

An Integrated Model of Human Motion Perception with Visual-Vestibular Interaction

Robert J. Telban*

Frank M. Cardullo†

State University of New York
Binghamton, New York

Abstract

The authors conducted a study, the purpose of which was to develop a model of human motion perception that includes both visual and vestibular sensation and incorporates the interaction between the two stimuli. The model includes the effects of visually induced self-motion, normally referred to as vection.

The general characteristics of visually induced self-motion in the absence of confirming vestibular stimuli as supported in the literature is summarized. Several experiments are also described that quantify the latency of the onset of vection.

The authors reviewed existing models that characterize the visual-vestibular interaction. It was revealed that a simple linear summation of the cues failed to predict the perceived response, suggesting that each cue must be weighted based on the amount of inter-cue conflict. It was also suggested that the visual estimate of self-motion is an optokinetic influence that is formed by filtering the difference between the cues through a first-order low-pass filter.

A visual-vestibular interaction model for rotational motion is proposed. The model incorporates mathematical models of the semicircular canals and visual sensory dynamics. The model also includes a conflict signal estimator that controls the optokinetic influence gain. This conflict estimator also models the latency to onset of vection. A model for translational motion is also proposed, incorporating a mathematical model of the otolith dynamics. Model responses with either separate visual or vestibular cues and responses to confirming cues are examined.

* PhD Candidate, Department of Mechanical Engineering, Student Member AIAA

† Associate Professor, Department of Mechanical Engineering, Associate Fellow AIAA

Introduction

The purpose of this study is to develop a model of human motion perception that includes both vestibular and visual motion sensation models and incorporates the non-linear interaction between the visual and vestibular stimuli. This model also includes the effects of visually induced self-motion, commonly known as vection. This integrated perception model will be used for implementation into a proposed motion cueing algorithm and will also be used to evaluate motion cueing responses. This study is based on the literature presented by several researchers who investigated both the characteristics of vection and the visual-vestibular interaction.

Visually Induced Self-Motion

Young¹ described the self-motion response to a full visual field surround rotating about a vertical axis:

“The response to a full field surround which suddenly begins to move at constant velocity is rather startling, although quite repeatable. At first, the veridical motion is sensed – the surround appears to be moving and the subject feels himself stationary. After a period of typically two to five seconds, the visual field appears to slow down, often to a stop, and the subject perceives himself as rotating in the opposite direction. The sensation of rotation builds to a maximum over a period of three to ten seconds, rising approximately as an exponential.”

Young noted that in order to achieve a complete “saturation” of this effect, in which the visual field is perceived to be entirely stationary, it is useful to have a wide, compelling field of view in the periphery, moving uniformly at speeds less than 60 degrees per second. He then notes that if the visual surround is allowed to accelerate smoothly to its final velocity, at accelerations comparable to the acceleration thresholds of the semicircular canals, then the self-motion is more likely to be perceived as a smooth, continuous development of circularvection.

The general characteristics of visually induced self-motion in the absence of confirming vestibular stimuli as reported by Young² and supported by other researchers is summarized. Young noted two distinct

classes of visual cues for flight simulation: the foveal cues, the high acuity, high information-dense central field cues that must be “read” to be interpreted, and the peripheral cues, the wide-field, lower acuity, rapidly moving cues that generate non-cognitive motion perception. These cues correspond respectively to the high static acuity, cone-filled fovea, and the high dynamic sensitivity, rod-filled periphery of the retina.

Brandt, et al.³ demonstrated that the peripheral visual field is of primary importance in stimulating self-motion over the central visual field. Brandt, et al.⁴ showed that background stimulation is dominant over foreground stimulation; movement in the background induces self-motion while if the foreground moved the stationary background inhibited circularvection. Held, et al.⁵ showed that the spatial frequency of the visual scene also determines its effectiveness in generating self-motion. Young² commented that the peripheral visual field display should have a sufficient number of borders such as stars, clouds, or ground features to induce the perceived self-motion. Young² also noted that the visual field velocity determines the magnitude of the self-motion up to an upper limit that most likely corresponds to the blurring of the visual field associated with increased visual acuity.

Young² found that the approximate frequency response for both circularvection and linearvection is flat from static inputs up to a frequency of 0.1 Hz, beyond which it decreases at least as rapidly as a first-order filter. Berthoz, et al.⁶ confirmed these results for forward linearvection, with similar results obtained by Van der Steen⁷ for yaw circularvection.

Latency to Onset of Vection

The latency of the onset of visually induced motion has an impact on motion perception in flight simulation. Young² observed that this latency is highly variable among individuals. Repeated exposures will reduce this latency, as does the development of the appropriate mental set, thus allowing for the development of vection. Several experimenters have quantified this phenomenon.

Brandt, et al.³ conducted experiments using a chair located in a closed cylindrical drum whose inner walls were painted with alternating black and white stripes subtending 7 degrees of visual angle. Both the chair

and the drum could be rotated separately or simultaneously in either the same or opposite directions. The test subjects were asked to fixate on a one-degree luminous spot presented straight ahead. Brandt, et al.³ observed the latency to onset of circularvection to be about three to four seconds and independent of the stimulus magnitude.

Young and Oman⁸ carried out experiments on a fixed-base simulator inside a forty-foot diameter sphere. Visual scenes were projected on the sphere wall interior by two servo-driven plastic hemispheres. A high-intensity light source at the center of each hemisphere projected a pattern of randomly spaced black and white rectangles against a white background. They observed a rapid decrease in the time to onset with increasing visual pattern speed (from 11 seconds at 5 deg/sec to 6 seconds at 10 deg/sec).

Howard and Howard⁹ demonstrated that the latency is reduced with the presence of stationary objects in view and with fixation of the subject’s gaze. With a stationary visual frame similar to a simulator cockpit video monitor and with fixation, they observed latencies of about 5 seconds that were relatively unchanged with stimuli from 5 to 25 deg/sec. Howard and Howard also yielded results within the ranges reported by Young and Oman with no fixation and the presence of vertical rods in the center of the display. A latency of 9.4 seconds was found for a stimulus of 5 deg/sec that decreased to 5.6 seconds for 25 deg/sec.

Berthoz, et al.⁶ tested the latency to onset of forward linearvection. They observed latencies of about 1 to 1.5 seconds for stimulus velocities between 0.2 and 1 m/sec. This significant difference in latency between linearvection and circularvection may be related to the convergence of visual signals with otolith afferents in the former case and semicircular canals in the latter case, taking into account their different response dynamics.

Visual-Vestibular Interaction

Zacharias¹⁰ reported that both psychophysical and neurophysiological studies support the theory that visual and vestibular cues are jointly processed to provide for a perceived sense of self-motion. The vestibular nucleus complex was identified as a possible interaction for the convergence of sensory inputs. Zacharias noted that experiments reported by Dichgans,

et al. on measurements from the vestibular nuclei of goldfish indicate that the majority of units respond to both vestibular and moving visual field inputs. When both stimuli were presented in opposing directions consistent with rotation in the presence of a physically stationary visual surround, the afferent firing rate was characterized by the faster response and greater sensitivity of vestibular stimulation combined with the non-adapting behavior of visual stimulation. The result was a signal that accurately indicated the perceived angular velocity.

Zacharias¹⁰ also reported a study by Young, et al., in which subjective velocity was measured in response to combined yaw-axis rotational cues. The study showed that the subjective velocity response was biased in the direction of the induced circularvection, but not to the extent of a simple summation of circularvection and expected vestibular response. These studies indicate that a simple linear summation of the visual and vestibular cues fails to predict the response when both cues are simultaneously presented.

Young² noted that visual motion cues dominate the perception of velocity in the steady state and at frequencies below 0.1 Hz. At higher frequencies, the vestibular cues will tend to dominate. Confirming vestibular cues, in the direction opposite to the visual field, can produce a rapid onset of self-motion that is sustained by vision after the vestibular cues have been washed out. When visual and vestibular motion cues are in conflict, either due to the direction of motion or to a difference in magnitude, the vestibular cues will initially dominate. Young¹ suggested that when both inputs are presented to a subject simultaneously, he would combine or “mix” the two cues in a non-linear manner, favoring the visual input for confirming cues and the vestibular input for conflicting cues.

Zacharias¹⁰ developed a cue conflict model for yaw perception. For low conflict, i.e. when the cues are confirming, the perceived motion is calculated from a weighted sum of the two signals. For high conflict, the weighting on the visual cue is reduced and that on the vestibular cue is increased until the conflict is reduced. Borah¹¹ later developed a visual-vestibular interaction model that involved the implementation of an optimal estimator as a “central processor” representation of sensory inputs that included a modified version of the cue conflict model proposed by Zacharias.

Van der Steen⁷ proposed a self-motion perception model in which vestibular and visual stimuli are combined to describe perceived self-motion. This model is shown in Figure 1. The model can describe perceived self-motion induced by either vestibular or visual stimuli alone, or a combination of both. However, unlike the model proposed by Zacharias¹⁰, cue conflict estimation is not considered.

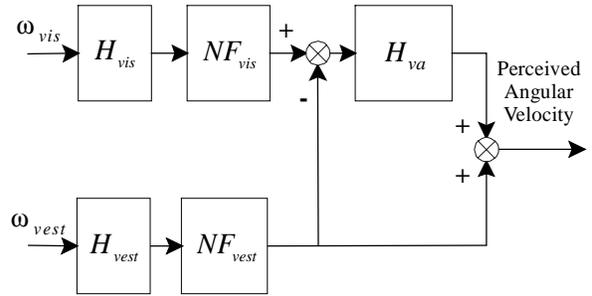


Figure 1. Model for Self-Motion Perception.

Van der Steen⁷ introduced the concept of a neural filter in the model. The neural filter transfers the afferent response of either the visual or vestibular sensor to a perceptual physical variable. The transfer function H_{vest} cascaded with the neural filter NF_{vest} represents the perceived self-motion estimate from vestibular stimuli. The visual receptors transfer function H_{vis} cascaded with the neural filter NF_{vis} represents the self-motion estimate from visual stimuli.

Van der Steen⁷ noted that psychophysical experiments concerningvection showed that the visual estimate of self-motion “attracts” the vestibular estimate. The difference between these cues is passed through a first-order low-pass filter H_{VA} as given in Eqn (1) that represents the gradual build-up of self-velocity, forming the optokinetic influence:

$$H_{VA} = \frac{1}{1 + \tau_{VA}s} \quad (1)$$

The perceived self-motion the model yields is then the sum of the optokinetic influence and the vestibular system’s estimates of perceived self-motion.

Visual Sensory Dynamics

Each visual-vestibular interaction model examined incorporates a model of the visual receptor dynamics that in turn produces a perceptual estimate of the visual scene motion. Zacharias¹⁰ did not model visual sensory dynamics due to the lack of experimental data for single channel visual response, and assumed that the visual system has a relatively wide-band response. Borah¹¹ modeled the dynamics of the visual sensor as unity, noting that the eye detects the visual field motion almost immediately after a short neural transmission delay. Van der Steen⁷ modeled the perceptual dynamics as a cascade of the visual receptor transfer function and neural filter with a unity gain and a delay τ_d :

$$NF_{vis} H_{vis} = -e^{\tau_d s} \quad (2)$$

where the change in sign reflects the fact that the visual field is opposite in direction to the perceived self-motion, i.e. a visual field moving to the left induces self-motion of the subject to the right.

Hosman¹² noted that τ_d is due to the delay of the visual receptors along with the delays due to both neural transmissions from the retina to the visual cortex and information processing during motion perception. From experiments in roll rate perception with visual displays only, Hosman found values for τ_d to be about 90 msec for peripheral visual field stimulation and about 150 msec for central visual field stimulation.

Proposed Rotational Model

A revised visual-vestibular interaction model will now be constructed for rotational motion. This model can be used to estimate perceived motion for yaw, roll, and pitch stimuli. As suggested by Borah¹¹ the visual motion cues considered will be limited to peripheral visual scenes provided by a flight simulator with a wide visual scene field. These peripheral cues would be equivalent to the passage of stars or clouds in a wide field simulation. The cues do not include any elements in the structure of the scene such as landmarks, orientation cues, or a visual horizon.

A visual-vestibular interaction model for rotational motion is proposed and is shown in Figure 2. A model of the semicircular canal afferent dynamics obtained by

Goldberg and Fernandez¹³ is used. Using the concept suggested by Van der Steen⁷, the vestibular model combines the afferent dynamics model with a neural filter gain, resulting in a model with a perceived response to vestibular stimuli. The optokinetic influence proposed by Van der Steen is also implemented. The time constant τ_{va} governing the optokinetic influence is chosen to be 1.592 seconds, which is equal to the 0.1 Hz low-pass filter break frequency noted by Young².

As first proposed by Zacharias¹⁰, the visual cue is passed through an internal model of the vestibular dynamics to produce an “expected” vestibular signal that is then subtracted from the actual vestibular signal. To allow for long-term resolution of steady-state conflict the absolute value of this error is passed through an adaptation operator to generate a conflict signal ω_{err} . The adaptation operator determines how long a steady-state inter-cue conflict should be allowed to continue by washing out the conflict signal. Zacharias suggested that the value for the adaptation time constant τ_w be chosen based on typical latencies observed in a simulator. For the rotational model a time constant $\tau_w = 8$ seconds is chosen to produce the latency responses noted in the literature.

From ω_{err} the weighting of the optokinetic gain K is then computed by a modified cosine bell function suggested by Borah¹¹ and shown in Figure 3. The gain K varies between zero and one. A large conflict signal greater than the conflict threshold \mathcal{E} will drive the gain to zero, whereas a small signal below the threshold value will drive the gain to a value between zero and one, approaching one as ω_{err} approaches zero. For ω_{err} less than zero, the gain remains at one. As proposed by Borah the vestibular path gain remains fixed at unity.

As first suggested by Zacharias¹⁰ the conflict threshold \mathcal{E} is chosen to equal the vestibular indifference motion threshold. From experiments on subjects seated on a motor-driven turntable conducted in darkness with angular velocity stimuli, Benson¹⁴ obtained thresholds of 1.6 deg/sec for yaw stimuli and 2.0 deg/sec for roll and pitch stimuli. These values were found consistent with other researchers who employed angular accelerations as the test stimuli and will be used in the proposed model.

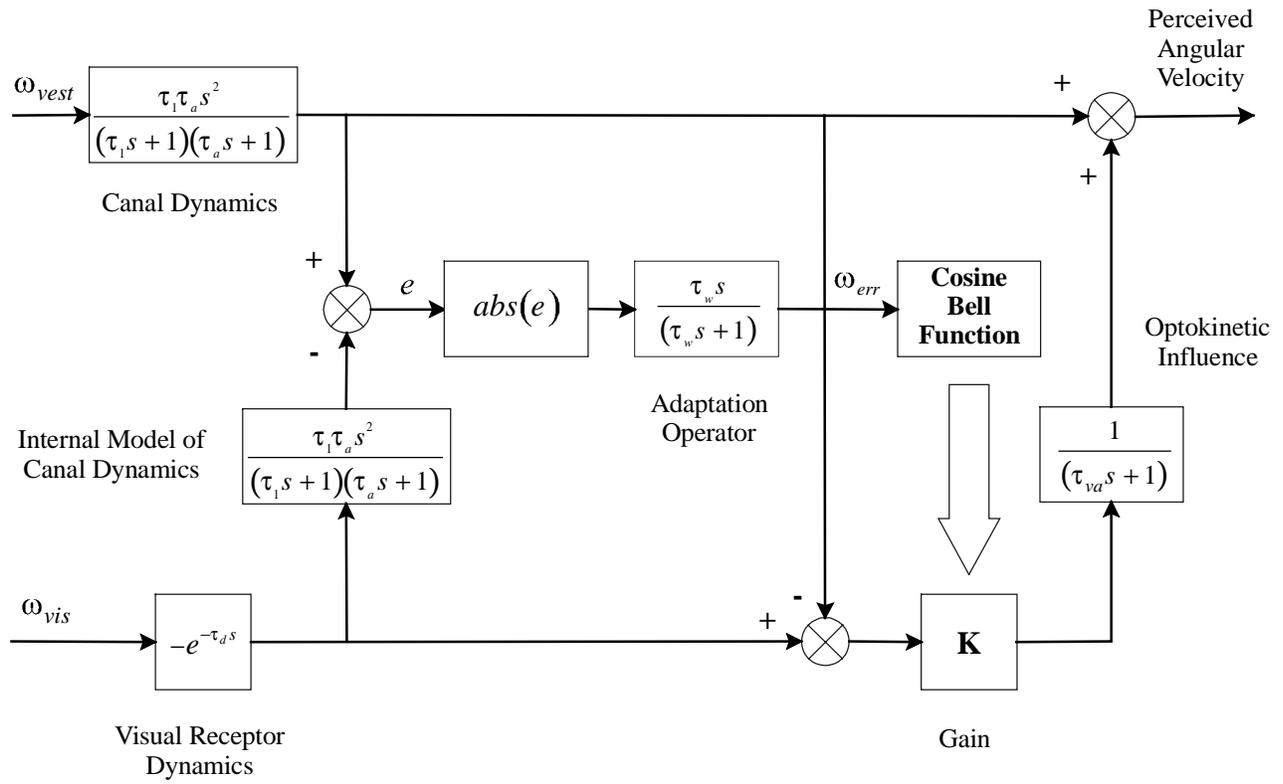


Figure 2. Proposed Visual-Vestibular Interaction Model for Rotational Motion.

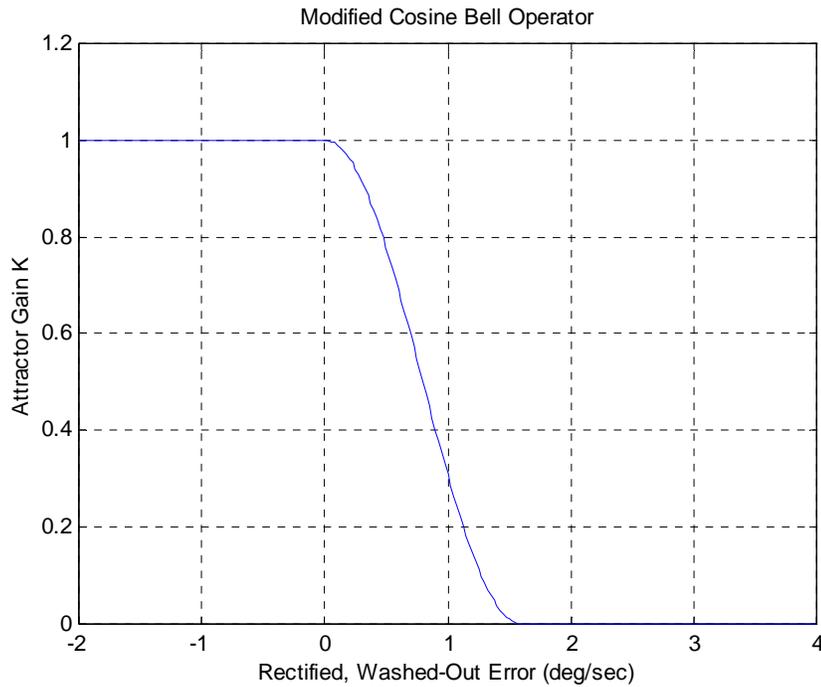


Figure 3. Modified Cosine Bell Operator for Optokinetic Influence Gain.

In order to examine model responses to various stimuli, a MATLAB/SIMULINK representation of the proposed rotational model shown in Figure 2 was developed. Model responses to yaw inputs with either visual or vestibular cues alone, or responses to confirming visual and vestibular cues were examined.

Figure 4 shows the responses to a visual field step input of 10 deg/sec. Since there is no vestibular input the rectified error is the magnitude of the visual input filtered through the internal model of the semicircular canals. The adaptation operator then generates the washout error ω_{err} . Due to the large value of ω_{err} the cosine bell function will produce a gain of zero for nearly five seconds. This results in a corresponding latency in the perceived angular velocity response. Once ω_{err} starts to decrease below the conflict threshold ϵ the optokinetic gain will start to vary between zero and one, resulting in the onset of perceived motion or circularvection. This gain will rapidly rise to a value of one once ω_{err} reaches zero. As ω_{err} becomes negative the gain K remains at one. If a cosine bell operator were applied to this negative response the gain would decrease back to zero, resulting in a large sag in the perceived response. The perceived motion reaches its maximum value with a rise time of about ten seconds, as governed by the time constant τ_{va} .

Various magnitudes of angular velocity inputs were examined in order to compare latency responses with those obtained from psychophysical experiments in the literature. The latencies resulting for the proposed rotational model are shown in Table 1. While the latencies generated by the model increase with increased stimulus magnitude, these values are a compromise between the results obtained by Brandt, et al.³, and by Howard and Howard⁹ with a stationary visual frame and fixation. The latency obtained for a step input of 5 deg/sec will result in a latency that is within the range of values (3 to 4 seconds) obtained by Brandt. For higher magnitudes the latencies are closer to those obtained by Howard (about 5 seconds). When the visual field input is less than or equal to the conflict threshold, the latency to onset of circularvection is reduced to zero.

Table 1. Model Results for Latency to Onset of Circularvection.

ω_{vis} (deg/sec)	Latency (sec)
1.6	0
5	3.05
10	4.3
15	4.85
20	5.15
25	5.325

The model responses to a vestibular step input of 10 deg/sec are shown in Figure 5. The rectified and washout errors are identical to those obtained in Figure 4 for the visual input; therefore the optokinetic gain response is the same as well. During the first five seconds the optokinetic path is gated out, and the decay in the first portion of the perceived response is due only to the semicircular canals. As the gain K increases to one and the optokinetic influence is weighted more heavily, the perceived response will decay more rapidly. Zacharias¹⁰ observed a similar “double exponential” response with responses to vestibular inputs. Borah¹¹ noted that this phenomenon might be due to the fact that for vestibular rotation in the presence of a fixed visual field, the visual surround is actively “denying” the presence of motion.

Figure 6 shows the model responses to a visual field step input of 10 deg/sec, along with a fully confirming vestibular step input of the same magnitude. Note that there is no visual-vestibular error, therefore the optokinetic gain always remains at one. For this special case the proposed model reduces to the Van der Steen⁷ model with the cue conflict estimator not contributing to the response. The perceived response shows a rapid onset due to the semicircular canals that quickly decays, with the optokinetic influence then gradually increasing until the maximum response is achieved.

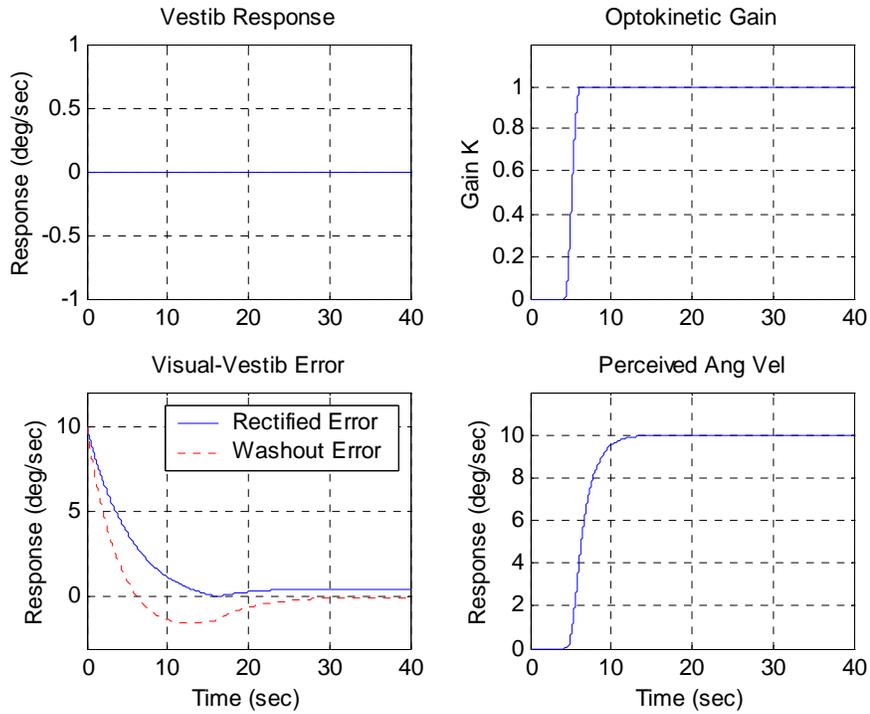


Figure 4. Rotational Perception Model Responses to Visual Field Step Input of 10 deg/sec.

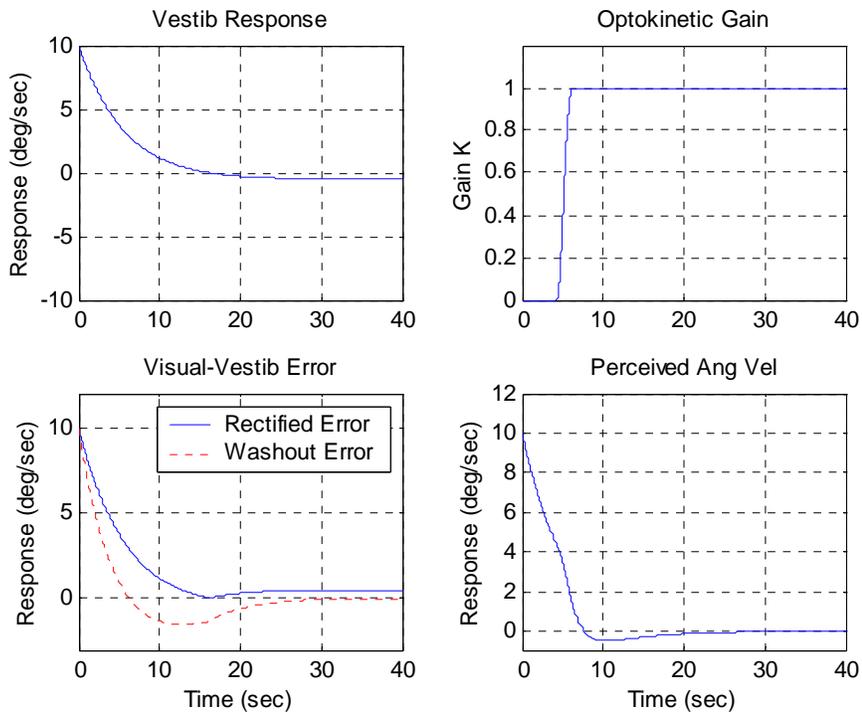


Figure 5. Rotational Perception Model Responses to Semicircular Canals Step Input of 10 deg/sec.

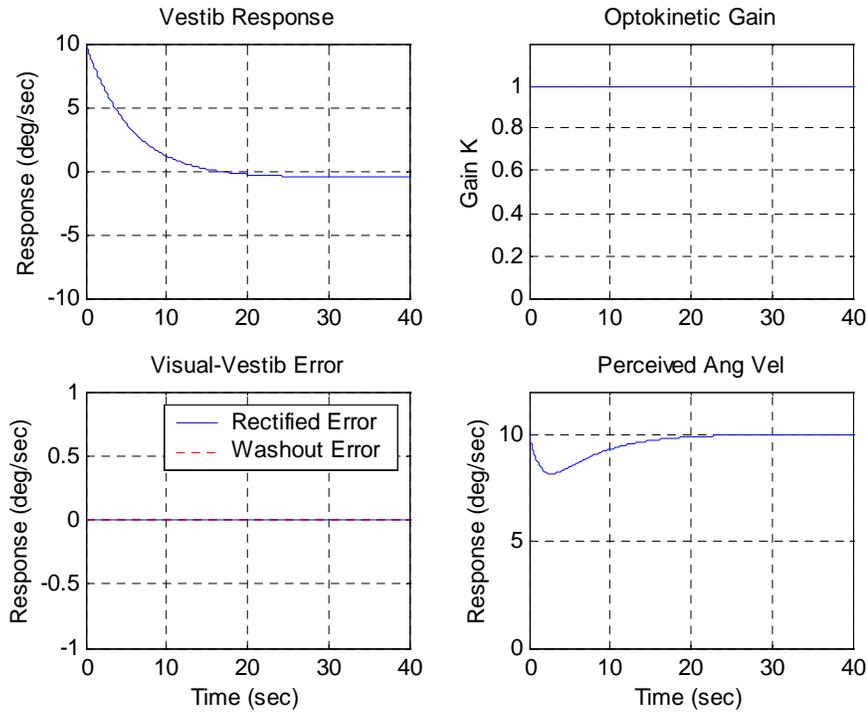


Figure 6. Rotational Perception Model Responses to Confirming Visual and Vestibular Inputs of 10 deg/sec.

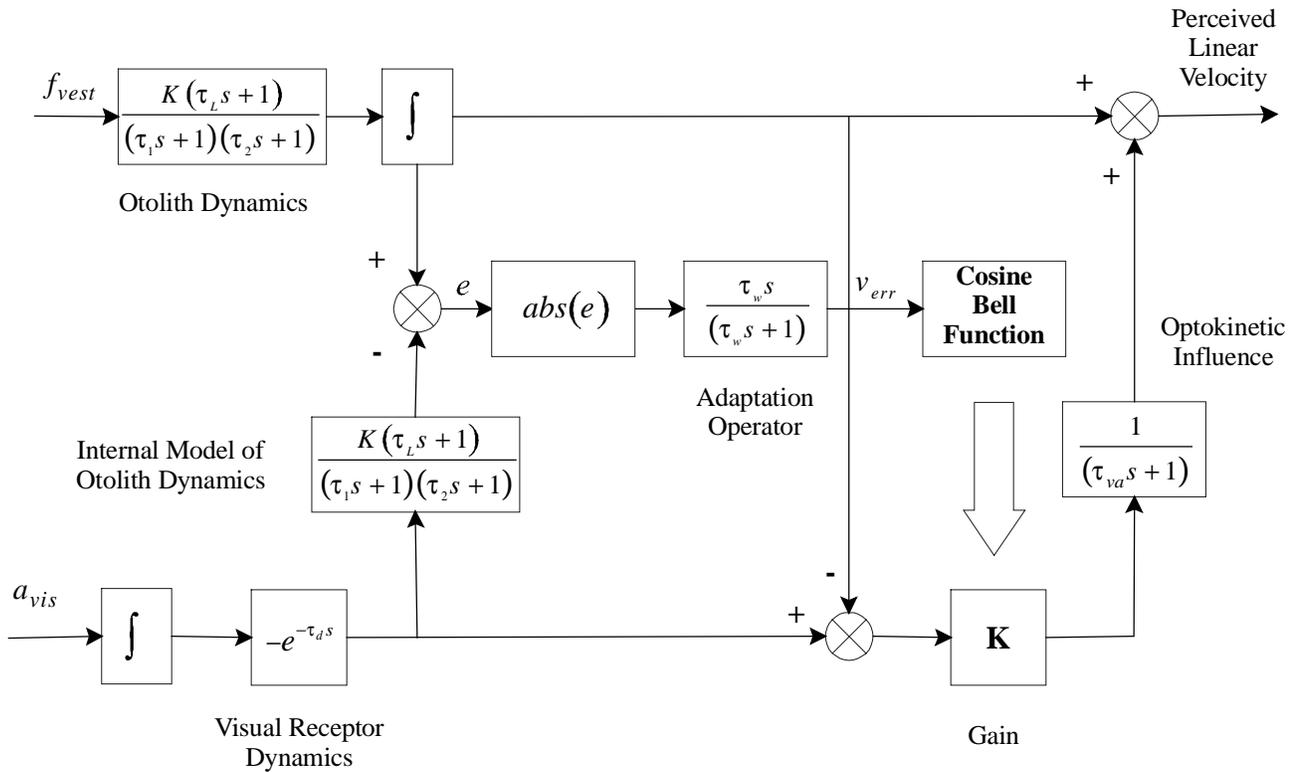


Figure 7. Proposed Visual-Vestibular Interaction Model for Translational Inputs.

Proposed Translational Model

A visual-vestibular interaction model can also be developed to estimate translational motion in the surge, sway, and heave directions. The same assumptions applied to the rotational cues will again be considered. The proposed visual-vestibular motion model for translational motion is shown in Figure 7. The model structure is similar to the rotational model. The otolith model with revised parameters proposed by Telban, et al¹⁵. is used. As proposed with the rotational model, the visual delay will be neglected and the same time constant governing the optokinetic influence will be used. In this model the washout error v_{err} is used to estimate the gain K . Vestibular and optokinetic responses are combined to produce perceived linear velocity, which can then be differentiated to give a perceived specific force response.

As with the rotational model, the conflict signal threshold ϵ is chosen to equal the vestibular indifference motion threshold. Zacharias¹⁰ estimated a threshold of 0.2 m/s from experiments performed in a linear motion simulator by Meiry in which linear acceleration thresholds were measured. The modified cosine bell operator functions similar to that used in the rotational model shown in Figure 2.

A MATLAB/SIMULINK representation of the model shown in Figure 7 was developed. Figure 8 shows responses to a visual field step input of 1 m/sec. An adaptation time constant $\tau_w = 1$ sec was chosen to generate latencies close to those obtained by Berthoz, et al.⁶. The rectified error is the magnitude of the visual velocity response filtered through the internal model of the otoliths. As a result of this faster time constant the washout error decays very quickly, resulting in a much shorter latency of about 1.5 seconds. The perceived linear velocity then reaches its maximum value in about ten seconds, as governed by the time constant τ_{va} .

Various magnitudes of linear velocity inputs were examined in order to compare latency responses with those obtained from psychophysical experiments in the literature. The latencies for $\tau_a = 1$ sec and $\epsilon = 0.2$ m/sec result in latencies that fall within the range of values noted by Berthoz, et al.⁶ (1 to 1.5 m/sec) for velocity inputs from 0.6 m/sec to 1 m/sec. Due to the conflict threshold, the latency decreases to zero when

the input equals the threshold. Table 2 lists the latencies the model generates for inputs from 0.2 to 1 m/sec.

Table 2. Model Results for Latency to Onset of Linearvection.

v_{vis} (m/sec)	Latency (sec)
0.2	0
0.4	0.8
0.6	1.1
0.8	1.275
1.0	1.425

Figure 9 shows the model responses to confirming visual and vestibular pulse inputs of 1 m/s² magnitude and 1 second duration, which produces a ramp to step velocity input of 1 m/s. Note that there is no visual-vestibular error, with the resulting optokinetic gain set at unity. The perceived velocity response shows a rapid onset due to the otolith dynamics that then gradually increases to the maximum response due to the contribution of the optokinetic influence.

Conclusions

An integrated model of human motion perception was proposed. This model includes mathematical models of both vestibular and visual motion sensation and incorporates the non-linear interaction between the vestibular and visual stimuli. The visual estimate of perceived self-motion is modeled as an optokinetic influence that filters the difference between the cues through a first-order low-pass filter that represents the gradual build-up of self-velocity. A conflict signal estimator is used to control the optokinetic influence gain. In this approach a large conflict signal greater than the vestibular indifference threshold will drive the gain to zero, whereas a sub-threshold signal will result in a gain from zero to one. The conflict signal estimator also models the latency to onset of vection as quantified experimentally in the literature.

Models for both rotational and translational motion were developed. Responses to both models yield satisfactory results that concur with the characteristics of self-motion observed in the literature.

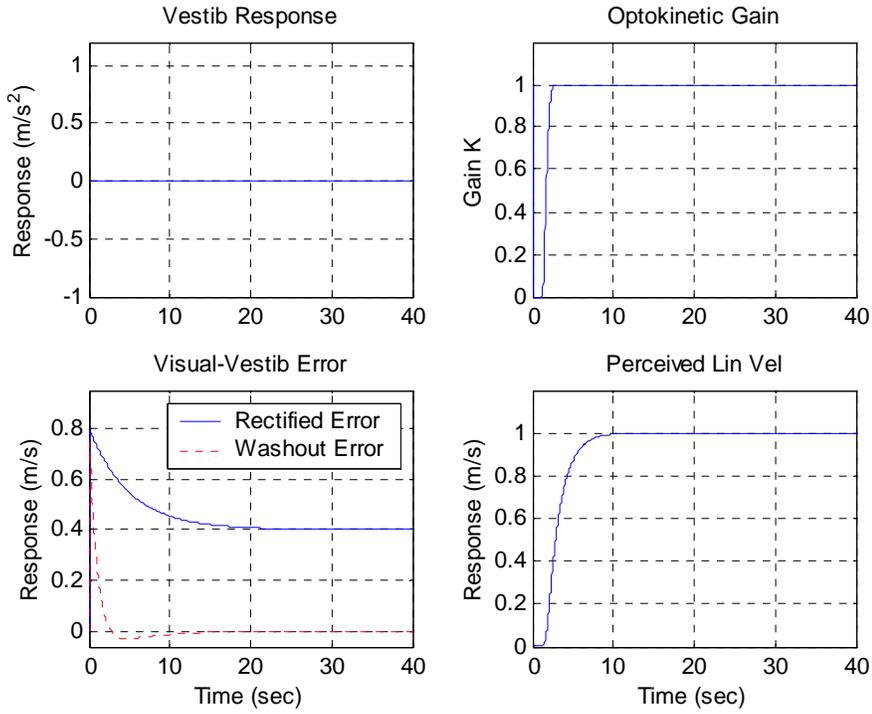


Figure 8. Translational Perception Model Responses to Visual Field Step Input of 1 m/sec.

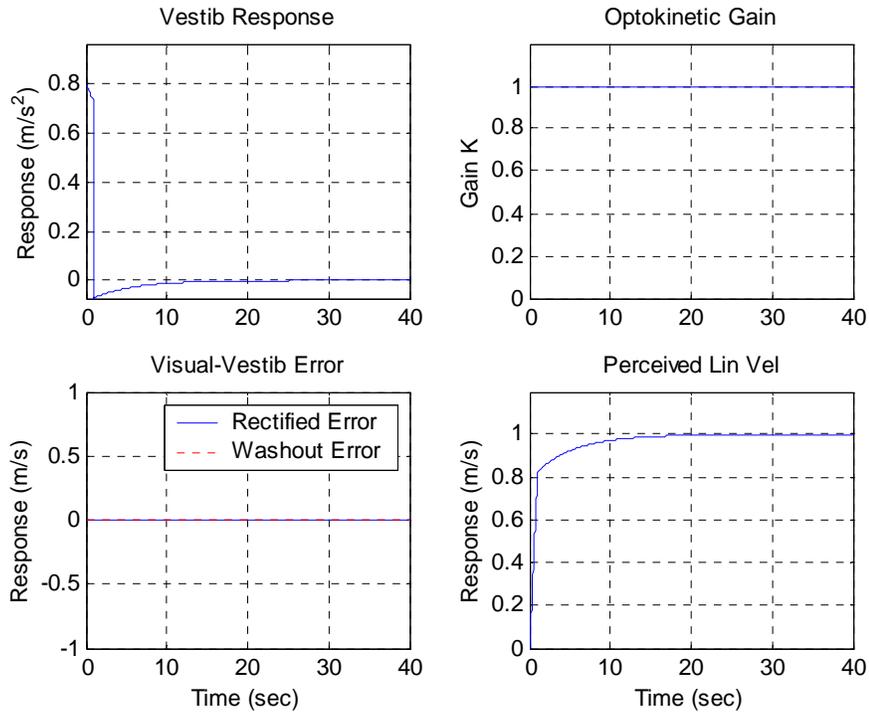


Figure 9. Translational Perception Model Responses to Confirming Visual and Vestibular Pulse Inputs of 1 m/s^2 Magnitude and 1 second Duration.

References

- ¹ Young, L.R. *Visual-Vestibular Interaction*. in *Sixth International Symposium on Biocybernetics*. 1977. Leipzig, East Germany.
- ² Young, L.R. *Visually Induced Motion in Flight Simulation*. in *AGARD Symposium on Flight Simulation*. 1978. Brussels, Belgium.
- ³ Brandt, T., Dichgans, J., and Koenig, E., *Differential Aspects of Central Versus Peripheral Vision on Egocentric and Exocentric Perception*. *Experimental Brain Research*, 1973. **16**(476-491).
- ⁴ Brandt, T., Wist, E. R., and Dichgans, J., *Foreground and Background in Dynamic Spatial Orientation*. *Perception & Psychophysics*, 1975. **17**(5): p. 497-503.
- ⁵ Held, R., Dichgans, J., and Bauer, J., *Characteristics of Moving Visual Scenes Influencing Spatial Orientation*. *Vision Research*, 1975. **15**: p. 357-365.
- ⁶ Berthoz, A., Pavard, B., and Young, L. R., *Perception of Linear Horizontal Self-Motion Induced by Peripheral Vision (Linearvection)*. *Experimental Brain Research*, 1975. **23**: p. 471-489.
- ⁷ Van der Steen, H., *Self-Motion Perception*. 1998, Ph.D. Thesis, Delft University of Technology: Delft, The Netherlands.
- ⁸ Young, L.R., and Oman, C. M. *Influence of Head Position and Field on Visually Induced Motion Effects in Three Axes of Rotation*. in *Proceedings of the 10th Annual Conference on Manual Control*. 1974.
- ⁹ Howard, I.P., and Howard, A., *Vection: The Contributions of Absolute and Relative Motion*. *Perception*, 1994. **23**: p. 745-751.
- ¹⁰ Zacharias, G.L., *Motion Sensation Dependence on Visual and Vestibular Cues*. 1977, Ph.D. Thesis, Massachusetts Institute of Technology: Cambridge, MA.
- ¹¹ Borah, J., Young, L. R., and Curry, R. E., *Optimal Estimator Model for Human Spatial Orientation: Representation of Three-Dimensional Space in the Vestibular, Oculomotor, and Visual Systems*, in *The New York Academy of Sciences*. 1988: New York, New York. p. 51-73.
- ¹² Hosman, R.J., and Van der Vaart, J. C. *Perception of Roll Rate from an Artificial Horizon and Peripheral Displays*. in *19th Annual Conference on Manual Control*. 1983. Cambridge, MA.
- ¹³ Goldberg, J.M., and Fernandez, C., *Physiology of Peripheral Neurons Innervating Semicircular Canals of the Squirrel Monkey. II. Response to Sinusoidal Stimulation and Dynamics of Peripheral Vestibular System*. *Journal of Neurophysiology*, 1971. **34**(4): p. 661-675.
- ¹⁴ Benson, A.J., Hutt, E. C. B., and Brown, S. F., *Thresholds for the Perception of Whole Body Angular Movement About a Vertical Axis*. *Aviation, Space, and Environmental Medicine*, 1989. **60**: p. 205-213.
- ¹⁵ Telban, R.J., Cardullo, F. M, and Guo., L. *Investigation of Mathematical Models of Otolith Organs for Human Centered Motion Cueing Algorithms*. in *AIAA Modeling and Simulation Technologies Conference*. 2000. Denver, CO.