Spatiotemporal Properties of Vestibular Responses in Area MSTd

Christopher R. Fetsch,1 Suhrud M. Rajguru,1 Anuk Karunaratne,1,2 Yong Gu,1 Dora E. Angelaki,1,2,* and Gregory C. DeAngelis3,*

1Departments of Anatomy and Neurobiology and 2Biomedical Engineering, Washington University School of Medicine, St. Louis, Missouri; and 3Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York

Submitted 21 November 2008; accepted in final form 10 July 2010

Fetsch CR, Rajguru SM, Karunaratne A, Gu Y, Angelaki DE, DeAngelis GC. Spatiotemporal properties of vestibular responses in area MSTd. J Neurophysiol 104: 1506–1522, 2010. First published July 14, 2010; doi:10.1152/jn.91247.2008. Recent studies have shown that many neurons in the primate dorsal medial superior temporal area (MSTd) show spatial tuning during inertial motion and that these responses are vestibular in origin. Given their well-studied role in processing visual self-motion cues (i.e., optic flow), these neurons may be involved in the integration of visual and vestibular signals to facilitate robust perception of self-motion. However, the temporal structure of vestibular responses in MSTd has not been characterized in detail. Specifically, it is not known whether MSTd neurons encode velocity, acceleration, or some combination of motion parameters not explicitly encoded by vestibular afferents. In this study, we have applied a frequency-domain analysis to single-unit responses during translation in three dimensions (3D). The analysis quantifies the stimulus-driven temporal modulation of each response as well as the degree to which this modulation reflects the velocity and/or acceleration profile of the stimulus. We show that MSTd neurons signal a combination of velocity and acceleration components with the velocity component being stronger for most neurons. These two components can exist both within and across motion directions, although their spatial tuning did not show a systematic relationship across the population. From these results, vestibular responses in MSTd appear to show characteristic features of spatiotemporal convergence, similar to previous findings in the brain stem and thalamus. The predominance of velocity encoding in this region may reflect the suitability of these signals to be integrated with visual signals regarding self-motion perception.

INTRODUCTION

The dorsal subdivision of the medial superior temporal area (MSTd) is well known as an area that processes complex optic flow patterns, such as those experienced during self-motion (Duffy and Wurtz 1995, 1991; Tanaka and Saito 1989; Tanaka et al. 1986; Warren 2003). Previous work has also shown that MSTd neurons are selective for the direction of physical motion in darkness (Bremmer et al. 1999; Duffy 1998; Page and Duffy 2003), suggesting that this area may contribute to the integration of visual and nonvisual self-motion cues. More recently, the responses of MSTd neurons to translation and rotation have been characterized in 3D (Gu et al. 2006; Takahashi et al. 2007). Labyrinthotomies experiments (Gu et al. 2007; Takahashi et al. 2007) strongly suggest that MSTd responses during physical motion originate from the vestibular labyrinths, and these responses have been linked to perceptual performance in monkeys trained to report their direction of self-motion (Gu et al. 2007, 2008, 2010). However, previous studies have quantified vestibular responses in MSTd based only on mean firing rate (MFR) within a particular stimulus interval (Duffy 1998; Fetsch et al. 2007; Gu et al. 2006; Page and Duffy 2003; Takahashi et al. 2007). A more complete understanding of the nature of vestibular signals in MSTd requires a detailed analysis of their spatiotemporal dynamics.

Primary ototh sensory afferents encode linear acceleration of the head in space (Fernandez and Goldberg 1976; Loe et al. 1973; Si et al. 1997). In contrast, central neurons as early as the brain stem vestibular nuclei (VN) and vestibulo-cerebellum show a diversity of response dynamics during translation. The majority of these neurons appear to encode velocity or combinations of acceleration and velocity (Angelaki and Dickman 2000; Dickman and Angelaki 2002; Musallam and Tomlinson 2002; Shiakh et al. 2005; Yakusheva et al. 2008). Response dynamics have also been shown to vary with the direction of motion (Angelaki and Dickman 2000; Musallam and Tomlinson 2002). Using position-transient stimuli, Musallam and Tomlinson (2002) observed spatially asymmetric temporal integration in a population of VN neurons. These neurons showed velocity-like responses to translation in one direction, but acceleration-like responses during motion in the opposite direction. The authors propose a cellular mechanism to explain this asymmetry, but its function remains unknown. Angelaki and Dickman (2000) have reported a similar variation of temporal dynamics across motion directions for VN neurons, suggesting a form of spatiotemporal filtering via convergence of afferents with different spatial and temporal response properties (Angelaki 1993a,b).

Beyond the VN, very little is known about the full range of vestibular response dynamics present in the brain and even less about the functional roles of these diverse signals for self-motion perception and spatial orientation. An important step toward addressing these questions is to characterize response dynamics at multiple sites along the vestibular sensory pathway. To this end, Meng et al. (2007) recorded translation responses in the primate thalamus, a probable source of vestibular inputs to cortical regions such as MSTd. Similar to the VN and cerebellar nuclei, thalamic neurons showed a broad distribution of response phases during motion in their preferred direction (Meng et al. 2007). A goal of the present study was to continue tracking the evolution of ototh-driven signals into the cortex, specifically in area MSTd.

Whether MSTd neurons encode velocity, acceleration, or a combination of kinematic variables remains unclear. Gu and colleagues (2006) reported that >90% of MSTd responses to translation in the preferred direction were better correlated with
velocity than acceleration. However, only responses in the preferred direction were characterized, leaving open the possibility that single MSTd neurons might encode both velocity and acceleration for different directions of motion. Moreover, the possibility that individual temporal response profiles could contain both velocity and acceleration components was not tested.

In this report, we analyzed the vestibular translation responses of MSTd neurons to investigate if they reflect primarily velocity or acceleration. For this purpose, we developed a method using Fourier analysis to characterize the temporal dynamics of responses to a transient motion stimulus. This method provides a metric of the overall temporal modulation that can be attributed to the stimulus as well as quantifying the velocity and acceleration components within each response. We compared this frequency domain method to a model-fitting approach that produced similar results. We further analyzed the spatial distribution of velocity and acceleration components (e.g., how they vary across motion directions for a given cell), to investigate the role that temporal response structure might play in direction selectivity.

**METHODS**

**Experimental setup, stimuli, and procedures**

The vestibular responses included in the present analysis were recorded in area MSTd of three male rhesus monkeys (*Macaca mulatta*). The spatial tuning of these neurons based on mean firing rate has previously been described (Fetsch et al. 2007; Gu et al. 2006; Takahashi et al. 2007). Here we have used data from these previous studies to perform temporal analyses, and thus we only provide a brief description of the experimental setup. Animals were chronically implanted with a halo-style head stabilization device and one or two scleral search coils for measuring eye position (Judge et al. 1980). A plastic recording grid was secured to the inside of the head cap such that it extended from the midline to the area overlying MSTd bilaterally. Vertical microelectrode penetrations were made via transdural guide tubes inserted in the grid holes. All procedures were approved by the Institutional Animal Care and Use Committee at Washington University and were in accordance with National Institutes of Health guidelines.

Spike times from well-isolated MSTd neurons were collected while animals fixated a central head-fixed target during translation on a motion platform (MOOG 6DOF2000E; MOOG, East Aurora, NY). In each trial, the animal was moved along 1 of 26 trajectories sampled uniformly on a sphere (45° apart in azimuth and elevation). The width of the fixation window was kept no larger than 2° (1° for 1 animal, 1.5–2° for the other 2) to minimize potential contributions of smooth eye movement signals. Motion directions were parameterized using two angles: azimuth and elevation (spherical coordinates). The stimuli had a Gaussian velocity profile and a corresponding biphasic acceleration profile with the following parameters: duration = 2 s, amplitude (total displacement) = 0.13 m, peak acceleration = 0.09 G (0.87 m/s²), peak velocity = 0.27 m/s (Fig. 1A). Neurons were included in the sample if each motion trajectory was successfully repeated at least three times.

**Data analysis**

All analyses were performed off-line using MATLAB Version 6.5 (Mathworks, Natick, MA). Peristimulus time histograms (PSTHs) were generated for each of the 26 directions by binning the spike counts (averaged across repetitions) in 20-ms bins. Repeating the analyses using a larger bin width (50 ms) did not substantially change the pattern of results. The frequency content of each PSTH was characterized using the discrete Fourier transform (DFT). Specifically, we computed the frequency components from the DFT performed on a selected 2-s window within each PSTH (see following text). This was accomplished by 1) subtracting the mean value across the selected window from each PSTH bin, 2) computing the DFT of the mean-subtracted PSTH using the Matlab function *fft*, 3) computing the amplitude of the frequency components by taking the absolute value (complex modulus) of the Fourier-transformed PSTH, and 4) computing the phase angle (θ) of each component using the Matlab function *angle*, which computes the four-quadrant arctangent of the real (x) and imaginary (y) parts of the transformed data.

We then introduced the DFT ratio (DFTR) as a measure of temporal modulation strength for a given direction of motion. The
DFTR was defined as the ratio of the average amplitude from 0.5 to 2.5 Hz to the average amplitude from 3 to 25 Hz:

\[
\text{DFTR} = \frac{\frac{1}{5} \sum_{i=1}^{5} \text{amp}_i}{\frac{1}{20} \sum_{i=1}^{20} \text{amp}_i}
\]

Here, \(\text{amp}_i\) is the amplitude of the \(i\)th frequency component, ranging from 0 to 24.5 Hz in increments of 0.5 Hz [a 2-s window with a sampling rate of 50 Hz (20-ms bins) gives a 100-point DFT and a Nyquist frequency of 25 Hz, yielding 50 components at 0.5-Hz intervals]. We chose these particular frequency ranges because most of the amplitude in the stimulus was below 3 Hz (99.7% for the velocity profile and 99.1% for the acceleration profile; Fig. 2, D–F). It is thus reasonable to assume that for a given response, any frequency component >2.5 Hz was not driven by the stimulus. Accordingly, the DFTR provides a measure of the stimulus-driven temporal modulation (low-frequency components) relative to the noise (high-frequency components) present in the response. In principle this ratio can range from 0 (pure high-frequency noise) to an arbitrarily high number (a noiseless low-frequency response), but in practice, the range for our data set was 0.29–7.80.

The statistical significance of the DFTR was assessed for each neuron and each stimulus direction using a permutation test based on 1,000 random reshufflings of binned spike counts. From visual inspection of hundreds of individual PSTHs from our sample (blind to the results of the permutation test), a significance criterion of \(P < 0.01\).

**FIG. 2.** Example temporal response profiles and discrete Fourier transform ratio (DFTR) analysis. A–C: example PSTHs for a velocity-like response (A), an acceleration-like response (B), and a weakly modulated response (C) with temporal envelopes (red traces) superimposed. The delay of each response relative to the stimulus was calculated as the center of mass of the temporal envelope, and the 2-s window centered on this delay was selected for the DFTR analysis (black bars). Acceleration (cyan) and velocity (green) traces are shown (normalized, vertically offset, and shifted horizontally by the delay) for display purposes, to illustrate their similarity (or lack thereof) with the neural response. D–F: Fourier amplitude spectra (black lines) for the responses shown in A–C, respectively. Amplitude spectra for stimulus velocity (green) and acceleration (cyan) profiles are shown for comparison (normalized to the peak value of the response amplitude spectrum in each plot). The DFTR is defined as the ratio \(A/B\), where \(A\) is the mean amplitude of the frequency components below and including 2.5 Hz (gray area), and \(B\) is the mean amplitude of the components >2.5 Hz (yellow area). G–I: phase spectra for the responses shown in A–C for low-frequency components only (0.5–2.5 Hz). Phase spectra for the velocity (green) and acceleration (cyan) stimulus profiles are also shown.
was chosen for this test because it produces a classification of significant temporal responses that generally agrees well with qualitative classification by eye. In addition, DFTR values were only considered significant if they were greater than the DFTR computed from “null” trials in which the motion platform remained stationary while the monkey was required to fixate the target. This additional criterion was designed to reject a small proportion of cases in which responses to inertial motion showed some structure but did not exceed the temporal modulation observed during fixation alone, for instance due to lingering visual responses to the fixation point or responses time locked to the onset of ocular fixation. Out of 10270 total directions tested (395 cells * 26 directions), 4,196 responses (40.9%) passed the permutation test, and 179 (1.7%) were eliminated by the “greater-than-null” criterion.

To quantify the degree to which stimulus velocity and/or acceleration were reflected in the neural responses, we computed velocity and acceleration DFTR “components” (DFTRvel and DFTRacc) as follows. The amplitude of each low-frequency component (0.5-2.5 Hz) was rescaled by multiplication with the negative cosine (velocity) or sine (acceleration) of its phase angle (θ). The velocity and acceleration DFTR components were then computed from the rescaled amplitudes

\[
\text{DFTR}_{\text{vel}} = \sum_{i} \frac{1}{5} (-\cos(\theta) \cdot \text{amp}_i)
\]

\[
\text{DFTR}_{\text{acc}} = \sum_{i} \frac{1}{44} \sin(\theta) \cdot \text{amp}_i
\]

The logic behind scaling the amplitudes by the cosine or sine of the phase angle can be summarized as follows. Note that the dominant phase of the Gaussian stimulus velocity profile is 180°, whereas the dominant phase of the acceleration profile is 90° (Fig. 1A). Thus frequency components that reflect stimulus velocity will have a phase near 180°. Such components will contribute near-maximally to the DFTRvel [-cos(180°) = 1] but contribute very little to the DFTRacc [sin(180°) = 0], and vice versa for components that reflect acceleration (phase near 90°). Using these definitions, negative DFTRvel or DFTRacc values will arise naturally when a cell shows an inhibitory (suppressed) response in correspondence with the velocity [phase near 0° (or 360°)] or acceleration [phase near –90° (or 270°)] stimulus profiles. Motion directions were considered to have statistically significant velocity or acceleration DFTR components if (1) their corresponding “total” DFTR was significant (see preceding criteria) and 2) they passed a separate permutation test at a threshold of P < 0.01.

Phase analysis is complicated by the fact that neuronal responses are always delayed to some degree with respect to the stimulus (i.e., due to sensory transduction, synaptic and conduction delays, and local processing time). If we had used a fixed analysis window (for instance, corresponding to the 2-s stimulus duration), phase angles computed for each response would be affected by response delays, which vary across neurons and even across stimulus directions. Thus to dissociate delay from actual response dynamics, the analysis window for the DFT was shifted such that it was centered on the midpoint of the response profile, defined as the center of mass of the temporal envelope. In signal processing, the envelope of a real-valued function (time series) is the magnitude of its analytic signal, or equivalently its amplitude-modulating waveform. To compute the envelope, we used a standard approach (Bracewell 1978; Gabor 1946) that has previously been applied to the analysis of visual receptive fields (DeAngelis et al. 1993; Field and Tolhurst 1986). Briefly, we first subtracted the baseline response (mean firing rate in a 200 ms window just prior to stimulus onset) and smoothed each PSTH using a 60-ms (3-bin) boxcar function. From the smoothed, baseline-subtracted version of the PSTH, denoted R(t), we computed the Hilbert transform H[R(t)] using the Matlab function hilbert (version 6.5, Signal Processing Toolbox), which shifts all frequency components by 90°. The temporal envelope E(t) was then computed as

\[
E(t) = \sqrt{R(t)^2 + H[R(t)]^2}
\]

To illustrate the procedure in more detail, consider the stimulus velocity and acceleration profiles themselves (Fig. 1A) as example “signals” from which we wish to extract the temporal envelope. In Fig. 1B, the velocity profile V(t) (dashed curve) is re-plotted along with its Hilbert transform H[V(t)] (dotted curve). These two functions can be considered as a pair of orthogonal vectors (known as a quadrature pair), where the angle between them represents the 90° phase shift due to the Hilbert transform. The magnitude of their vector sum (i.e., the square root of the sum of squares; Eq. 4) is the envelope E(t), depicted as the solid curve. Similarly, the acceleration profile of the stimulus and its Hilbert transform (Fig. 1C, dashed and dotted curves, respectively) form a quadrature pair whose vector sum comprises the temporal envelope of acceleration (solid curve, Fig. 1C). Note that the temporal envelopes of the velocity and acceleration stimulus profiles are very similar in shape: the envelope peak coincides with the maximum velocity and the zero-crossing of acceleration, both at time t = 1.0 s. Thus centering the analysis window using the temporal envelope serves to capture the relevant temporal modulation in the response, while being largely unbiased with respect to velocity versus acceleration components.

For a given neuronal response profile (PSTH), the difference (in units of time) between the envelope center of mass and time t = 1.0 s was defined as the delay. This procedure was applied to each PSTH separately (e.g., see Fig. 2, A–C), allowing the responses to different motion directions to have independent delays. It is important to note that delay, by this definition, does not correspond to neuronal latency per se. Traditional measures of latency are difficult to compute in the case of a gradual stimulus onset (i.e., the early tail of our Gaussian velocity profile), and in any case such a measure would not be useful for our purposes. Rather the DFTR method required us to quantify the delay of the overall response profile (temporal modulation of firing rate) relative to the stimulus, without assuming a particular shape or frequency content of the response. The Hilbert transform approach is well suited for this purpose.

As mentioned in the preceding text, significance of the DFTR velocity and acceleration components was determined using a permutation test in which the temporal delay derived from the Hilbert transform was constant across permutations. We also tested significance using a bootstrap test in which the temporal delay was determined for each resampling of the data. This produced similar proportions of PSTHs with significant velocity and acceleration components, and all of the results were similar when responses were classified in this fashion.

Spatial tuning

Spatial tuning maps were generated for each cell by plotting the mean response as a function of the azimuth and elevation of the motion trajectory. Here, “response” could be defined as MFR, DFTRvel, or DFTRacc. MFR was computed within a fixed 2-s window beginning with stimulus onset. Computing MFR in the 2-s window used for the DFT (i.e., centered on the temporal envelope) did not appreciably change the pattern of results. The statistical significance of spatial tuning was quantified by a one-way ANOVA performed on the MFR/DFTR values derived from each individual trial. Computing DFTR based on single-trial PSTHs (rather than across-trial averaged PSTHs as described in the preceding text) was necessary to provide single-trial metrics for the ANOVA. Note that when computing single-trial versions of DFTRvel and DFTRacc, we shifted the single-
trial PSTHs by the delay computed from the original temporal envelope (from the averaged PSTH) to reduce uncertainty in the delay estimate.

A preferred direction was computed for all four response metrics (MFR, DFTR, DFTRe, and DFTRa) as follows. Each MFR/DFTR value was considered to represent the magnitude of a 3D vector the direction of which was defined by the azimuth and elevation of the corresponding motion stimulus. These vectors were summed, and the direction of the resultant vector was taken as the preferred direction. Tuning maps were also compared for each cell across response metrics (e.g., DFTRe vs. MFR, DFTRa vs. MFR, and DFTRe vs. DFTRa) via a simple correlation coefficient.

To quantify spatial tuning strength, we calculated a direction discrimination index (DDI) (Nguyenkim and DeAngelis 2003; Prince et al. 2002; Takahashi et al. 2007; Uka and DeAngelis 2003), which measures peak-to-trough response modulation (due to changes in motion direction) relative to response variability

$$\text{DDI} = \frac{R_{\text{max}} - R_{\text{min}}}{R_{\text{max}} - R_{\text{min}} + 2\sqrt{\text{SSE}/(N - M)}}$$

where $\text{SSE}$ is the sum of squared errors, $N$ is the total number of trials, $M$ is the number of motion directions tested (26), and $\text{SSE}$ is the total sum of squared error around the mean responses for all motion directions. Thus computing DDI based on DFTR values also required the analysis of single-trial PSTHs, as was the case for the spatial tuning ANOVA (see preceding text).

Validation of the DFTR method using simulated data

We evaluated the accuracy and robustness of the DFTR method for estimating velocity and acceleration components by testing it with simulated neuron response profiles for which the true mixture of velocity and acceleration components was known. These simulated profiles, denoted $R(t)$, were computed as the weighted sum of a normalized Gaussian (mimicking the velocity profile of the stimulus) and its derivative (the biphasic acceleration profile)

$$r_{\text{vel}}(t) = e^{-\frac{-(t - \mu)^2}{2\sigma^2}}$$

$$r_{\text{acc}}(t) = \frac{d}{dt} r_{\text{vel}}(t)$$

$$R(t) = [w_{\text{vel}} r_{\text{vel}}(t) + w_{\text{acc}} r_{\text{acc}}(t) + DC + cv]_+$$

The Gaussian velocity component, $r_{\text{vel}}(t)$, was assigned a mean ($\mu = 1.0$ s) and SD ($\sigma = 0.2$ s) approximating the stimulus velocity profile (see Fig. 1A). The values of $r_{\text{vel}}(t)$ ranged from 0 to 1, and its derivative, $r_{\text{acc}}(t)$, was normalized into a range from −1 to 1. The weighted sum, $R(t)$, of these two normalized components was then computed. The weights $w_{\text{vel}}$ and $w_{\text{acc}}$ also ranged from −1 to 1, with negative weights representing inhibitory responses and positive weights representing excitatory responses (relative to baseline). The two weights were allowed to have different sign, but to maintain proper normalization the sum of their absolute values was constrained to be 1. The DC offset term represents the baseline response (i.e., spontaneous activity), ranging from 0 to 1. The final term represents Gaussian noise ($\nu$) drawn from a standard normal distribution, multiplied by a scaling factor $c$. Varying this scaling factor had little effect on the main results of the simulation. For the results presented here, we fixed the value of $c$ at 0.14, which typifies the noise level observed for a typical well-modulated neuron from our sample (e.g., Fig. 2, A–C) as quantified by the denominator of the DFTR (see Eq. 1). Last, the subscript “+” indicates half-wave rectification, simulating the “clipping” effect at zero firing rate.

We varied $w_{\text{vel}}$ (and consequently $w_{\text{acc}}$) in fine steps from −1 to 1 and subjected each simulated response profile to the DFTR analysis described in the preceding text (Eqs. 1–3). We then compared the relative magnitudes of the resulting DFTR components (DFTRe and DFTRa, Eqs. 2 and 3) to the corresponding weights that were used to produce each simulated profile. This comparison can be thought of as the input-output relationship of the DFTR method: the “inputs” are $w_{\text{vel}}$ and $w_{\text{acc}}$ from Eq. 6, and the “outputs” are the normalized velocity and acceleration DFTR components, defined as follows

$$\text{DFTR}_{\text{vel}}(\text{norm}) = \frac{\text{DFTR}_{\text{vel}}}{\sqrt{\text{DFTR}_{\text{vel}}^2 + \text{DFTR}_{\text{acc}}^2}}$$

$$\text{DFTR}_{\text{acc}}(\text{norm}) = \frac{\text{DFTR}_{\text{acc}}}{\sqrt{\text{DFTR}_{\text{vel}}^2 + \text{DFTR}_{\text{acc}}^2}}$$

Note that these normalized DFTR components should be comparable to $w_{\text{vel}}$ and $w_{\text{acc}}$ from Eq. 6 because they also range from −1 to 1 and their absolute values sum to 1. Thus if the DFTR analysis perfectly measures the relative strength of velocity and acceleration components in the response, then the weights from Eq. 6 should match the normalized DFTR values from Eq. 7. This comparison will be discussed in the following text in RESULTS (Fig. 10).

Curve fitting analysis

As a complement to the DFTR approach, we performed a curve fitting analysis in which each PSTH was modeled as a weighted sum of velocity and acceleration components. The fitting procedure minimized the mean-squared difference between the data and the values of the following equation

$$R(t) = \left[A \ast w_{\text{vel}} r_{\text{vel}}(t) + \text{sign}_{\text{acc}} \ast w_{\text{acc}} r_{\text{acc}}(t) + \text{DC} \right]_+$$

This equation is similar to Eq. 6 but with a few key differences. First, the mean and SD of the Gaussian velocity component $r_{\text{vel}}(\mu$ and $\sigma$ in Eq. 6, respectively) were free parameters rather than being fixed to the values of the velocity profile itself. The noise term $cv$ was removed, and an amplitude parameter $A$ was added to scale the responses appropriately (PSTHs were not normalized prior to fitting). The velocity weight ($w_{\text{vel}}$) varied from −1 to 1, $w_{\text{acc}}$ was constrained to be equal to $(1 - w_{\text{vel}})$, and the parameter $\text{sign}_{\text{acc}}$ was used to generate fits with positive or negative acceleration components. We fit two versions of the model, one with $\text{sign}_{\text{acc}} = 1$ and the other with $\text{sign}_{\text{acc}} = -1$, and the version with the smaller error was taken as the best fit for each PSTH. In preliminary testing, we found that a model in which both weights ranged from −1 to 1 was too poorly constrained for many PSTHs, causing convergence problems to arise during the fitting procedure (i.e., frequent local minima and/or large correlations between parameters).

RESULTS

Temporal response properties

The analyses presented here were performed on responses from 395 MSTd neurons recorded from three monkeys. We recorded from any well-isolated neuron in MSTd that was spontaneously active or responded to a large flickering field of random dots (for details, see Fetsch et al. 2007; Gu et al. 2006; Takahashi et al. 2007); neurons were not prescreened for vestibular responsiveness before inclusion in the current data set. For each motion direction, we computed an average PSTH using 20-ms bins. Figure 2, A–C, shows three example PSTHs: a velocity-like response (A), an acceleration-like response (B), and a poorly modulated (nonsignificant) response profile (C).
components, termed DFTRvel and DFTRacc, by incorporating 
carried by MSTd neurons, the DFTR was separated into two 
respectively.

DFTR calculation (Eqs. 2) is well aligned with that of the velocity profile (near 180°), 
(positive) and inhibitory (negative) responses relative to the 
in the example responses in Fig. 2, A–C, along with the stimulus 
v and acceleration spectra. Note that all three of these 
responded from a single neuron (in response to different 
directions of motion)

Each temporal response profile was characterized by its 
discrete Fourier transform ratio (DFTR; Eq. 1 in METHODS). For 
the amplitude spectra in Fig. 2, D–F, the horizontal dashed line 
labeled A marks the mean amplitude of the frequency components 
in the yellow shaded area (3–25 Hz). The DFTR (A/B) is an index of the strength of 
temporal modulation relative to the noise level. For the three 
example responses in Fig. 2, A–C, this ratio was 5.9 (P < 0.001 
permutation test), 3.5 (P < 0.001), and 1.6 (P = 0.05), 
respectively.

To quantify the strength of velocity and acceleration signals 
carried by MSTd neurons, the DFTR was separated into two 
components, termed DFTRvel and DFTRacc, by incorporating 
the phase information from the Fourier transform into the 
DFTR calculation (Eqs. 2 and 3 in METHODS). These quantities 
are signed, unlike the original DFTR described in the preceding 
text, such that they can adequately describe both excitatory 
(positive) and inhibitory (negative) responses relative to the 
stimulus. The phase of the velocity-like response (Fig. 2, A and 
G) is well aligned with that of the velocity profile (near 180°),

at least for the large-amplitude components [0.5–1.5 Hz; the 
2.0- and 2.5-Hz components have very small amplitudes and 
contribute little to DFTRvel (Eq. 2)]. As a result, the DFTRvel 
of this example response was 5.70 (P < 0.001) and the 
DFTRacc was 0.81 (P = 0.03). In contrast, the phase of the 
acceleration-like response (Fig. 2, B and H) is better aligned 
with the acceleration profile of the stimulus (near 90°). This 
yields a DFTRvel of −0.12 (P = 0.74) and DFTRacc of 3.23 
(P < 0.001). Neither the DFTRvel (-0.28, P = 0.46) nor the 
DFTRacc (-0.48, P = 0.16) were significant for the weak 
response shown in Fig. 2C.

To utilize the DFT phase to compute velocity and acceleration 
components, it was necessary to shift the time window 
for the DFT to account for the temporal delay of each response. 
Delay was estimated from the temporal envelope of the 
response as described in METHODS. A histogram of these delays 
for all significant responses is shown in Fig. 3A. Importantly, 
this metric does not correspond to the latency of response onset 
but rather the delay of the midpoint of the response relative to 
the midpoint of the stimulus. Thus negative delays are plausible 
because the temporal envelope can be asymmetric (i.e., 
skewed toward earlier times). The broad distribution of delays 
suggests a population consisting of diverse combinations of 
temporal response dynamics and neuronal latencies, while also 
highlighting the fact that the envelopes themselves (and hence 
the delays) are a somewhat noisy measurement (e.g., red traces 
in Fig. 2, A and B). Although the phases of all low-frequency 
components were used to compute DFTRvel and DFTRacc (Eqs. 
2 and 3), we took the phase of the maximum-amplitude 
component as an approximation of the overall response phase. 
These response phases are summarized as a histogram in Fig. 
3B. The large peak near 180° suggests a preponderance of 
velocity-like responses.

Across all neurons in our sample, 37.4% (3,843/10,270) of 
motion directions were classified as having significant DFTRs 
(see METHODS for significance criteria). As a fraction of the 
3,843 significant motion directions, 41% showed significant 
DFTR components for velocity only, 16% for acceleration 
only, 15% for both, and 28% for neither. Table 1 summarizes 
the number of motion directions that fell into each possible 
significance category. Responses in the latter category (“neither”) were typically weak and only marginally significant for 
DFTR; thus the overall DFTR reached significance but the 
v and acceleration components did not in these cases. In 
addition, many of the neither directions had very low firing 
rates: median MFR for these directions was significantly less 
than for directions with significant DFTRvel or DFTRacc values 
(19.7 vs. 27.4 spike/s; Mann-Whitney U test, P < 0.0001).

We also computed the number of motion directions with 
significant temporal response modulation (of a possible 26) for 
each cell, as summarized in Fig. 4. The average number of 
directions with a significant DFTR was 7.1 per neuron (Fig.

---

**Table 1. Percent of motion directions with significant temporal modulation**

<table>
<thead>
<tr>
<th>DFTR</th>
<th>DFTRvel</th>
<th>DFTRacc</th>
<th>DFTRvel Only</th>
<th>DFTRacc Only</th>
<th>Both</th>
<th>Neither</th>
</tr>
</thead>
<tbody>
<tr>
<td>37.5</td>
<td>21.2 (56.7)</td>
<td>11.6 (30.9)</td>
<td>15.5 (41.4)</td>
<td>5.9 (15.6)</td>
<td>5.7 (15.2)</td>
<td>10.4 (27.7)</td>
</tr>
</tbody>
</table>

Percent of all motion directions (n = 395 cells * 26 directions = 10,270) passing the discrete fourier transform ratio (DFTR) significance criteria in each possible category. 'Only' refers to directions with exclusively one component (velocity or acceleration) significant and not the other. Numbers in parentheses indicate percentages computed as a fraction of directions with a significant DFTR and hence the percentages in the last four columns add up to 100.

---

**Fig. 3.** Delay and phase distributions. A: histogram of response delays for all motion directions having significant temporal modulation (i.e., significant DFTR values; n = 3,843/10,270). Zero delay corresponds to a temporal response envelope that is aligned with the peak of stimulus velocity. B: histogram of response phase for the same population as in A where response phase was defined as the phase of the largest amplitude DFT component.
4A; median = 5), as compared with 3.3 directions that were significant for velocity only (B; median = 1), 1.1 for acceleration only (C; median = 0), and 1.4 for both (D; median = 0).

This suggests broader spatial tuning for velocity compared with acceleration, a result we will revisit in the spatial tuning analyses in the following text. Taken together, the results discussed so far indicate that the majority of MSTd responses reflect stimulus velocity, while a substantial minority signal acceleration alone or both velocity and acceleration.

To compare velocity and acceleration components for each direction, we plotted DFTRacc versus DFTRvel as a scatter plot in Fig. 5A. This figure includes only those directions with significant DFTRs (n = 3,843); the proportion of data points falling into each color-coded significance category corresponds to the numbers in parentheses in the last four columns of Table 1. There are three main results evident in Fig. 5A: first, there is no significant correlation between velocity and acceleration DFTR values across the whole population (Pearson r = 0.02, P = 0.27). Second, velocity DFTR values were significantly greater than acceleration DFTR values (Wilcoxon matched pairs test, P < 0.0001). Note, however, that this does not by itself indicate that velocity components were stronger, because DFTRacc and DFTRvel are signed quantities. It is clear from the distributions of green and blue points that the velocity responses in our sample were much more likely to be positive (excitatory) instead of negative (inhibitory), whereas acceleration components were more evenly split between positive and negative. Thus to compare the strength of velocity and acceleration components on an equal footing, we computed their absolute values, shown in Fig. 5B. The median absolute value of DFTRvel (1.12, left) was significantly greater than that of DFTRacc (0.64, right; Mann-Whitney U test, P < 0.01), reinforcing the conclusion that the velocity components across the population of MSTd responses tend to be stronger than the acceleration components.

Here we should emphasize that the analysis of Fig. 5 treats each motion direction (PSTH) as the unit of replication, rather than each neuron. Thus the greater number of directions per cell having only a significant velocity component (the longer tail of the distribution in Fig. 4, B vs. C) may exaggerate the apparent velocity bias in a population plot such as Fig. 5. However, the number of neurons having at least one motion direction with a significant DFTR component (i.e., the cumulative sum of bins in Fig. 4, B and C) still favors velocity, by a margin of 260 (66%) to 194 (49%).

Spatial tuning properties

Having established the trend for responses to be more velocity-like overall, we wanted to examine how the temporal components varied with stimulus motion direction. For each cell, we first normalized the DFTRvel and DFTRacc values for all 26 directions by dividing them by their corresponding DFTR. We then removed the directions for which the DFTR was not significant and sorted the remaining values from lowest to highest. For many cells, both velocity and acceleration DFTR components show a striking amount of variation across motion directions. We quantified this variation by computing, for each cell, the SD of the normalized DFTR component values. The average of these SDs across all cells was 0.32
and 0.29 for velocity and acceleration, respectively. Similarly, the mean range (max–min) across cells was 0.88 (velocity) and 0.85 (acceleration). This degree of within-cell variation in the velocity and acceleration DFTR components indicates that the temporal dynamics of MSTd vestibular responses cannot adequately be classified based on one or even a few motion directions.

The preceding analysis raises several questions: is there a pattern to the spatial variation in temporal dynamics, and more specifically, what is the relationship between the temporal and spatial information contained in these responses? To address these questions, we compared the 3D spatial tuning of velocity and acceleration components to that of mean firing rate (MFR). Figure 6 shows the spatial tuning of two example neurons (A–D and E–H). The contour plots show response strength as a function of azimuth and elevation, where response strength is defined by (from top to bottom): MFR (A and E), DFTRvel (B and F), or DFTRacc (C and G). Figure 6, D and H, illustrates average PSTHs for the cells’ responses to translation in the horizontal plane, which corresponds to a cross section through the 3D tuning profiles at zero elevation. As in Fig. 2, stimulus velocity and acceleration profiles are overlaid for visual comparison.

Example cell 1 (Fig. 6, A–D) shows a strong velocity-like (Gaussian) response in one direction (azimuth 45°), but a more acceleration-like (biphasic) response in the opposite direction (azimuth 225°). This opposite directional tuning for velocity and acceleration response components can be observed in the PSTHs (Fig. 6D) and is clearly seen in the 3D tuning maps (compare Fig. 6, B and C). Also note that the directional tuning for mean firing rate (Fig. 6A) was well matched to that of the velocity component but not the acceleration component.

For example cell 2 (Fig. 6, E–H), all three spatial tuning maps (MFR, velocity, acceleration) are fairly well aligned. This can occur when the response to a given direction of motion carries both velocity and acceleration signals. The PSTHs (Fig. 6H) for 0 and 45° azimuth can be described as mixtures of velocity and acceleration, showing an initial peak followed by a modest suppression below baseline. Indeed, the DFTRvel and DFTRacc of the 0° response were 1.7 and 1.4, respectively (P < 0.001 for both), and for the 45° response they were 2.0 and 1.8 (P < 0.001 for both). Moreover, responses in the opposite directions (180 and 225°) resemble an inverted acceleration-like profile, resulting in a negative DFTRacc (−1.6 and −2.0, P < 0.001 for both). DFTRvel for the 180° response was also negative (−0.8, P = 0.01). In summary, this cell shows positive and negative velocity/accel-

---

**Figure 6.** Spatial tuning of different temporal response components. Two example cells illustrating the relationship between 3-dimensional (3D) spatial tuning for the velocity and acceleration components of response, as well as for mean firing rate (MFR). 3D spatial tuning maps for each cell were calculated using MFR (A and E), DFTRvel (B and F), and DFTRacc (C and G). D and H: average peristimulus time histograms (PSTHs) for responses to translation in the horizontal plane, corresponding to a slice through the contour maps at elevation = 0. Stimulus velocity (solid black) and acceleration (dashed black) profiles are overlaid as in Fig. 2 (i.e., not fits to the data, but shifted by the measured response delay).
We then found the angular difference defined as the direction of the vector sum of responses around the preferred direction in 3D space for each tuning map, correlations represented by filled black bars. Second, we found tuning maps (Fig. 7, summarized in Fig. 7, acceleration vs. MFR, and acceleration vs. velocity). These are computed a Pearson correlation coefficient between each of the four response components. For the both acceleration and velocity response components, respectively. For the both acceleration only (filled gray symbols), acceleration only (gray filled symbols), velocity and acceleration (black open symbols), or neither (gray open symbols). Across all cells, velocity DDIs were modestly but significantly larger than their acceleration counterparts (Wilcoxon matched-pairs test, \( P < 0.0001 \); velocity mean = 0.574, acceleration mean = 0.548), suggesting that velocity-based spatial tuning is stronger than that based on acceleration. This trend also held for the subset of neurons with significant spatial tuning for both components (open black symbols; \( P < 0.0001 \)).

Finally, to investigate differences in the breadth of spatial tuning for velocity and acceleration, we re-examined the number of significant directions per cell (i.e., Fig. 4) in light of the spatial tuning data. There instead of pooling all cells as in Fig. 4, we divided cells into the four groups mentioned in the preceding text (velocity only, acceleration only, both, and neither) based on the significance of their spatial tuning. For the groups labeled velocity only and acceleration only, the abscissa represents the number of directions with a significant velocity or acceleration component, respectively. For the both preferred directions metric: velocity- and MFR-based direction preferences tended to be very similar (angular differences clustering toward zero, Fig. 7D), while the distributions for acceleration versus MFR and acceleration versus velocity were roughly uniform. Although a weak tendency toward bimodality can be seen in Fig. 7, B, E, and F, none of these distributions was found to be significantly bimodal (Hartigan’s dip test, \( P > 0.19 \) for each) (Hartigan and Hartigan 1985; Hartigan 1985). This result suggests that MFR captures mainly the spatial tuning for velocity signals in MSTd, while the tuning of the acceleration components bears no systematic relationship to either velocity or MFR.

We also compared the strength of spatial tuning for the different response components, using a direction discrimination index (DDI; Eq. 5). The DDI ranges from 0 to 1, with larger DDIs indicating stronger selectivity for stimulus direction. Figure 8A plots the DDI values computed for velocity-based versus acceleration-based spatial tuning maps. Cells are separated into four groups based on whether they showed significant spatial tuning (1-way ANOVA, \( P < 0.05 \)) for velocity only (black filled symbols), acceleration only (gray filled symbols), velocity and acceleration (black open symbols), or neither (gray open symbols). Across all cells, velocity DDIs were modestly but significantly larger than their acceleration counterparts (Wilcoxon matched-pairs test, \( P < 0.0001 \); velocity mean = 0.574, acceleration mean = 0.548), suggesting that velocity-based spatial tuning is stronger than that based on acceleration. This trend also held for the subset of neurons with significant spatial tuning for both components (open black symbols; \( P < 0.0001 \)).

Overall, spatial tuning for velocity showed strong positive correlations with spatial tuning calculated using MFR (Fig. 7A). In contrast, tuning based on acceleration could be negatively correlated (e.g., example cell 1 in Fig. 6), positively correlated (e.g., example cell 2 in Fig. 6), or uncorrelated with MFR and velocity tuning (Fig. 7, B and C, open bars). A similar pattern of results was obtained using the difference of spatial tuning data. Here instead of pooling all cells as in Fig. 4, we divided cells into the four groups mentioned in the preceding text (velocity only, acceleration only, both, and neither) based on the significance of their spatial tuning. For the groups labeled velocity only and acceleration only, the abscissa represents the number of directions with a significant velocity or acceleration component, respectively. For the both preferred directions metric: velocity- and MFR-based direction preferences tended to be very similar (angular differences clustering toward zero, Fig. 7D), while the distributions for acceleration versus MFR and acceleration versus velocity were roughly uniform. Although a weak tendency toward bimodality can be seen in Fig. 7, B, E, and F, none of these distributions was found to be significantly bimodal (Hartigan’s dip test, \( P > 0.19 \) for each) (Hartigan and Hartigan 1985; Hartigan 1985). This result suggests that MFR captures mainly the spatial tuning for velocity signals in MSTd, while the tuning of the acceleration components bears no systematic relationship to either velocity or MFR.

We also compared the strength of spatial tuning for the different response components, using a direction discrimination index (DDI; Eq. 5). The DDI ranges from 0 to 1, with larger DDIs indicating stronger selectivity for stimulus direction. Figure 8A plots the DDI values computed for velocity-based versus acceleration-based spatial tuning maps. Cells are separated into four groups based on whether they showed significant spatial tuning (1-way ANOVA, \( P < 0.05 \)) for velocity only (black filled symbols), acceleration only (gray filled symbols), velocity and acceleration (black open symbols), or neither (gray open symbols). Across all cells, velocity DDIs were modestly but significantly larger than their acceleration counterparts (Wilcoxon matched-pairs test, \( P < 0.0001 \); velocity mean = 0.574, acceleration mean = 0.548), suggesting that velocity-based spatial tuning is stronger than that based on acceleration. This trend also held for the subset of neurons with significant spatial tuning for both components (open black symbols; \( P < 0.0001 \)).

Finally, to investigate differences in the breadth of spatial tuning for velocity and acceleration, we re-examined the number of significant directions per cell (i.e., Fig. 4) in light of the spatial tuning data. Here instead of pooling all cells as in Fig. 4, we divided cells into the four groups mentioned in the preceding text (velocity only, acceleration only, both, and neither) based on the significance of their spatial tuning. For the groups labeled velocity only and acceleration only, the abscissa represents the number of directions with a significant velocity or acceleration component, respectively. For the both preferred directions metric: velocity- and MFR-based direction preferences tended to be very similar (angular differences clustering toward zero, Fig. 7D), while the distributions for acceleration versus MFR and acceleration versus velocity were roughly uniform. Although a weak tendency toward bimodality can be seen in Fig. 7, B, E, and F, none of these distributions was found to be significantly bimodal (Hartigan’s dip test, \( P > 0.19 \) for each) (Hartigan and Hartigan 1985; Hartigan 1985). This result suggests that MFR captures mainly the spatial tuning for velocity signals in MSTd, while the tuning of the acceleration components bears no systematic relationship to either velocity or MFR.
and neither groups, the number of directions with at least one significant component was used. These distributions are plotted in Fig. 8B, showing that cells with significant spatial tuning for acceleration only (gray filled bars) had relatively few responses with significant DFTR\textsubscript{acc} values, compared with cells that showed significant spatial tuning for velocity only (black filled bars). The mean number of significant directions for the velocity-only category was 5.9, compared with 2.9 for the acceleration-only category (medians = 4 and 3, respectively; Mann-Whitney U test, \( P = 0.005 \)). These data suggest that MSTd neurons encode stimulus acceleration over a narrower range of motion directions compared with the encoding of stimulus velocity. Note also that cells with both velocity and acceleration components tend to have the largest number of directions with significant response modulation (mean = 9.8, median = 8).

**Limitations of the frequency-domain analysis method:**
**analysis of simulated data**

The frequency-domain analysis presented here represents one possible approach for estimating spatiotemporal properties of neuronal responses, and like any method, it is not without shortcomings. Perhaps the greatest pitfall when using phase information to estimate the relative strength of velocity and acceleration components is the potential for misestimating delay, and hence phase, due to nonlinearities in the response profiles. For example, if the baseline activity is relatively low, the trough of an acceleration-like response may be “clipped” at zero, causing it to appear more velocity-like.

To assess the accuracy of the DFTR method and its robustness to response clipping, we performed the temporal analyses (Eqs. 1–4) on simulated response profiles that were designed to reflect the diversity of temporal dynamics we observed in area MSTd (see METHODS). The simulated responses were modeled (Eq. 6) as weighted sums of normalized velocity and acceleration profiles, plus a DC offset term (representing baseline activity) and a Gaussian noise term. Some examples of simulated temporal responses are shown in Fig. 9, where each row corresponds to a different set of weights (\( w\textsubscript{vel} \) and \( w\textsubscript{acc} \)), and each column a different value of the DC offset term, resulting in various degrees of response clipping. For all combinations of \( w\textsubscript{vel} \) and \( w\textsubscript{acc} \), we computed the DFTR components according to Eqs. 2 and 3, then normalized them according to Eq. 7 to enable a direct comparison with the true weights that were assigned to the velocity and acceleration components in the simulated responses. The inset text in each panel of Fig. 9 gives the delay in milliseconds (estimated from the temporal response envelope, Eq. 4) and the normalized velocity and acceleration components estimated using the DFTR method (Eq. 7; abbreviated vel and acc in Fig. 9). A quantitative comparison between the estimated DFTR components and the actual weights used to create the synthesized data are shown in Fig. 10 for velocity (\( A \)) and acceleration (\( B \)). Separate curves are plotted for each level of the DC offset parameter in Eq. 6 (i.e., the amount of potential clipping). For simplicity of display, Figs. 9 and 10, \( A \) and \( B \), do not show the symmetric combinations of negative velocity components with positive acceleration components, or vice versa; these results were similar and are shown in \( B \).

The first result to take away from this analysis is that the DFTR method is nearly perfect at estimating the relative strength of velocity and acceleration components (i.e., the weights) when there is little or no clipping in the responses (leftmost column of Fig. 9; red and orange traces in Fig. 10). Interestingly, this accurate estimation of weights occurs despite variations in the estimated response delay that are not actually present in the simulated data (e.g., \( -44 \) ms when \( w\textsubscript{vel} = w\textsubscript{acc} = 0.5 \); Fig. 9). The simulated response profiles are assigned an actual delay of zero (i.e., the response for \( w\textsubscript{vel} = 1.0 \) peaks at the same time, \( t = 1.0 \) s, as the stimulus velocity profile), and hence one might expect that this discrepancy in the actual versus estimated delays would cause a misestimation of the velocity and acceleration components. In fact, the accuracy of the DFTR method depends on these “misestimated” delays: artificially fixing the delay at zero rather than using the temporal envelope method, causes systematic errors in the DFTR analysis even in the absence of clipping (data not shown).

Second, even for more substantial amounts of response clipping (e.g., Fig. 9, middle), the DFTR method remains fairly accurate (Fig. 10, green trace). For example, a simulated acceleration-like response profile (\( w\textsubscript{vel} = 0, w\textsubscript{acc} = 1 \)) with a DC offset of 0.5 results in clipping of 50% of the trough of the response, yet the relative magnitude of the DFTR components still indicates a largely acceleration-like response \([DFTR\textsubscript{vel}(\text{norm}) = 0.15, DFTR\textsubscript{acc}(\text{norm}) = 0.85]\). In contrast, for the case of 100% clipping (DC offset = 0), the DFTR method estimates this same pure acceleration response as nearly a 50/50 mixture of velocity and acceleration components \([DFTR\textsubscript{vel}(\text{norm}) = 0.47, DFTR\textsubscript{acc}(\text{norm}) = 0.53]\). This error arises from a misestimation of the delay based on the temporal envelope, which is shifted from the true value (0) by \( -109 \) ms due to clipping. Shifting the window for the DFT by this value amounts to a phase shift of \( \pm 20^\circ \) at 0.5 Hz, \( \pm 40^\circ \) at 1.0 Hz, etc. This phase shift of the low-frequency components drives DFTR\textsubscript{vel} upward and DFTR\textsubscript{acc} downward (see Eqs. 2 and 3), resulting in a substantial discrepancy between the true weights of velocity and acceleration in the synthetic data and the estimated DFTR components. However, even this extreme case is somewhat encouraging: in the near-complete absence of a trough in the response profile (due to clipping), the acceleration DFTR component remains slightly larger than the velocity component (0.53 vs. 0.47).

Figure 10 summarizes the effect of response clipping across all values of \( w\textsubscript{vel} \) and \( w\textsubscript{acc} \), showing that the DFTR method becomes progressively less accurate as the DC offset parameter decreases (i.e., potential clipping increases). Note that deviations from the diagonal in Fig. 10A signify different things in different quadrants: in the upper-right quadrant, moving upward off the diagonal indicates a velocity bias for positive-going (excitatory) responses, whereas in the lower-left quadrant, the same deviation indicates an acceleration bias for negative-going (inhibitory) responses. Deviations in the upper-left quadrant—including the point of greatest departure from the diagonal (\( w\textsubscript{vel} = -0.5 \))—mainly indicate a sign reversal of the velocity component. Negative-going responses that are largely (but not completely) acceleration-like, when subjected to clipping, are interpreted by the DFTR as a mixture of a negative-going acceleration response with a positive-going velocity response, resulting in a (correctly) negative acceleration DFTR component but an (incorrectly) positive velocity response.
DFTR component. This can be seen more explicitly in Fig. 10B, which plots the acceleration weights in the same format (see the corresponding point at $w_{\text{acc}} = 0.7$, where the black trace crosses the unity diagonal). Overall, the same pattern of results is evident in Fig. 10, both A and B. For example, the velocity bias in the upper-right quadrant of Fig. 10A is seen as a downward deviation from the diagonal in B. Figure 10C displays the same results in a different format, where the abscissa represents $w_{\text{vel}}$ or DFTR$_{\text{vel}}$(norm), the ordinate $w_{\text{acc}}$ or DFTR$_{\text{acc}}$(norm), and the difference between the two methods (error) is represented by drawing a vector connecting the corresponding values. For clarity, only the three largest amounts of clipping (50, 75, and 100%) are shown as the smaller two amounts are nearly error-free. This format permits displaying the symmetric combinations of positive velocity weights with negative acceleration weights (bottom-right quadrant), and vice versa (top-left quadrant).

The simulation results of Figs. 9 and 10 show that the DFTR method systematically misestimates the velocity and acceleration components when a substantial portion of the temporal...
response is clipped at zero. How might this have affected our MSTd results and their interpretation? Unlike in the simulations, we do not have access to the true shape of an apparently clipped response, and thus the underlying velocity and acceleration components in such a response profile cannot be known with certainty. However, our simulations suggest that typical degrees of clipping should only modestly affect the estimation of the components using DFTR, except for neurons with very low levels of spontaneous activity relative to maximum evoked activity. Such neurons were relatively rare in our sample. One way to estimate the potential for clipping in the real data is to compute the ratio of the baseline firing rate (measured during the 600 ms prior to stimulus onset) to the peak evoked firing rate, which we term the baseline-to-peak ratio (BPR). Higher values of BPR suggest less potential for clipping; for example, the minimum BPR for the simulated data in the leftmost column of Fig. 9 is 0.5 (~1.0/2.0), whereas it decreases to 0.33 (~0.5/1.5) for the middle column, and is near zero for the rightmost column.

The distribution of BPR values for all responses with significant DFTR values is shown in Fig. 11A. The median BPR was 0.36, slightly higher than the minimum BPR for the 50% cutoff scenario in the simulations (Fig. 9, middle; Fig. 10, green curve). To examine the relationship between BPR and the relative strength of velocity and acceleration components, we separated the data shown in Fig. 5B (absolute value of DFTR components) depending on whether BPR was less than or greater than 0.33. The medians of these separate distributions are plotted in Fig. 11B, alongside the medians from the full data set (reproduced from Fig. 5B). The results indicate that the median $|\text{DFTR}_{vel}|$ was lower for responses with greater BPR, although it remained significantly greater than the median $|\text{DFTR}_{acc}|$ across all BPR categories. This reduction in $|\text{DFTR}_{vel}|$ may not be solely due to clipping; it could also simply reflect lower peak activity of responses with high BPR or perhaps a saturation effect when baseline activity is high. However, even if we assume that clipping accounts for most of the effect, velocity components in MSTd are still significantly stronger than acceleration components for the largest BPR category (Wilcoxon matched pairs test, $P < 0.0001$). Importantly, the median $|\text{DFTR}_{acc}|$ remained stable across BPR categories, suggesting that clipping did not greatly influence the strength of estimated acceleration components. Furthermore, as discussed in the preceding text, clipping does not always bias the component estimates toward velocity; for negative-going velocity and acceleration components, the bias can be in the direction of acceleration instead. Thus our main conclusion—the relative prevalence of velocity signals in MSTd—is unlikely to be strongly influenced by the effect of clipping on measured temporal responses.

**Comparison with curve fitting approach**

As an additional validation of the DFTR method for measuring velocity and acceleration response components, we applied a model-based curve fitting procedure to each PSTH in our sample ($n = 10,270$). The fitting function was a weighted sum of a Gaussian velocity profile and its derivative (Eq. 8), with the magnitudes of the velocity and acceleration weights summing to 1. The best-fitting weights were determined by minimizing the sum-squared error between the PSTH data and the fitted function (see methods for details).
Figure 12 shows three examples of well-modulated PSTHs and their corresponding model fits. It is clear that the model does an excellent job of fitting the data when the responses are strong (Pearson correlation between model and data, \( R > 0.95 \), \( P < 0.0001 \) for each example). In contrast, the fitting procedure often failed badly when the temporal response profile was weak or noisy, which was the case for the majority of PSTHs in our dataset. This was not surprising, given that our sample included many nonpreferred directions of motion as well as neurons without any significant vestibular responses. After attempting many fits to weak responses, we found that imposing criteria of DFTR \( > 3.0 \) and at least one significant DFTR component (\( P < 0.01 \) for DFTR\(_{vel}\) and/or DFTR\(_{acc}\) by bootstrap test) was typically required to have a reasonable chance of successful curve fitting, and thus we excluded PSTHs that did not meet these criteria.

Figure 13A shows the distribution of Pearson correlation coefficients for fits to the accepted response profiles (\( n = 752 \)). While many of these responses were reasonably well fit, others were not, despite having filtered the dataset with the preceding criteria. This is evident from the distribution of the mean parameter (\( \mu \)), which is plotted in Fig. 13, B and C, relative to the midpoint of the stimulus, thus giving a measure of response delay. For model fits with \( R \) values \( < 0.8 \) (Fig. 13B), this distribution is extremely broad with many delay values that are unreasonably large or small relative to the timing of the stimulus. Inspection of the results revealed that these outliers often resulted from fits that were poorly constrained by the data. This suggests a potential advantage of the DFTR method over curve-fitting: for all but the strongest, best-fit responses, velocity and acceleration components (weights) based on model fits may be less reliable than those based on the DFTR. For the best model fits (\( R > 0.8 \)), the distribution of response delays from curve fits (Fig. 13C) was narrower but was still broader than the delay distribution computed from the Hilbert-based temporal envelope (Fig. 3A; \( P < 0.0001 \), 2-sample Kolmogorov-Smirnov test). We considered this more restricted subset of responses (\( R > 0.8, n = 202 \)) to form a reasonable

![Figure 12](image-url)  
**FIG. 12.** Examples of model-based curve fitting analysis. A–C: 3 exemplars of well-modulated PSTHs are plotted along with the best fits of the model (thicker, smooth curve) defined by Eq. 8. Velocity weights derived from the model fit (\( w_{vel} \)) and the DFTR method (DFTR\(_{vel}(\text{norm})\)) are shown for each example, along with the correlation coefficient between the fit and the data.

![Figure 13](image-url)  
**FIG. 13.** Distributions of model correlation coefficients and delay parameter values from a subset of significantly modulated responses. A: histogram of \( R \) values for the model fits, including only those PSTHs with DFTR \( > 3.0 \) and at least one significant DFTR component (\( P < 0.01 \) for DFTR\(_{vel}\) and/or DFTR\(_{acc}\) by bootstrap test) was typically required to have a reasonable chance of successful curve fitting, and thus we excluded PSTHs that did not meet these criteria. B: histogram of response delay metric based on curve fitting (\( \mu \) parameter of Gaussian velocity component in Eq. 6, expressed relative to the peak of stimulus velocity) for the responses included in A but with \( R \) values \( < 0.8 \) (\( n = 550 \)). C: delay histogram for model fits with \( R \) values \( > 0.8 \) (\( n = 202 \)).
sample with which to compare the curve-fitting approach with the DFTR method.

We compared the velocity weight (w_{vel}) from these 202 high-quality fits to the corresponding metric based on DFTR (DFTR_{vel}n, normalized). For each of the examples in Fig. 12, both w_{vel} and DFTR_{vel}(norm) were >0.5, suggesting a general correspondence between the two methods in classifying these responses as velocity dominant. This rough correspondence was also evident in the population (Fig. 14): the median absolute value of w_{vel} for this sample was 0.7 (Fig. 14A), as compared with 0.82 for |DFTR_{vel}(norm)| (B). The two metrics were significantly correlated (R = 0.29, P < 0.0001), although their medians were significantly different (P < 0.0001, Wilcoxon matched pairs test) and the shapes of their distributions also differed significantly (P < 0.0001, K-S test). These discrepancies probably reflect the different sensitivities of the two methods to some combinations of velocity and acceleration. For example, when the PSTH had a small dip on one side of a much larger peak (e.g., Fig. 12C), the DFTR-based velocity weight tended to be much closer to 1 than the velocity weight from the curve fit. Importantly, however, both methods of analysis yield a consistent conclusion: the majority of temporal responses of MSTd neurons are dominated by velocity with a relatively small minority that is dominated by the acceleration component.

Finally, we examined the performance of our curve-fitting approach using simulated data to test its robustness to clipping. We fit the model (Eq. 8) to the same simulated profiles used to test the DFTR method (e.g., Fig. 9) and plotted the velocity weight used to generate the simulated data (abscissa) against the weight derived from the best model fit (ordinate) for five different levels of clipping (Fig. 14C). The traces are noisy due to variability across runs of the fitting routine, and this variability increased substantially with greater amounts of clipping. For the 100% clipping case (black trace), the velocity weights from the model fits fluctuate wildly, as particular combinations of weights and instances of random noise interacted to produce inconsistent results. However, the underlying trend appears to match the simulated data weights well, and there is little systematic bias in w_{vel} for all but the largest amounts of clipping. This result, in comparison to Fig. 10, suggests that model fits are more accurate, on average, than the DFTR method when substantial clipping is present. This result also counters the potential concern that the apparent velocity dominance in MSTd is an artifact of clipping, since the curve fitting results also support velocity dominance (Fig. 14A). On the other hand, both methods fail, in different ways, to cope adequately with the largest amounts of clipping, and the curve-fitting approach is clearly less robust to noise than the DFTR method. The DFTR method is also applicable to single-trial responses whereas the curve fitting approach is less amenable to single trials. Thus both methods have strengths and weaknesses, but they yield consistent conclusions regarding the velocity dominance of MSTd responses.

DISCUSSION

We have developed a frequency-domain analysis to characterize the temporal properties of MSTd responses during dynamic translational motion stimuli, showing that the majority of well-modulated responses are velocity-like. A smaller subset of responses encodes acceleration only, and another subset encodes both velocity and acceleration simultaneously. Spatial tuning based on velocity components is well aligned with tuning computed from mean firing rate (MFR), whereas tuning for acceleration showed no consistent relationship with tuning for either velocity or MFR. On average, spatial tuning for velocity is stronger than that for acceleration, with acceleration components being carried over a narrower range of motion directions than velocity. Finally, most cells show considerable variation in re-

FIG. 14. Comparison of model-based and DFTR-based velocity weights, and evaluation of the curve fitting approach with simulated data. A: distribution of the velocity weight parameter (w_{vel}) from model fits for 202 PSTHs with R > 0.8. B: distribution of velocity “weight” based on DFTR (DFTR_{vel}(norm)) for the same 202 responses. Both metrics show substantial velocity dominance (weights >0.5). C: model-based velocity weights for fits of simulated data, plotted in the same format as Fig. 10A. The unity diagonal represents perfect performance by the fitting procedure in estimating the known weights used to generate the simulated profiles, and colors represent different levels of clipping (DC offset).
response dynamics across directions of motion. These results suggest that the diversity of response dynamics observed in brainstem, cerebellum and thalamus is preserved to some extent in cortex, supporting the hypothesis that this diversity is functionally relevant for perception and behavior (Klam and Graf 2003).

To our knowledge, this is the first study to use frequency analysis to quantify temporal response properties for transient vestibular stimuli. Our method complements previous studies that utilized traditional sinusoidal motion stimuli (Chen-Huang and Peterson 2006; Dickman and Angelaki 2002; Liu and Angelaki 2009; Meng et al. 2007; Shaikh et al. 2005; Yakushin et al. 2006; Zhou et al. 2006). Indeed a recent study (Liu and Angelaki 2009) showed that MSTd responses show predominately velocity-like dynamics during sinusoidal translations, consistent with our findings. However, quantifying response dynamics using steady-state sinusoidal stimuli is complicated by the fact that central translation responses exhibit nonminimum phase behavior (i.e., phase spectra that exhibit larger or smaller frequency-dependent variation than would be expected from the dependence of amplitude on frequency) (Angelaki and Dickman 2000; Dickman and Angelaki 2002; Shaikh et al. 2005). As a result, response phase during sinusoidal stimulation cannot be used to conclusively determine whether cells encode velocity or position/acceleration. Using position transients (Gaussian velocity profiles) instead of sinusoids allows for a more direct evaluation of the temporal dynamics of translation responses in the CNS.

The use of the Hilbert transform to compute the temporal envelope allowed us to distinguish response phase from response delay (e.g., due to neuronal latency). Consequently, the phase information from the DFT provides a relatively unbiased way to measure the relative strength of velocity and acceleration signals in the observed MSTd responses. An advantage of the DFTR method is that it involves no free parameters and is model-independent, unlike curve-fitting approaches. However, a potential drawback to the DFTR approach is that the response delay may not be estimated correctly if there is substantial response clipping at zero firing rate (e.g., Fig. 6H, PSTHs at 135 and 180° azimuth). Such a misestimation of delay has implications for the phase of the DFT frequency components, and thus the DFTRvol and DFTRacc values computed from Eqs. 2 and 3. However, based on the analysis of simulated data (Figs. 9 and 10) and the separation of the data set based on baseline-to-peak ratio (Fig. 11), this problem is unlikely to have substantially altered our main conclusions. Indeed the dominance of velocity coding in MSTd was confirmed by application of a model-based curve fitting approach to a subset of the data (Fig. 14A). The curve-fitting approach is less biased by clipping (Fig. 14C) but is also less robust to noise (Fig. 13B) and therefore not well suited to single-trial analyses. Thus both methods have strengths and weaknesses but they yield the same basic conclusions regarding vestibular responses in MSTd.

In a similar study, Klam and Graf (2003) developed a random stimulation method to analyze response kinematics in the medial and ventral intraparietal areas (MIP and VIP) during yaw rotation. Because head position, velocity, and acceleration were only weakly correlated during their random motion paradigm, they were able to plot firing rate in a 2D (position vs. velocity) or 3D (position, velocity, and acceleration) space. The resulting “functional vestibular receptive fields” showed that most MIP and VIP neurons encoded multiple kinematic variables simultaneously, similar to our findings in MSTd. Note that the temporal response modulation in our analyses was only decomposed into velocity and acceleration components because we didn’t observe clear indications of higher-order derivatives (e.g., jerk) in well modulated responses. However, for a few motion directions the “baseline” firing rates immediately before and after the stimulus were clearly different (e.g., see 45 and 90° azimuth PSTHs in example cell 2, Fig. 6H). Because our stimuli were position transients, this suggests a possible influence of position for some motion directions. The possibility of position coding in MSTd deserves further study, especially in light of a previous report of “path” and “place” responses in this region (Froehler and Duffy 2002).

**Spatial variation in temporal dynamics**

One of our main findings is that response dynamics in MSTd are not constant for a given cell but instead vary depending on the direction of motion. This is consistent with findings in the vestibular nuclei and cerebellum (Angelaki and Dickman 2000; Musallam and Tomlinson 2002; Shaikh et al. 2005; Yakushin et al. 2006) but not with the behavior of otolith afferents (Angelaki and Dickman 2000; Fernandez and Goldberg 1976; Si et al. 1997). Otolith afferents are characterized by a single response vector, corresponding to the direction of maximal sensitivity. In contrast, the spatial tuning of many MSTd neurons could be described as having multiple response vectors. For example, some neurons clearly encoded velocity and acceleration along different directions of motion (e.g., example cell 1 in Fig. 6, A–D; note the gradual transition from velocity to acceleration going from 45 to 225° azimuth in Fig. 6D).

What then is the origin of this spatial variation in temporal response dynamics, and what function(s) could it serve? Previous theoretical (Angelaki 1992, 1993b; Angelaki et al. 1992, 1993) and neurophysiological (Angelaki and Dickman 2000; Chan et al. 2002; Chen-Huang and Peterson 2006; Dickman and Angelaki 2002; Yakushin et al. 2006) studies have proposed that VN neurons receive convergent inputs from otolith afferents that differ in both spatial and temporal properties. This convergence could include afferents from both labyrinths or across the striola within a given labyrinth (Uchino et al. 2005). This process, known as spatiotemporal convergence (STC), results in a type of spatio-temporal filtering that is consistent with the dependence of dynamics on motion direction that we observed in MSTd. Evidence for STC has been found as early as the second-order neurons in the VN (Angelaki and Dickman 2000) as well as in the vestibulocerebellum (Shaikh et al. 2005) and thalamus (Meng et al. 2007). Although differences in the stimuli preclude a direct comparison, it seems likely that the diverse spatial and temporal properties conferred by STC at lower processing stages have been carried forward to MSTd.

The functional significance of spatiotemporal convergence is unknown, but such a broad range of temporal dynamics within and across neurons may reflect the diversity of roles that linear acceleration signals play in perception and motor control. These roles include reflexive eye movements (translational VOR) (Angelaki 2004; McHenry and Angelaki 2000; Paige and Tomko 1991; Schwarz and Miles 1991; Schwarz et al. 1989), spatial updating (Klier et al. 2008), perception of head orientation with respect to gravity (Angelaki et al. 2004; Green
and Angelaki 2004, 2007; Lewis et al. 2005; Merfeld and Zupan 2002), and perception of self-motion direction or heading (Gu et al. 2007, 2008; Telford et al. 1995). One intriguing speculation is that having several kinematic variables multiplexed within the same population is computationally advantageous (Klam and Graf 2003), perhaps analogous to the idea that a mixture or “hybrid” spatial representation confers flexibility in sensorimotor coordinate transformations (Pouget et al. 2002).

However, this diversity can be a double-edged sword, as such a representation may complicate the readout of signals by downstream regions. For instance, MSTd is well known as a strong candidate for the neural basis of heading perception, both from visual and vestibular cues (Britten and van Wezel 1998; Gu et al. 2007, 2008; Page and Duffy 2003). According to some models of perceptual discrimination, optimal performance of a particular task is achieved by preferentially “reading out” (pooling) the most sensitive neurons in a population (Parker and Newsome 1998; Purushothaman and Bradley 2005; but see Shadlen et al. 1996). In this context, a neuron is most sensitive if the base stimulus value (e.g., motion direction) around which the discrimination is performed lies on the steep part of the cell’s tuning curve. Thus the sensitivity of a given cell depends on its tuning properties, such as its preferred direction and tuning width. However, these properties themselves can depend strongly on the chosen metric of response strength. Assuming that the temporal information in MSTd spiking activity is retained to some extent in downstream areas, the present results suggest that these areas must contend with the fact that a given cell can have different preferred directions and/or tuning widths (and hence different regions of maximal sensitivity) depending on whether the important variable is velocity, acceleration, or some other metric. Thus models of heading discrimination that posit any kind of selective readout or pooling may need to account for the variation in temporal structure found in MSTd responses.

Spatial tuning of velocity and acceleration components

It is important to point out that, unlike brain stem and cerebellar cells, MSTd neurons encode primarily velocity along most motion directions. Thus it is not surprising that spatial tuning for velocity was strongly correlated with spatial tuning calculated using MFR (Fig. 7A) and preferred directions computed from these two metrics were well aligned (D). This suggests that previous studies that computed spatial tuning from MFR (Duffy 1998; Fetsch et al. 2007; Gu et al. 2006; Page and Duffy 2003; Takahashi et al. 2007) were mainly capturing the velocity response of MSTd neurons. In contrast, MFR analysis does not appropriately capture the properties of the acceleration component of the responses. This may not pose a major problem for understanding the function of a velocity-dominant area such as MSTd, but it could constitute a problem for other areas with stronger acceleration tuning [e.g., parieto-insular vestibular cortex (PIVC)] (Chen et al. 2010).

Interestingly a small number of neurons had significant temporal modulation for the majority of motion directions (Fig. 4A, tail of the distribution). This means that even neurons that are not spatially selective may have strong temporal modulation to inertial motion. The possible role of this type of neuron is unclear, but some evidence indicates that a larger proportion of neurons in PIVC have this property as well (Chen et al. 2010).

Suitability for integration with visual (optic flow) signals

The velocity-like Gaussian shape of the most common MSTd vestibular responses appears similar to the dynamics of visual (i.e., optic flow) responses in this area (Gu et al. 2006) as well as visual motion responses in area MT (Lisberger and Movshon 1999; Rodman and Albright 1987), a primary source of input to MSTd. Similar response dynamics for both vestibular and optic flow responses in MST might be useful for the integration of these two complementary self-motion cues, and indeed there is evidence that MSTd neurons may be involved in this sensory integration (Gu et al. 2008). For integration to occur, one might assume that responses to different cues would signal motion with similar temporal dynamics. Interestingly the spatial variation in temporal dynamics that we have observed in the present study suggests that visual and vestibular signals in many MSTd neurons may only be temporally matched for some directions of motion. Future studies will be needed to assess whether visual and vestibular dynamics are matched on a cell-by-cell and direction-specific basis, and whether the degree of temporal similarity predicts other signatures of sensory integration such as improved sensitivity under combined visual-vestibular stimulation (Gu et al. 2008).

ACKNOWLEDGMENTS

We thank K. Kocher, E. White, and A. Turner for animal care and training. Special thanks to Dr. Katsumasa Takahashi for assisting with data collection.

GRANTS

The work was supported by National Institutes of Health grants ROI EY-017866 and EY-019087 to D. E. Angelaki and EY-016178 to G. C. DeAngelis.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES


J Neurophysiol • VOL 104 • SEPTEMBER 2010 • www.jn.org


Frohler MT, Duffy CJ. Cortical neurons encoding path and place: where you go is where you are. Science 295: 2462–2465, 2002.


J Neurophysiol • VOL. 104 • SEPTEMBER 2010 • www.jn.org