Rate-dependent perceptual interactions between vibrotactile and visual pulses

Maxim A. Bushmakin & Robert Sekuler
Volcan Center for Complex Systems and
Department of Psychology, Brandeis University

Abstract
Many everyday tasks demand an ability to attend selectively to input from one modality while ignoring input(s) from another modality. We examined this ability for concurrent visual and vibrotactile inputs. Stimuli of both kinds were delivered via a handheld computer tablet. Subjects categorized the rate at which a visual target pulsed, while attempting to ignore concurrent task-irrelevant vibrotactile pulsations delivered to their hands. The visual stimulus was a luminance patch whose size alternated at either 4.28 or 6 Hz. Vibrotactile pulses were generated within the tablet, also at 4.28 or 6 Hz, and were delivered to subjects’ hands. Throughout, subjects categorized the rate of visual alternation as “Fast” or “Slow.” Responses were accurate and faster when the frequencies of concurrent vibrotactile and visual stimuli were matched than when they were mismatched. To characterize the rate at which sensory information accumulated under various conditions, subjects’ response speed and accuracy were analyzed jointly with a Linear Ballistic Accumulator model. The model revealed two distinct effects: facilitation (when visual and vibrotactile frequencies matched, accumulation of sensory information was sped up), and interference (when the frequencies were mismatched accumulation of sensory information was retarded). Of these two opposing effects, interference was somewhat stronger than facilitation. Our results confirm the benefits of using an information accumulation framework to interpret results from selective attention tasks, particularly tasks that entail multisensory perception.

Many everyday activities put a premium on attending selectively to inputs from one sensory modality while discounting inputs from another modality. In various circumstances, perception benefits when signals from multiple modalities are combined, but only when such inputs complement one another. For example, in audio-visual speech, observing a speaker’s mouth movements can greatly enhance the intelligibility of what is heard (Crosse,

We thank Paul Dizio for valuable assistance and encouragement at several stages of our project. MAB was supported in part by NIH Training Grant, T32 NS007292.
The potential benefits of multi-sensory combination are not limited, of course, to speech intelligibility. For example, perception of stimulus intensity can be affected by a concurrent stimulus in another modality (Stein, London, Wilkinson, & Price, 1996). Further, an auditory signal can speed localization of some target in a visual search array (Ngo & Spence, 2010; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008, 2009; Vroomen & Gelder, 2000), or can alter the perception of visual motion (Sekuler, Sekuler, & Lau, 1997; Zhou, Wong, & Sekuler, 2007). Moreover, multisensory facilitation effects have been considered for use in warning systems designed to boost speed of response (Santangelo, Ho, & Spence, 2008; Selcon, Taylor, Shadrake, & Farmer, 1991; Spence & Ho, 2008a).

Although powerful and beneficial multisensory interactions have been known since the earliest work on the topic (Wundt, 1896), potentially harmful interactions have attracted less attention. Generally, with uncorrelated inputs from multiple sensory modalities, selective attention must intervene in order to prioritize one signal stream over another. When such prioritization is weak, perceptual quality can be degraded. Spatial as well as temporal ventriloquism are good examples. In spatial ventriloquism, speech sounds emanating from one person are captured and misattributed to an altogether person or to the ventriloquist’s dummy. In the case of temporal ventriloquism, the temporal characteristics of a stimulus in one modality affects the perceived temporal characteristics of a concurrent stimulus in a different modality (Keetels & Vroomen, 2008). Effective selective attention can reduce or even eliminate inappropriate multisensory combination.

Although audio-visual interaction has been a frequent target of multisensory research (Calvert, Spence, & Stein, 2004), visual-vibrotactile interaction has attracted relatively little attention. Recently, the importance of understanding visual-vibrotactile interactions has increased because of common devices that exploit vibrotactile signals (e.g., cell phones and smart watches). For example, while steering a car, a human driver may need to devote attention to visual signals, that is, traffic conditions and pedestrians, while ignoring vibration of a cellphone that is in their pocket. However, some alerting and warning systems are designed to pulsate or vibrate in order to capture a user’s attention. Recognizing that the timing, duration, and other characteristics of vibrotactile notification signals could impact their effectiveness, we set out to examine how information from such stimuli might be accumulated over time.

Our study investigated the impact of vibrotactile pulse rate on the ability to process and categorize the rate of concurrent visual pulses. Subjects attempted to categorize the rate at which a visual stimulus pulsed, while they tried to ignore concurrent, task-irrelevant vibrotactile pulses. Instead of examining general, non-specific distraction effects, we devised a categorization task and complex stimuli that could reveal finer-grained, quantitative details of multisensory combination. Previous work with auditory and visual signals showed that accuracy in categorizing visual or auditory stimuli improves when the categorized stimuli are accompanied by frequency-matched stimuli from the other modality (Sun, Hickey, Shinn-Cunningham, & Sekuler, 2017).

While many previous studies focused primarily on the facilitatory effects of multisensory congruency, we were especially interested in the possible interference that might be produced by discrete signals whose rates were incongruent or mismatched. Moreover, most previous studies of multisensory interaction treated response accuracy and speed separately,
CONCURRENT VIBROTACTILE AND VISUAL STIMULI

forgoing the additional theoretical advantage that might come from a joint examination of the two response measures. We opted to analyze response accuracy and response speed together, using a Ballistic Accumulation model, a theoretical framework that characterizes the rate at which information is accumulated. This coordinated treatment of multiple dependent variables can bealogized to the way that signal detection theory (Green & Swets, 1966) allows for the joint analysis of hits and false alarms.

Recently, Varghese et al. (2017) used a general diffusion model to describe participants’ response times and choices in an audio-visual selective attention game. Their detailed account was largely limited to showing how various model parameters varied with experimental condition. We had a different objective; we sought a mechanistic account of the rate at which timing information from successive pulsed stimuli accumulates. In particular, we want to characterize the accumulation process in a way that allows for the possibility that stimulus conditions alter the process qualitatively. Specifically, we applied an information accumulation framework to model the perception of pulsating multisensory stimuli. Stimulus pulses, within either modality, visual or vibrotactile, were identical to one another. The pulses therefore were utterly uninformative about the rate at which pulses were delivered, the very variable subjects had to categorize. Therefore, our model assumes that subjects categorize pulse rate by drawing on a noisy representation of the intervals between pulses (see Fig. 1), rather than any characteristic of individual pulses themselves.

Various evidence-accumulation frameworks (Brown & Heathcote, 2008; Ratcliff & Rouder, 1998; Ratcliff, Thapar, & McKoon, 2001) have been useful in understanding perceptual decision making. In common with other general models of decision making (Green & Swets, 1966; Luce, 1986), the evidence accumulation framework assumes that a categorization response is based on perceptual evidence for or against that response. However, the evidence accumulation framework assumes the signal is not instantly extracted from the stimulus, but rather accumulates over time. In such models, the first accumulator to reach a decision threshold determines the ultimate response. Accumulator models can not only account for speed-accuracy trade-offs in many different tasks, but can also account for how sensory information is accumulated over time in various conditions of a single task, and the importance of individual differences in decision making (Brown & Heathcote, 2008; Ratcliff, Smith, Brown, & McKoon, 2016). Such models have been successfully used to understand decision making in visual perception, memory recall, and psychiatric disorders (Ho et al., 2014; Weigard, Huang-Pollock, & Brown, 2016), as well as in alcohol intoxication (van Ravenzwaaij, Dutilh, & Wagenmakers, 2012) and sleep deprivation (Ratcliff & Van Dongen, 2011).

Methods

Participants

Thirteen subjects, 20 to 29 years old, were recruited from the Brandeis University community. All had normal or corrected-to-normal visual acuity, as measured by a Snellen chart. Each gave written informed consent to a protocol approved by Brandeis University’s Committee for the Protection of Human Subjects, and received $10 for participation.
Apparatus & Stimuli

Visual and vibrotactile stimuli were delivered on a hand-held Samsung Note 10.1 tablet, using the Android 4.0 mobile operating system. This same system had been previously successfully used in a game-based study of audiovisual integration (Goldberg, Sun, Hickey, Shinn-Cunningham, & Sekuler, 2015).

![Figure 1](image_url)

**Figure 1.** (A) Photograph of a subjects’ hands holding the Samsung Note 10.1 tablet with which visual and vibrotactile stimuli were presented. (B) Schematic representations of visual and vibrotactile pulses. Visual pulses, presented at the center of the tablet, were generated by displaying a Gaussian disk whose size switched between large and small, either at 4.28 or 6 Hz. Vibrotactile pulses, generated inside the tablet, were each 30 ms in duration. (C) Schematic representation of the six experimental conditions: Control (4.28 and 6 Hz), M (4.28 and 6 Hz), and MM (4.28 and 6 Hz) conditions. In this panel, the two sizes of visual stimuli are represented by black and gray bars, for larger and smaller disks, respectively.

The experiment was coded in HTML5 and Javascript (Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016); data were saved locally on the Samsung tablet and then downloaded securely for off-line analysis. Subjects were fitted with supraural headphones that
played continuous white noise (56 dB$_{SPL}$) throughout the experiment to mask the sound generated by the tablet’s vibration.

During a trial, the tablet displayed a circular Gaussian blob whose size alternated between two clearly discriminable sizes, 0.25 (SD = 0.063) and 0.50 visual angle (SD = 0.12). Alternation was at either 4.28 or 6 Hz. The peak luminance of the Gaussian blob was 200 cd/m$^2$. The tablet screen was maintained at a constant, uniform luminance of 92 cd/m$^2$. For notational simplicity, hereafter the rates of size alternation will be described as 4 and 6 Hz.

Vibrotactile stimulation was generated by the tablet’s built-in, motor (model number GH59-11990A) whose rotating eccentric load causes the tablet to vibrate. The motor, located behind the tablet’s touch screen, rotates at 250 Hz, a frequency well above humans’ resolution threshold (Gescheider, 1966). To produce a vibrotactile stimulus pulse, the tablet’s motor was turned on and off. Each vibrotactile stimulus pulse was presented for 30 msec, and was well above detection threshold. The 30-msec long vibrotactile pulses were presented at either 4.28 or 6 Hz (again, we will refer to these as 4 and 6 Hz). These two rates was chosen because the literature suggested they would be readily, though imperfectly discriminated (Fujisaki & Nishida, 2009), which was confirmed by our own pilot studies. At a trial’s start, the visual and vibrotactile pulses were synchronized.

**Procedure**

Testing was conducted in a dimly lit, sound attenuated room. Subjects were instructed to hold the tablet in “landscape” orientation with the hands grasping opposing sides of the tablet (see 1A). Subjects held the tablet at an average distance of 38 cm from their eyes. Concurrent with the visual stimulation, the tablet delivered (i) vibrotactile pulses matched in rate to the visual pulses, (ii) vibrotactile pulses mismatched in rate to the visual pulses, or (iii) no vibrotactile pulses at all. Subjects were instructed to categorize the pulsing visual stimulus (as either slow or fast), while ignoring vibrations delivered to the hands. The stimuli on each trial continued until a response was made. Subjects signaled their responses by tilting the hand-held tablet, either away or toward themselves. To register as a response, the tablet had to be tilted 17 degrees or more relative to its orientation at the start of the trial. The next trial started 500 ms after the response.

Subjects were allowed brief breaks every 50 trials. All trial conditions were randomized and interleaved within and across all subjects. Each subject’s testing began with 150 practice trials. During practice, feedback was given after each response. This was followed by 900 test trials presented without feedback.

**Results**

Accuracy and speed of categorization were compared for conditions whose visual and vibrotactile stimulus rates were **Matched** (either both 4 Hz or both 6 Hz) or **Mismatched** (one rate 4 Hz, the other 6 Hz), and in a **Control** condition with visual pulses (either 4 or 6 Hz) but no vibrations. Figure 2 shows the categorization accuracy and reaction times for **Control, Matched** and **Mismatched** conditions. A repeated-measures ANOVA showed a significant difference in accuracy among **Matched**, **Mismatched**, and **Control** conditions ($F(2,24) = 26.49$, $p = <0.0001$, $\eta^2 = .39$). Planned comparisons confirmed a significant
Figure 2. Mean accuracy of responses in each of the six conditions (upper panel), and mean response time in those conditions (lower panel). Error bars represent standard error of the mean.

difference between Matched and Mismatched conditions ($t(38.28) = 4.39, p < .001$) as well as between Mismatched and Control conditions ($t(45.14) = 3.01, p = 0.004$), but not between Control and Matched conditions ($t(46.48) = 1.44, p = 0.155$). Repeated-measures also ANOVA showed a significant difference in reaction times among Matched, Mismatched and Control conditions ($F(2,24) = 12.6, p = 0.0002, \eta^2 = 0.073$). Planned comparison confirmed a significant difference between Matched and Mismatched conditions ($t(49.95) = 2.13, p = 0.0384$), but not between Mismatched and Control conditions ($t(49.94) = 1.62, p = 0.111$), and not between Control and Matched conditions ($t(49.71) = .57, p = 0.571$). Additionally to evaluate any practice effect we compared first half and the second half of the trials; there were no significant differences in either accuracy ($p = .34$) or response time ($p = .48$).

Modeling

To better understand the implications of relationship between our two dependent variables, accuracy and response time, results were modeled using an evidence accumulation framework. Specifically, we assessed the rate of evidence accumulation across different kinds of trials using the Linear Ballistic Accumulator Model (LBA; Brown & Heathcote,
The Linear Ballistic Accumulator model was chosen for its computation tractability and because it estimates evidence accumulation rates separately for correct and incorrect responses. These two estimates allow us to determine the amount of facilitation, expediting progress toward a correct response, and interference, slowing down that progress.

\[ v = \text{Rate for correct frequency} \]
\[ \sim v = \text{Rate for the incorrect frequency} \]

**Figure 3.** In the LBA model, information about stimulus rate accumulates over time (x-axis) until the information reaches the decisional threshold \( b \) (horizontal line). On Matched trials, the evidence accumulation favoring the correct choice of the rate is boosted by a parameter \( (f) \) and is reduced for the incorrect choice of the rate; on Mismatched trials, evidence accumulation for the correct choice of the rate is reduced by a parameter \( (h) \) and is increased for the incorrect choice of the rate.

Our model assumes that estimates of the interval between successive stimulus pulses is the evidence that accumulates about the rate of stimulus pulsation. Previous research has demonstrated that for low frequency vibrotactile stimuli, it was not the spike rate, but the time between spike bursts that coded the perceived frequency (Birznieks & Vickery, 2017). Thus at each moment, a subject can use the accumulated inter-pulse intervals as the basis for categorizing the rate of visual oscillation as either 4 Hz or at 6 Hz. If each sampled inter-pulse interval were an imperfect, noisy sample from a distribution centered on the true mean rate, subjects could gain from aggregating multiple samples over time. Recognizing the value of aggregating information over successive samples, our model defines \( v \) as the rate of evidence accumulation favoring a correct choice, and \( \sim v \) as the rate of evidence accumulation in support of a choice that is incorrect. If the trial had co-occurring, rate-
matched vibrotactile pulses, those pulses could boost the rate at which evidence accumulates in favor of the correct response $v = v + f$, and slow the rate of accumulation in favor of the incorrect response $-v = -v - f$, where $f$ is the amount of facilitation (Figure 3). If the concurrent visual and vibrotactile pulses were mismatched in rate, the vibrotactile pulses could slow accumulation of evidence toward the correct response $v = v - h$, but increase the rate of evidence accumulation toward the incorrect response $-v = -v + h$, where $h$ is the amount of interference (Selcon et al., 1991).

The model also takes into account the amount of initial evidence variability, designated in LBA as $A$, and the threshold amount of evidence needed in order for the subject to commit to a choice, designated as $b$ in the model. The initial evidence variability $A$ and threshold $b$ allow for individual variability among subjects independent of our experimental conditions, and thus will be used in fitting procedure but not be analyzed further.

Models were fit to each subject’s categorization responses and response times using maximum-likelihood estimation. Figure 5 shows model fits for the rate of accumulation, facilitation, and interference effects. The facilitation effects $f$ and interference effects $h$ were both significantly above zero, showing that both effects were present in the data ($t(12) = 5.68, p = .0001; t(12) = 3.13, p = .008$). The facilitation and interference parameters of the model signify how much does the visual evidence accumulation rate is aided or impeded by the co-occurring vibration pulses. Moreover, the interference effect was greater than the facilitation effect ($t(22.24) = 2.65, p = .014$). This suggests that the categorization of visual rate in visual-vibrotactile paradigm is much more affected by the presence of mismatched stimuli than was found in audio-visual tasks (Goldberg et al., 2015; Sun et al., 2017).

We also examined whether the facilitation and interference effects were related to each other. Such a relationship might help to explain individual differences among our subject. In particular, we wondered whether subjects differed in sensitivity to the effects of the vibrotactile stimulation. Such people, we reasoned, would suffer the greatest vibrotactile interference in Mismatched conditions, and the greatest vibrotactile facilitation in Matched conditions. Figure 6 shows that this was not the case. In fact, there was no appreciable relationship between individual subjects’ interference and facilitation parameters from the model. The correlation between the two parameters was not statistically reliable ($p = 0.93$). Future research might examine whether the obvious individual differences in the two model-derived parameters are related to a variable such as an individual’s need for closure (Evans, Rae, Bushmakin, Rubin, & Brown, 2017) or sensitivity to time pressure (van Maanen, Fontanesi, Hawkins, & Forstmann, 2016).

**Discussion**

Our study showed that concurrent vibrotactile stimulation can facilitate as well as interfere with categorization of the rate at which a visual stimulus pulses. Accuracy and latency in categorizing visual rate were consistently negatively affected when the rate of vibrotactile pulses did not match the rate of the visual pulses, showing that the Mismatched condition interfered with the accumulation of visual information. In contrast, when visual and vibrotactile rates were matched, accuracy and latency in categorizing visual rate were facilitated. By analyzing both reaction times and accuracies we are able to account for speed-accuracy trade-offs across subjects and build a more complete picture of decision making in the face of multisensory stimuli.
Note that our modeling approach takes a mechanistic approach to the way that subjects arrive at their perceptual judgments. The model incorporates two main processes: the accumulation of information from intervals separating successive visual pulses, and the way that such accumulation is affected by concurrent vibrotactile inputs. Importantly, our approach made it possible to examine results from Control, Matched and Mismatched conditions within a single model framework. That single model allowed us to better understand the relationship among factors entailed in the accumulation of information over time. Note that had response accuracy or response times been considered in isolation from one another, the analysis would have obscured the fact that vibrotactile stimulation can facilitate as well as interfere with judging visual rate.

These results seem to implicate some form of *temporal ventriloquism* where timing of a perceptual event in one sensory domain is associated with timing of a perceptual event in another sensory irrespective of their spatial locations. Similar results have been found with pairs of visual pulses whose perceived timing was shifted toward the timing of concurrent auditory or vibrotactile pulses (Keetels & Vroomen, 2008, 2012). This suggests that the brain resolves inconsistencies in timing of concurrent stimuli within different modalities by shifting the perceived temporal properties of events within one modality toward the perceived temporal properties of events in the other modality. Demonstrated cross-adaptation between perceived rates of visual and auditory stimulation (Levitan, Ban, Stiles, & Shimojo, 2015), and between auditory and tactile stimulation (Pérez-Bellido, Barnes, Crommett, & Yau, 2017) have led some researchers to hypothesize that a common, supramodal mechanism is important for processing timing information. Interestingly, that hypothesis is supported by a recent functional magnetic resonance imaging (fMRI) result showing that auditory frequency is broadly represented in human cerebral cortex, including in classically defined somatosensory cortex (Pérez-Bellido et al., 2017).

In this experiment we focused on the effect of temporal ventriloquism, however it is also important to consider is spatial location of visual and vibrotactile stimuli. In our experiment visual and vibrotactile pulses were generated by the tablet held by the participant. It has been shown visual and haptic integration relies on spatial proximity (Gepshtein, Burge, Ernst, & Banks, 2005) thus future work will be needed to examine the size of the effect as a function of spatial proximity.

It has been suggested that perception of a stimulus’ temporal properties in auditory and tactile sensory modalities is superior to that in vision (Keetels & Vroomen, 2012). This could bias people toward greater reliance on and trust of timing information in those perceptual domains. That difference in temporal processing among sensory modalities could explain why auditory and vibrotactile inputs can so strongly influence visual perception of temporal attributes (Shams & Kim, 2010).

The task our participants performed was designed to reveal how task-irrelevant vibrotactile signals influence perceived visual rate. In the future it would be useful to investigate the reciprocal effect, that is, how vibrotactile perception is, in turn, influenced by the rate of concurrent visual signals. The possibility of such effects is important for practical as well as theoretical reasons. Vibrotactile stimulation has great potential as a silent and invisible communication channel whose signals are reliable and easy-to-interpret (Cincotti et al., 2007). This potential is likely to be modulated by the attentional demands placed on the competing modalities, moreover the relative strength of influence in each direction is likely...
to be function of multiple variables, including the reliability of signals within each modality (Ernst & Banks, 2002).

In conclusion, we found that concurrent vibrotactile stimulation can facilitate as well as interfere with judgments of visual rate. The direction of the outcome depends upon the temporal relationship between signals in the two modalities. We believe that further study should be given to this bidirectional multisensory interaction whose sign switches with the frequency relationship between inputs. After all, vibrotactile signals play increasingly important roles in technologies and devices such as smartphones, smartwatches, and driver-assist systems (Ege, Cetin, & Basdogan, 2007; Spence & Ho, 2008b). Such studies would bring a better understanding of the perceptual and cognitive mechanisms involved in multisensory perception, and will guide engineering and development of more effective multimodal communication systems for everyday applications.

References


CONCURRENT VIBROTACTILE AND VISUAL STIMULI


Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization* (No. 8). Oxford University Press on Demand.


Figure 4. Cumulative Density Functions demonstrating model fits to reaction times across subjects. Symbols represent observed data, and lines show model fits (green for correct responses, red for incorrect responses).
Figure 5. A) Evidence accumulation rates estimated by the sequential sampling model, B) interference and facilitation of accumulation rates for congruent and incongruent conditions.
Figure 6. Values of the accumulator model’s Interference parameter ($h$) plotted against its Facilitation ($f$) parameter (in arbitrary units) for individual subjects. Note that there was no reliable correlation between values of the two parameters.