**Importance of spike timing in touch: an analogy with hearing?**
Hannes P Saal¹, Xiaqin Wang² and Sliman J Bensmaia¹

Touch is often conceived as a spatial sense akin to vision. However, touch also involves the transduction and processing of signals that vary rapidly over time, inviting comparisons with hearing. In both sensory systems, first order afferents produce spiking responses that are temporally precise and the timing of their responses carries stimulus information. The precision and informativeness of spike timing in the two systems invites the possibility that both implement similar mechanisms to extract behaviorally relevant information from these precisely timed responses. Here, we explore the putative roles of spike timing in touch and hearing and discuss common mechanisms that may be involved in processing temporal spiking patterns.

**Introduction**

Touch has traditionally been conceived as a spatial sense, drawing compelling analogies to vision. Indeed, both modalities involve a two-dimensional sensory sheet tiled with receptors (the retina and the skin) that each respond to local stimulation (radiant or mechanical). In both modalities, the spatial configuration of the stimulus is reflected in the spatial pattern of activation across the receptor sheet. In both modalities, higher order stimulus representations — of object shape and motion — are remarkably analogous [1–4]. The similarity between tactile and visual representations has been used as powerful evidence for the existence of canonical computations: the nervous system seems to implement similar computations to extract similar information about the environment, regardless of the source of this information [5].

As compelling as the visual analogy is, however, there are aspects of touch that flout it, in particular its temporal precision and the putative functional role thereof. Indeed, cutaneous mechanoreceptive afferents respond to skin stimulation with sub-millisecond precision, and the relative latencies of the spikes evoked across afferents are highly informative about contact events [6**]. Furthermore, afferents respond to skin vibrations up to about 1000 Hz in a precisely phase-locked manner. Their responses to sinusoids, for example, are restricted to a small fraction of each stimulus cycle over the range of tangible frequencies [7,8,9*,10**,11]. This temporal pattern underlies our ability to distinguish the frequency of skin vibrations and even to discern fine surface texture. At a first approximation, these aspects of touch are more similar to hearing than they are to vision.

In the present essay, we examine the role of spike timing in the processing of tactile stimuli and draw analogies to hearing. Hearing, like touch, involves a highly temporally precise stimulus representation at the periphery: relative spike latencies across cochlea play a role in sound localization and the phase locking of auditory afferents contributes to pitch and timbre perception. First, we discuss potential analogies between the use of delay lines and coincidence detectors for auditory localization and for the tactile coding of contact events. Second, we explore parallels in the way the somatosensory and auditory systems extract information about the frequency composition of skin vibrations and sound waves, respectively.

**Computing from differences in spike latency**

One of the most remarkable examples of the role of spike timing in extracting information about the environment is in sound localization. Indeed, the relative time at which an acoustic stimulus reaches the two ears depends on the azimuthal location of the source. The small temporal disparities in the relative arrival of the stimulus at each eardrum — measured in the tens to hundreds of microseconds — are exploited to compute its azimuth using precisely timed excitatory and inhibitory interactions (in mammals) [12]. Specifically, neurons in the medial superior olive receive excitatory input from both cochleae, and strong and precisely timed inhibitory input from the contralateral one. As the relative timing of all excitatory and inhibitory inputs depends on azimuth, so does the strength of the response, which confers to it a selectivity for location [13**]. This circuit implements a form of coincidence detection based on excitatory and inhibitory interactions (Figure 1a).
Exploiting first spike latencies in hearing and touch. (a) Precise spike timing is used in hearing to localize sound sources. Sound from a source towards the left will excite hair cells in the left ear (L) before hair cells in the right ear (R). Precisely timed excitatory and inhibitory inputs will reach an output cell (O) at different times, determining the strength of the response. (b) Potential use of delay lines in touch. Touching an object of a given curvature will excite some tactile afferents earlier than others. Nerve fibers can be exploited as delay lines to detect the specific sequence of afferent firing by neurons in cuneate nucleus.

The sense of touch may deploy an analogous mechanism to rapidly determine the properties of objects upon first contact. Indeed, tactile afferents have been shown to encode information about object curvature [14], the direction at which forces are applied to the skin [15], and the torque applied to the skin [16] (among other features), in the relative latency of their initial responses. Thus, changes in object properties lead to robust and repeatable patterns of relative latencies across afferents with spatially displaced receptive fields [6*]. As with azimuthal location in hearing, then, the information is carried in the relative timing of responses across spatially displaced receptors. Thus, to the extent that these highly informative latency patterns are exploited to extract feature information, it is likely that a mechanism akin to intraural time difference detection is involved. A population of coincidence detectors in the cuneate nucleus, the first synapse for touch signals, could in principle extract feature information from patterns of first spike latencies. In fact, a simpler mechanism than that for mammalian auditory localization might be at play in touch. Because the conduction velocities of individual tactile nerve fibers vary over a range [17,18], they naturally act as delay lines; tactile features could therefore be extracted even without the need for precisely timed inhibition, a computation similar to that underlying auditory localization in birds [19,20] (Figure 1b).

**Temporal spiking patterns and frequency coding**

**Acoustic stimuli and skin vibrations**

A primary function of the peripheral hearing organ (the cochlea) is to extract information from time varying pressure oscillations, transmitted from the ear drum, through the ossicles, and ultimately transduced in the inner ear. In touch, the transduction and processing of skin vibrations is, at least at first glance, analogous in that these too carry behaviorally relevant information in their time-varying waveforms [21]. First, just as we can discriminate the pitch of a pure or complex tone, we can distinguish the frequency of a sinusoidal skin vibration, though at a much lower resolution — with Weber fractions of 0.2–0.3 versus 0.003 — and over a much narrower range — up to 1000 Hz versus 20 000 Hz. Second, in both hearing and touch, sinusoidal stimuli can be detected without evoking a pitch percept at low amplitudes, over a range known as the atonal interval [22,23]. The atonal interval coincides with the range of amplitudes over which phase locking of afferents in the nerve is either weak or absent [8]. Third, in both hearing and touch, low frequency stimuli (<50 Hz) elicit a sensation of flutter that is devoid of pitch and in which individual stimulus cycles are discriminable [24,25], while high-frequency stimuli evoke percepts in which individual pulses are fused into a singular percept, to which a pitch can be ascribed [26].

**A spatial code for frequency?**

A remarkable aspect of hearing is the frequency decomposition engendered by the mechanics of the cochlea: Because the resonance frequency of the basilar membrane progresses systematically along its length, the surface wave produced by a pure tone peaks in amplitude in a different region of the membrane depending on its frequency, with low frequency tones peaking near the apex and high frequency tones peaking near the base. Hair cells in the resonating region respond more than others elsewhere, so different frequencies of stimulation maximally excite different populations of hair cells. Stimulus frequency is thus encoded spatially at the very first stage of auditory processing.
The sense of touch comprises three different populations of mechanoreceptive afferents, each of which responds to different aspects of skin deformations (excluding one population, which responds predominantly to skin stretch). One well-documented difference across afferent types is in their frequency selectivity (Figure 2a). Indeed, slowly-adapting type 1 (SA1) afferents are most sensitive below 5 Hz, rapidly adapting (RA) afferents respond best between 10 and 50 Hz (in the so-called flutter range), and Pacinian (PC) afferents peak in sensitivity at around 250 Hz [7,8]. While it is tempting to draw parallels between this frequency tuning and its counterpart in the auditory periphery, careful examination suggests that it plays little role in the frequency decomposition of skin vibrations. First, the three relevant tactile afferent classes, which are only very broadly tuned for frequency, would lead to a far poorer frequency resolution than is observed if relative activation level were used to decode frequency (using a scheme akin to that used for visual color perception [27]). Second, complex stimuli with distinct frequency components that excite the three populations of afferents equally are nonetheless distinguishable perceptually [9]. Third, frequency differences can be discerned in pairs of stimuli that engage only a single class of afferents [28]. The available information about frequency is therefore not encoded spatially in the relative activations of the three populations of afferents.

The role of spike timing in frequency coding

Cutaneous mechanoreceptive afferents respond with precisely timed spiking patterns to sinusoidal vibrations. They produce phase-locked responses to sinusoidal vibrations, systematically spiking within restricted phases of each stimulus cycle (Figure 2b) [7,8,11]. They also exhibit temporally patterned responses to complex, naturalistic skin vibrations, such as those elicited when we run our fingers across a finely textured surface (Figure 2c), and this temporal patterning underlies our capability to discern the frequency composition of simple and complex skin vibrations [9].

In hearing, the extraction of frequency composition also relies, at least in part, on a temporal mechanism. Indeed, like their tactile counterparts, auditory afferent responses are phase locked to acoustic stimuli, be they pure tones [29] or complex sounds [30], up to several thousands of Hertz. Is this high-precision temporal patterning epiphenomenal, reflecting the process of mechanotransduction, or does it play a role in encoding frequency composition?

Circumstantial evidence suggests that a spatial code cannot account for our remarkable auditory sensitivity to differences in frequency. First, as the loudness of a tone increases, (a) the surface wave on the basilar membrane grows wider, which leads to the progressive recruitment of hair cells with increasingly different best frequencies, and (b) afferents most sensitive to the center frequency saturate [31]. If frequency coding relied on a spatial code, our capacity for pitch discrimination would get worse with increasing loudness. That this is not the case suggests that the phase locking, which only gets stronger with increased loudness, plays a role in resolving pitch [32,33]. Furthermore, harmonic tones with missing lower harmonics elicit a pitch percept at their fundamental frequency [34]. The higher harmonics of such tones

---

**Figure 2**

Coding of skin vibrations in the somatosensory nerves. (a) Absolute threshold for three types of cutaneous mechanoreceptive afferents over a range of frequencies. Different afferent types are most sensitive at different frequencies. (b) Neural responses (red) of a PC afferent to five repeated presentations of a 400 Hz sinusoidal skin oscillation (black). Responses are repeatable and tightly locked to the stimulus waveform. (c) Responses of a PC afferent to 30 presentations of a texture (nylon, see inset showing the surface profile of a 7 x 7 mm patch), scanned across the skin at 80 mm/s. The fine structure of the texture is reflected in precise and repeatable temporal spiking patterns.
cannot be resolved spatially on the basilar membrane but can
be extracted from the beat frequency in the phase-locked
responses (that is, the rate at which these wax and wane) [35].
A subpopulation of neurons in auditory cortex encodes the
fundamental frequency of harmonic complex tones, even if
it is missing, demonstrating an explicit extraction of the
fundamental has been completed at this stage [36**].

In summary, while it has not been conclusively demon-
strated that phase locking in afferents plays a major role in
the coding of auditory frequency, it is unlikely that our
ability to distinguish changes in auditory frequency as
small as 0.3% is mediated by the relatively coarse spatial
coding of frequency over the basilar membrane. Rather,
this spatial code is thought to be complemented by a
temporal one, in which high-precision patterning in au-
ditory afferent responses convey precise information
about the frequency composition of acoustic stimuli.
Moreover, studies on speech coding have long suggested
that temporal coding in the auditory nerves plays a crucial
role in representing vowels across a wide range of sound
levels and in noisy environments [30,37].

**Transformation from temporal to rate code**

While auditory afferents produce high-precision temporal
spiking patterns, this patterning becomes less and less
prevalent as one ascends the auditory neuraxis. Indeed,
while auditory afferents can follow at frequencies up to
5000 Hz, the upper limit of entrainment in the inferior
colliculus is around 1000 Hz, and drops to below 100 Hz
in auditory cortex [38]. The precise temporal information
is successively converted into a rate code [39]: In cortex,
high frequency tones are encoded in firing rates, not
phase-locked spiking patterns.

While neurons in primary auditory cortex can phase-lock
only up to about 100 Hz, neurons in Brodmann’s area 3b of
primary somatosensory cortex exhibit phase locking up to at
least 800 Hz [40,41*], an observation that might be inter-
preted as evidence that different principles are at play in the
two modalities. However, area 3b is only three synapses
away from afferents (via the cuneate nucleus and the
thalamus), whereas primary auditory cortex is five synapses
away (via the cochlear nucleus, olivary nucleus, inferior
colliculus, and the thalamus), so temporal patterns in touch
might not have been subject to the gradual conversion to a
rate code to the degree that their auditory counterparts have
[39]. In fact, the peak frequency for phase-locking drops
abruptly in Brodmann’s area 1 (the next stage of processing
in somatosensory cortex) and phase locking is almost
completely abolished in area 2 [40], so a gradual conversion
is also observed along the somatosensory neuraxis.

**Figure 3**

Potential mechanisms for exploiting precise spike timing in pitch processing. (a) Delay lines can be used to detect periodic spiking activity in the
nerve at a given frequency. In this illustration, I denotes the input neuron, while O denotes the output neuron, which receives input from I through
a fast connection (1) as well as a delay line (2), which temporally shifts the neural responses by a constant delay. Coincidence detection will
confer to O a selectivity for a frequency determined by the delay. (b) Coincidence detection on several inputs can be used in order to extract the
fundamental frequency from complex harmonic tones. Here, the output neuron receives input from two neurons. Coincident spikes in the input
(indicated in red) elicit output spikes, while other spikes will not. (c) Lateral inhibition mediated by detecting phase differences in neighboring
neurons can be used to sharpen spatial representations of spectral frequency composition. Inhibitory interneurons are denoted by L. The spatially
distributed response at the periphery results in a broad spatial pattern of activation which is converted to a sparser spatial representation through
precisely timed inhibition.
Mechanisms for processing temporal information

Several putative mechanisms have been proposed to explain how precise spike timing contributes to pitch coding.

First, pitch might be extracted with delay lines and coincidence detectors: If a coincidence detector receives input from two afferents, with the second input delayed relative to the first, the neuron will only receive coincident spikes when the delay corresponds to the cycle duration, which will confer to it a preference for the corresponding frequency [42–44] (Figure 3a). Because this approach relies on calculating autocorrelation in the signal, it can extract periodicity that might not be evident from a spectral decomposition. Consistent with this idea, listeners can assign a pitch to short repeated bursts of white noise, even though the repetition rate is obscured in its spectrum [25]. While models based on delay lines can explain some aspects of pitch perception (including missing fundamentals), they imply the existence of delay lines lasting up to tens of milliseconds to code for low frequencies. Neurons exhibiting such long delays have been elusive, however.

Second, neurons might simply detect spikes that arrive within a short time window. Suppose a neuron has a short integration time window of less than 5 ms. If this neuron receives input from one afferent phase-locked at 400 Hz and another neuron phase-locked at 600 Hz, it will experience coincident spikes every 5 ms, and therefore phase lock at 200 Hz, the fundamental frequency (Figure 3b). In this way, coincidence detectors can exploit temporal information to extract pitch without the need for delay lines [33*]. Furthermore, coincidence detectors can easily be realized in neural circuits and provide a powerful framework for computation across different senses [46].

Figure 4

(a) Timbre

(b) Texture

(c) Power

(d) Power

Auditory timbre and tactile texture. (a) Spectrograms of two tones (F3 and C3, different rows), played by two different instruments (horn and viola, different columns). (b) Spectrograms of skin oscillations elicited by two different textures (vinyl and silk) when scanned at different speeds across the fingertip skin (80 and 120 mm/s). (c,d) Time-averaged power spectra of sound waves (c) and skin vibrations (d) when correcting for fundamental frequency/scanning speed by shifting the spectra of the lower tone/speed (red trace) to the higher one (black trace). The general spectral composition is preserved across pitch/speed and can be used to resolve different instruments or textures.

Source: [Sound files were obtained from the University of Iowa Electronic Music Studios database, http://theremin.music.uiowa.edu].

Current Opinion in Neurobiology 2016, 40:142–149
Third, spike timing might play a role in sharpening or refining spatial representations of pitch. Nearby neurons along the tonotopic axis are known to inhibit each other. One possibility is that this lateral inhibition is modulated by the degree of synchronized phase locking in the responses of mutually inhibitory neurons, with high synchrony leading to less inhibition and low synchrony leading to high inhibition [32, 47] (Figure 3c). Because the different harmonics of a complex tone are shifted in phase, individual harmonics will become more distinct despite the fact that their initial spatial representation is overlapping on the basilar membrane.

**Timbre and texture**

Auditory timbre is an elusive construct, one that is associated with the harmonic structure of acoustic stimuli and underlies the perceptual differences between musical instruments or voices [48]. A key property of timbre is that it remains relatively constant over changes in pitch: different notes played on a piano sound like a piano; a voice remains recognizable across intonations. This invariance implies that some aspect of the harmonic structure is preserved when the fundamental frequency changes, and that the auditory system is able to extract this invariant structure [49].

A similar phenomenon underlies the tactile perception of texture. When we run our fingers across a textured surface, small vibrations are produced in the skin [50–52] and the spectral composition of these vibrations reflects the surface microstructure. In turn, these vibrations evoke highly repeatable temporal spiking patterns which, as discussed above, convey information about the frequency composition of the texture-elicted vibrations. The frequency composition of the texture vibrations is also dependent on the scanning speed: The power spectrum translates right or left along the frequency axis with increases or decreases in speed [51]. As a result, afferent responses to texture dilate or contract with decreases or increases in scanning speed (which amounts to left or right shifts of the spectrum along the frequency axis) [10**]. While afferent responses to textured surfaces are highly dependent on scanning speed, however, the perception of texture is completely invariant with respect to speed [53, 54].

Thus, timbre reflects the harmonic structure of an acoustic stimulus independently of the left and right shifts along the frequency axis that accompany changes in the fundamental, and texture invariance reflects the harmonic structure of skin vibrations independently of the left and right shifts along the frequency axis that are caused by changes in scanning speed (Figure 4). Texture information is conveyed in temporal spiking patterns evoked in mechanoreceptive afferents [10**], and it is likely that timbre information is, at least in part, conveyed in temporal spiking patterns evoked in auditory afferents [30].

The remarkable parallels between auditory timbre and tactile texture suggest that they rely on similar neural mechanisms, which have yet to be elucidated [55*, 56].

**Spike timing in touch and hearing**

While precise timing in tactile afferent responses was documented long ago, it has only recently been ascribed an ecological role, namely in the perception of contact events and of surface texture. The role of spike timing in hearing has been somewhat obscured by the spatial decomposition of the frequency that is implemented in the cochlea. Indeed, the role of timing in auditory representations of frequency is hard to isolate given that temporal codes coexist with spatial ones throughout the auditory neuraxis. Perhaps, as we discover how temporal representations in touch are transformed into rate-based representations, we will gain a better understanding of the role of spike timing in hearing.

**Conflict of interest statement**

Nothing declared.

**Acknowledgements**

We would like to thank Jeffrey M. Yau for comments on a previous version of this manuscript. This work was supported by National Science Foundation (NSF) grant IOS 1150209 and National Institutes of Health (NIH) grant NS 095162 to SJB, as well as National Institutes of Health (NIH) grant DC 003180 to XW.

**References and recommended reading**

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


This study showed that the first spike latencies evoked in a population of tactile afferents while touching an object carry considerable information about the properties of the touched surface.


9. Mackeivichis EL, Best MD, Saal HP, Bensaia SJ: Millisecond precision spike timing shapes tactile perception. J Neurosci 2012, 32:15309-15317. This study was among the first to demonstrate the behavioral relevance of precise spike timing in the processing of skin oscillations.


36. Bendor D, Wang X: The neuronal representation of pitch in primate auditory cortex. Nature 2005, 436:1161-1165. This study showed that the auditory cortex in primates contains pitch-sensitive neurons which respond to the fundamental frequency of a sound even in the absence of that fundamental from the harmonic stack.


41. Saal HP, Harvey MA, Bensaia SJ: Rate and timing of cortical responses driven by separate sensory channels. Elite 2015, 4:7250-7257. This study demonstrated that precise spike timing is preserved through precisely timed inhibition and through the interplay of different tactile submodalities in primary somatosensory cortex.


55. Town SM, Bizley JK: Neural and behavioral investigations into timbre perception. Front Syst Neurosci 2013, 7:88. This paper provides a comprehensive review of our current understanding of auditory timbre perception.