Perceptual timing precision with vibrotactile, auditory, and multisensory stimuli

Rachel F. Sussman
Department of Psychology and Program in Neuroscience
Brandeis University, Waltham MA USA

Mercedes B. Villalonga
Volen Center for Complex Systems and Department of Psychology
Brandeis University, Waltham MA USA

Robert Sekuler
Volen Center for Complex Systems and Department of Psychology
Brandeis University, Waltham MA USA

Author Note

Corresponding author: Robert Sekuler, Volen Center for Complex Systems, Brandeis University, Waltham, MA, USA. Email: vision@brandeis.edu
Abstract

It is important to understand the perceptual limits on vibrotactile information-processing because of the increasing use of vibrotactile signals in common technologies like cell phones. To advance such an understanding, we examined vibrotactile temporal acuity and compared it to auditory and bimodal (synchronous vibrotactile and auditory) temporal acuity. In a pair of experiments, subjects experienced a series of empty intervals, demarcated by stimulus pulses from one of the three modalities. One trial contained up to 5 intervals, where the first intervals were isochronous at 400 ms, and the last interval varied from 400 by ±1-80 ms. If the final interval was < 400 ms, the last pulse seemed “early”, and if the final interval was > 400 ms, the last pulse seemed “late”.

In Experiment One, each trial contained four intervals, where the first three were isochronous. Subjects judged the timing of the last interval by describing the final pulse as either “early” or “late”. In Experiment Two, the number of isochronous intervals in a trial varied from one to four. Psychometric modeling revealed that vibrotactile temporal processing was less acute than auditory or bimodal temporal processing, and that auditory inputs dominated bimodal perception. Additionally, varying the number of isochronous intervals did not affect temporal sensitivity in either modality, suggesting the formation of memory traces. Overall, these results suggest that vibrotactile temporal processing is worse than auditory or bimodal temporal processing, which are similar. Also, subjects need no more than one isochronous reminder per trial for optimal performance.

*Keywords*: Temporal sensitivity, reminder task, auditory perception, vibrotactile perception, multisensory perception, memory trace
Perceptual timing precision with vibrotactile, auditory, and multisensory stimuli

Introduction

Our sensory systems are the source of speedy and reliable warnings when consequential changes occur in our surroundings. Many now-common technologies like cell phones piggyback on this important sensory function as a way to provide their own environment-altering notifications. One way to create salient signals is to modulate the timing of stimuli, for example, through generating rhythmic patterns of light, sound, or vibrations. Growing reliance on vibrotactile signaling devices makes it important to understand the capabilities and limitations of vibrotactile perception as a communication channel.

A thoughtful essay by von Békésy (1959) provides a valuable introduction to the skin’s advantages and disadvantages in communicating temporal information. Comparing the temporal responsiveness of audition and somatosensation, von Békésy wrote that “all the phenomena in which time patterns are involved can be discriminated by the skin only if the changes are slow relative to changes that the ear can recognize” (p. 6). Putting this comparison quantitatively, when discriminating the frequency of sinusoidal vibration applied to the skin, the Weber fraction is \( \approx 0.2-0.3 \), but with sinusoidal stimulation of the ear, the Weber fraction is approximately one hundred times better, just 0.003 (Saal, Wang, & Bensmaia, 2016).

How might the skin’s seemingly poor temporal sensitivity limit the use of vibrotactile stimuli as an alerting mechanism? Rather than detecting the temporal characteristics of the vibrations themselves, could one strength of this sensory modality instead lie in discriminating the timing of empty intervals demarcated by vibrotactile pulses? To answer this question, we measured human observers’ sensitivity to subtle changes in the timing of intervals between brief vibrotactile pulses, and compared that sensitivity to performance with sequences of brief auditory and bimodal (concurrent,
TACTILE, AUDITORY, AND BIMODAL TEMPORAL PRECISION

synchronous vibrotactile and auditory pulses.

**Experiment One**

In our first experiment, subjects received sequences of four empty intervals, defined by brief pulses of either vibrotactile, auditory, or bimodal stimulation. The first three intervals were isochronous, each 400 ms long. The duration of the fourth interval, however, was randomly varied to be either shorter or longer than 400 ms. Subjects judged that final interval, categorizing it as shorter or longer than its predecessors. This arrangement is shown schematically in Fig. 1.

The task was based on one used in a study of mismatch negativity, an

![Diagram of Experiment One](attachment:image.png)

**Figure 1.** Sample “early” (A) and “late” (B) trials. Stimulus pulses are represented by solid vertical lines. The time at which the 5th pulse would occur had it followed the preceding pulse by 400 ms is depicted by a dashed vertical line. The variables x and y represent different amounts of change from 400 ms, x for an “early” trial (interval < 400 ms) and y for a “late” trial (interval > 400 ms).
event-related potential generated by an atypical or unexpected event (Näätänen, Paavilainen, Rinne, & Alho, 2007). To generate such a potential, Lumaca, Trusbak Haumann, Brattico, Grube, and Vuust (2019) introduced unexpected temporal changes into a sequence of regular auditory pulses. We modified that approach in several key ways, most notably by capturing subjects’ behavioral responses rather than event-related potentials, and by assessing the effects of a wide range of timing changes, not just one or two large changes.

**Methods**

**Subjects.** Twenty-one subjects (13 female; mean age = 20.8 years, SD = 2.1 years) participated in this experiment. Earlier work from our laboratory showed the sufficiency of such sample sizes (Bushmakin & Sekuler, 2016). Subjects provided written informed consent to experimental procedures that had been approved by Brandeis University’s Institutional Review Board. All work was conducted in accordance with the Declaration of Helsinki. All subjects denied having difficulty in both hearing and sensing vibrations on the skin.

**Apparatus and stimuli.** Stimulus presentation and response collection were controlled by a Macintosh computer running MATLAB and extensions from the Psychophysics Toolbox (Brainard, 1997). These worked in tandem with an Arduino UNO microprocessor that directly controlled the stimulus generation devices (Fig. 2). Auditory (hereafter, A) stimulation was produced by a piezo active buzzer, which sounded at -70 dB. Tactile (hereafter, T) stimulation was produced by a linear resonant actuator vibrotactor (Engineering Acoustics, Inc., Casselberry, FL).

The vibrotactor’s active element was 7.87 mm in diameter, and produced vibrations at 250 Hz with < 2 ms rise time and 0.5 mm peak displacement. Vibrations were delivered to the ventral surface of the subject’s left index finger, which rested on the vibrotactor’s active element. To enforce a constant position of the finger, the hand was held within a close-fitting custom 3D-printed container tailored to individual subject’s
Figure 2. A test subject whose left hand is nestled within a 3D-printed support that positions the left index finger on the vibrotactor’s active element. Subjects used their right hands to communicate their binary judgments on the computer keyboard. In the photo, the computer is displaying one of the experiment’s instruction screens. The Arduino microprocessor and associated circuitry are enclosed in the purple 3D-printed box. Photo used with permission.

hands. This container also included insulation for the vibrotactor, to prevent unexpected auditory emissions resulting from vibrations against the plastic. Bimodal auditory-vibrotactile (hereafter, AT) stimuli were created by activating the vibrotactor and the buzzer simultaneously, with a negligible (< 10 µs) delay between their onsets.

Stimulus intensity calibration. Throughout the experiment, the intensities of vibrotactile and auditory stimuli were held constant at levels well above threshold. Despite the insulation provided by its 3D-printed container, the vibrotactor produced an audible, low-frequency sound when activated. To prevent this sound from contaminating the conditions in which the vibrotactor was used, it was masked by continuous Brownian
noise whose level was customized for each subject using a standard calibration procedure as described below. Subjects heard the continuous noise with supraural headphones, and, at subjects' option, also wore in-ear, sound-absorbing earplugs.

The calibration procedure used a modified method of limits to find a level of Brownian noise that completely masked the sound of the vibrotactor and still allowed the piezo buzzer to be heard. At each level of masking noise, five A stimuli were presented with the buzzer; the subject then reported how many pulses they heard. Brownian noise amplitude was decreased following incorrect responses. This procedure allowed us to find the maximal level of noise through which subjects could still hear the A stimuli. With this level of noise, we repeated the procedure with T stimuli, increasing noise amplitude if subjects verbally noted they could hear the sound of the vibrotactor. This iterative process produced a level of noise that allowed the subject to hear A stimuli but not the sound generated by the vibrotactor.

**Experimental task.** On each trial, subjects received five 50 ms-pulses of A, T, or AT stimulation. These five pulses marked off four empty intervals, each defined as the time from the offset of one pulse to the onset of the next. The first three intervals were isochronous, at 400 ms each. We refer to these as “reminder” intervals because they were meant to remind the subject of the duration of the isochronous interval (Ulrich, 2010; Rammsayer, 2014). However, the fourth interval (before the fifth and final pulse) was equally likely to be shorter or longer than 400 ms. The difference between the duration of the final interval and the duration of each isochronous reminder interval defined what we will call the trial’s **rhythmic deviancy**. If the fourth interval had a **negative rhythmic deviancy** (shorter than 400 ms; Fig. 1A), the last pulse seemed “early”, and if the fourth interval had a **positive rhythmic deviancy** (longer than 400 ms; Fig. 1B), the last pulse seemed “late”. Correspondingly, the subject’s task was to characterize the arrival of the fifth pulse as either “early” or “late” compared to its predecessors.

Preliminary testing confirmed that a range of absolute rhythmic deviances of 1 to
80 ms produced response accuracies which ranged from chance to ceiling. So, rhythmic
deviances for the experiment were drawn from two random uniform distributions:
320-399 ms ("early" trials) and 401-480 ms ("late" trials). To facilitate our planned
analysis of the data, each of the two ranges was subdivided into four time bins, each
spanning 20 ms (e.g., 320 - 339 ms). In total, for each modality, a subject received 20
trials with rhythmic deviancies drawn randomly and equally from each of the eight (four
"early" and four "late") bins. This yielded a total of 480 trials per subject (8 [rhythmic
deviancy bins] × 20 [trials per bin] × 3 [modality conditions]).

At the start of a trial, a white fixation cross appeared at the center of a black
computer screen and remained visible for 500 ms. Then, the stimulus sequence began.
After the fifth pulse, subjects signaled their responses with their right hands, using the “1”
and “2” keys on a numeric computer keyboard. The mapping of the keys onto response
categories ("early" and "late") was counterbalanced across subjects. Following each
response, immediate feedback (correct or incorrect) was displayed on the screen for
1000 ms, along with that subject’s average accuracy so far during that block of trials. So
that subjects could monitor their performance throughout the block, the displayed
accuracy was updated after every trial. Subjects had up to 1000 ms to enter their
responses; if they failed to respond within 1000 ms, the trial was categorized as incorrect
for the purposes of subject feedback, and was subsequently excluded from further
analysis.

During the consent process, subjects were informed that they would receive
$12.00 (USD) for participation, with a bonus of $0.25 for each block on which their
accuracy exceeded 80%. Subjects then completed the calibration procedure described
above, which was followed by an instruction phase. During the instruction phase,
subjects first experienced 2 example trials (1 “early” and 1 “late”), and then completed 6
practice trials (3 “early” and 3 “late”) for each modality (3 rounds total). The absolute
rhythmic deviancy on all pre-experimental trials was 60-80 ms. All trials in the instruction
TACTILE, AUDITORY, AND BIMODAL TEMPORAL PRECISION

Phase followed the general trial structure described above.

Following the instruction phase, each subject completed 12 blocks of 40 trials per block. The experiment followed a block-randomized design, with all experimental trials in a single block devoted exclusively to one modality. Subjects were permitted ad lib breaks between blocks. During breaks, an on-screen text display informed the subject which modality would be tested in next block of trials. Subjects completed the entire experiment, including instructions and practice, within ~45-60 minutes.

Data curation. First, we checked for response bias in each condition by assessing the number of trials without a response. Overall, subjects failed to respond before the 1000 ms maximum response time on just 0.8% of trials. The number of trials that subjects missed was unrelated to stimulus modality (p = 0.17), experiment counterbalancing (p = 0.884), or trial type (“early” or “late”, p = 0.358). The subsequent analysis excluded trials on which a response was missed.

Psychometric modeling. Psychometric functions (PFs) provided the main analytical method for characterizing subjects’ temporal sensitivity. Each PF modeled the likelihood that a subject would categorize a trial’s final pulse as “late” given the trial’s rhythmic deviancy value and its modality. Hierarchical Bayesian PFs were fit using JAGS (Plummer et al., 2003), as implemented in the Palamedes toolbox for MATLAB (Prins & Kingdom, 2018). We used uninformative, flat priors for all model parameters, with one chain, 5000 (group-level) / 1000 (subject-level) samples per chain, and 2000 discarded samples per parameter. The slope parameter estimated by each model indexed temporal sensitivity. When we fit a group-level model, data were aggregated across all subjects. This was done in addition to fitting subject-level models, where each subject received an individual fit.

To test for differences in temporal sensitivity between modalities, we used the group-level posteriors to obtain a distribution of differences between each pair of modalities. That is, a difference distribution was generated for each comparison, A vs. T,
TACTILE, AUDITORY, AND BIMODAL TEMPORAL PRECISION

\( \text{AT vs. T, and A vs. AT} \). We then generated summary statistics on these posterior difference distributions to determine whether each difference distribution differed significantly from 0. Subject-level PFs were used in the generation and analysis of just-noticeable difference (JND) estimates. For each subject, we defined the “early” and “late” JND as the rhythmic deviancy levels corresponding to the 0.25 and 0.75 points on each subject’s psychometric function, respectively. The resulting pairs of JND values for each modality were then averaged across subjects. Paired comparisons of the average JND among modalities were performed using paired Bayesian \( t \)-tests implemented using R’s BayesFactor package (Morey et al., 2018). These tests yield Bayes Factors (BFs), which represent how many times more likely the alternate hypothesis (that a true difference exists) is compared to the null hypothesis (that no difference exists). Customarily, BF values \( \geq 3 \) are taken as reliable evidence in favor of the alternate hypothesis (Jeffreys, 1961).

**Cue combination.** For a quantitative account of how subjects combined unimodal (A and T) cues when judging AT stimuli, we evaluated a pair of classical models (To, Baddeley, Troscianko, & Tolhurst, 2011). Specifically, we tested whether the combination of unimodal cues followed linear addition (both cues weighted equally) or a MAX metric (strong dominance by one of the two inputs). We also tested whether audio-tactile cue combination in our task mirrored the mode of audio-visual cue combination reported by To et al. (2011), and we also analytically optimized the cue combination metric for each subject in each condition. We report summed squared error (SSE) values as an estimate of model fit.

**Results**

Analysis of group-level PFs (Fig. 3A) revealed that subjects were least sensitive to rhythmic deviancy in the T condition. Estimated T slopes were significantly lower than slopes in both the A \( (p < 0.001, 95\% \text{ CI} = [0.155, 0.468]) \) and AT conditions \( (p < 0.001, 95\% \text{ CI} = [0.276, 0.641]) \). Sensitivity to rhythmic deviancy was similar in A and AT.
conditions, and PF slopes for the A and AT conditions did not significantly differ, $p = 0.053$, 95% CI = [-0.034, 0.337].

**Figure 3.** Experiment One: Psychometric modeling. (A) Group-level psychometric functions fit via Bayesian estimation for the three modalities show that subjects were least sensitive to rhythmic deviancy in the T condition (red). The intersections of each curve with the 25% and 75% choice proportion lines (grey, dashed) represent the group-level “early” and “late” JNDs, respectively, for each modality. (B) Boxplots of subject-level JND estimates for the three modalities depict differences in JND among modalities. T JNDs were found to be significantly larger than AT JNDs (comparison 1, BF > 100) and A JNDs (comparison 2, BF > 100). AT JNDs were not different from A JNDs (comparison 3, BF = 0.535). (C, D) Representative subject-level psychometric functions showing good and poor model fit, respectively.

We obtained subject-level JNDs from PFs fit to individual subjects. The best and worst of these model fits are shown in Figs. 3C & 3D, respectively. Averaging across
For Review Only

TACTILE, AUDITORY, AND BIMODAL TEMPORAL PRECISION

“early” and “late” trials (Fig. 3B), the JND for A trials was 19.32 ms (SD = 4.90 ms), for T trials was 27.45 ms (SD = 5.80 ms), and for AT trials was 18.03 ms (SD = 1.05 ms). At the subject level, A JNDs were similar to AT JNDs (BF = 0.535). Both A and AT JNDs were smaller than T JNDs (BF > 100 for both comparisons). In other words, as shown in Fig. 3B, subjects were able to detect smaller changes more reliably in the duration of the final interval on A and AT trials than they were on T trials. Weber fractions for A, AT, and T conditions were 0.048, 0.045, and 0.069, respectively. All results suggest that subjects are least sensitive to rhythmic deviancy with unimodal T signals, and more sensitive to rhythmic deviancy with either A or AT stimuli.

Cue combination was used to assess the relative contributions of each unimodal input to bimodal perception. We determined how subjects combined unimodal (A and T) cues by finding the parameter $m$ from the following equation, as described in To et al. (2011):

$$ Slope_{AT} = (Slope_A^m + Slope_T^m)^{1/m} $$

where $m = 1$ reflects linear addition and $m = \infty$ represents the MAX rule, or dominance by one input alone. Additionally, $m = 2.56$ was found to be the optimized summation of bimodal auditory and visual stimuli in another discrimination task, so it was tested as well (To et al., 2011).

We found that AT slopes were best predicted from individual A and T slopes using a MAX metric, mean SSE < 0.001. The MAX rule outperformed both linear addition (SSE = 0.915) and $m = 2.56$ (SSE = 0.473). When the cue combination parameter was optimized, it returned a MAX metric for every subject. This result suggested that one of the two unimodal inputs strongly dominated the bimodal condition. The slope and JND comparisons among modalities, together with the similarity between A and AT conditions, points to A signals as the dominant input in the bimodal condition.
Experiment Two

Experiment One required subjects to judge the duration of a trial's final interval relative to its isochronous predecessors. To make these comparisons, subjects must have formed some internal representation of the isochronous ("reminder") interval (i.e., 400 ms) against which they could judge that final interval. We cannot tell, though, what information subjects actually used to construct that representation. For example, did subjects incorporate each and every one of the preceding intervals into an internal representation against which to compare the final interval, or just some of them?

It is well known that multiple exposures to a stimulus promotes formation of a memory trace, or perceptual memory, of that stimulus (Drake & Botte, 1993), and it is known that such a trace can persist over many trials (Dyjas, Bausenhart, & Ulrich, 2012; Tong, Dubé, & Sekuler, 2019). That persistent trace could have allowed subjects in Experiment One to form a robust 400-ms trace from trials early in the experiment, rendering reminder intervals on subsequent trials superfluous. Subjects could then use that persistent trace as a basis for their judgments, rather than, or in addition to, using the multiple reminders presented on each trial (e.g., Morgan, Watamaniuk, & McKee, 2000; Nachmias, 2006). If something like this were the case, the number of reminder intervals on a trial would be inconsequential. However, there is some evidence that repetitions of reminder stimuli do aid subjects' performance (e.g., Grondin & McAuley, 2009). That result could lend itself to a simple statistical explanation: just as additional samples make estimates of a distribution more precise, additional samples of reminder stimuli can sharpen subjects' representation of that stimulus. If this explanation applied to Experiment One, as additional reminder intervals were presented on each trial, the subject's sensitivity to the rhythmic deviancy of the final interval would increase.

These alternative, conflicting views led us to conduct a second experiment in which we varied the number of reminder intervals that preceded the final test interval. If subjects utilized all of the reminder repetitions in their judgments, performance would
increase as additional isochronous intervals were presented before the comparison interval. Alternatively, if subjects instead developed and based their judgments on some persistent memory trace of the reference interval, then the number of isochronous intervals presented before the comparison interval in a trial should not impact performance.

Methods

Except where explicitly noted, the methods were the same as in Experiment One.

Subjects. Fifteen new subjects (8 female, mean age = 19.33 years, SD = 1.25 years) participated in this experiment. No subject reported issues related to processing auditory or vibrotactile signals. Experimental procedures were approved by Brandeis University’s Institutional Review Board and were conducted in accordance with the Declaration of Helsinki. All subjects provided written informed consent prior to participation.

Apparatus and stimuli. With the exception of some additional sound-absorbing insulation inserted beneath the vibrotactor, the apparatus was unchanged from Experiment One. Note that this additional sound-absorbing material did not affect the intensity of T stimulation. Also, the strong similarity between performance with AT and A stimuli in Experiment One led us to drop the bimodal condition here.

Stimulus intensity calibration. Although we had no reason to suspect that subjects could hear the T stimulus in Experiment One (our PF analysis revealed significant differences between A PFs and T PFs), we made our calibration procedures more rigorous for Experiment Two. We used a 2AFC task to find an optimal level of Brownian noise needed to mask the auditory sounds of the vibrotactor for each subject. Subjects did not touch the vibrotactor during the calibration task. Subjects were told that the vibrotactor would emit an auditory artifact, and were instructed to listen for any such artifact on the calibration trials. Each trial consisted of two components presented in random order. During one component, the vibrotactor was actively vibrating, while in the
other component, it was inactive. Subjects responded by indicating which trial component, the first or the second, contained the auditory artifact. Accuracy \( \geq 50\% \) constituted chance performance, and was taken as an indication that the subject could not hear the sounds produced by the vibrotactor. Each block contained 24 trials. After each block, until accuracy fell in the middle 90% of a binomial distribution, the amplitude of the masking noise was increased and another block was presented.

**Task.** We modified Experiment One’s task by varying the number of reminder (isochronous) intervals delivered on each trial. Now, any number between one and four reminder intervals could precede the final test interval.

This experiment also used a block randomized design. In any block, all trials used stimuli from the same modality, and all trials in the block employed the same number of reminder intervals (either one, two, three, or four). At the beginning of each block, subjects completed six practice trials with that block’s modality and number of reminder intervals. To make practice judgments relatively easy, practice trials had a rhythmic deviancy \( \pm 60 \text{ ms} \). After the practice trials, subjects completed 48 test trials per block with randomized rhythmic deviancy values, as in Experiment One.

Each testing session lasted around \( \sim 1.5 \text{ hours} \), and consisted of 16 blocks. This made a total of 768 trials per subject (12 [trials of a rhythmic deviancy bin] \( \times 8 \) [rhythmic deviancy bins] \( \times 2 \) [modality condition] \( \times 4 \) [interval number conditions]). As in Experiment One, subjects were allowed short, self-timed breaks between blocks. We also used the same feedback structure as in Experiment One, except now, subjects were awarded credits towards a university requirement for their participation, and could earn additional credits with accuracy scores of 80% on 8 or more of the 16 blocks.

**Data curation.** Except where noted, we used the same analytic procedures as in Experiment One. Although 15 subjects started the experiment, data from two subjects was excluded from analysis: one subject withdrew before completing the experiment, and data from another were lost because of a computer error. Like Experiment One,
subjects failed to respond before the 1000 ms maximum response time on 0.9% of all trials; these trials were excluded from analysis. Trials with missing responses were not disproportionately represented by modality ($p = 0.293$), number of reminder intervals ($p = 0.769$), counterbalancing condition ($p = 0.52$), or whether the trial was “early” or “late” ($p > 0.99$).

**Psychometric modeling.** Four psychometric models were fit: two models comparing group- and subject-level modality trends, using data aggregated across all reminder interval conditions, and two models comparing group- and subject-level reminder interval trends, using data aggregated across the modality conditions. Estimated group- and subject-level model parameters had the same analytical purposes as in Experiment One.

**Results**

Replicating a result from Experiment One, at the group level and aggregating across all reminder interval conditions, the posterior slope distribution for the T modality fell significantly below that for the A modality, $p = 0.002$, 95% CI $= [0.116, 0.464]$ (Fig. 4A). This result confirms what we previously observed: subjects were less sensitive to the timing of vibrotactile pulses than to the timing of auditory pulses. Additionally, the A JNDs (21.26, SD = 8.45) were smaller than the T JNDs (35.25, SD = 16.19), BF = 77.5 (Fig. 4B).

As can be seen in Fig. 5, averaged over modalities, the number of reminder intervals did not significantly affect temporal sensitivity. Specifically, all comparisons of posterior slope distributions were nonsignificant, and no JND comparison yielded a BF $\geq 3$. These results suggest that the number of reminder intervals did not affect temporal sensitivity.

**Discussion**

**Review of main findings**

Our two experiments revealed three key findings:
Figure 4. Experiment Two: Psychometric model fits by modality. (A) Group-level psychometric functions fit via Bayesian estimation show that, like Experiment One, T temporal sensitivity was worse than A temporal sensitivity, \( p = 0.002 \). (B) Boxplots of JNDs derived from hierarchical subject-level models show that auditory JNDs were smaller than vibrotactile JNDs (BF = 77.5), confirming this result.

1. Processing of temporal information carried by vibrotactile stimuli is more precise than previous research suggested.

2. Auditory temporal processing is more precise than its vibrotactile counterpart, and bimodal processing seems to be dominated by auditory signals.

3. Presenting additional reminder intervals does not improve the precision of temporal processing.
Figure 5. Experiment Two: Psychometric model fits by number of reminder intervals. (A) Group-level psychometric functions fit via Bayesian estimation show that the number of reminder intervals presented to subjects on each trial did not affect temporal sensitivity. (B) Boxplots of JNDs derived from hierarchical subject-level models show that no significant differences exist among the JNDs from the four conditions, confirming this result.

The next few sections discuss each of these three findings in turn.

**Processing of temporal information carried by vibrotactile stimuli is more precise than previous research suggested**

Previous comparisons of vibrotactile and auditory frequency discrimination showed that vibrotactile frequency discrimination is as much as $100 \times$ poorer than that for auditory stimuli (Saal et al., 2016). This large disadvantage in vibrotactile temporal
processing (compared to auditory temporal processing) differs considerably from the comparatively modest disadvantage seen in our experiments. For example, in Experiment One the auditory Weber fraction was 0.048, while the vibrotactile Weber fraction was 0.069, less than twofold difference, differing drastically from the estimates reported by Saal et al.. What might account for this substantial discrepancy between previous results and what we found? One possibility is that the discrepancy was caused, at least in part, by the characteristics of our stimuli and experimental task.

In addition to lower-level sensory judgments, participants utilized higher cognitive functions like memory to complete this task, which aided their performance. Previous reports of a substantial vibrotactile disadvantage resulted from tasks with judgments based only on basic stimulus properties, where the stimuli would have been processed at early stages in the respective sensory systems (Saal et al., 2016). While the judgments in our task also utilized these mechanisms, they also required memory and other higher-order cognitive functions (Romo & Rossi-Pool, 2020). That memory had a central role in our task is reinforced by the results of Experiment Two, which suggested that the memory of events' timing was sufficiently robust to persist over trials. Previous research has shown that temporal regularity promotes predictive processing and builds strong expectations for upcoming events (de Lange, Heilbron, & Kok, 2018; Lumaca et al., 2019), and these expectations in turn promote temporal sensory discrimination (Rajendran, Harper, Abdel-Latif, & Jan, 2016; Lawrance, Harper, Cookie, & Schnupp, 2014). So, engaging memory processes could have facilitated change detection in our task, beyond information that would have been available from early somatosensory processing, allowing for better vibrotactile discrimination than found in other experiments.

**Auditory temporal processing is more precise than its vibrotactile counterpart, and bimodal processing seems to be dominated by auditory signals**

As expected from previous research, short unfilled intervals bounded by auditory stimuli are judged with greater precision than empty intervals bounded by vibrotactile
stimuli (*e.g.* Rammsayer, 2014). This difference between hearing and touch could have arisen from any number of factors, for example from underlying differences in the temporal responsiveness of sensory organs and central somatosensory processing regions (Recanzone, Merzenich, & Schreiner, 1992; Recanzone, Schreiner, & Merzenich, 1993).

It is well-understood that the spatial and temporal characteristics of different modalities are optimized for different challenges, an idea called the modality appropriateness hypothesis (Welch & Warren, 1980; Welch & Duttonhurt, 1986). For example, vision is purportedly specialized for processing spatial information, while audition is similarly specialized for temporal processing (*e.g.* Repp & Penel, 2002). Evidence from this and other studies from our lab suggests that when interval judgments are at stake, somatosensory processing approximates and in some cases surpasses that of vision (Villalonga, Sussman, & Sekuler, 2020). Taken together with the results of this study, these findings suggest that temporal acuity in the vibrotactile domain may fall between what vision and audition can each support.

Interestingly, we found that only one of the unimodal inputs strongly dominates responses to bimodal stimulation. Others have found different effects; namely, bimodal processing reflects some degree of integration of *both* unimodal inputs, which allows performance to surpass what would have been achieved with either modality alone (Stevenson, Wilson, Powers, & Wallace, 2013). However, the outcome can vary when two unimodal contributors to some bimodal mixture differ considerably in their precision (Ernst & Banks, 2002; Repp & Penel, 2002). Our results are consistent with the latter outcome: as unimodal auditory processing was considerably more precise than tactile processing, the dominant input to the bimodal condition was audition. Future research with auditory and vibrotactile stimuli could manipulate the variance of pulse timing within each modality, which presumably would encourage heavier reliance on information from whichever modality had less temporal variance, and see if under certain conditions,
vibrotactile sensation could dominate bimodal inputs instead.

**Presenting additional reminder intervals does not improve the precision of temporal processing**

Experiment Two showed that the number of isochronous (400 ms) reminder intervals did not seem to affect subjects’ performance, consistent with previous results from auditory and visual studies (Miller & McAuley, 2005; Grondin & McAuley, 2009). As we previously discussed, reminder tasks with a constant standard interval duration allow for the creation of a memory trace, leading to expectations about the timing of future events (Lumaca et al., 2019). Importantly, we saw this null result in both the auditory and vibrotactile conditions. If our results are evidence that subjects exploit a memory trace, our experiment extends that point from what was already known about vision and audition to the case of vibrotactile stimulation, replicating results from Kang, Lancelin, and Pressnitzer (2018).

**Conclusions**

The experiments presented here measured and compared auditory and vibrotactile temporal sensitivity, and explored the impact of repeated standard intervals in a reminder task. We found that temporal sensitivity was lower for vibrotactile than for auditory signals, with little difference between the auditory and bimodal conditions. Additionally, the number of reminder intervals had no effect on temporal sensitivity, pointing to the contribution of a robust memory trace in both the auditory and vibrotactile conditions.

Auditory and vibrotactile cues are ubiquitous in everyday life, for example in cell phones that deliver notifications to users by means of tones or vibrations, either separately or simultaneously. Our results show that, if the rhythmic nature of the notification were varied in order to communicate specific information, such auditory notifications would be processed more accurately than vibrotactile ones, and that an alert involving both auditory and vibrotactile cues may be no better than alerts using auditory cues alone. Future experiments would do well to explore the neural correlates
of temporal processing in the vibrotactile and auditory sensory systems.

**Data Availability**

Datasets generated during and analyzed during the current study are available in the Open Science Framework (OSF) repository, found at https://osf.io/q8y6j/?view_only=76df6923d10a498b84ec44883a75b525.

**Acknowledgments**

RFS was supported by the NIH grant R90 DA033463 for Undergraduate and Graduate Training in Computational Neuroscience. MBV was supported by the NIGMS grant T32 GM132498. Adam Fleishaker and Jacob Smith provided coding support. The content is solely the responsibility of the authors and does not represent the official views of any funding agency. The authors declare they have no conflicts of interest.
References


Figure 1. Sample "early" (A) and "late" (B) trials. Stimulus pulses are represented by solid vertical lines. The time at which the 5th pulse would occur had it followed the preceding pulse by 400 ms is depicted by a dashed vertical line. The variables x and y represent different amounts of change from 400 ms, x for an "early" trial (interval < 400 ms) and y for a "late" trial (interval > 400 ms).
Figure 2. A test subject whose left hand is nestled within a 3D-printed support that positions the left index finger on the vibrotactor’s active element. Subjects used their right hands to communicate their binary judgments on the computer keyboard. In the photo, the computer is displaying one of the experiment’s instruction screens. The Arduino microprocessor and associated circuitry are enclosed in the purple 3D-printed box. Photo used with permission.
Figure 3. Experiment One: Psychometric modeling. (A) Group-level psychometric functions fit via Bayesian estimation for the three modalities show that subjects were least sensitive to rhythmic deviancy in the T condition (red). The intersections of each curve with the 25% and 75% choice proportion lines (grey, dashed) represent the group-level “early” and “late” JNDs, respectively, for each modality. (B) Boxplots of subject-level JND estimates for the three modalities depict differences in JND among modalities. T JNDs were found to be significantly larger than AT JNDs (comparison 1, BF > 100) and A JNDs (comparison 2, BF > 100). AT JNDs were not different from A JNDs (comparison 3, BF = 0.535). (C, D) Representative subject-level psychometric functions showing good and poor model fit, respectively.