Feeling the Beat (and Seeing It, Too): Vibrotactile, Visual, and Bimodal Rate Discrimination

Mercedes B. Villalonga^{1,*}, Rachel F. Sussman² and Robert Sekuler²

¹ Department of Psychology, Brandeis University, Waltham, MA, USA

² Volen National Center for Complex Systems, Brandeis University, Waltham, MA, USA

Received 16 January 2019; accepted 9 July 2019

Abstract

Beats are among the basic units of perceptual experience. Produced by regular, intermittent stimulation, beats are most commonly associated with audition, but the experience of a beat can result from stimulation in other modalities as well. We studied the robustness of visual, vibrotactile, and bimodal signals as sources of beat perception. Subjects attempted to discriminate between pulse trains delivered at 3 Hz or at 6 Hz. To investigate signal robustness, we intentionally degraded signals on two-thirds of the trials using temporal-domain noise. On these trials, inter-pulse intervals (IPIs) were stochastic, perturbed independently from the nominal IPI by random samples from zero-mean Gaussian distributions with different variances. These perturbations produced directional changes in the IPIs, which either increased or decreased the likelihood of confusing the two pulse rates. In addition to affording an assay of signal robustness, this paradigm made it possible to gauge how subjects' judgments were influenced by successive IPIs. Logistic regression revealed a strong primacy effect: subjects' decisions were disproportionately influenced by a trial's initial IPIs. Response times and parameter estimates from drift-diffusion modeling showed that information accumulates more rapidly with bimodal stimulation than with either unimodal stimulus alone. Analysis of error rates within each condition suggested consistently optimal decision making, even with increased IPI variability. Finally, beat information delivered by vibrotactile signals proved just as robust as information conveyed by visual signals, confirming vibrotactile stimulation's potential as a communication channel.

Keywords

Beat perception, vibrotactile stimulation, temporal information, multisensory stimulation, decision making, drift-diffusion modeling, decision criterion

* To whom correspondence should be addressed. E-mail: vision@brandeis.edu

1. Introduction

When a brief stimulus repeats periodically on a short time scale, its repetition can generate perceptual experiences known as beats. Although the experience of a beat is most commonly associated with repetition of musical sounds, a comparable experience can also be reliably generated by repeated presentation of brief visual stimuli (Grahn, 2012; Guttman *et al.*, 2005; Patel *et al.*, 2005). However, beat perception with modalities other than vision and audition has attracted little attention. For example, with a few notable exceptions (e.g., Frings and Spence, 2010, 2011; Meng *et al.*, 2015), researchers have ignored beat perception with vibrotactile stimulation. This neglect is somewhat surprising. After all, the human skin, a target for vibrotactile stimulation, is the largest of our sensory receptor surfaces, $\sim 1.7 \text{ m}^2$ in average adults (Bender and Bender, 1999). As a result, vibrotactile beats afford a potentially valuable communication channel.

Although vibrotactile signals are widely used in cellphones and vehicles, the ability of such signals to transmit information more nuanced than simple alerts remains to be examined. Additionally, little is known about the effect of vibrotactile stimulation when paired with stimulation from one or more other sensory modalities. For example, although rate perception is enhanced when auditory and visual stimuli are paired (Levitin *et al.*, 2015; Maddox *et al.*, 2015; Recanzone, 2003), would a comparable benefit arise from a stimulus pairing that includes vibrotactile stimulation?

To investigate a sensory signal's information-carrying capacity, one can measure the robustness of information transmission when the signal is perturbed by some random variable. Under real world conditions, random variation can diminish signal reliability and undermine perception; for example, variation in the skin's contact with a vibrating source introduces variability into signals received by the skin. In controlled research settings, the addition of random variation, or noise, to a stimulus can yield valuable insights into the computations that support perception (Allard *et al.*, 2015). Measurements of visual detection or identification in the presence of noise have been used to quantify vision's efficiency (Gold *et al.*, 1999) and to identify the particular features that subjects use to perform a task (Allard *et al.*, 2015; Sekuler *et al.*, 2004). Would systematic degradation help us understand more about the robustness of vibrotactile signals?

To address these questions, we had subjects discriminate between sequences of pulses delivered at two different mean rates, 3 Hz and 6 Hz. Pulses were presented in one of three modes: as vibrotactile stimuli delivered to the palms and fingers, as visual stimuli generated by turning on and off a spot of light, or as concurrent stimuli in both vibrotactile and visual modalities. To test signal robustness, external time-domain noise was added to the pulses in

a sequence. Because we focused on the perception of beats, a purely temporal feature, we employed a form of noise that was defined in the temporal domain. Beats within any modality were generated by identical pulses, which forced subjects to base perceptual judgments solely on the temporal gaps between pulses, the inter-pulse intervals (IPIs). We inserted random temporal noise into our stimuli by adding a zero-mean Gaussian random variate to each IPI in a pulse sequence. This rendered the temporal separations between successive pulses stochastic. For these stochastic stimuli, IPIs in a sequence varied independently. This same basic operation was applied to stimuli from each of the three different modalities.

We had theoretical as well as applied reasons for examining how rate discrimination was affected by inserting random intervals into the stimuli. Changes in behavioral measures associated with increased variability inform the robustness of sensory signals (Pelli and Farell, 1999). If we observed different effects of temporal variability on task performance among modalities, we could infer differences in signal robustness among modalities as well. We wanted to benchmark the robustness of information carried by vibrotactile signals against the robustness of visual information, in particular. With signal detection efficiencies as high as 83% (Burgess *et al.*, 1981), visual information provides a ‘gold standard’ for signal robustness and serves as a baseline for comparisons. Additionally, as explained later, we wanted to exploit random variation in the IPIs for a detailed analysis of how successive pulses were processed. Sequential sampling models of decision making vary with respect to their assumptions about evidence accumulation, some assuming a constant rate, and others emphasizing the role of later evidence, akin to a recency effect (Evans *et al.*, 2017). Specifically, we asked whether all IPIs in a stochastic pulse sequence contributed equally to a response, as assumed by drift-diffusion models, or whether specific IPI positions carried additional weight. Finally, we wanted to assess whether designers of signaling devices should consider combining visual and vibrotactile stimuli to optimize information transfer. In particular, we wanted to determine whether processing of visual signals would benefit from coordination with vibrotactile signals.

2. Methods

2.1. Subjects

Twenty-five subjects (13 female, 10 male, 2 declined to identify; mean age = 19.1 years, $SD = 1.2$) served in the experiment. Previous rate perception work from our laboratory (Bushmakina and Sekuler, 2016) demonstrated that such sample sizes are sufficient for observing significant effects. All subjects had best-corrected Snellen acuity 20/40 or better. Experimental procedures

Figure 1. Photograph showing a subject holding the Samsung tablet that delivered visual stimuli, vibrotactile stimuli, and concurrent stimuli from both modalities.

were approved by Brandeis University's Institutional Review Board and were conducted in accordance with the Declaration of Helsinki. All subjects gave written informed consent prior to participation.

2.2. Apparatus

All stimuli were delivered *via* a handheld computer tablet (Samsung Note 10; Fig. 1). Subjects held the tablet bimanually so that on average it was ~ 41 cm from their eyes.

2.3. Stimuli

Stimuli were sequences of pulses delivered either as visual (V), vibrotactile (vT), or concurrent visual and vibrotactile (V-vT) pulses. In all three of these conditions, each pulse in a sequence was 33 ms in duration, but the IPIs, defined as the time elapsed from the end of one pulse to the onset of the next, varied as explained below. Stimuli with two different mean pulse rates, 3 Hz and 6 Hz, were randomly intermixed; pilot testing showed that these two rates would be reliably, though imperfectly, discriminated. Subjects categorized the rate of each stimulus as either 'slow' (3 Hz) or 'fast' (6 Hz). Hereafter, we refer to each variable (MODALITY, NOISE, and RATE) in small capitals to differentiate the experimental variable from the broader meaning of each term.

Vibrotactile stimulation was produced by the tablet's built-in motor whose rotating eccentric load delivered vibrations to the subjects' fingers and palms. Located behind the tablet's touch screen, the tablet's motor rotated at and produced vibrations at 250 Hz. Each 33-ms vibrotactile stimulus pulse was generated by gating the motor's rotation on and off. The strength of each vibrotactile stimulus pulse was well above detection threshold and did not vary throughout the task. Visual pulses were generated by turning on and turning off a small

circular Gabor patch presented at the center of the tablet screen. At the average viewing distance, each Gabor subtended $\sim 1.4^\circ$, and had a peak luminance of 125 cd/m^2 on a steady background luminance of 52 cd/m^2 . As with vibrotactile pulses, visual pulse strength was constant throughout the experiment. For V-vT stimuli, the concurrent visual and vibrotactile pulses were synchronized at the start of each trial. Previous work from our laboratory showed that vibrotactile input can either aid or interfere with processing of concurrent visual stimulation, depending upon whether information received from one modality reinforced or contradicted information received from the other (Bushmakin and Sekuler, 2016). To control for such complications, concurrent visual and vibrotactile pulses in our bimodal stimuli were always presented at the same rates and in synch.

Nominal IPIs for the two stimulus RATES were set at 300 ms for 3-Hz sequences (33 ms pulse + 300 ms IPI = 333 ms cycle length) and 133 ms for the 6-Hz sequences (33 ms pulse + 133 ms IPI = 166 ms cycle length). As mentioned before, each IPI in a sequence could be independently perturbed by a new random variate, ω_i . If the nominal mean IPI were X_i ms, each IPI in a train of pulses would be changed to $X_i + \omega_i$ ($X_i + 33$) (Fig. 2B). Random variates were generated by sampling a zero-mean Gaussian distribution, $\omega_i \sim N(0, \sigma^2)$, where σ was set either to 0, 0.2, or 0.4. This produced temporally stochastic pulse sequences with 0%, 20%, or 40% NOISE, respectively.

Our choice to vary ω_i randomly and independently created the potential to observe directional effects of IPI perturbation *post hoc*. We expected that values of $\omega_i < 0$ and values of $\omega_i > 0$ would have opposite effects on discriminability when (i) ω_i was added to the mean IPI of 3-Hz stimuli, and when (ii) ω_i was added to the mean IPI of 6-Hz stimuli. As an example, if $\omega_i > 0$ were added to the mean IPI from a 6-Hz stimulus, the resulting IPI would increase, shifting it toward the mean IPI from a 3-Hz stimulus. Larger positive values of ω_i would produce larger shifts toward the mean IPI from a 3-Hz stimulus. This shift toward the mean IPI of a 3-Hz stimulus would promote confusion and lead subjects to mistake the nominally 6-Hz stimulus for one of 3 Hz. A comparable but opposite effect would occur if $\omega_i < 0$ and it were added to the mean IPI from a 3-Hz stimulus. Here, the reduced IPI would be more like an IPI from a 6-Hz stimulus, causing a subject to mistake the sample of a 3-Hz stimulus for a 6-Hz stimulus. Table 1 summarizes how the addition of positive and negative ω_i values to IPIs for 3- and 6-Hz stimuli was expected to affect subjects' responses.

2.4. Task

To control for the effect of switching attention between modalities (Boulter, 1977), V, vT, and V-vT stimuli were presented in separate test blocks, of 130 trials each. For V-vT stimuli, concurrent visual and vibrotactile pulses were

Figure 2. Representation of trial structure. (A) On a single trial, subjects were presented with a fixation cross for 33 ms. After a pre-stimulus waiting period of 500 ms, a pulse sequence began. The first pulse was presented for 33 ms, followed by the first inter-pulse interval (IPI), whose duration could vary depending on the NOISE level. Following the first IPI, the next pulse was presented for 33 ms, followed by the second IPI. Pulse and IPI presentations repeated in this way until the subject responded. (B) NOISE level was randomized from trial to trial. If a given trial had 0% NOISE, all IPIs in that pulse sequence were held constant at a nominal IPI $X_i = 300$ ms or 133 ms, depending on the pulse RATE. If a trial had 20% NOISE, each nominal IPI X_i in the sequence was independently perturbed to have a value of $X_i + \omega_i$ ($X_i + 33$), where $\omega_i \sim N(0, 0.2^2)$. For trials with 40% NOISE, each nominal IPI X_i in the sequence was independently perturbed similarly, but the distribution from which ω_i was drawn became $\omega_i \sim N(0, 0.4^2)$. These independent perturbations created stochastic temporal variability scaled by the nominal pulse RATE within each pulse sequence.

Table 1.
Effect of ω value on IPI and on the likelihood of a correct response

ω	IPI change	Rate change	Expected effect on accuracy	
			3-Hz stimulus	6-Hz stimulus
>0	Longer	Slower	Increased	Decreased
<0	Shorter	Faster	Decreased	Increased

IPI: inter-pulse interval.

always synchronized with one another. Departing from the approach others have taken with bimodal stimuli (Varghese *et al.*, 2017), we allowed subjects to base their decisions about V-vT pulse rate on whichever modality they preferred: visual or vibrotactile. The 130 trials in each block were randomized by mean pulse RATE (3 Hz or 6 Hz) and NOISE level ($\sigma = 0, 0.2$, or 0.4). Each of the three MODALITY conditions, V, vT, and V-vT, was repeated twice in a

block-randomized design. This resulted in a total of 780 (6×130) trials per subject.

Figure 2A depicts overall trial structure. A fixation cross, displayed at the tablet's center for 33 ms, alerted the subject that a trial was about to begin. Stimulus pulses began 500 ms later and continued until the subject made a categorization response by tilting the handheld tablet either downward, away from themselves, or upward, toward themselves. A downward tilt $\geq 17^\circ$ meant that pulse train was categorized as 'slow' (3 Hz) by the subject; the same amount of tilt upward meant that the pulse train was categorized as 'fast' (6 Hz). Differences in latency and speed of movement of the wrist's flexor and extensor muscles caused the average times to rotate the tablet in each direction to differ. The failure to counterbalance how judgment categories ('slow' and 'fast') mapped onto directions of tablet rotation, upward or downward, foreclosed comparing judgment times for 3- and 6-Hz stimuli, but had no other effect on data analysis or interpretation. On each trial, subjects received pulses until they responded by tilting the tablet. While there was no limit to the number of pulses subjects could receive, they were encouraged to respond as quickly as possible. Whenever a response time exceeded 1600 ms, the tablet screen displayed a message encouraging more rapid responses on subsequent trials.

Response time (RT) was defined as the time elapsed from the onset of the trial's pulse sequence to the time of the subject's response. For the entire experiment, a reminder remained on the screen telling subjects how to tilt the tablet for the two different responses. Subjects received feedback for their responses only during each test block's first 10 trials. To prevent any auditory signals generated by the tablet's vibrations from impacting subjects' responses, throughout the experiment subjects heard a masking sound, white noise delivered over noise-canceling headphones.

2.5. Data Analysis

2.5.1. Pre-Processing

Of the 25 subjects who were tested, data from two were excluded. One excluded subject consistently performed at or below chance (percent correct $\leq 50\%$); the second excluded subject consistently gave abnormally long response times (that is, up to five seconds long). For the remaining 23 subjects, data for the first 10 trials in each block were treated as practice and were discarded. We analyzed RTs only from trials on which the subject's response was correct. In addition, we excluded RTs that were outliers relative to a subject's entire set of correct RTs. Specifically, we identified each subject's first and third quartiles as well as their interquartile range (IQR), and then excluded trials when $RT < Q1 - 1.5 \times IQR$, or when $RT > Q3 + 1.5 \times IQR$ (Tukey, 1977). Using these criteria, 4.35% of all correct trials were removed. After such trials were

excluded, we summarized a subject's RTs in each condition by that subject's median RT. Data from all test trials were used for analyses of accuracy.

2.5.2. Accuracy and Response Time Analyses

Separate within-subject factorial ANOVAs were used to test how accuracy and response time were affected by NOISE and MODALITY. Each ANOVA incorporated two sets of contrasts for planned comparisons: polynomial coefficients were used to dissect the ordered effect of NOISE, and Helmert coefficients were used to dissect the unordered categorical variable of MODALITY.

2.5.3. Equivalence Testing

Signal detection theory defines an optimal decision criterion as one that elicits equal error rates across experimental conditions (Macmillan and Creelman, 2005). In our study, an optimal decision criterion would have yielded equal probabilities for responding 'slow' when the stimulus RATE was fast [$\Pr(\text{'slow'} \mid \text{fast stimulus})$] and for responding 'fast' when the stimulus RATE was slow [$\Pr(\text{'fast'} \mid \text{slow stimulus})$]. We did not expect to observe a response bias in our study, as both stimulus RATES occurred equally often and carried equal payoff incentives (Macmillan and Creelman, 2005). The theoretical criterion IPI that would have fulfilled the equation $\Pr(\text{'fast'} \mid \text{slow stimuli}) = \Pr(\text{'slow'} \mid \text{fast stimulus})$ differed among the three NOISE levels, and each is shown by a dashed vertical line in the panels of Fig. 3.

To determine whether subjects actually used the optimal criteria described in the preceding paragraph, we tested the proposition that pairs of error rates, one for 3-Hz stimuli and one for 6-Hz stimuli, would be sufficiently similar

Figure 3. Probability density functions of inter-pulse intervals (IPIs) when NOISE = 0% (A), 20% (B), and 40% (C). In each panel, red curves show the distributions of IPIs for 6-Hz stimuli, and blue curves show distributions for 3-Hz stimuli. In the NOISE = 0% panel (A), the vertical red and blue lines indicate that IPIs for each stimulus RATE were fixed, at either 300 ms (for 3 Hz) or 133 ms (for 6 Hz). Unlike in panels B and C, the Y-axis in panel A extends to 1.0 to reflect the fixed nature of IPIs in 0% NOISE condition. The dashed vertical line in each panel defines the IPI that would produce equal proportions of the two kinds of categorical errors, $\Pr(\text{'slow'} \mid \text{fast stimulus})$ and $\Pr(\text{'fast'} \mid \text{slow stimulus})$.

for each NOISE level as to be reasonably considered equivalent. Equivalence testing was performed using the Two One-Sided Tests (TOST) method (Lakens, 2017) implemented in the *TOSTER* package in R. The TOST procedure determines whether an observed effect is small enough to be considered statistically equivalent to zero. The approach involves specifying a set of lower (Δ_L) and upper (Δ_U) bounds that define a null effect region, i.e., a range of effect sizes considered to be small enough to constitute no worthwhile effect. Two one-sided tests are then performed to jointly test the null hypotheses H_{01} : $\Delta \leq \Delta_L$ and H_{02} : $\Delta \geq \Delta_U$. In our implementation, Δ was the mean difference between the two kinds of error rates for each NOISE condition, or $\Delta = \text{Pr}(\text{'slow' response} \mid 6\text{-Hz stimulus}) - \text{Pr}(\text{'fast' response} \mid 3\text{-Hz stimulus})$.

2.5.4. Drift-Diffusion Modeling

The drift-diffusion model (DDM) framework provides a comprehensive analysis of subjects' decision-making processes, portraying decisions as the product of a stochastic process that accumulates information over time toward one of two response criteria (Ratcliff and McKoon, 2008). In our case, the two response criteria were either 'slow' or 'fast'. The basic model includes three parameter estimates that characterize the two-choice decision-making process: (i) drift rate, the rate at which evidence accumulates in favor of one response; (ii) decision criterion, the amount of evidence needed to make a response; and (iii) non-decision time, described as the portion of response time allocated to both stimulus encoding and motor response. We used the Python-based toolbox HDDM (see Wiecki *et al.*, 2013) to estimate these parameters.

Using Markov Chain Monte Carlo sampling, we generated 10 000 samples from the joint posterior distribution of all model parameters and discarded the first 1000 samples as burn-in. For each parameter, Bayesian methods in HDDM compared the posterior distributions estimated for each MODALITY and NOISE condition (Kruschke, 2014). To assess model fit, we performed posterior predictive checks at the individual subject level. For each subject and each of the nine conditions, 500 parameter values were sampled from the posterior distributions and used to simulate a different data set for each parameter value. To judge the model fit in terms of accuracy as well as RT, both simulated and empirical RT distributions were error-coded such that RTs on correct trials were given a positive sign and RTs on incorrect trials were given a negative sign. Simulated RT distributions were compared to empirical RT distributions using two-sample Kolmogorov–Smirnov tests for each subject and condition (Siegel, 1956).

2.5.5. Modeling the Influence of Individual IPIs

The idea that subjects accumulate information over successive samples or 'looks' (Holt and Carney, 2005; Moore, 2003) is not novel; it is actually central to evidence-accumulation models. However, there is no principled reason

to assume that successive samples are integrated without regard to their serial order. In fact, quite the opposite may be true: sequential sampling decision strategies that give extra weight to early samples can outperform their fixed sample-size counterparts (Wald, 1947). The power of a sequential sampling strategy and its demonstrated usefulness in diverse decision-making domains (Ratcliff *et al.*, 2016) led us to examine how subjects processed the IPIs that comprised a stimulus sequence. Our analysis exploited the fact shown in Table 1: with stochastic stimuli, each individual IPI should have a directionally predictable influence on subjects' responses, either promoting correct judgments or promoting errors. Knowing how each random variate might influence subjects' judgments allowed us to test whether all IPIs in a series of IPIs contributed equally to the ultimate response. Such a test is important because theories of evidence accumulation in cognitive tasks typically assume that successive stimulus samples are given equal weight in the decision-making process (Evans *et al.*, 2017).

We used mixed logistic regression to model the impact of each IPI in a pulse sequence on the likelihood of a correct response. With the *lme4* package in R, we regressed accuracy on the first three IPIs for each trial, controlling for stimulus MODALITY. We focused our analysis exclusively on 6-Hz stimuli because their shorter IPIs meant that many responses would have been made after at least four pulses (and therefore three IPIs) occurred (see Section 3. Results; mean number of IPIs before response in the 6-Hz condition: 4.92 IPIs, SD: 1.69). In contrast, with longer IPI, 3-Hz stimuli, response times on nearly one-half of all trials would not have permitted at least four pulses (and three IPIs) to be delivered before the response (see Section 3. Results; mean number of IPIs before response in the 3-Hz condition: 2.46 IPIs, SD = 1.03).

We included the first three IPIs in the analysis, reasoning that three pulse cycles would afford a clear, though incomplete, picture of how successive IPIs influenced subjects' decisions. While we expected subjects to base their decisions on more than the first three IPIs in a sequence, we did not include these additional IPIs individually at the risk of overfitting the models. We fit models for 20% NOISE trials and 40% NOISE trials separately; observations from trials with deterministic stimuli (NOISE = 0%) were omitted because all IPIs were the same in this condition.

For the reason explained already, these logistic regression models could not be applied to the 3-Hz data. To get around this limitation, we took another approach, examining only the impact of the initial IPI on each trial, for both 3- and 6-Hz stimuli. First, we subsetting the data into four groups for each NOISE level in way that corresponded to our predictions in Table 1: (i) 3-Hz trials with a shortened first IPI, (ii) 3-Hz trials with a lengthened first IPI, (iii) 6-Hz trials with a shortened first IPI, and (iv) 6-Hz trials with a lengthened first IPI. Given the large impact of the first IPI and the small effect of MODALITY

in our logistic regression analysis, we focused on the first IPI and collapsed data across MODALITY conditions. Then, to test for directional effects of IPI perturbations, we used paired Student's *t*-tests to compare subjects' mean accuracy on trials with negatively perturbed (shortened) IPIs to that on trials with positively perturbed (lengthened) IPIs. We performed separate *t*-tests on 20% NOISE and 40% NOISE data for each RATE condition.

3. Results

3.1. Accuracy

Figure 4A shows variation in mean response accuracy with NOISE level and MODALITY. As the ANOVA summary in Table 2 confirms, NOISE and MODALITY each impacted accuracy significantly, but the two variables' interaction did not, $p = 0.692$, $\eta_p^2 = 0.02$. Specifically, random variation in stimulus IPI decreased the accuracy with which subjects categorized stimulus RATE, $p < 0.001$, $\eta_p^2 = 0.85$. The effect of MODALITY was also significant, but accounted for a considerably smaller proportion of variance, $p = 0.020$, $\eta_p^2 = 0.16$. Polynomial contrast analysis showed statistically significant linear and quadratic effects of NOISE, $F(1, 44) = 247.56$, $p < 0.001$, $\eta_p^2 = 0.85$, and $F(1, 44) = 11.34$, $p = 0.002$, $\eta_p^2 = 0.21$, respectively. Helmert contrasts showed that accuracy in the V condition was significantly different from accuracy in the vT condition, $F(1, 44) = 4.08$, $p = 0.0496$, $\eta_p^2 = 0.08$, and that mean unimodal performance (average of V and vT conditions) was significantly different from performance with bimodal stimuli (V-vT), $F(1, 44) =$

Figure 4. Mean response accuracy (A) and response time on correct trial (B) with various levels of NOISE added to the inter-pulse intervals. Parameter for the family of curves is MODALITY of stimulus: V = visual, vT = vibrotactile, V-vT = combined visuo-tactile. Error bars represent 95% confidence intervals. Y-axes scaled to reflect the effect size of NOISE on each parameter: $\eta_p^2 = 0.85$ (A) and $\eta_p^2 = 0.44$ (B).

Table 2.
ANOVA on response accuracy

Effect	DF	SumSq	MeanSq	<i>F</i>	<i>p</i> value
MODALITY	2, 44	0.03	0.01	4.26	0.020*
NOISE	1.32, 29.12 [†]	0.59	0.29	129.45	<0.001***, [†]
MODALITY:NOISE	4, 88	0.004	0.001	0.56	0.692

DF: degrees of freedom; MeanSq: mean squares; SumSq: sums of squares.
[†]Corrected for sphericity, Greenhouse–Geisser $E = 0.66$. * $p < 0.05$; *** $p < 0.001$.

Table 3.
ANOVA on response times

Effect	DF	SumSq	MeanSq	<i>F</i>	<i>p</i> value
Modality	1.43, 31.54 [†]	0.09	0.04	6.16	0.011*, [†]
Noise	2, 44	0.02	0.01	17.15	<0.001***
Modality:Noise	4, 88	0.004	0.001	1.61	0.178

DF: degrees of freedom; MeanSq: mean squares; SumSq: sums of squares.
[†]Corrected for sphericity, Greenhouse–Geisser $E = 0.72$. * $p < 0.05$; *** $p < 0.001$.

4.44, $p = 0.041$, $\eta_p^2 = 0.09$. While each of these main effects was statistically significant, MODALITY’s overall small effect size suggests that changes in NOISE level were the primary source of variation in accuracy.

3.2. Response Time

Figure 4B displays mean RT as a function of stimulus NOISE and MODALITY. As Table 3 shows, RT was significantly impacted by both NOISE, $p < 0.001$, $\eta_p^2 = 0.44$, and MODALITY, $p = 0.011$, $\eta_p^2 = 0.22$, but not by their interaction, $p = 0.178$, $\eta_p^2 = 0.07$. Contrast analysis with polynomial coefficients showed that overall, RT increased linearly with NOISE, $F(1, 44) = 33.88$, $p < 0.001$, $\eta_p^2 = 0.44$, but there was no significant quadratic trend, $F(1, 44) = 0.42$, $p = 0.522$, $\eta_p^2 = 0.009$. Contrast analysis with Helmert coefficients showed that RTs in the two unimodal V and vT conditions did not significantly differ from one another, $F(1, 44) = 0.02$, $p = 0.904$, $\eta_p^2 = 0.0003$. However, as the lowest curve in Fig. 4B shows, response times to the V-vT stimulus were significantly faster than the mean RT of the two unimodal conditions, $F(1, 44) = 12.30$, $p = 0.001$, $\eta_p^2 = 0.22$.

3.3. Subjects Adopt Optimal Criteria

To determine whether subjects used optimal criteria as defined by Macmillan and Creelman (2005), we tested the proposition that the error rate for 3-Hz stimuli [Pr(‘fast’ | slow stimulus)] and the error rate for 6-Hz stimuli

Figure 5. Mean differences between error rates ($\text{Pr}(\text{'slow'} \mid \text{fast stimulus}) - \text{Pr}(\text{'fast'} \mid \text{slow stimulus})$; black squares) and 90% TOST confidence intervals (horizontal black lines) with equivalence bounds $\Delta_L = -0.0768$ and $\Delta_U = 0.0768$. In all NOISE levels, the observed mean difference was statistically equivalent to zero, indicating that the average proportion of errors made on 6-Hz trials was statistically equivalent to the average proportion of errors made on 3-Hz trials.

$[\text{Pr}(\text{'slow'} \mid \text{fast stimulus})]$ would be sufficiently similar within each NOISE level as to be reasonably considered equivalent.

Using equivalence bounds of $(-0.0768, 0.0768)$, calculated with $n = 23$ subjects at 80% power and $\alpha = 0.05$, Two One-Sided Tests (TOST) for equivalence were significant in all NOISE levels at $\alpha = 0.05$ [0% NOISE: $t(22) = -1.93$, $p = 0.034$; 20% NOISE: $t(22) = -2.52$, $p = 0.01$; 40% NOISE: $t(22) = -2.88$, $p = 0.004$], allowing us to conclude that the observed difference between proportions of the two types of errors was statistically equivalent to zero. Figure 5 depicts these results as mean difference values (black squares) in each NOISE level, with 90% TOST confidence intervals falling within the established equivalence bounds. These results suggest that subjects showed no statistically significant response bias in any of the NOISE levels, and that they used a decision criterion indistinguishable from one that was statistically optimal.

3.4. Evidence Accumulation Varies With Condition

As Fig. 4 shows, results with our two dependent variables, accuracy and response time, revealed an interesting, clear divergence. Response accuracy showed little or no reliable differences among V, vT, and V-vT conditions; in contrast, response times to V-vT stimuli were shorter than response times to stimuli in either of the unimodal conditions. We used DDM to reconcile this divergence. Effects of NOISE and MODALITY on the posterior distributions of each model parameter are shown in Fig. 6. Each panel shows the values of an individual parameter associated with a particular NOISE level and MODALITY. Results of group-level posterior comparisons are reported below as Bayesian

Figure 6. Posterior distributions of HDDM parameters as a function of stimulus NOISE level. Left column: drift rate, middle column: decision criterion, right column: non-decision time. Top row: 0% NOISE, middle row: 20% NOISE, bottom row: 40% NOISE. Curves correspond to the MODALITY of stimulus presentation: V = visual (red), vT = vibrotactile (blue), V-vT = combined visuo-tactile (purple). In each panel, greater separation of distributions indicates an effect of MODALITY. Within a single column, a shift in a given curve along the x-axis indicates an effect of NOISE. AU: arbitrary units.

probabilities that describe differences between pairs of conditions. To facilitate interpretation, $P = 0.50$ indicates a null effect, while $P > 0.50$ and $P < 0.50$ indicate differences between conditions.

Both MODALITY and NOISE had a reliable effect on drift rate (Fig. 6). For all MODALITIES, increased levels of NOISE were associated with lower drift rates [V: $P(0\% > 20\%) = 0.992$, $P(20\% > 40\%) = 0.999$; vT: $P(0\% > 20\%) = 0.985$, $P(20\% > 40\%) = 0.997$; V-vT: $P(0\% > 20\%) = 0.993$, $P(20\% > 40\%) = 0.999$]. With 0% NOISE, drift rate on V-vT trials was greater than that for either V [$P(V-vT > V) = 0.916$] or vT [$P(V-vT > vT) = 0.911$] trials. At this NOISE level, drift rates for V and vT trials did not differ [$P(vT > V) = 0.519$]. With 20% NOISE, drift rate on V-vT trials was still greater than that for either V [$P(V-vT > V) = 0.899$] or vT [$P(V-vT > vT) = 0.855$] trials, and drift rates for V and vT trials were still very similar [$P(vT > V) = 0.592$]. With 40% NOISE, however, drift rate on vT trials resembled that on V-vT trials [$P(V-vT > vT) = 0.617$]. Drift rates for vT and V-vT trials were both greater than drift rates for V trials at this NOISE level [$P(V-vT > V) = 0.853$; $P(vT > V) = 0.773$].

Only NOISE reliably affected the decision criterion parameter (Fig. 6). For all MODALITIES, increased levels of NOISE were associated with lower decision criteria [V: $P(0\% > 20\%) = 0.940$; $P(20\% > 40\%) = 0.963$; vT: $P(0\% > 20\%) = 0.895$, $P(20\% > 40\%) = 0.961$; V-vT: $P(0\% > 20\%) = 0.948$, $P(20\% > 40\%) = 0.848$]. The effects of MODALITY on decision criterion were less clear. With 0% NOISE, the decision criterion used on V-vT trials was lower than that for either V [$P(V-vT > V) = 0.369$] or vT [$P(V-vT > vT) = 0.397$] trials. At this NOISE level, drift rates for V and vT trials did not differ [$P(vT > V) = 0.471$]. With 20% NOISE, the average decision criterion on V-vT trials was still smaller than that for either V [$P(V-vT > V) = 0.329$] or vT [$P(V-vT > vT) = 0.253$] trials, and decision criteria for V and vT trials were still very similar [$P(vT > V) = 0.598$]. With 40% NOISE, the decision criterion on vT trials resembled that on V trials [$P(V-vT > vT) = 0.504$]. Decision criteria for vT and V-vT trials were both higher than the average criterion used on V trials at this NOISE level [$P(vT > V) = 0.627$; $P(V-vT > V) = 0.634$].

Both MODALITY and NOISE had a reliable effect on non-decision time (Fig. 6). For the visual MODALITY, increased levels of NOISE were associated with longer non-decision time values [$P(0\% < 20\%) = 0.734$, $P(20\% < 40\%) = 0.669$]. On vT trials, non-decision time in the 0% NOISE condition was similar to that in the 20% NOISE condition [$P(0\% < 20\%) = 0.540$], but comparing the 40% NOISE condition to the 20% NOISE condition showed an increase like that seen for V trials [$P(20\% < 40\%) = 0.806$]. For V-vT trials, non-decision time increased as NOISE increased from 0% to 20% [$P(0\% < 20\%) = 0.817$], but non-decision times appeared to decrease in this MODALITY condition as NOISE increased from 20% to 40% [$P(20\% < 40\%) = 0.399$]. As with the drift rate parameter, unimodal non-decision time seemed to differ from bimodal non-decision time across NOISE levels. With 0% NOISE, non-decision time on V-vT trials was less than that for either V [$P(V-vT < V) = 0.919$] or vT [$P(V-vT < vT) = 0.905$] trials. At this NOISE level, non-decision time for V and vT trials did not differ [$P(vT < V) = 0.525$]. With 20% NOISE, non-decision time on V-vT trials was still less than that for either V [$P(V-vT < V) = 0.880$] or vT [$P(V-vT < vT) = 0.709$] trials, and non-decision time for vT trials was less than that for V trials [$P(vT < V) = 0.719$]. With 40% NOISE, non-decision time on V-vT trials was less than that for either V [$P(V-vT < V) = 0.969$] or vT [$P(V-vT < vT) = 0.954$] trials. Non-decision time for vT and V trials did not differ at this NOISE level [$P(vT < V) = 0.545$].

Assessment of model fit showed that the DDM fit our data well overall. Figures 7 and 8 show how model-simulated RT distributions fit the empirical RT distributions for two representative subjects. The subject whose data

Figure 7. Empirical (red) and HDDM-simulated (blue) response time (RT) distributions in each experimental condition for an exemplary subject. Correct RTs are plotted as positive values and error RTs are plotted as negative values, with dashed vertical lines plotted at zero. Kolmogorov–Smirnov test results are reported for each distribution comparison. For all nine conditions, the RT distributions simulated by HDDM fit the observed data well.

are shown in Fig. 7 was the ‘best’ fit; for all nine conditions, Kolmogorov–Smirnov tests showed no significant differences between simulated and empirical RT distributions (all $p > 0.05$). Figure 8 shows the ‘worst’ fit. The simulated RT distributions were significantly different from this subject’s observed RT distributions in three of the nine experimental conditions: V, 0% NOISE: $p = 0.044$; vT, 0% NOISE: $p = 0.038$; V–vT, 0% NOISE: $p = 0.035$. This subject was the only subject for whom the model fit was poor. For six other subjects, only in one condition each did simulated and observed RT distributions differ significantly, and the majority of subjects had no significant differences between distributions (Table 4). In total, only 4.3% (9/207) of the

Figure 8. Empirical (red) and HDDM-simulated (blue) response time (RT) distributions in each experimental condition for subject with the poorest model fit. Correct RTs are plotted as positive values and error RTs are plotted as negative values, with dashed vertical lines plotted at zero. Kolmogorov–Smirnov test results are reported for each distribution comparison. For three of the nine conditions, the RT distributions simulated by HDDM were significantly different from the observed RT distributions. * $p < 0.05$.

HDDM-simulated RT distributions significantly differed from the empirical RT distributions.

3.5. First Impressions (and First IPIs) Matter Most

To examine successive IPIs' influence on subjects' categorization of pulse RATE, we applied multiple logistic regression to the first three IPIs for each trial with stochastic stimuli in the 6-Hz condition. Separate analyses were done for 20% NOISE and 40% NOISE levels. We focused our analysis exclusively on 6-Hz stimuli as their shorter IPIs meant that many responses were made after

Table 4.
Number of subjects with poor HDDM fit, by condition

Noise	Modality		
	Visual (V)	Vibrotactile (vT)	Bimodal (V-vT)
0%	3	1	2
20%	1	0	1
40%	1	0	0

Number of subjects (out of 23) for whom HDDM-simulated RT distributions were significantly different from empirical RT distributions in each experimental condition. To assess goodness of the model fit for each subject, a two-sample Kolmogorov–Smirnov test was performed for each of the nine conditions. Only one subject, whose RT distributions are shown in Fig. 8, had a poor model fit in multiple conditions. For all other subjects, the model fit was poor for only one of the nine experimental conditions.

at least four pulses (and therefore three IPIs) occurred (mean number of IPIs before response in the 6-Hz condition: 4.92, SD: 1.69; 3-Hz condition: 2.46, SD = 1.03). For each stochastic stimulus sequence with a 6-Hz pulse rate, we regressed decision accuracy (a binary variable) against predictors including the MODALITY of testing (V, vT, and V-vT) and the values of each of the first three IPIs in the stimulus. The nested models shown in Table 5 were each fit sequentially to determine the impact of each parameter on the likelihood of a correct response. These models omit interaction terms because preliminary tests that included those terms produced unstable model fits. In the 20% NOISE condition, interactions between the first and second IPIs [$\chi^2(1) = 5.15$, $p = 0.023$], the second and third IPIs [$\chi^2(1) = 12.53$, $p < 0.001$], and the first and third IPIs [$\chi^2(1) = 11.34$, $p < 0.001$] were all statistically significant; the same models, however, failed to converge using the 40% NOISE data. Given the instability of models with interaction terms, we opted for the simpler, additive full model (boldface model in Table 5).

We fit the mixed logistic regression models listed in Table 5 using the *lme4* package in R. Likelihood ratio testing was performed using the deviance (-2LL) statistics for each nested model. The table shows that as expected, deviance, sometimes described as ‘badness of fit’, shrinks as additional parameters are included in successive nested models. Comparing the values of Δ -2LL for successive models yielded the p values presented in Table 5 and gave a sense of each additional parameter’s impact. With this criterion, we found that for both NOISE levels, the first and second IPIs each consistently significantly affected response accuracy. The effects of the third IPI and of stimulus MODALITY were statistically significant only for the 40% NOISE condition. Additionally, the third IPI and the stimulus MODALITY each had

Table 5.
Mixed logistic regression model specification and results for each level of NOISE

Model specification	NOISE = 20%			NOISE = 40%		
	AIC	Δ -2LL (DF)	<i>p</i> value	AIC	Δ -2LL (DF)	<i>p</i> value
Acc. $\sim$ (1 subj)	1939			2643		
Acc. $\sim$ Modality + (1 subj)	1941	2.06 (2)	0.36	2640	6.83 (2)	0.033*
Acc. $\sim$ Modality + <i>IP1</i> + (1 subj)	1903	40.28 (1)	<0.001***	2525	116.92 (1)	<0.001***
Acc. $\sim$ Modality + <i>IP1</i> + <i>IP12</i> + (1 subj)	1878	26.25 (1)	<0.001***	2467	60.50 (1)	<0.001***
Acc. $\sim$ Modality + <i>IP1</i> + <i>IP12</i> + <i>IP13</i> + (1 subj)	1879	1.43 (1)	0.23	2458	10.99 (1)	<0.001***

Model specification is given in the notation used by R's lme4 package. Inclusion of error term (1 | subj) allowed model intercept to vary between subjects. The full model used for final results is indicated in **bold**. -2LL: deviance; Acc.: accuracy; AIC: Akaike information criterion; DF: degrees of freedom. * $p < 0.05$; *** $p < 0.001$.

a considerably smaller impact on model deviance than did the first and second IPIs.

The odds ratio (OR) for each parameter estimate in the full model (Table 5) expresses the magnitude of each IPI's effect on the odds of responding correctly. With 20% NOISE added to 6-Hz stimuli, increasing the first IPI by just 10 ms decreased the likelihood of a correct response by 11.2% (OR = 0.888, 95% CI = [0.856, 0.921], $p < 0.001$). An analogous change of IPI in the 40% NOISE condition decreased accuracy by 8.4% (OR = 0.916, 95% CI = [0.902, 0.931], $p < 0.001$). The same change in the second IPI in a sequence had smaller effects, decreasing the likelihood of a correct response by 9.2% for the 20% NOISE level (OR = 0.908, 95% CI = [0.874, 0.942], $p < 0.001$), and by 5.9% for the 40% NOISE level (OR = 0.941, 95% CI = [0.926, 0.955], $p < 0.001$). A 10 ms increase in the third IPI decreased the odds of a correct response by just 2.2% with 20% NOISE (OR = 0.978, 95% CI = [0.943, 1.014], $p = 0.230$), and by 2.5% with 40% NOISE (OR = 0.975, 95% CI = [0.960, 0.990], $p < 0.001$). While changes in the third IPI produced similar effect sizes for both of the two NOISE levels, it is important to note that this effect was only significant in the 40% NOISE model.

These results suggest that subjects' decisions were based disproportionately on the first two IPIs in a pulse sequence, and that only in the 40% NOISE condition did subjects make appreciable use of subsequent IPIs. It is important to consider the additive nature of the results presented: subjects undoubtedly based their decisions on sequences of IPIs rather than on any one IPI in isolation, and many or all IPIs prior to the decision could be taken into consideration by the subject. For example, if both the first and second IPIs in a 6-Hz stimulus with 20% NOISE deviated from the nominal IPI by +10 ms [$\Pr(\text{IPI}_1 \geq 143 \text{ ms} \cap \text{IPI}_2 \geq 143 \text{ ms}) = 14.6\%$], holding all other IPIs in the trial constant, the likelihood that a subject would respond correctly on that trial would decrease by more than 19%.

As predicted, directional changes in the first IPI affected accuracy differentially for 3-Hz and 6-Hz stimuli (Fig. 9). Figure 9A shows that, with 20% NOISE, subjects were more accurate when the first IPI on a 3-Hz trial was lengthened compared to when the first IPI was shortened, $t(20) = 8$, $p < 0.001$, $d = 1.7$ (mean difference: 0.145, 95% CI [0.106, 0.184]). In contrast, for 6-Hz trials in the same NOISE condition, subjects were *less* accurate when the first IPI was lengthened than when the first IPI was shortened, $t(20) = -4$, $p = 0.002$, $d = 0.48$ (mean difference: -0.059, 95% CI [-0.094, -0.024]). The same trend was observed with 40% noise (Fig. 9B), which produced even larger effects (3 Hz: $t(20) = 10$, $p < 0.001$, $d = 3.02$, mean difference: 0.295, 95% CI [0.246, 0.343]; 6 Hz: $t(20) = -7$, $p < 0.001$, $d = 1.23$, mean difference: -0.13, 95% CI [-0.168, -0.0924]).

Figure 9. Mean accuracy as a function of changes in the first IPI for 3-Hz and 6-Hz stimuli in 20% noise (A) and 40% noise (B). Red bars correspond to mean accuracy levels for 3-Hz stimuli while blue bars correspond to mean accuracy levels for 6-Hz stimuli. Opacity of each bar indicates the direction of the change in the first IPI: light red/blue for trials with shortened IPIs, and dark red/blue for trials with lengthened IPIs. Error bars reflect within-subject standard error. For 3-Hz stimuli, longer IPIs improved accuracy compared to shorter IPIs. For 6-Hz stimuli, longer IPIs decreased mean accuracy compared to trials with shortened first IPIs. We observed these differences in both the 20% and 40% NOISE conditions. ** $p < 0.01$, *** $p < 0.001$.

4. Discussion

4.1. Review of Key Findings

The preceding sections expanded a basic examination of accuracy and response time results by deploying three different, but complementary analytic approaches. As expected, the multiple analytic approaches produced a variety of findings. So, before discussing our study's implications, it will be useful to summarize what we consider our principal findings:

1. The results of our accuracy and response time analyses showed that, compared to either unimodal condition, bimodal stimulus presentation was associated with improved task performance, particularly faster response speed on correct trials (Fig. 4B). Performance did not differ substantially between the two unimodal conditions. The analyses of accuracy and response times also showed no significant NOISE $\times$ MODALITY interactions, indicating that time-domain noise did not affect performance differently among the V, vT, and V-vT conditions.
2. Error rate analysis showed that subjects used statistically optimal decision criteria, adapting in real time to the random noise present in a stimulus. Drift-diffusion modeling confirmed this result.
3. Drift-diffusion modeling also offered one potential explanation of why MODALITY had a larger effect on RT ($\eta_p^2 = 0.22$) than on accuracy ($\eta_p^2 =$

0.09): subjects accumulated evidence more quickly with and took less time to encode, process, and respond to bimodal stimuli, compared to either V or vT stimuli (Fig. 6).

4. Importantly, logistic regression revealed that early evidence had the most weight in subjects' responses, and later evidence only informed decisions when signals were degraded by noise (Table 5). The results of the logistic regression analysis suggest that each IPI in a sequence provided subjects a different amount of evidence, but this is in direct conflict with the assumptions of the drift-diffusion model.

We discuss our most notable findings as well as our study's limitations in further detail below.

4.2. *An Assay of Signal Robustness*

Of all three MODALITY conditions in our experiment, subjects were least accurate with visual stimuli. Multiple studies have demonstrated that sensory systems are specialized for processing different stimulus attributes (the 'modality appropriateness hypothesis'; Welch and Warren, 1980). When experiments combine auditory and visual stimuli, the modalities' relative influences vary with task demands: when judgments must be based on temporal information, auditory cues are afforded more weight than visual cues; when judgments must be based on spatial information, visual cues are given more weight than auditory cues (Gebhard and Mowbray, 1959; Michalka *et al.*, 2015; Recanzone, 2003; Welch *et al.*, 1986). Analogous relationships have been reported between visual and haptic cues (Ernst, 2007; Ernst and Banks, 2002). The relatively poorer performance that we observed with visual stimuli may, therefore, be related to vision's specialization for processing spatial information rather than the temporal information demanded by our task.

Our results may actually underestimate the potential effect of temporal variation in the IPIs. Specifically, difference thresholds for duration of vibrotactile stimuli have been reported to be as high as 13% (Francisco *et al.*, 2015). With the Gaussian distributions from which we drew random variates, for the 20% noise condition many of the IPI perturbations would have been below threshold in vT and V-vT conditions. However, for the 40% noise condition, fewer IPI perturbations were likely below threshold (Fig. 10). Future work could apply temporal noise that extends the narrow range of noise we examined, possibly drawing noise from distributions that are less heavily weighted toward the mean.

Because we applied NOISE relative to the nominal IPI in a pulse sequence, the widths of IPI sampling distributions at any one NOISE level differed between the two RATES (Fig. 10). Adding variability to the IPIs in this way meant that the amount of nominal NOISE (in ms) varied between the two pulse

Figure 10. Inter-pulse interval (IPI) sampling distributions used in the 20% NOISE (A) and 40% NOISE (B) conditions. In each panel, red curves show the distributions of IPIs for 6-Hz stimuli, and blue curves show distributions for 3-Hz stimuli. The dashed vertical lines in each panel indicate a $\pm 10\%$ perturbation from the nominal IPI for each stimulus type, and shaded regions indicate the proportion of each distribution that would have no effect on subject performance. While Francisco *et al.* (2015) reported a 13% difference threshold for duration of vibrotactile stimuli, we examined IPI distributions using a 10% difference threshold in order to account for the unknown difference threshold for visual stimuli in this kind of task.

RATES, but this approach scaled the amount of temporal NOISE present in a pulse sequence relative to that sequence's other temporal features (namely, its frequency). Our failure to counterbalance the way the two response types were signaled, however, prevented us from testing for differences in the effect of NOISE on 3-Hz and 6-Hz trials.

4.3. The Benefit of Bimodality

Bimodality had a beneficial impact on both response speed and on drift-diffusion model estimates of rate at which evidence is accumulated (Figs 4 and 6). This result is consistent with earlier suggestions that multisensory stimuli are processed by supramodal cortical mechanisms (Crommett *et al.*, 2017; Levitan *et al.*, 2015). Interestingly, this hypothesis is supported by a recent functional magnetic resonance imaging (fMRI) demonstration that auditory frequency is broadly represented in human cerebral cortex, including in classically defined somatosensory cortex (Pérez-Bellido *et al.*, 2018).

As mentioned earlier, with V-vT combinations, subjects could base their decisions on whichever modality they chose. With concurrent stimuli in multiple modalities, previous work showed that a subject's preferred modality depends upon the modalities' relative reliabilities (Bresciani and Ernst, 2007; Bresciani, Dammeier and Ernst, 2008; Ernst and Banks, 2002). However, the close equivalence in accuracy for V and vT stimuli in our experiment meant neither MODALITY offered a clear advantage in terms of reliability. Despite this, subjects may have consistently relied upon one of the two unimodal cues that comprised a V-vT stimulus, inadvertently rendering cues from one modality

task-irrelevant and undermining the true bimodal nature of the V-vT condition. It is worth noting that task-irrelevant stimuli still can affect perception and become automatically integrated with target stimuli (Bresciani *et al.*, 2008; Maddox *et al.*, 2015; Varghese *et al.*, 2017). Additionally, because the separate concurrent components of V-vT stimuli were perfectly correlated and synchronized, the two components might have been bound perceptually, an effect likely amplified by their shared spatial location (Badde *et al.*, 2018; Locke and Landy, 2017). These speculations could be tested in experiments that added independent noise samples to the separate V and vT components of bimodal V-vT stimuli, and/or varied their spatial relationship.

4.4. Investigating the Decision-Making Process

Drift-diffusion modeling yielded deeper insight into subjects' decision-making process than did our separate analyses of accuracy and response speed. The diffusion model analysis showed that information was extracted from noisier stimuli more slowly than from less noisy stimuli. This result implies that in noisy sequences, individual IPIs provided less evidence toward a decision. Similar changes in drift rate were seen when diffusion modeling was applied to tasks that vary in difficulty (Voss *et al.*, 2004; Wagenmakers, 2009).

Our results diverge from standard drift-diffusion accounts in two important ways. First, our diffusion modeling results, as well as the results of equivalence testing, imply that subjects adjusted their decision thresholds as a function of the NOISE level they experienced within a trial. Subjects' ability to adjust decision thresholds in real time, during the course of a trial, has special theoretical interest. Diffusion models of decision making typically assume fixed boundaries set by the subject at trial onset (Ratcliff and McKoon, 2008; Ratcliff *et al.*, 2016). In our study, however, NOISE level was randomized, guaranteeing that subjects had no prior knowledge of how much IPI variability they might encounter until the trial was underway. If subjects set and held a decision threshold at the beginning of each trial, we would have seen no differences in the decision criterion parameter between NOISE levels. Second, classic diffusion models also assume a constant rate of evidence accumulation during a trial (Ratcliff and McKoon, 2008; Ratcliff *et al.*, 2016). In our study, where the IPI was the unit of evidence, a constant drift rate would imply that each IPI in a single trial provided the same amount of evidence toward the decision criterion threshold. Our logistic regression analysis, however, suggested a different narrative: that each IPI in a pulse sequence yielded a different amount of evidence, with the first and second IPIs in a sequence providing the most information and later IPIs providing comparatively less.

While our implementation of the drift-diffusion model yielded good model fits (Figs 7 and 8), the departures from the model assumptions described above may have detracted from the utility of fitting such a model. Holmes and

Trueblood (2018) have reported on the limited ability of diffusion models to account for non-stationary decision criteria in particular, and suggested fitting a piecewise variant of the classic drift-diffusion model to account for time-varying parameters. Additionally, alternative sequential sampling models may have been better suited for our task. The accumulator model, for example, assumes that values of decision criteria vary exponentially across trials (Ratcliff and Smith, 2004). That model also assumes that evidence is sampled at discrete time points, reflecting the trial structure in our task more accurately.

We used our regression analysis to investigate the weight of each discrete unit of evidence in a trial, but what we have called the full model (Table 5) was not meant to be a complete account of subjects' responses. Our analysis focused mainly on the relative importance of the first few IPIs that a subject experienced during a trial; an improved approach would account for not only the first three but *all* of the IPIs in a pulse sequence. We chose not to include each IPI experienced before a response as individual predictors to avoid overfitting the model. An alternative full model could have included predictors for each of the first three IPIs, as in Table 5, as well as an additional predictor that captured the mean of the remaining IPIs experienced before a response. We opted not to use this approach, however, because of the amount of data that we would have excluded to accommodate such an analysis. Our full model undoubtedly omitted other potentially consequential variables as well, such as interactions among IPIs, trial-to-trial variation in attention (e.g., Chambers and Pressnitzer, 2014; Parise and Ernst, 2017; Schwiedrzik *et al.*, 2014), and variability in the decision criterion (Cabrera *et al.*, 2015; Mueller and Weidemann, 2008), as discussed above.

4.5. *Conclusions and Future Work*

Our experiment set out to (i) benchmark the robustness of information conveyed by vibrotactile signals compared to visual signals, (ii) investigate the decision-making process in rate discrimination tasks with stochastic temporal sequences, and (iii) assess the potential of combined visual and vibrotactile stimuli for use in signaling devices. Subjects extracted the mean rate equally well from either of the two unimodal conditions; from this, we conclude that vibrotactile signals are *as robust as* visual signals as vehicles for transmitting information about the rate at which stimuli occur. Our results also support the notion that early information is given the most weight when making speeded decisions, and show that processing of rate information is expedited when information is presented bimodally.

Vehicles, mobile devices, and medical equipment now utilize pulsatile cues, usually auditory or visual, to convey information. Our results suggest that adding vibrotactile stimuli to the mix could promote faster, more accurate responses. Currently, vibrotactile signals are used in mobile devices primarily to

alert users to some event, such as an incoming call. The robustness of beat-like vibrotactile signals demonstrated in our study confirms that various attributes of these cues could be manipulated to provide more information than a simple alert. We concur with the suggestions made by Meng and Spence (2015) about the considerable information-carrying potential of vibrotactile stimulation, particularly when its temporal, spatial, and intensity dimensions can all be varied. Future studies would do well to ascertain the limits on the ability of vibrotactile stimuli to convey meaningful, timely information, perhaps in combination with concurrent stimuli in other sensory modalities.

Author Contributions

MBV and RS were responsible for study concept and design. MBV and RFS collected and analyzed the data under the supervision of RS. MBV and RS drafted the manuscript; RFS provided critical revisions. All authors approved the final version of this manuscript for submission.

Acknowledgements

This work was supported by NIGMS Training Grant T32GM132498 (MBV) and NIH Training Grant R90DA033463 (RFS). Publication of this open access article was funded by the Brandeis Library Open Access Fund. We thank Xiaodong Liu for statistical advice, Paul DiZio for constructive comments on an earlier version of the manuscript, and Maxim Bushmakina for programming contributions early in the project.

Declaration of Conflicting Interests

The authors declared that there were no conflicts of interest with respect to the authorship or publication of this article.

References

- Allard, R., Faubert, J. and Pelli, D. G. (2015). Using noise to characterize vision, *Front. Psychol.* **6**, 1707. DOI:10.3389/fpsyg.2015.01707.
- Badde, S., Navarro, K. and Landy, M. S. (2018). Vision and touch are not automatically integrated, *J. Vis.* **18**, 95. DOI:10.1167/18.10.95.
- Bender, D. A. and Bender, A. E. (1999). Body surface area, in: *Benders' Dictionary of Nutrition and Food Technology*, D. A. Bender (Ed.), p. 61. CRC Press, Boca Raton, FL, USA.
- Boulter, L. R. (1977). Attention and reaction times to signals of uncertain modality, *J. Exp. Psychol. Hum. Percept. Perform.* **3**, 379–388.
- Bresciani, J.-P. and Ernst, M. O. (2007). Signal reliability modulates auditory–tactile integration for event counting, *Neuroreport* **18**, 1157–1161.
- Bresciani, J.-P., Dammeier, F. and Ernst, M. O. (2008). Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events, *Brain Res. Bull.* **75**, 753–760.

- Burgess, A., Wagner, R., Jennings, R. J. and Barlow, H. B. (1981). Efficiency of human visual signal discrimination, *Science* **214**, 93–94.
- Bushmakina, M. A. and Sekuler, R. (2016). Congruency counts in visual-vibrotactile interaction, in: *Meeting of the Psychonomic Society*, Boston, MA.
- Cabrera, C. A., Lu, Z.-L. and Dosher, B. A. (2015). Separating decision and encoding noise in signal detection tasks, *Psychol. Rev.* **122**, 429–460.
- Chambers, C. and Pressnitzer, D. (2014). Perceptual hysteresis in the judgment of auditory pitch shift, *Atten. Percept. Psychophys.* **76**, 1271–1279.
- Crommett, L. E., Pérez-Bellido, A. and Yau, J. M. (2017). Auditory adaptation improves tactile frequency perception, *J. Neurophysiol.* **117**, 1352–1362.
- Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch, *J. Vis.* **7**, 7. DOI:10.1167/7.5.7.
- Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion, *Nature* **415**, 429–433.
- Evans, N. J., Hawkins, G. E., Boehm, U., Wagenmakers, E. J. and Brown, S. D. (2017). The computations that support simple decision-making: a comparison between the diffusion and urgency-gating models, *Sci. Rep.* **7**, 16433. DOI:10.1038/s41598-017-16694-7.
- Francisco, E. M., Holden, J. K., Nguyen, R. H., Favorov, O. V. and Tommerdahl, M. (2015). Percept of the duration of a vibrotactile stimulus is altered by changing its amplitude, *Front. Syst. Neurosci.* **9**, 77. DOI:10.3389/fnsys.2015.00077.
- Frings, C. and Spence, C. (2010). Crossmodal congruency effects based on stimulus identity, *Brain Res.* **1354**, 113–122.
- Frings, C. and Spence, C. (2011). Increased perceptual and conceptual processing difficulty makes the immeasurable measurable: negative priming in the absence of probe distractors, *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 72–84.
- Gebhard, J. W. and Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter, *Am. J. Psychol.* **72**, 521–529.
- Gold, J. M., Murray, R. F., Sekuler, A. B., Bennett, P. J. and Sekuler, R. (2005). Visual memory decay is deterministic, *Psychol. Sci.* **16**, 769–774.
- Grahn, J. A. (2012). See what I hear? Beat perception in auditory and visual rhythms, *Exp. Brain Res.* **220**, 51–61.
- Guttman, S. E., Gilroy, L. A. and Blake, R. (2005). Hearing what the eyes see: auditory encoding of visual temporal sequences, *Psychol. Sci.* **16**, 228–235.
- Holmes, W. R. and Trueblood, J. S. (2018). Bayesian analysis of the piecewise diffusion decision model, *Behav. Res. Meth.* **50**, 730–743.
- Holt, R. F. and Carney, A. E. (2005). Multiple looks in speech sound discrimination in adults, *J. Speech, Lang. Hear. Res.* **48**, 922–943.
- Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: a Tutorial With R, JAGS, and Stan*, 2nd edn. Academic Press, London, UK.
- Lakens, D. (2017). Equivalence tests: a practical primer for *t* tests, correlations, and meta-analyses, *Soc. Psychol. Personal. Sci.* **8**, 355–362.
- Levitani, C. A., Ban, Y.-H. A., Stiles, N. R. B. and Shimojo, S. (2015). Rate perception adapts across the senses: evidence for a unified timing mechanism, *Sci. Rep.* **5**, 8857. DOI:10.1038/srep08857.
- Locke, S. M. and Landy, M. S. (2017). Temporal causal inference with stochastic audiovisual sequences, *PLoS One* **12**, e0183776. DOI:10.1371/journal.pone.0183776.

- Macmillan, N. A. and Creelman, C. D. (2005). *Detection Theory: a User's Guide*, 2nd edn. Lawrence Erlbaum Associates, Mahwah, NJ, USA.
- Maddox, R. K., Atilgan, H., Bizley, J. K. and Lee, A. K. C. (2015). Auditory selective attention is enhanced by a task-irrelevant temporally coherent visual stimulus in human listeners, *eLife* **4**, e04995. DOI:10.7554/eLife.04995.
- Meng, F. and Spence, C. (2015). Tactile warning signals for in-vehicle systems, *Accid. Anal. Prev.* **75**, 333–346.
- Meng, F., Gray, R., Ho, C., Ahtamad, M. and Spence, C. (2015). Dynamic vibrotactile signals for forward collision avoidance warning systems, *Hum. Fact.* **57**, 329–346.
- Michalka, S. W., Kong, L., Rosen, M. L., Shinn-Cunningham, B. G. and Somers, D. C. (2015). Short-term memory for space and time flexibly recruit complementary sensory-biased frontal lobe attention networks, *Neuron* **87**, 882–892.
- Moore, B. C. J. (2003). Temporal integration and context effects in hearing, *J. Phon.* **31**, 563–574.
- Mueller, S. T. and Weidemann, C. T. (2008). Decision noise: an explanation for observed violations of signal detection theory, *Psychon. Bull. Rev.* **15**, 465–494.
- Parise, C. V. and Ernst, M. O. (2017). Noise, multisensory integration, and previous response in perceptual disambiguation, *PLoS Comput. Biol.* **13**(7), e1005546. DOI:10.1371/journal.pcbi.1005546.
- Patel, A. D., Iversen, J. R., Chen, Y. and Repp, B. H. (2005). The influence of metricality and modality on synchronization with a beat, *Exp. Brain Res.* **163**, 226–238.
- Pelli, D. G. and Farell, B. (1999). Why use noise?, *J. Opt. Soc. Am. A* **16**, 647–653.
- Pérez-Bellido, A., Barnes, K. A., Crommett, L. E. and Yau, J. M. (2018). Auditory frequency representations in human somatosensory cortex, *Cereb. Cortex* **28**, 3908–3921. DOI:10.1093/cercor/bhx255.
- Ratcliff, R. and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks, *Neural Comput.* **20**, 873–922.
- Ratcliff, R. and Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time, *Psychol. Rev.* **111**, 333–367.
- Ratcliff, R., Smith, P. L., Brown, S. D. and McKoon, G. (2016). Diffusion decision model: current issues and history, *Trends Cogn. Sci.* **20**, 260–281.
- Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception, *J. Neurophysiol.* **89**, 1078–1093.
- Schwiedrzik, C. M., Ruff, C. C., Lazar, A., Leitner, F. C., Singer, W. and Melloni, L. (2014). Untangling perceptual memory: hysteresis and adaptation map into separate cortical networks, *Cereb. Cortex* **24**, 1152–1164.
- Sekuler, A. B., Gaspar, C. M., Gold, J. M. and Bennett, P. J. (2004). Inversion leads to quantitative, not qualitative, changes in face processing, *Curr. Biol.* **14**, 391–396.
- Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. McGraw Hill, New York, NY, USA.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA, USA.
- Varghese, L., Mathias, S. R., Bensussen, S., Chou, K., Goldberg, H. R., Sun, Y., Sekuler, R. and Shinn-Cunningham, B. G. (2017). Bi-directional audiovisual influences on temporal modulation discrimination, *J. Acoust. Soc. Am.* **141**, 2474. DOI:10.1121/1.4979470.
- Voss, A., Rothermund, K. and Voss, J. (2004). Interpreting the parameters of the diffusion model: an empirical validation, *Mem. Cogn.* **32**, 1206–1220.

- Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy, *Eur J. Cogn. Psychol.* **21**, 641–671.
- Wald, A. (1947). *Sequential Analysis*. J. Wiley and Sons, New York, NY, USA.
- Welch, R. B. and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy, *Psychol. Bull.* **88**, 638–667.
- Welch, R. B., DuttonHurt, L. D. and Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception, *Percept. Psychophys.* **39**, 294–300.
- Wiecki, T. V., Sofer, I. and Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python, *Front. Neuroinform.* **7**, 14. DOI:10.3389/fninf.2013.00014.