Journal of Experimental Psychology: Human Perception and Performance

Multimodal Sensory Integration: Diminishing Returns in Rhythmic Synchronization

Vinith Johnson, Wan-Yu Hsu, Avery E. Ostrand, Adam Gazzaley, and Theodore P. Zanto

Online First Publication, July 30, 2020. http://dx.doi.org/10.1037/xhp0000833

American Psychological Association

Journal of Experimental Psychology: Human Perception and Performance

ISSN: 0096-1523

http://dx.doi.org/10.1037/xhp0000833

Multimodal Sensory Integration: Diminishing Returns in Rhythmic Synchronization

Vinith Johnson
University of California, San Francisco

Wan-Yu Hsu
University of California, San Francisco, and Halo Neuroscience, San Francisco, California

Avery E. Ostrand, Adam Gazzaley, and Theodore P. Zanto
University of California, San Francisco

Synchronizing movements with events in the surrounding environment is a ubiquitous aspect of behavior. Experiments studying multimodal integration and rhythmic synchronization tend to focus on how bimodal (e.g., audio-visual) stimuli enhances synchronization performance (i.e., reduced variability) compared with synchronization with its unimodal constituents. As such, it is unclear whether trimodal (i.e., audio-visual-tactile) stimuli may yield additional performance benefits. To address this, we developed a multimodal sensorimotor synchronization assessment that incorporates audio, visual, and vibrotactile stimuli. Results replicate performance improvements with bimodal compared with unimodal stimuli. However, trimodal stimuli yields less, or in some cases no advantage compared with bimodal stimuli. These results demonstrate that in this case, increasing the amount of sensory information beyond bimodal stimuli yields little or no additional performance benefits.

Public Significance Statement

This study contributes to the increasing research in multisensory integration by revealing that an increased amount of sensory information does not necessarily lead to improved performance in rhythmic synchronization.

Keywords: synchronization, rhythm, multimodal, music, performance

Every day we are subjected to multisensory stimuli from the environment surrounding us. Most of these daily events are rarely unimodal; they are multisensory experiences, deriving from the integration of information acquired through several sensory modalities, which provide access to numerous types of information on the surrounding environment. The brain integrates multisensory information to provide a complete and coherent representation of what is being perceived and consequently for appropriate behavioral responses to be generated (Dionne-Dostie, Paquette, Lassonde, & Gallagher, 2015).

Sensorimotor synchronization refers to the coordination of movement with an externally presented rhythmic stimulus. During sensorimotor synchronization with unimodal stimuli (e.g., auditory or visual), performance is generally better when the stimuli consist of auditory beeps rather than flashes of light (Repp, 2003; Repp & Penel, 2004). Yet, when visual stimuli are not discretely flashing, but rather continuously oscillating (as in a bouncing ball), synchronization performance may be comparable with performance with auditory stimuli (Iversen, Patel, Nicodemus, & Emmorey, 2015). Although less research has been conducted into how our

Vinith Johnson, Neuroscape and Department of Neurology, University of California, San Francisco; Wan-Yu Hsu, Neuroscape and Department of Neurology, University of California, San Francisco, and Halo Neuroscience, San Francisco, California; Avery E. Ostrand, Neuroscape and Department of Neurology, University of California, San Francisco; Adam Gazzaley, Neuroscape and Departments of Neurology, Physiology, and Psychiatry, University of California, San Francisco; Theodore P. Zanto, Neuroscape and Department of Neurology, University of California, San Francisco.

This work was supported by the Hart Foundation for Deep Rhythm and a grant from the UCSF Resource Allocation Program. Vinith Johnson, Wan-Yu Hsu, Adam Gazzaley, and Theodore P. Zanto designed the experiment. Vinith Johnson, Wan-Yu Hsu, and Avery E. Ostrand collected the data. Vinith Johnson, Avery E. Ostrand, and Theodore P. Zanto analyzed the data. Vinith Johnson and Theodore P. Zanto drafted the manuscript. All authors discussed the results and contributed to editing the manuscript. We thank Di Zhao and Tiffany Jane Ford for assistance in data collection and Peter Wais and Courtney Gallen for their valuable comments and feedback.

Correspondence concerning this article should be addressed to Vinith Johnson, Department of Neurology, University of California, San Francisco, Sandler Neuroscience Center, Room NS 505, 675 Nelson Rising Lane, San Francisco, CA 94158. E-mail: vinith.johnson@ucsf.edu or vinith.johnson@gmail.com

somatosensory system may enable sensorimotor synchronization, research suggests that the sense of touch performs well in sensory synchronization tasks. Notably, if the saliency of the stimuli are comparable, the accuracy of synchronization with a tactile metronome can equal synchronization with an auditory metronome (Ammirante, Patel, & Russo, 2016).

Studies assessing the influence of multisensory stimuli on sensorimotor synchronization ability generally indicate that synchronization performance is enhanced (i.e., variability is reduced) when bimodal stimuli are presented compared with performance when presented with either of the constituent unimodal stimuli. These results are observed regardless of whether the bimodal stimulus is audio-tactile (Kelso, Fink, DeLaplain, & Carson, 2001; Wing, Doumas, & Welchman, 2010), audio-visual (Armstrong & Issartel, 2014; Elliott, Wing, & Welchman, 2010), or visual-tactile (Elliott et al., 2010). This performance improvement observed by integrating two sensory modalities has previously been studied using Bayesian models of prediction. Specifically, performance variability is lowered when integrating two redundant sensory signals, in line with predictions from a maximum likelihood estimation (MLE) model (Alais & Burr, 2004; Bresciani, Dammeier, & Ernst, 2008; Ernst & Banks, 2002; Ernst & Bülthoff, 2004). Although stimulus perception from three simultaneous modalities similarly follows statistically optimal integration (Wozny, Beierholm, & Shams, 2008), it remains unclear whether performance gains may be achieved when three modalities are used for sensorimotor synchronization. To address this, we designed an experiment to ascertain whether trimodal integration of auditory, visual and vibriotactile stimuli yield significant gains in rhythmic synchronization.

Despite numerous studies assessing the impact of multisensory stimuli on synchronization performance (Armstrong & Issartel, 2014; Elliott et al., 2010; Kelso et al., 2001; Wing et al., 2010), to our knowledge, there has not been a systematic assessment of all three types of bimodal stimulus combinations or a concomitant assessment of whether trimodal (audio-visual-tactile) stimuli may yield greater benefits than bimodal stimuli during sensorimotor synchronization. During stimulus detection, there is evidence that performance improves (i.e., speeded response times) when trimodal stimuli are presented, compared with bimodal or unimodal stimuli (Colonius & Diederich, 2006; Diederich & Colonius, 2004). Furthermore, vision, touch, and audition are automatically integrated during the perception of the sequence of events (Bresciani et al., 2008). Specifically, distractors reduced the ability to count target (attended modality) events and this effect was larger during bimodal distraction (i.e., trimodal stimulus presentation). Additionally, results from the trimodal integration of visual, tactile, and auditory signals indicated that the three modalities do not contribute equally to combined perception, where the contribution of each modality depends on its relative reliability (i.e., the inverse of variability during unimodal presentation). Yet, it is unclear whether multisensory integration of trimodal stimuli may facilitate sensorimotor synchronization beyond the known improvements bimodal stimuli offers over unimodal stimuli. Therefore, we conducted an assessment of sensorimotor synchronization with audio, visual, and vibrotactile stimuli to systematically assess the role of unimodal, bimodal, and trimodal stimuli on task performance. It is hypothesized that the additional redundant timing signals afforded by trimodal stimuli will serve to improve sensorimotor synchronization performance beyond the performance with only one (unimodal) or two (bimodal) input streams.

Method

Participants

Participants in the study were mainly recruited among staff and students from University of California, San Francisco, and the University of California, Berkeley. There were 21 adult participants in total ($N = 12$ women) in the age range of 19–30 years. All participants gave informed consent to participate in the study. Based on previous data assessing differences between stimulus types during the same sensorimotor synchronization task (Zanto, Padgaonkar, Nourishad, & Gazzaley, 2019), 20 participants corresponds to a 90% (1-β) power estimate at a 95% (1-α) significance for a repeated measures analysis of variance (ANOVA) with three levels. We collected data from 21 participants in case one would need to be discarded (e.g., not complete the task, complete the task incorrectly, data corruption, etc.), but this was not the case and all participants were included in the analysis. All participants had normal or corrected to normal vision and hearing. Participants had no known history of neuromuscular deficits that could affect performance in the tasks.

Stimuli

The multimodal sensorimotor synchronization assessment was programmed in Unity, executed on a Microsoft Surface Pro 3 Tablet (60 Hz refresh rate) and designed to assess rhythmic capabilities as measured by the ability to tap to different metronome-like sequences. In the assessment, rhythmic ability was measured across 63 levels, which consisted of parametrically manipulating three variables: *tempo* of the metronome (three tempos), the audio-visual-tactile *stimulus* provided (seven stimuli), and the rhythmic *task* performed (three tasks). Out of which 27 levels were fixed and played with tactile stimulus (audio-visual-tactile, visual-tactile, and audio-tactile), iterating through the three tasks (on-beat, off-beat, and continuation), then iterating through the stimulus types, then the tempos (slow, medium, and fast). Similarly, 27 levels were fixed and played without tactile stimulus (audio-visual, visual, and audio) and nine levels with tactile only (tactile). The three sets of the levels were then randomized and counterbalanced. The tempo varied between slow (750 ms), medium (525 ms), and fast (350 ms) interonset intervals (IOIs). Each level was 30 s long. Therefore, the number of trials for slow, medium and fast levels were 40, 57, and 85, respectively. For purposes of the current study, task and tempo was not assessed, and all data were averaged over the different tasks and tempos. This was done because we are interested in the effects of multimodal integration that are generalizable across musical tasks and tempos.

There were a total of seven different stimuli, which were grouped under three categories—Unimodal, Bimodal, and Trimodal. The stimuli presented to the participants were as follows, trimodal: audio-visual-tactile; bimodal: visual-tactile, audio-tactile, audio-visual and unimodal—auditory only, visual only, and tactile only. The visual stimulus had a blue circle with a small vertical line through it moving horizontally from one side of the screen to the other, passing between larger vertical lines, a pair on the left side of the screen and a pair on the right—each equidistant from the center of the screen—such that when the circle was in the middle of the vertical lines it changed direction, indicating the beat (see Figure 1). The auditory stimulus consisted of a 50 ms long 500 Hz pure tone. Sound intensity was set to a comfortable listening level and was delivered to both the ears through KOSS UrR/29 headphones. The tactile stimulus consisted of a vibration on the participant's back (see Figure 2) at a fixed intensity (75% of max intensity possible) generated by a SUBPAC M2 (StudioFeed), which was triggered by an inaudible 20 Hz auditory waveform that was 50 ms long. For bimodal and trimodal stimuli, the constituent stimulus components were presented simultaneously. Headphones were used to attenuate the sound produced by the tactile stimulus during visual-tactile and tactile only trials.

Experimental Procedure

Participants were asked to perform three tasks to characterize the robustness of the effects of multisensory synchronization. These tasks are as follows: (a) On-beat: tap along with each stimulus event (i.e., beat: sound onset, tactile onset, and/or when the ball was centered with the lines at either side of the screen); (b) Off-beat: tap half-way between each stimulus event; or (c) Continuation: after four stimulus events (i.e., four beats), the stimuli were discontinued and participants had to continue the metronomic rhythm by tapping for four beats without disrupting the tempo. After the four-beat "silent period" where participants were to tap, stimuli were resumed for another four beats and followed by another four-beat silent period where participants were instructed to tap. The stimuli and silent periods continued to alternate for the duration of the level. Each level lasted approximately 30 s. Tempo progressed from slow to medium to fast, whereas task progressed from On-beat to Off-beat to Continuation. All three tasks were conducted before changing tempo. Two fingers were used for response (left index and right index) during the on-beat and continuation tasks. One finger (dominant index) was used during the off-beat task.

Before the experiment, participants were able to practice each task and demonstrated they understood task instructions. During the experiment, after each level, participants were provided with their average absolute offset (asynchrony) in milliseconds to provide feedback on their performance and were then provided instructions for the next level to complete. During game play, task instructions remained in the upper right corner of the screen (i.e., "on-beat," "off-beat," or "continuation"). Moreover, the lower left corner of the screen indicated which level the participant was on and a measure of tap offset was displayed to provide online feedback (see Figure 2). Additional feedback was provided in the form of a vertical dashed line to indicate when the screen was touched and was located where the visual ball was (or would have been during audio-only, tactile-only, and audio-tactile) at the time of tap onset; thereby, providing a visualization of the tap asynchrony.

Analysis

All analyses were conducted using MATLAB, SPSS, and JASP. Accuracy, Absolute Offset, and Standard Deviation of the Offset were calculated. *To capture multiple aspects of* synchronization performance, a composite Rhythm Score was also calculated (Zanto et al., 2019). To account for touchscreen input lag, all recorded tap data were adjusted by 72.28 ms (Deber et al., 2016). Accuracy was determined by calculating the number of taps per stimulus. One tap within ±IOI/2 of stimulus onset was considered correct, while an incorrect tap involves misses or more than one tap per stimulus. Absolute offset and the corresponding standard deviation were calculated from correct taps only. The absolute offset was calculated as the absolute value of the difference between the tap time and the time at which they were supposed to tap (i.e., on-beat: stimulus onset, continuation: where stimulus onset

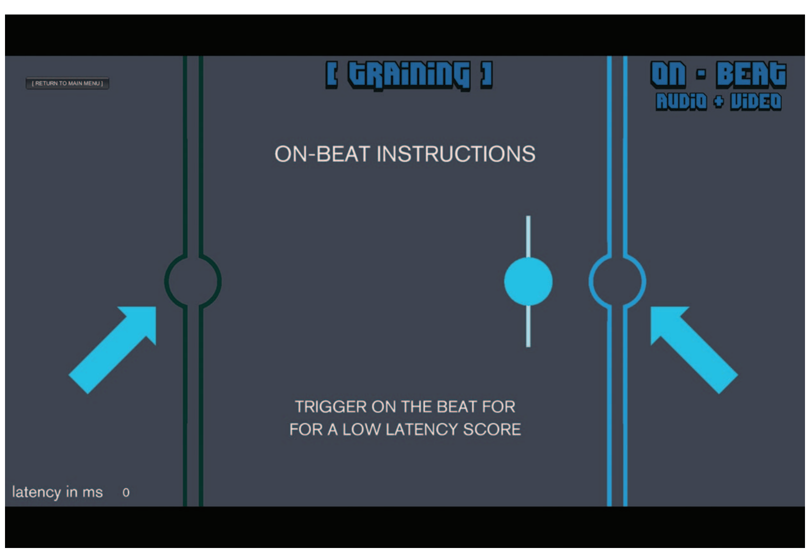

Figure 1. Screenshot of the multimodal sensorimotor synchronization assessment. See the online article for the color version of this figure.

Figure 2. Positioning of StudioFeed's SUBPAC for tactile stimulation. See the online article for the color version of this figure.

should be, off-beat: half-way between stimuli). Standard deviation was calculated from the actual offset times (i.e., standard deviation of the differences between tap time and stimulus onset). Rhythm Scores were calculated by *z*-scoring the accuracy, absolute offset, and standard deviation of the offset separately across all conditions for each participant, and then averaging the three *z*-scores together per participant and level. Before averaging the *z*-scores together, *z*-scored absolute offset, and standard deviation values were multiplied by -1, so that positive values indicate better performance. Furthermore, because the Rhythm Score is an average of *z*-scores, a value of zero represents mean performance and negative values are worse performance. Data were averaged together based on stimulus types: unimodal, bimodal, or trimodal. Because trimodal stimuli had fewer data than bimodal and unimodal, a bootstrap procedure was conducted to calculate the bimodal and unimodal means. A one-way repeated measures ANOVA with three levels (Unimodal, Bimodal, and Trimodal) was conducted. A Greenhouse-Geisser correction was applied as needed, and follow-up *t* tests were used to interrogate main effects according to our a priori hypothesis. Multiple comparisons were corrected by the false discovery rate method.

Results

To address our main hypothesis (i.e., trimodal > bimodal > unimodal performance), the data were collapsed across tempos and tasks and then compared on the basis of stimulus type as described above. To assess whether multisensory stimuli yielded performance gains at the group level, the means of each unimodal, bimodal, and trimodal performance measures were assessed. To confirm multisensory integration, it was necessary to account for individual biases for a particular stimulus type. Therefore, the best bimodal and the best unimodal performances were selected for each participant and used for comparisons (Stevenson et al., 2014). First, repeated measures ANOVAs were conducted separately for the means of the Standard Deviations and the Rhythm Scores. Main effects were observed for the Standard Deviations ($F(2, 40) = 32.76$, $p < .001$, $\eta^2 = 0.26$, $BF_{incl} = 244e\hat{6}$; Figure 3A gray bars), and Rhythm Scores ($F(2, 40) = 15.48$, $p < .001$, $\eta^2 = 0.40$, $BF_{incl} = 127254.94$; Figure 3B gray bars). Further, repeated measures ANOVAs were conducted comparing the best unimodal performance, best bimodal performance, and the trimodal performance, as measured by Standard Deviations (minimum values) and Rhythm Scores (maximum values) separately. Main effects were observed for the Standard Deviations ($F(2, 40) = 4.31$, $p < .05$, $\eta^2 = 0.26$, $BF_{incl} = 2.20$; Figure 3A orange bars), and Rhythm Scores ($F(2, 40) = 7.69$, $p = .001$, $\eta^2 = 0.22$, $BF_{incl} = 120.45$; Figure 3B orange bars).

T tests were then used to interrogate these main effects. For the mean Standard Deviations (Figure 3A gray bars), unimodal stimuli yielded lower performance than bimodal, $t(20) = 15.54$, $p < .001$, Cohen's $d = 3.39$ and trimodal stimuli, $t(20) = 5.23$, $p < .001$, Cohen's $d = 1.14$; Figure 3B. However, no difference was observed between bimodal and trimodal performance, $t(20) = 0.49$, $p = .62$, Cohen's $d = 0.10$. Comparing the best unimodal and bimodal conditions (minimum values; Figure 3A orange bars), bimodal stimuli yielded better performance than unimodal, $t(20) = 4.40$, $p < .001$, Cohen's $d = 0.96$. Participants performed better in the best bimodal condition than the mean trimodal condition, $t(20) = 2.12$, $p < .05$, Cohen's $d = 0.13$. However, no difference was observed between the best unimodal condition and the trimodal condition, $t(20) = 0.03$, $p = .97$, Cohen's $d = 0.008$.

Similarly, for the Rhythm Score, mean unimodal stimuli yielded lower performance than bimodal, $t(20) = 7.73$, $p < .001$, Cohen's $d = 1.61$ and trimodal stimuli, $t(20) = 4.35$, $p < .001$, Cohen's $d = 1.01$; Figure 3B gray bars. No significant difference was observed between bimodal and trimodal performance, $t(20) = 0.53$, $p = .60$, Cohen's $d = 0.16$. Comparing the best unimodal and bimodal conditions (Figure 3B orange bars), bimodal stimuli yielded better performance than unimodal, $t(20) = 4.15$, $p < .001$, Cohen's $d = 0.90$. The best bimodal condition also performed

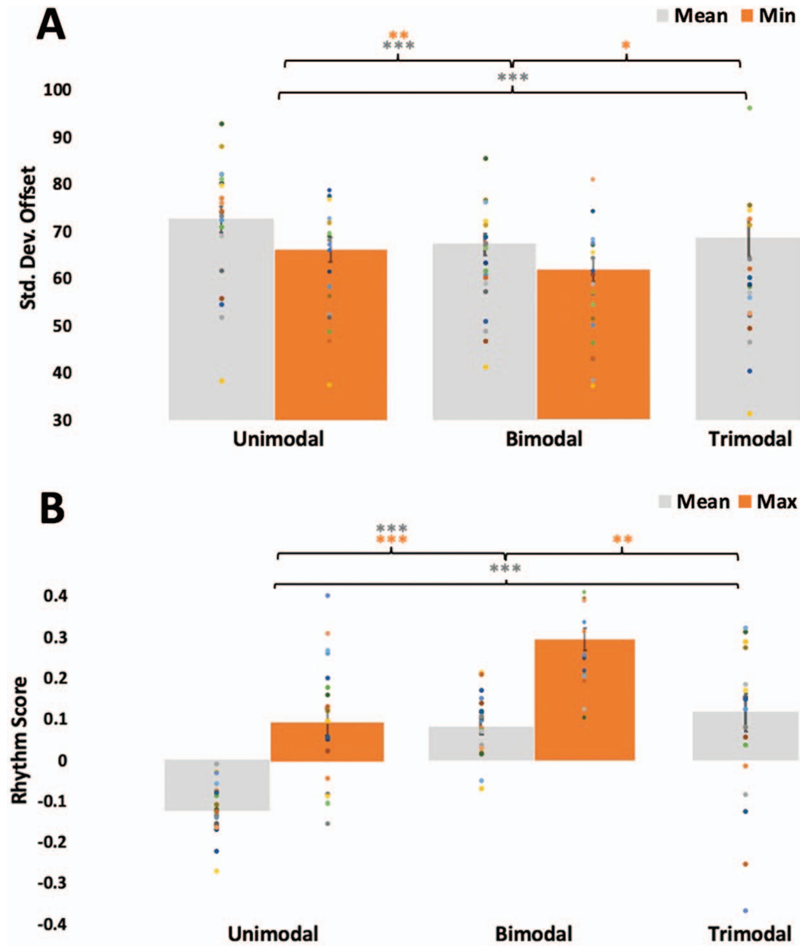

Figure 3. Performance measures for (A) Standard Deviations and (B) Rhythm Scores. Gray bars indicate average performance (i.e., unimodal = mean of each unimodal stimulus type; bimodal = mean of each bimodal stimulus type). Orange bars indicate the best unimodal and bimodal performance from each participant. Orange asterisk indicates significance between orange bars and gray asterisk indicates significance between gray bars. Error bars indicate standard error of the mean (*SEM*). Dots are individual data points. * $p < .05$. ** $p < .01$. *** $p < .001$.

better than the trimodal performance, $t(20) = 2.99$, $p < .01$, Cohen's $d = 0.65$. However, no difference was observed between best unimodal condition and trimodal performance, $t(20) = 0.44$, $p = .66$, Cohen's $d = 0.09$. Together, these results converge to show that bimodal stimuli enables improved sensorimotor synchronization performance compared with unimodal stimuli. Yet, trimodal stimuli do not provide any additional benefit above bimodal stimuli. Of note, accuracy did not show any significant difference across the modalities, while absolute offset showed a similar effect as the standard deviation.

Bimodal Versus Unimodal

To assess differences in performance between bimodal stimuli and its unimodal constituents, *t* tests were conducted. Assessment of the Standard Deviations showed that bimodal stimuli generally elicited better performance than unimodal stimuli (Figure 4A). Specifically, bimodal audio-video stimuli led to better performance than unimodal auditory, $t(20) = 9.64$, $p < .001$ Cohen's $d = 2.10$, and unimodal visual stimuli, $t(20) = 3.98$, $p < .001$, Cohen's $d = 0.86$. Similarly, bimodal visual-tactile stimuli resulted in better performance than unimodal visual, $t(20) = 3.36$, $p < .01$, Cohen's $d = 0.73$ and unimodal tactile stimuli, $t(20) = 4.79$, $p < .001$, Cohen's $d = 1.04$. However, bimodal audio-tactile stimuli did not exhibit better performance than unimodal audio, $t(20) = 1.34$, $p = .19$, Cohen's $d = 0.29$ or unimodal tactile stimuli, $t(20) = 1.03$, $p = .31$, Cohen's $d = 0.22$. *T* tests were then conducted to compare mean bimodal performance with its constituently best unimodal condition. Audio-visual stimuli yielded better performance measures than A,V (min; $t(20) = 3.98$, $p < .001$, Cohen's $d = 0.86$). Similarly, visual-tactile performed better than V,T (min; $t(20) = 1.83$, $p = .08$, Cohen's $d = 0.4$). However, audio-tactile did not yield better results than A,T (min; $t(20) = 2.0$, $p = .05$, Cohen's $d = 0.43$).

As depicted in Figure 4B, bimodal stimuli generally elicited better Rhythm Scores than unimodal stimuli. Specifically, bimodal audio-video stimuli led to better performance than unimodal au-

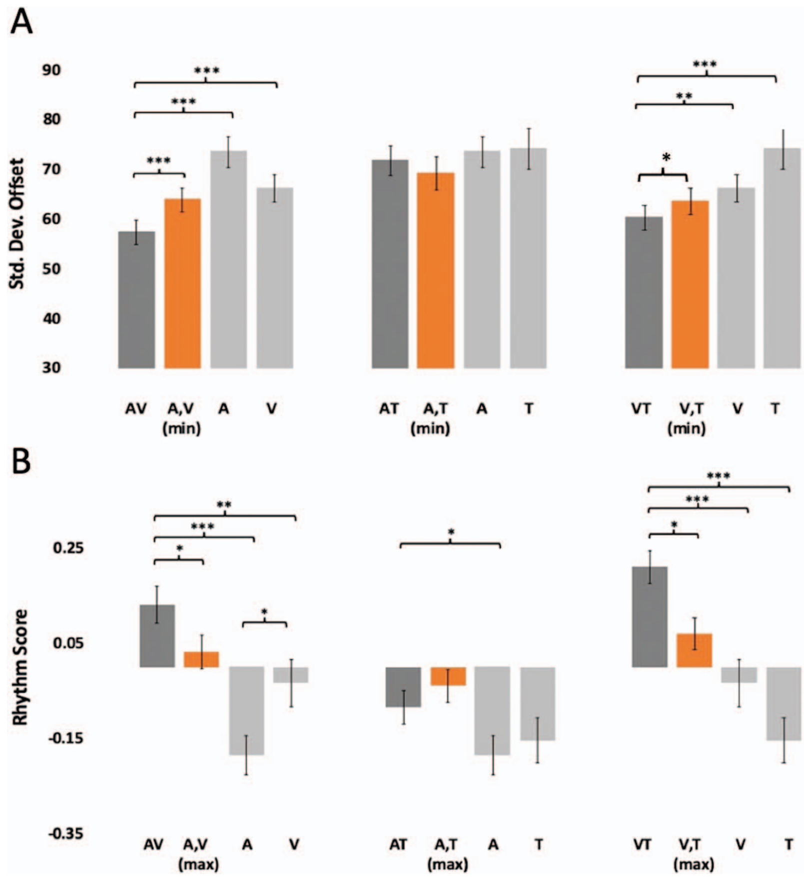

Figure 4. Difference in performance measures for (A) Standard Deviations and (B) Rhythm Score between bimodal and unimodal stimuli. Gray bars indicate average performance (i.e., unimodal = mean of each unimodal stimulus type; bimodal = mean of each bimodal stimulus type). Orange bars indicate the best unimodal performance from each participant. A = auditory; V = video; T = tactile; AV = audio-visual; VT = video-tactile; AT = audio-tactile. Error bars indicate standard error of the mean (*SEM*). $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

ditory, $t(20) = 5.59$, $p < .001$ Cohen's $d = 1.22$, or visual stimuli, $t(20) = 3.06$, $p < .01$, Cohen's $d = 0.66$. Similarly, bimodal visual-tactile stimuli resulted in better performance than unimodal visual, $t(20) = 4.04$, $p < .001$, Cohen's $d = 0.88$ or tactile stimuli, $t(20) = 5.74$, $p < .001$, Cohen's $d = 1.25$. While bimodal audio-tactile stimuli yielded better performance than unimodal audio, $t(20) = 2.18$, $p < .05$, Cohen's $d = 0.47$, it was not significantly better than unimodal tactile, $t(20) = 1.02$, $p = .31$, Cohen's $d = 0.22$. *T* tests were then conducted to compare bimodal Rhythm Scores with its constituently best unimodal condition. Results showed that audio-visual stimuli yielded better performance measures than A,V (max; $t(20) = 2.19$, $p < .05$, Cohen's $d = 0.48$). Similarly, visual-tactile yielded better performance than V,T (max; $t(20) = 2.62$, $p < .05$, Cohen's $d = 0.57$). However, audio-tactile did not yield better performance than A,T (max; $t(20) = 0.86$, $p = .41$, Cohen's $d = 0.18$). Together, bimodal stimuli generally yielded greater performance than unimodal stimuli regardless of the stimulus type. However, some variability in the magnitude of the effect was observed, such that the addition of visual information tended to produce the greatest benefits.

Trimodal Versus Bimodal

To assess differences in performance between trimodal and its bimodal constituents, *t* tests were conducted. Assessment of the Standard Deviations (Figure 5A) showed that trimodal yielded better performance than audio-tactile, $t(20) = 4.46$, $p < .001$, Cohen's $d = 0.90$. While no significant difference was observed between trimodal and audio-visual, $t(20) = 1.56$, $p = .13$, Cohen's $d = 0.34$ and visual-tactile, $t(20) = 0.61$, $p = .54$, Cohen's $d = 0.13$. Comparison of the Rhythm Scores between the trimodal and bimodal stimuli showed that trimodal yielded better performance than audio-tactile, $t(20) = 3.21$, $p < .01$, Cohen's $d = 0.71$. No significant difference was observed between audio-visual, $t(20) = 0.19$, $p = .85$, Cohen's $d = 0.04$ and visual-tactile, $t(20) = 1.65$, $p = .113$, Cohen's $d = 0.36$.

Visual Tracking

Results demonstrated that the addition of visual stimuli yielded the greatest performance benefits of bimodal over unimodal stimuli, as well as trimodal over bimodal stimuli. Although this was an

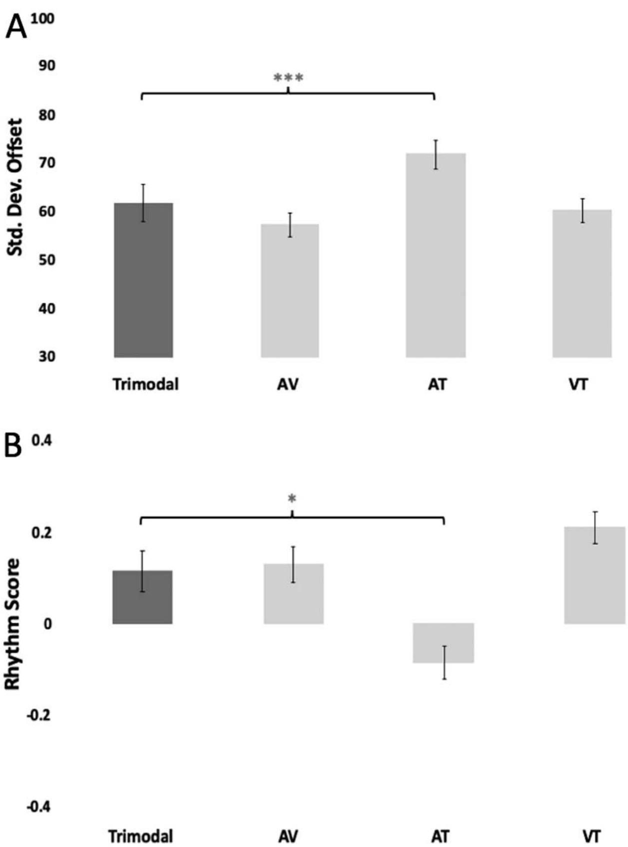

Figure 5. Difference in performance measures for (A) Standard Deviations and (B) Rhythm Score between trimodal and bimodal stimuli. AV = audio-visual; VT = visual-tactile; AT = audio-tactile. Error bars indicate standard error of the mean (*SEM*). * $p < .05$. *** $p < .001$.

unexpected result, it may be hypothesized that this could be a consequence of the continuous feedback that the moving ball provided. In other words, participants may have relied on visual tracking more heavily than the discrete timing information provided by the auditory or tactile stimuli. To understand the influence of visual tracking, we averaged together the Rhythm Scores from both the On-beat and Off-beat tasks, because each task allowed participants to continually track the visual stimulus while engaged in sensorimotor synchronization. Next, we used *t* tests to compare those results to the Rhythm Scores from the continuation task—where no visual tracking was available during sensorimotor synchronization. As in previous analyses, data was averaged across tempos. Results from this exploratory analysis showed that when visual tracking was possible (i.e., On-beat and Off-beat tasks with a visual stimulus), performance was better than when visual tracking was not possible (i.e., continuation task with a visual stimulus). These results are summarized in Figure 6 (audio-visual-tactile: $t(20) = 1.87$, $p = .07$, Cohen's $d = 0.40$; visual-tactile: $t(20) = 2.79$, $p < .01$, Cohen's $d = 0.60$; audio-visual: $t(20) = 5.48$, $p < .001$, Cohen's $d = 1.19$; visual-only: $t(20) = 4.79$, $p < .001$, Cohen's $d = 1.04$). Conversely, no such task differences were observed when the stimuli was tactile only, $t(20) = 1.025$, $p = .31$, Cohen's $d = 0.22$. Although task differences were observed with auditory-only stimuli, $t(20) = 2.45$, $p < .05$, Cohen's $d = 0.53$ and with audio-tactile stimuli, $t(20) = 5.37$, $p < .001$, Cohen's $d = 1.17$, performance was greater during the Continuation task, which is opposite of what was observed when visual stimuli are tracked. Together, these results indicate that visual tracking played a strong role in the overall rhythm performance and importantly, in multisensory integration.

Hypothesis Revisited

In the previous section, we demonstrated that visual tracking played an important role in the performance. Although our main focus is on assessing the effects of unimodal, bimodal, and trimodal stimuli on synchronization performance, we conducted a final analysis to confirm that the results are independent of task. Repeated measures ANOVAs were conducted on the Standard Deviations (minimum values) and Rhythm Scores (maximum values) separately, with factors Modality (best unimodal, best bimodal, and trimodal) and Task (Continuation, Average of On-Beat, and Off-Beat) as factors. Results from the Standard Deviation ANOVA showed main effects for Modality, $F(2, 40) = 7.98$, $p = .001$, $\eta^2 = 0.054$, $BF_{incl} = 259.8$ and Task, $F(2, 40) = 16.07$, $p < .001$, $\eta^2 = 0.09$, $BF_{incl} = 15.97$, but no interaction, $F(2, 40) = .33$, $p = .71$, $\eta^2 = 0.001$, $BF_{incl} = 0.702$). Similarly, analysis of Rhythm Scores indicated main effects of Modality, $F(2, 40) = 12.36$, $p < .001$, $\eta^2 = 0.12$, $BF_{incl} = 251.43$ and Task, $F(2, 40) = 3.69$, $p = .06$, $\eta^2 = 0.06$, $BF_{incl} = 15.75$, but no interaction, $F(2, 40) = 0.934$, $p = .401$, $\eta^2 = 0.006$, $BF_{incl} = 0.612$. Although our hypotheses were about the effects of multimodal integration that are generalizable across musical tasks and tempos, for reference, Table 1 summarizes all the data for each modality, task, and tempo.

Discussion

Contrary to our hypothesis, the addition of stimulus modalities does not continually benefit sensorimotor synchronization ability, such that trimodal stimuli does not necessarily yield an advantage over bimodal stimuli. Yet, our findings show that bimodal stimuli generally elicits better rhythmic performance than unimodal stimuli, which converges with previous results (Ammirante et al., 2016; Armstrong & Issartel, 2014; Bauer, Oostenveld, & Fries, 2009; Elliott et al., 2010; Kelso et al., 2001; Wing et al., 2010). Specifically, audio-visual and visual-tactile stimuli yielded better performance than their unimodal constituents.

It is interesting that performance with bimodal audio-tactile stimuli did not yield the predicted performance improvement beyond its unimodal constituents. Although we hypothesized a bimodal audio-tactile advantage, prior research using this stimulus combination has seen conflicting results with notable differences in the method of applying tactile stimuli. For example, performance with bimodal audio-tactile stimuli is better than with unimodal stimuli when the tactile stimuli is applied to the nondominant hand (Elliott et al., 2010; Kelso et al., 2001; Wing et al., 2010). Yet, when tactile stimuli is applied to the back, as done here, no bimodal advantage is observed (Ammirante et al., 2016), in line with our results. This discrepancy could reflect differences in how the brain integrates tactile information from different bodily regions. Using MLE models of multisensory integration, it

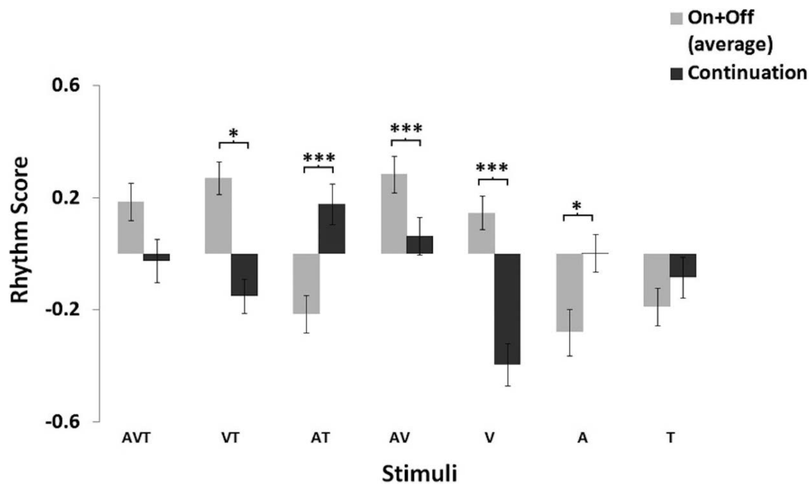

Figure 6. Comparison between Continuation task and On + Off (average) tasks. Error bars indicate standard error of the mean (*SEM*). A = auditory; V = video; T = tactile; AV = audio-visual; VT = video-tactile; AT = audio-tactile; AVT = audio-visual-tactile. * $p < .05$. *** $p < .001$.

has been suggested that if the processing noise associated with each of the two signals are highly correlated, then their combination would provide little to no advantage above either stimulus alone (Ammirante et al., 2016; Wing et al., 2010). Therefore, the implication would be that auditory and tactile stimuli on the back have highly correlated processing noise, which prevents bimodal stimuli from providing a performance gain. Conversely, auditory and tactile stimuli to a nondominant hand have less correlated processing noise, which enables bimodal stimuli to enhance performance. Although additional research will be needed to test the hypothesis that different regions of the body elicit differing levels of somatosensory noise during sensorimotor synchronization, it seems plausible given that the sensitivity of mechanoreceptors differ greatly between fingers and back (Verrillo, 1992). Yet, it remains to be seen why audio and tactile signals from the fingers would be less correlated than audio and tactile signals from the back.

Table 1
Top Row Numbers Indicate Rhythm Scores and SEM, Bottom Row Numbers Indicate Standard Deviation and SEM

Task	Unimodal	Bimodal	Trimodal
		SLOW	
ON	0.04 (.09)	0.40 (.06)	0.57 (.06)
	62.56 (3.86)	48.80 (2.53)	33.45 (2.03)
OFF	−0.44 (.07)	0.06 (.08)	−0.26 (.04)
	88.78 (6.19)	66.34 (4.55)	76.23 (17.21)
CONT	−0.51 (.08)	−0.32 (.06)	−0.54 (.12)
	85.72 (4.05)	81.27 (3.74)	82.04 (6.34)
		MEDIUM	
ON	0.35 (.05)	0.46 (.05)	0.59 (.05)
	47.89 (2.79)	43.52 (1.79)	42.55 (1.94)
OFF	−0.21 (.11)	−0.15 (.11)	0.17 (.13)
	84.69 (6.8)	75.38 (6.2)	72.04 (7.8)
CONT	−0.04 (.07)	0.13 (.06)	0.13 (.09)
	69.2 (3.12)	62.18 (2.9)	66.32 (3.19)
		FAST	
ON	0.33 (.06)	0.51 (.07)	0.49 (.06)
	53.21 (3.2)	4.87 (2.29)	41.22 (2.09)
OFF	−0.73 (.09)	−0.60 (.11)	−0.04 (.15)
	86.26 (3.11)	86.6 (4.55)	85.57 (5.47)
CONT	0.33 (.08)	0.27 (.07)	0.33 (.08)
	62.21 (0.08)	58.64 (.07)	57.73 (.08)

Note. CONT = continuation; ON = on beat; OFF = off beat.

Surprisingly, synchronization with trimodal stimuli was generally worse than the best performance with bimodal stimuli, and comparable with mean bimodal performance, suggesting diminishing returns with increasing sensory information from unimodal to bimodal to trimodal stimuli. Unfortunately, it is unclear why this is the case. One possibility is that previous experience with trimodal stimuli was a contributing factor. It is known that multisensory integration is an acquired ability that is not present at birth (Wallace & Stein, 2001). Therefore, it is possible that our participants do not have much exposure to trimodal rhythmic stimulation, which may limit their ability to integrate, and hence benefit from, trimodal stimuli. Future research should compare musicians to nonmusicians to ascertain whether musical experience may yield greater benefits from multisensory integration.

Another possibility (though not necessarily distinct) is that multisensory integration is a nonlinear process, so that the inclusion of additional stimuli does not additively benefit sensorimotor synchronization. In support of this, multisensory integration exhibits stochastic resonance (Lugo, Doti, & Faubert, 2008), which is an inherently nonlinear phenomenon. With stochastic resonance, an optimal level of noise exists, so that when a specific amount of noise is applied, it will lead to improved performance. However, too much or too little noise yields worse performance. Although it is unclear whether resonance could serve as a plausible physiological mechanism underlying the observed effects with our structured (i.e., nonstochastic) stimuli, at least two interesting parallels exist. First, stochastic noise is automatically integrated with structured sensory stimuli for enhanced performance (Lugo et al., 2008), just as structured bimodal or trimodal stimuli are automatically integrated to enhance performance (Bresciani, Dammeier, &

Ernst, 2006; Bresciani et al., 2008). Second, and more importantly, the current data indicates that there is an optimal level of sensory information for maximal sensorimotor synchronization performance, just as would be expected from a nonlinear resonant system. As can be seen in Figure 5, performance with trimodal stimuli can be better or worse than bimodal stimuli. One of the determining factors appears to be the amount of information that is provided, and it is largely contingent on the visual stimuli, which carries not only temporal information (as do auditory and tactile stimuli), but it also provides spatial information as well. Of note, the spatial and temporal information provided by the visual stimuli, coupled with additional temporal information from one other modality (i.e., bimodal audio-visual or bimodal visual-tactile) yields the best performance values (see Figure 4). Yet, the addition of a third modality (i.e., trimodal audio-visual-tactile) or the subtraction of a modality (i.e., unimodal stimuli), results in lowered performance. Thus, there appears to be an optimal level of information that can be used for sensorimotor synchronization performance. Additional research will be necessary to determine whether this is related to resonant phenomena. When assessing contributions from individual modalities, as hypothesized, the greatest bimodal and trimodal gains were from the modality with the best unimodal performance.

In rare occasions, the best unimodal performance was not part of the best bimodal performance. However, it was unexpected that the best performance occurred with visual stimuli. This advantage of visual information was limited to cases where continuous visual tracking was possible during performance. Therefore, the continuous timing information provided by the moving visual stimuli enabled enhanced sensorimotor synchronization ability and facilitated multisensory integration.

When comparing the effects of unimodal auditory and visual stimuli on sensorimotor synchronization performance, previous studies have demonstrated that auditory rhythmic sequences are reproduced more accurately than visual rhythmic sequences (Goldstone & Lhamon, 1972; Grondin, 1993; Stauffer, Haldemann, Troche, & Rammsayer, 2012). Similarly, better sensorimotor synchronization is observed with auditory more than visual stimuli (Repp, 2005; Repp & Su, 2013). When auditory and visual modalities convey conflicting temporal information, auditory dominance is usually observed (Burr, Silva, Cicchini, Banks, & Morrone, 2009; Fendrich & Corballis, 2001; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Yet, these studies utilized discrete (i.e., flashing) visual stimuli, whereas our study implemented continuous visual information.

As hypothesized, the modality that yielded the greatest unimodal synchronization performance (vision), also yielded the greatest bimodal synchronization performance. As can be seen in Figure 6, stimuli with visual information yielded better performance than stimuli without visual information—but only when visual tracking was possible (i.e., On-beat and Off-beat tasks). These results are in line with previous reports suggesting that continuous visual stimuli elicit comparable or better performance than discrete auditory stimuli (Armstrong & Issartel, 2014; Iversen et al., 2015). Experiments utilizing visual stimuli have found that the addition of spatial information; for example, a stimulus oscillating horizontally or vertically, improves synchronization compared with a stimulus containing only temporal information (Armstrong & Issartel, 2014; Buekers, Bogaerts, Swinnen, & Helsen, 2000; Hove, Spivey, & Krumhansl, 2010). As such, it is likely that participants in our study capitalized on the additional spatial information afforded by the visual modality to enhance performance. This advantage was not present when the spatial information was removed and participants had to rely on an internal representation of timing from the previously observed visual stimuli (i.e., Continuation task). Additional support for the notion that spatial information is the likely contributor to enhanced performance comes from research demonstrating that a continuously changing (but not moving) visual stimulus results in comparable performance to auditory stimuli (Varlet, Marin, Issartel, Schmidt, & Bardy, 2012). Together, these results support the conclusion that continuous visual stimuli, and the spatial information it imparts, enhances synchronization ability but not temporal perception (Silva & Castro, 2016).

In a separate study with a different group of participants, we recently used the same sensorimotor synchronization paradigm with auditory and visual stimuli (no tactile) and exhibited comparable performance between unimodal auditory and unimodal visual stimuli, even when visual tracking was possible during task performance (i.e., On-beat and Off-beat tasks; Zanto et al., 2019). Similarly, we previously observed comparable gains (i.e., lowered standard deviations) in performance with bimodal stimuli compared with unimodal auditory or visual stimuli. Therefore, it appears that the advantage of visual stimuli from this study is at odds with our previous study. However, this current study featured one critical difference—visual feedback during auditory only levels. In the previous study, no visual feedback was presented during auditory only. Therefore, we speculate it could be that the spatial-temporal visual feedback might have been a distractor and, therefore, affected the synchronization performance (Booth & Elliott, 2015). This may also explain why audio-tactile stimuli did not improve performance beyond its unimodal constituents. Additional research will be required to systematically manipulate how feedback from different modalities affect sensorimotor synchronization and multisensory integration.

As for why visual feedback was provided, it was an attempt to control for overall sensory feedback. During the task, taps could be felt by touching the screen, which provides sensory feedback as to when a tap was executed. Therefore, the difference between that tactile feedback on the finger and the sensory stimulation on the back provided participants with an online metric for tactile synchronization performance. Similarly, each tap produced a natural (not artificially generated) "thud" sound by striking the tablet, similar to the sound of a finger tapping on a table, which was audible despite the use of headphones. The difference between the tap sound and the auditory stimulus served as another online metric for participants to gauge performance. During pilot testing, we observed that participants would often position their fingers above the "strike" zone, which is the area that indicates when to synchronize with the visual stimulus. Unfortunately, the fingers could occlude the visual stimulus and make it difficult to judge the timing between the visual tap onset and the exact position of the visual stimulus. Therefore, a visual dashed line was inserted after each tap to provide visual feedback indicating when a tap occurred relative to when it should have occurred. This way, the dashed line served as a visual metric to assess performance in a similar way as the other forms of sensory feedback.

Previous research on sensorimotor synchronization handle sensory feedback in different ways (e.g., Armstrong & Issartel, 2014; Elliott et al., 2010; Kelso et al., 2001; Wing et al., 2010). Generally, visual stimuli is presented on a monitor that is placed above the participant's hands. As such, visual feedback from observing a button response is typically not available. In terms of auditory feedback from a button response, some will provide a masking noise, while others rely on headphones to attenuate the sound. In our study, the headphones were not sufficient to prevent hearing the tap sound, and we did not want to include noise because of the uncertainty of how such distraction could affect performance. These previous studies (and others) did not remove the tactile feedback from button responses, thereby creating some imbalance in sensory feedback. Although we attempted to equate sensory feedback across all three modalities, it is unclear what influence such feedback may have had on task performance. As noted above, additional research will be needed to systematically manipulate sensory feedback to ascertain its role in multimodal integration during sensorimotor synchronization.

Conclusion

Our experiment demonstrates that performance with bimodal stimuli is better than unimodal stimuli, but trimodal stimuli enables performance that is largely comparable with (sometimes better, sometimes worse than) performance with bimodal stimuli. Additionally, we show that synchronization benefits most from continuous spatial tracking provided by visual stimuli, especially when coupled with one modality that provides additional temporal information (i.e., auditory and/or tactile). Finally, our study contributes to the increasing research in multisensory integration by revealing that an increased amount of sensory information does not necessarily lead to improved performance in rhythmic synchronization. We suggest a potential role for experience-based benefits of multisensory integration, which may arise as a product of nonlinear interactions between modalities.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14,* 257–262. http://dx.doi.org/10.1016/j.cub.2004.01.029

Ammirante, P., Patel, A. D., & Russo, F. A. (2016). Synchronizing to auditory and tactile metronomes: A test of the auditory-motor enhancement hypothesis. *Psychonomic Bulletin & Review, 23,* 1882–1890. http://dx.doi.org/10.3758/s13423-016-1067-9

Armstrong, A., & Issartel, J. (2014). Sensorimotor synchronization with audio-visual stimuli: Limited multisensory integration. *Experimental Brain Research, 232,* 3453–3463. http://dx.doi.org/10.1007/s00221-014-4031-9

Bauer, M., Oostenveld, R., & Fries, P. (2009). Tactile stimulation accelerates behavioral responses to visual stimuli through enhancement of occipital gamma-band activity. *Vision Research, 49,* 931–942. http://dx.doi.org/10.1016/j.visres.2009.03.014

Booth, A. J., & Elliott, M. T. (2015). Early, but not late visual distractors affect movement synchronization to a temporal-spatial visual cue. *Frontiers in Psychology, 6,* 866. http://dx.doi.org/10.3389/fpsyg.2015.00866

Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision, 6,* 554–564. http://dx.doi.org/10.1167/6.5.2

Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2008). Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events. *Brain Research Bulletin, 75,* 753–760. http://dx.doi.org/10.1016/j.brainresbull.2008.01.009

Buekers, M. P., Bogaerts, H., Swinnen, S., & Helsen, W. (2000). The synchronization of human arm movements to external events. *Neuroscience Letters, 290,* 181–184. http://dx.doi.org/10.1016/S0304-3940(00)01350-1

Burr, D., Silva, O., Cicchini, G. M., Banks, M. S., & Morrone, M. C. (2009). Temporal mechanisms of multimodal binding. *Proceedings of the Royal Society B: Biological Sciences, 276,* 1761–1769.

Colonius, H., & Diederich, A. (2006). The race model inequality: Interpreting a geometric measure of the amount of violation. *Psychological Review, 113,* 148–154. http://dx.doi.org/10.1037/0033-295X.113.1.148

Deber, J., Araujo, B., Jota, R., Forlines, C., Leigh, D., Sanders, S., & Wigdor, D. (2016). Hammer time! A low-cost, high precision, high accuracy tool to measure the latency of touchscreen devices. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16; pp. 2857–2868). New York, NY: ACM. http://dx.doi.org/10.1145/2858036.2858394

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics, 66,* 1388–1404. http://dx.doi.org/10.3758/BF03195006

Dionne-Dostie, E., Paquette, N., Lassonde, M., & Gallagher, A. (2015). Multisensory integration and child neurodevelopment. *Brain Sciences, 5,* 32–57. http://dx.doi.org/10.3390/brainsci5010032

Elliott, M. T., Wing, A. M., & Welchman, A. E. (2010). Multisensory cues improve sensorimotor synchronisation. *European Journal of Neuroscience, 31,* 1828–1835. http://dx.doi.org/10.1111/j.1460-9568.2010.07205.x

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433. http://dx.doi.org/10.1038/415429a

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8,* 162–169. http://dx.doi.org/10.1016/j.tics.2004.02.002

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63,* 719–725. http://dx.doi.org/10.3758/BF03194432

Goldstone, S., & Lhamon, W. T. (1972). Auditory-visual differences in human temporal judgment. *Perceptual and Motor Skills, 34,* 623–633. http://dx.doi.org/10.2466/pms.1972.34.2.623

Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & Psychophysics, 54,* 383–394. http://dx.doi.org/10.3758/BF03205274

Hove, M. J., Spivey, M. J., & Krumhansl, C. L. (2010). Compatibility of motion facilitates visuomotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1525–1534. http://dx.doi.org/10.1037/a0019059

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. http://dx.doi.org/10.1016/j.cognition.2014.10.018

Kelso, J. A. S., Fink, P. W., DeLaplain, C. R., & Carson, R. G. (2001). Haptic information stabilizes and destabilizes coordination dynamics. *Proceedings Biological Sciences, 268,* 1207–1213. http://dx.doi.org/10.1098/rspb.2001.1620

Lugo, E., Doti, R., & Faubert, J. (2008). Ubiquitous crossmodal Stochastic Resonance in humans: Auditory noise facilitates tactile, visual and proprioceptive sensations. *PLoS ONE, 3,* e2860. http://dx.doi.org/10.1371/journal.pone.0002860

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17,* 154–163. http://dx.doi.org/10.1016/S0926-6410(03)00089-2

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,* 355–370. http://dx.doi.org/10.1080/00222890309603156

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12,* 969–992. http://dx.doi.org/10.3758/BF03206433

Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychological Research, 68,* 252–270. http://dx.doi.org/10.1007/s00426-003-0143-8

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review, 20,* 403–452. http://dx.doi.org/10.3758/s13423-012-0371-2

Silva, S., & Castro, S. L. (2016). Moving stimuli facilitate synchronization but not temporal perception. *Frontiers in Psychology, 7,* 1798. http://dx.doi.org/10.3389/fpsyg.2016.01798

Stauffer, C. C., Haldemann, J., Troche, S. J., & Rammsayer, T. H. (2012). Auditory and visual temporal sensitivity: Evidence for a hierarchical structure of modality-specific and modality-independent levels of temporal information processing. *Psychological Research, 76,* 20–31. http://dx.doi.org/10.1007/s00426-011-0333-8

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., . . . Wallace, M. T. (2014). Identifying and quantifying multisensory integration: A tutorial review. *Brain Topography, 27,* 707–730. http://dx.doi.org/10.1007/s10548-014-0365-7

Varlet, M., Marin, L., Issartel, J., Schmidt, R. C., & Bardy, B. G. (2012). Continuity of visual and auditory rhythms influences sensorimotor coordination. *PLoS ONE, 7,* e44082. http://dx.doi.org/10.1371/journal.pone.0044082

Verrillo, R. T. (1992). Vibration sensation in humans. *Music Perception, 9,* 281–302. http://dx.doi.org/10.2307/40285553

Wallace, M. T., & Stein, B. E. (2001). Sensory and multisensory responses in the newborn monkey superior colliculus. *The Journal of Neuroscience, 21,* 8886–8894. http://dx.doi.org/10.1523/JNEUROSCI.21-22-08886.2001

Wing, A. M., Doumas, M., & Welchman, A. E. (2010). Combining multisensory temporal information for movement synchronisation. *Experimental Brain Research, 200,* 277–282. http://dx.doi.org/10.1007/s00221-009-2134-5

Wozny, D. R., Beierholm, U. R., & Shams, L. (2008). Human trimodal perception follows optimal statistical inference. *Journal of Vision, 8*(3), 24. http://dx.doi.org/10.1167/8.3.24

Zanto, T. P., Padgaonkar, N. T., Nourishad, A., & Gazzaley, A. (2019). A tablet based assessment of rhythmic ability. *Frontiers in Psychology, 10,* 2471. http://dx.doi.org/10.3389/fpsyg.2019.02471

Received May 10, 2019
Revision received May 18, 2020
Accepted May 19, 2020 ■