



**HAL**  
open science

# Multimodal Fusion in Self-Motion Perception using Wavelets and Quaternions

Nizar Ouarti, Alain Berthoz

► **To cite this version:**

Nizar Ouarti, Alain Berthoz. Multimodal Fusion in Self-Motion Perception using Wavelets and Quaternions. Deuxième conférence française de Neurosciences Computationnelles, "Neurocomp08", Oct 2008, Marseille, France. hal-00331621

**HAL Id: hal-00331621**

**<https://hal.science/hal-00331621>**

Submitted on 17 Oct 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# MULTIMODAL FUSION IN SELF-MOTION PERCEPTION USING WAVELETS AND QUATERNIONS

Nizar OUARTI  
LPPA

UMR 7152 College de France /CNRS  
11, place Marcelin Berthelot  
Paris, France  
email: nizar.ouarti@college-de-france.fr

Alain BERTHOZ  
LPPA

UMR 7152 College de France /CNRS  
11, place Marcelin Berthelot  
Paris, France  
email: alain.berthoz@college-de-france.fr

## ABSTRACT

In this new model of self-motion perception we focused on two main questions that are central in this field. The first one is how the central nervous system (CNS) merges information coming from different sensory modalities (as vision, touch, vestibular system,...). Another important question is how the system uses translation and rotation information properly to give a correct estimation of motion in three dimensions. These two questions have many ramifications like gravito-inertial force problem or the explanation of the non-linearity of multimodal fusion. We propose that the multimodal fusion is realized in the time/frequency space. In addition, we propose that the CNS uses a population of quaternion-like cells to process rotational and translational information. This model can be considered as a general framework based as much on theoretical concerns as on experimental results of different studies.

## KEY WORDS

self-motion, multimodal, fusion, wavelet, quaternion.

## 1 Introduction

The ability to perceive one's own motion in three dimensions is fundamental in everyday life. Whether it is to stabilize images on retina, to realize a complex movement or to keep a stable posture, the perception of self-motion is involved and vital. Some theoretical problems are central for the modeling of self-motion perception. Among them we assume that the most important are based on fusion of information of different kinds. Indeed the particularity of self-motion perception is to use different modalities as vision, vestibular system and different proprioceptive systems. The second kind of information that the central nervous system (CNS) have to merge properly, is the information of translation and rotation that can come from all the modalities that we quoted. A model of perception of self-motion have to answer to these two questions to be able to explain how the CNS combines the different sources of information.

To answer to the question of multimodal fusion, we can firstly ask why the neural system need to use different modalities, and not only utilizes, for instance, vestibular

information. It has been suggested that one possible answer is linked to the fact that the different sensory modalities are complementary in term of frequency [1, 2]. For instance the visual system is known to be more accurate in very low frequencies [3, 4] approximately in the interval [0.14Hz; < 0.008Hz] and the semi-circular canals have an higher frequency passband, approximately [0.05Hz;30 Hz]. On the contrary, a proprioceptive sensory system as touch has a very high pass filter behavior with a passband interval of [70Hz;1000Hz] (for Pacini's receptors). It seems that the different captors involved in self motion perception are complementary in the frequency domain. Considering how neurons code information, a frequency coding of information can be considered as a realistic possibility. In addition, in recent years, increasing works emphasized the role of time in self motion perception [5, 6, 7]. For all these reasons, we propose that CNS uses a coding based on time/frequency representation for self-motion, and we consider that this assumption is realistic and biologically plausible.

Moreover, given that CNS uses a time/frequency coding, from a signal processing theoretical point of view, the choice of a proper time/frequency transform is crucial. For any system which has to analyze a signal in the frequency domain, the problem of the accurate localization in time of a given frequency is essential. We assume that the CNS, as a system which have to detect different combinations of frequencies in the appropriate time window, has exactly the same constraints. It is clearly understandable why the introduction of a finite temporal windows of observation for the CNS is very important to determine the temporal localization of given non-stationary frequencies. One possible implementation is to analyze the signal in a fix window of observation. It is exactly what the Gabor Transform proposes:

$$G_f(b, w) = \int_{\mathbb{R}} f(t) \overline{g(t-b)} e^{-iw(t-b)} dt \quad (1)$$

, where  $g(t)$  is the time window function and  $\|g\| = 1$  and  $w$  the angular frequency.

A principle called the inequality of Eisenberg states that there are a limit between the possible precision of the observation time window and the precision of the frequency

that can be detected by a given system. In changing the size of this window, it is possible to be closer to this limit for a given frequency. This is a complicated question because ecological signals are not usually pure frequency but a melting of many. It comes that having a unique size of observation window or to have sequentially different size of window do not give an optimal result. The solution to this problem is to analyze the signal with different windows of different sizes but in parallel. One possibility is to work with many windows at different scales. This is very close to the definition of wavelets.

Another important question is linked to geometrical problems. How the CNS uses and combines rotation and translation information to create an adequate percept of motion in 3D. We propose that specialized neurons of the vestibular nuclei have the ability to change the reference frame in real time. These neurons combine many inputs and we assume that they use quaternion-like activation function to compute the rotation of the reference frame. To be more accurate we assume that these neurons could perform a kind of quaternions multiplication. We differentiate two cases, the first one multiplication between a unit quaternion and an acceleration vector which make a rotation of frame, and a second one between a unit quaternion and a second unit quaternion, which represents a combination of rotation. Indeed, it is known that VN neurons can fusion information coming from canal and otoliths conveying translation information [8]. In our model translation information are considered as pure quaternion. But we know also that otoliths can give an information of rotation in the same VN neurons [9], in this case we considered that these rotating vectors coming from otoliths are unit quaternions.

## 2 Gravito-inertial Problem

Our approach is very low level because the goal is to model a population of neurons but it is important that it fits with psychophysics results. One of the most popular problem is the gravito-inertial force (GIF) problem, it comes from the observation that otoliths can be considered, in first approximation, as accelerometers. Thus like all accelerometers they cannot differentiate between linear acceleration and the tilting (gravitational acceleration) [2, 10, 11, 12, 13]. This can be represented mathematically by :

$$\vec{f} = \vec{g} + \vec{a} \quad (2)$$

where  $\vec{f}$  is the gravito-inertial force per unit mass transduced by the otoliths afferents,  $\vec{g}$  is the gravitational force per unit mass and  $\vec{a}$  the linear acceleration.

However tilt [14] and translation [15, 16, 17] can be perceived independently by humans. Thus the conclusion is that other information are needed to disambiguate otoliths information.

Former studies [2, 18] argued that two different populations of otolithic primary afferents activated during motion serve to identify the two parts of the signal :linear motion and

tilt. These populations of neurons are tonic units (low pass filters) and phasic units (high pass filters), tonic units being linked with gravitational acceleration and phasic with linear acceleration. More recently a determinant physiological study ([12]) showed in monkeys that, under a condition where resultant gravity for otoliths where 0G, the specific linear vestibulo-ocular reflex (LVOR) induced by the translation of the animal disappears only when the semicircular canals are inactivated. This result showed that semicircular canals information is used to compute the estimate of linear acceleration.

It is important here to note that there is no opposition between the two approaches. Indeed the CNS can use the information provided by the two types of otolithic units and at the same time, use the information coming from semicircular canals. Moreover it is exactly what Mayne (1974) [18] had proposed in his model.

We will subsequently explain how our new approach can be considered as a possible solution of the GIF problem.

## 3 The Model

### 3.1 Quaternion

A quaternion can be define as follow:

$$q = a+bi+cj+dk \quad \{a, b, c, d\} \in \mathbb{R}, \{i, j, k\} \in \mathbb{Z} \quad (3)$$

If we consider the unit quaternion  $\|q\| = 1$ , which represents a rotation in three dimensions, we can define the rotation of a three dimensional vector  $V$  as:

$$V_{rot} = V_{init} + 2Q(\theta).V_{init} \quad (4)$$

$$Q(\theta) = \begin{pmatrix} -(c^2 + d^2) & bc + ad & bd - ac \\ bc - ad & -(b^2 + d^2) & cd + ab \\ bd + ac & cd - ab & -(b^2 + c^2) \end{pmatrix} \quad (5)$$

$$V_t = V_{t-1} + 2Q((w(t) * \Delta t)).V_{t-1} \quad (6)$$

where  $w$  is the angular velocity.

We assume that quaternions-like cells can be useful to rotate the frame of observation of translations. Thus in this kind of computation, the direction of gravity is theoretically always known. This can facilitate the computation of acceleration. Because it can be inferred that if the resultant force is not aligned with the direction of gravity, it means that an additional acceleration occurs. One can argue that a limitation of the model happens when inertial acceleration is aligned with gravity. But it is exactly what can be observed, indeed many studies report a limitation to evaluate direction of acceleration for Z axis (earth vertical) stimulations [19, 20]. Thus quaternion-like cells could be a first answer to the resolution of the GIF problem.

### 3.2 Von Mises-Fisher distribution

In this article we used the Von Mises-Fisher distribution to represent the activation of canal/otoliths cells by linear acceleration only. The reason of the use of this distribution is

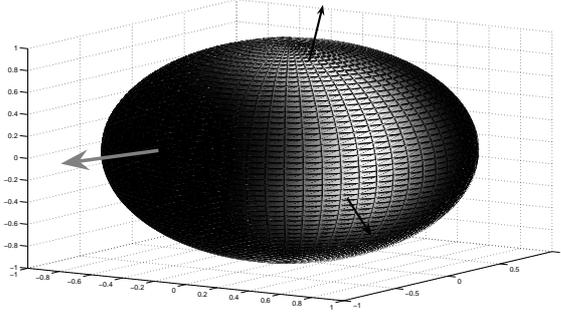


Figure 1. Von Mises-Fisher distribution on the unit sphere during a rotation about the gray axis.  $\mu = 15$  and the initial direction of acceleration was  $v_{init} = [0, 0.2, 1]$ , the quaternion axis was  $[0.2, 1, 0.2]$  and the rotation was 120 deg. The final direction of acceleration in the rotated frame was  $v_{final}[-0.69, 0.62, -0.42]$ .

linked to the non-cosine activation of these cells (contrary to primary afferent cells) which is never negative. Von Mises Fisher distribution in dimension three can be defined as :

$$VMF(x) = \frac{\mu^{1/2}}{(2\pi)^{3/2} Bes_{1/2}(\mu)} e^{(\mu v x)} \quad (7)$$

with  $Bes_{1/2}$  represents the Bessel function of the first kind and order  $1/2$ ,  $\mu$  is the concentration parameter and  $v$  the mean direction vector.

Our method is illustrated in the figure 1. It shows the initial vector of acceleration (black arrow on the top) and the final vector after the rotation of the frame (black arrow on the bottom). This graphic also shows the axis of rotation (gray arrow). In this example we showed the path of activation of the cell during acceleration and rotation. This path representation corresponds to the addition of activations during few seconds. Indeed at a given time the activation function corresponds to a Von Mises-Fisher distribution on the sphere but of course rotated. However, the path representation can give a better idea of the dynamic of the spatial tuning of our quaternion-like cells.

### 3.3 Wavelet

A wavelet transform can be defined as follows [21]:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad s \in \mathbb{R}^{+*}, \tau \in \mathbb{R} \quad (8)$$

with  $s$  the scaling factor and  $\tau$  the translation factor. The transform of a signal  $f(t)$  is defined as :

$$C_f(s, \tau) = \int_{\mathbb{R}} f(t) \overline{\psi_{s,\tau}(t)} dt \quad C_f \in \mathbb{R}^2 \quad (9)$$

And the reverse transform is :

$$f(t) = \frac{1}{K_\psi} \int_{\mathbb{R}^{+*} \times \mathbb{R}} C_f(s, \tau) \psi_{s,\tau}(t) \frac{ds d\tau}{s^2} \quad t \in \mathbb{R} \quad (10)$$

The admissibility constant  $K_\psi$  is defined as:

$$K_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega = \int_{-\infty}^0 \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < +\infty \quad (11)$$

which implies :

$$\text{For } \psi \in \mathbb{R}, \psi \in L^1 \cap L^2, \int_{\mathbb{R}} \psi(t) dt = 0 \quad (12)$$

The minimal admissibility constraints on wavelet functions are expressed by equation (12).

One of the limitation of the classic definition of a continuous wavelet is his lake of accuracy in low frequencies. It is the reason why we propose a function that complement the wavelet in low frequencies. Basically this continuous function have the same role than the scaling function. And the most important, this kind of function is biologically plausible because the CNS can process very low frequencies and some cells involved in self motion perception have a low pass filter behavior, like tonic otoliths units. The most important constraint for this function  $\phi$  is :  $\int_{\mathbb{R}} \phi(t) dt = 1$ .

We argued that the CNS probably work in the time/frequency domain. But we still do not explained how the information coming from different sources are merged to give the unique percept. In this article we propose a non linear fusion. This fusion is the  $max(f(t), g(t), h(t), \dots)$  in the time/frequency domain. The interest of the  $max$  is to be a very adaptive way to merge data without a priori about the variance of the different sources. Indeed the variance can change in different situations, for instance the detection of visual motion depend dramatically of the different cues used. And in a natural scene, the cues can change very quickly. Thus an algorithm based on the variance of the different captors has to find a compromise if the variance parameter is stored and to recompute it each time is computationally expensive. With the max algorithm, the CNS find the better source for a given time and frequency and with a weak computational cost. For instance the max function in figure (2) would be the upper curve at each moment. Moreover this algorithm is non linear as it is well known that it is the case for the fusion in self motion perception. The kind of result of the figure (2) with the use of the  $max$  function gives results very close to VN responses in alert monkey which are exposed to visual and vestibular stimulations [3].

In addition, our time/frequency approach give a natural framework for frequency segregation for solving the GIF problem. The VN neurons have variable frequency behaviors from low pass to high pass filters which is coherent with our assumption about time/frequency decomposition. This also allows a frequency segregation between gravity

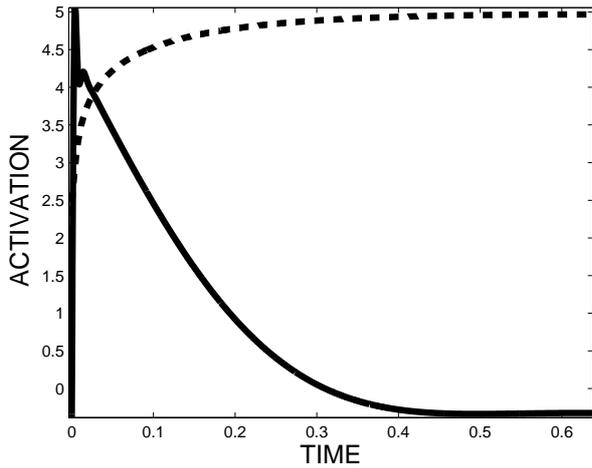


Figure 2. Step of acceleration process with a wavelet (solid line) and the low pass complementary filter of a wavelet (dotted line). In this example the wavelet is a Laplacian of Gaussian and the low pass filter function is a Gaussian. The curve are the result of a continuous decomposition and a reverse transform.

and inertial acceleration. Gravity being extracted from the low frequencies part of the signal and the inertial acceleration from the high frequencies part. One can argue that a limitation of the model is that constant inertial acceleration could be considered by the CNS as gravity. But it is exactly what can be observed, indeed many studies report a limitation to evaluate constant acceleration that is called the somatogravic illusion. Thus multi-resolution approach could be another appropriate answer to the resolution of the GIF problem.

#### 4 Application: Path Completion

We applied our model on data coming from a triangular path completion task. We used some sensors (accelerometers and gyrometers) to store the data related to the motion of the head of the subject (fig. 3). The activation of the population of quaternion-like neurons is shown in figure (5). This figure represent the activation of a population of neurons at a given time.

For the time-frequency aspect of our study, we showed the effect of the width of the window of integration. We used different sizes of window on the signal. And as expected, the larger the windows is, better is the accuracy of the result (see fig. 6 A and B). We also show that the "scaling function" is important for the accuracy of the restitution of the signal. In the figure (4), we showed the decomposition of the signal in the time-frequency domain with the help of the wavelet and the "scaling function". We computed the PSNR before and after the addition of the reversed transform coming from the "scaling function". It

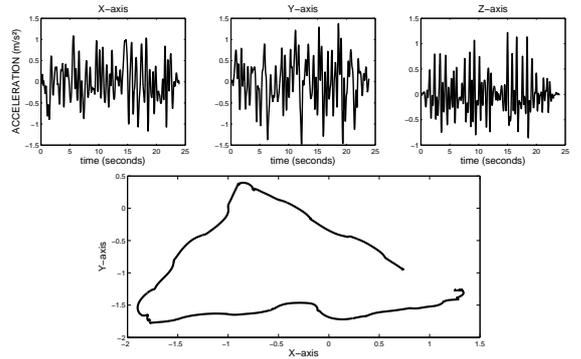


Figure 3. Top : Acceleration in three axes (orthogonal axes). Bottom : Path of the subject, double integration of the acceleration.

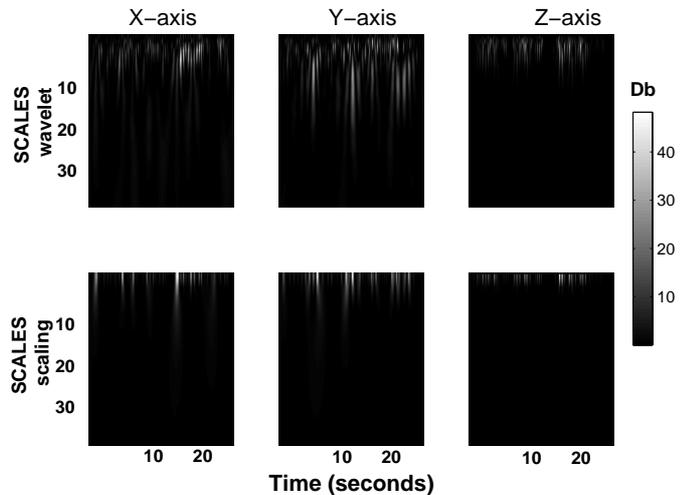


Figure 4. Time-frequency decomposition in the three axes. Top : wavelet decomposition; Bottom "scaling function" decomposition. The additional accuracy is evaluated with the PSNR. Respectively for the X,Y and Z-axis. The PSNR without "scaling function" (with a maximal windows of 25 seconds) were: 23.06, 24.67, 28.34 and with the "scaling function" : 24.65, 26.39, 28.35

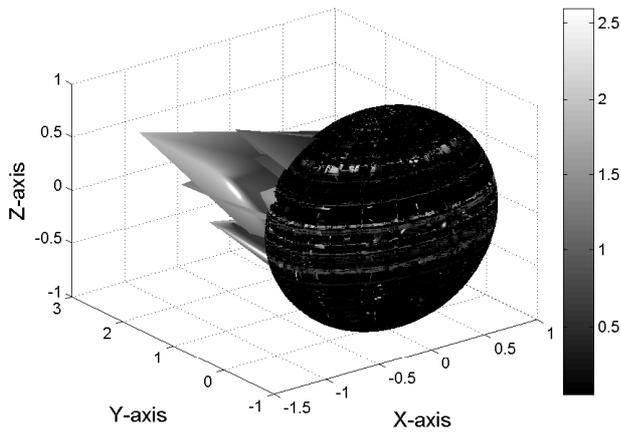


Figure 5. Activation of the population of neurons by an acceleration and a rotation. It can be seen that the most important part of the population had a resting discharge near from zero except some neurons which answered to the activation in a given direction of motion.

can be observed that the accuracy is always increased with the "scaling function" for different axes.

Another very important aspect of the model is the number of neurons in the population (see fig. 6 C and D). With 3 neurons equivalent of an orthogonal frame, the precision of the restitution of the curve is poor. But with 90 neurons, the accuracy can be considered as correct. To properly evaluate the effect of the number of neurons, we tested different numbers and we compared these results with the original data with the help of the PSNR. The results are summarized in the figure(7).

We can conclude that population of neurons are a good alternative to classical orthogonal frame. Population coding is also more biologically plausible than cartesian coding. The other important point in this study is to point out that the ideal maximal size of analysis window is difficult to determine. A large window will abusively use the information of future and past and the small window could provoke a loss in accuracy. It seems that the CNS find a kind of trade-off which results in a "real-time" and "accurate" perception of self-motion.

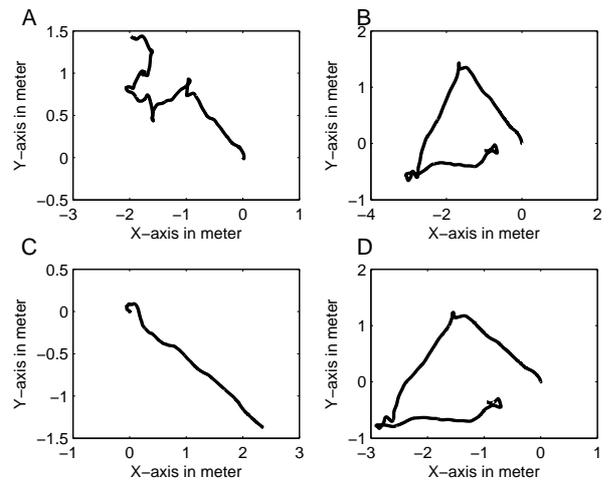


Figure 6. A: Reconstruction of the path after a wavelet-scaling decomposition with a maximum width window of 25 seconds. B: Reconstruction of the path after a wavelet/scaling decomposition with a maximum width window of 145 seconds. C: Reconstruction of the path after the computation with a population of 3 neurons. D: Reconstruction of the path after the computation with a population of 90 neurons.

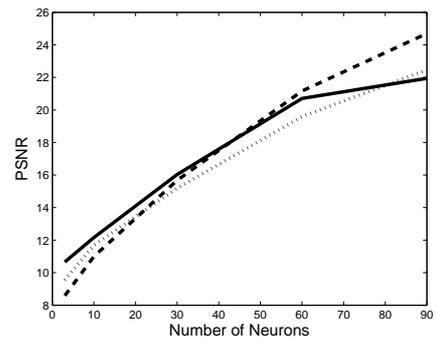


Figure 7. Effect of the increasing number of neurons in the population on the accuracy of the signal, evaluated with the PSNR. The dotted curved represent the X-axis, the dashed curve the Y-axis and the solid curve the Z-axis.

## References

- [1] L. H. Zupan, D. M. Merfeld, and C. Darlot. Using sensory weighting to model the influence of canal, otolith and visual cues on spatial orientation and eye movements. *Biol Cybern*, 86(3):209–230, Mar 2002.
- [2] J. Droulez and C. Darlot. *Attention and performance XIII*, chapter The geometric and dynamic implications of the coherence constraints in three dimensional sensorimotor coordinates, pages 495–526. Laurence Erlbaum Ass., New Jersey, 1989.
- [3] W. Waespe and V. Henn. Neuronal activity in the vestibular nuclei of the alert monkey during vestibular and optokinetic stimulation. *Exp Brain Res*, 27(5):523–538, Apr 1977.
- [4] J. Dichgans, R. Held, L. R. Young, and T. Brandt. Moving visual scenes influence the apparent direction of gravity. *Science*, 178(66):1217–1219, Dec 1972.
- [5] I. Israel, I. Siegler, S. Rivaud-Pchoux, B. Gaymard, P. Leboucher, M. Ehrette, A. Berthoz, C. Pierrot-Deseilligny, and T. Flash. Reproduction of self-rotation duration. *Neurosci Lett*, 402(3):244–248, Jul 2006.
- [6] Stefan Glasauer, Erich Schneider, Renato Grasso, and Yuri P Ivanenko. Space-time relativity in self-motion reproduction. *J Neurophysiol*, 97(1):451–461, Jan 2007.
- [7] A. Capelli and I. Isreal. One second interval production task during post-rotatory sensation. *Journal of Vestibular Research*, 2008.
- [8] J. David Dickman and Dora E Angelaki. Vestibular convergence patterns in vestibular nuclei neurons of alert primates. *J Neurophysiol*, 88(6):3518–3533, Dec 2002.
- [9] A. J. Benson, F. E. Guedry, and G. M. Jones. Response of semicircular canal dependent units in vestibular nuclei to rotation of a linear acceleration vector without angular acceleration. *J Physiol*, 210(2):475–494, Sep 1970.
- [10] T. Vieville and O. D. Faugeras. Cooperation of the inertial and visual systems. pages 339–350, 1990.
- [11] S. Glasauer. Interaction of semicircular canals and otoliths in the processing structure of the subjective zenith. *Ann N Y Acad Sci*, 656:847–849, May 1992.
- [12] D. E. Angelaki, M. Q. McHenry, J. D. Dickman, S. D. Newlands, and B. J. Hess. Computation of inertial motion: Neural strategies to resolve ambiguous otolith information. *Journal of Neuroscience*, 19:316–327, Jan 1999.
- [13] D. M. Merfeld, L. H. Zupan, and C. A. Gifford. Neural processing of gravito-inertial cues in humans. ii. influence of the semicircular canals during eccentric rotation. *Journal of Neurophysiology*, 85:1648–1660, Apr 2001. Article.
- [14] B. Clark and A. Graybiel. Perception of the visual horizontal in normal and labyrinthine defective subjects during prolonged rotation. nsam-936. *Res Rep U S Nav Sch Aviat Med*, pages 1–7, Jun 1965.
- [15] A. J. Benson, M. B. Spencer, and J. R. R. Stott. Thresholds for the detection of the direction of whole-body, linear movement in the horizontal plane. *Aviation Space and Environmental Medicine*, 57:1088–1096, 1986.
- [16] G. Melvill Jones and L. R. Young. Subjective detection of vertical acceleration: a velocity- dependent response. *Acta Otolaryngologica (Stockholm)*, 85:45–53, 1978.
- [17] I. Israel and A. Berthoz. Contribution of the otoliths to the calculation of linear displacement. *Journal of Neurophysiology*, 62(1):247–263, 1989.
- [18] R. Mayne. *A systems concept of the vestibular organs*, pages 493–580. Springer Verlag, Berlin, Heidelberg, New-York, 1974.
- [19] G. M. Jones and L. R. Young. Subjective detection of vertical acceleration: a velocity-dependent response? *Acta Otolaryngol.*, 85:45–53, Jan 1978. DA - 19780417IS - 0001-6489 (Print)LA - engPT - Journal ArticleSB - IMSB - S.
- [20] R. Malcolm and G. Melvill Jones. Erroneous perception of vertical motion by human seated in the upright position. *Acta Otolaryngologica (Stockholm)*, 77:274–283, 1974.
- [21] Y. Meyer. *Les ondelettes : algorithmes et applications*. Armand Colin, 1992.