ollen, J. E.: A Criterion-related Validity Test of Selected Indicators of Musical Sophistication Using Expert Ratings. Doctoral Dissertation, The Ohio State University (2006)
31. Bello, J.B., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., Sandler, M.B.: A tutorial on onset detection in music signals, IEEE Transactions on Speech and Audio Processing **13**(5), 1035–1047 (2008). https://doi.org/10.1109/TSA.2005.851998
32. Wallach, H., Newman, E. B., Rosenzweig, M. R.: The Precedence Effect in Sound Localization, The American Journal of Psychology **62**, 315–336 (1949). https://doi.org/10.2307/1418275
33. Vallet, G. T., Shore, D. I., Schutz, M., Exploring the Role of the Amplitude Envelope in Duration Estimation. Perception **43**(7), 616–630 (2014). https://doi.org/10.1068/p7656