

Evaluation of Head Movements in Short-term Measurements and Recordings with Human Subjects using Head-Tracking Sensors

Gy. Wersényi¹, J. Wilson²

¹Széchenyi István University, Department of Telecommunications
Egyetem tér 1, H-9026, Győr, Hungary
E-mail: wersenyi@sze.hu

²Georgia Institute of Technology, Interactive Media Technology Center, Atlanta
E-mail: jeff.wilson@imtc.gatech.edu

Abstract: Measurements for spatial hearing research, binaural recordings, virtual reality techniques etc. often rely on Head-Related Transfer Functions of human subjects and head-tracking techniques. Individually measured HRTFs and recordings on human heads may result in more accurate localization and sound field rendering. On the other hand, the measurement and recording procedure raise new problems such as decreased signal-to-noise ratio or subject comfort. For measurements with human subjects lots of methods are used from free heads to different head fixation methods. In this paper, we report an experiment that was conducted using commercially available sensors with the goal of characterizing the range of subject head movements in various postures under various circumstances. The study analyses the range of unwanted head movements during measurements using two sensors, 3-min sessions and four body positions based on the circular angle variance, errors in yaw-pitch-roll directions and magnitude of standard deviation. Results of 16 participants show errors about 2 degrees and magnitudes of standard deviation of 2-8 cm depending on the situation as well as a preference for sitting instead of standing posture.

Keywords: *head movement, binaural recording, HRTF, visual tracking*

1. Introduction

Measurements with human individuals that require stability of the subject are part of several research areas. These include medical applications, engineering and information technology approaches. The common problem is that subjects have to be instructed to be motionless, because even unwanted small movements of the body can distort data

capture. Therefore, different fixing methods are usually applied from simple head-rests to full-body fixations. Fixing installations, however, can interact with the measurement equipment (reflections, electromagnetic effects etc.), and fixation itself can increase discomfort of the subjects.

1.1. Measurements in spatial hearing research with human subjects

From the sound engineering point of view individual measurements with human subjects are dealing with this problem. Virtual audio simulators, auditory displays use the human Head-Related Transfer Functions (HRTFs) for rendering soundscapes with proper directional information [1, 2, 3, 4]. Sound sources are filtered with the left and right ears' HRTFs respectively. Localization performance is usually decreased in a virtual audio environment due to headphone induced errors (in-the-head localization, front-back reversals etc.) and due to lack of head motion [5, 6, 7, 8, 9]. Furthermore, individually measured HRTFs generally increase localization and overall performance in a virtual audio environment [10-13]. In free-field environments even small head movements of about 1-2 degrees can lead to interaural differences and thus, resolving in-the-localization problems. This can be beneficial in a simulated virtual environment as well [14].

Measurements of individual HRTFs require two-channel recordings within or at the entrance of the blocked earcanals [13]. Subjects are usually seated on a chair in the anechoic chamber with or without head fixation. Multichannel loudspeaker arrays from different spatial directions deliver broadband excitation signals, such as white noise, MLS signals or impulses. In general, the overall signal-to-noise ratio, accuracy, repeatability and spatial resolution are low due to the relatively short measurement time [12]. In contrast to dummy-head measurement techniques, this procedure is quite uncomfortable for the subjects.

1.2. Overview of Literature Survey on Head and Body Motion Effects during Measurements

There is only a few measurement data about the extent of unwanted head and body movement of subjects instructed to be stationary. Lot of these data can be found in the literature of medical sciences.

In case of Functional connectivity MRI (fcMRI) head motion is a confounding factor. Children move more than adults, older adults more than younger adults, and patients more than controls. Head motion varies considerably among individuals within the same population [15]. Mean head displacement, maximum head displacement, the number of micro movements (> 0.1 mm), and head rotation were estimated in 1000 healthy, young adult subjects. Head motion had significant, systematic effects on several network measures and was associated with both decreased and increased metrics.

In functional magnetic resonance imaging (fMRI) head motion can corrupt the signal changes induced by brain activation. For reducing motion-induced effects a full three-dimensional rigid body estimation of head movement was obtained by image-based motion detection to a high level of accuracy [16]. A high level of consistency (rotation $< 0.05^\circ$) was demonstrated for detected motion parameters.

Another experiment included 40 subjects [17]. Volunteers were examined lying still and performing two separate head movements to assess detection and compensation of in-plane motion during MRI. Head rotation and translation was detected in all subjects. Values less than 1 degree were measured in lying position.

Methods can be developed to correct for motion artifacts in head images obtained by positron emission tomography (PET). The methods are based on six-dimensional motion data of the head that have to be acquired simultaneously during scanning [18, 19]. The data are supposed to represent the rotational and translational deviations of the head as a function of time, with respect to the initial head position. Motion data were acquired with a volunteer in supine position, immobilized by a thermoplastic head holder, to demonstrate the effects of the compensation methods. PET images can be justified and upgraded with post processing algorithms where serious head motion was present.

Similarly to MRI experiments, engineering approaches often include pattern analysis, video capture or accelerometers [20, 21, 22, 23]. A method for tracking of rigid head motion from video using a 3D ellipsoidal model of the head was proposed that is robust to large angular and translational motions of the head [20]. The method has been successfully applied to heads with a variety of shapes, hair styles, and also has the advantage of accurately capturing the 3D motion parameters of the head. This accuracy is shown through comparison with a rendered 3D animation of a model head. Due to its consideration of the entire 3D aspect of the head, the tracking is very stable over a large number of frames. This robustness extends even to sequences with very low frame rates and noisy camera images.

In our case, the focus is on acoustic measurements and virtual audio display technologies where human subjects are essential part of the procedure, most likely at the stage of individual HRTF acquisition.

In Blauert's early study subjects had to localize a 300 ms sinusoidal signal. The localization blur was not influenced by the fact whether the head was fixed or not. The experiments concluded that if the head should be kept stable without fixings, probability of head movements greater than 1° is less than 5%. It was suggested that for signals shorter than 1 s head fixation is not required [24]. Table 1 shows head movements of ten subjects without head fixation (probability and value).

Table 1. Extent and probability of unwanted head movements of ten subjects without head fixing after Blauert [24]. Mean value of the movements was only 0,22 degrees.

0 - 0,2°	0,2 – 0,4°	0,4 – 0,6°	0,6 – 0,8°
53%	34%	9%	4%

The impact of head tracking on localization is well known in the literature [1, 2, 10, 11]. A study of sound localization performance was conducted using headphone-delivered virtual speech stimuli, rendered via HRTF-based auralization, and

blocked ear-canal HRTF measurements [11]. The independent variables were chosen to evaluate commonly held assumptions in the literature regarding improved localization: inclusion of head tracking, individualized HRTFs, and early and diffuse reflections. Significant effects were found for azimuth and elevation error, reversal rates, and externalization.

One of the fundamental limitations on the fidelity of interactive virtual audio display systems is the delay that occurs between the time a listener changes his or her head position and the time the display changes its audio output to reflect the corresponding change in the relative location of the sound source. In an experiment, the impact of difference head-tracker latency values were examined on the localization of broadband sound sources in the horizontal plane [25]. Results suggested that head-tracker latency values of less than 70 ms are adequate to obtain acceptable levels of localization accuracy in virtual audio displays.

Although, there exist measurement data about the effect of head motion in virtual audio simulation and localization tasks, there is no data about the extent of unwanted head motions in case of a measurements where subjects have to be stable.

Our goal with this study was to determine the extent of head movements and instability during different environmental conditions (fixation methods) using different state-of-the-art tracking sensors. Results should support both selection of the appropriate tracking sensor for a given application as well as to determine accuracy range for measurement setups with human subjects mostly for audio engineering applications and spatial hearing research.

2. Measurement setups

The measurement setup included two different, state-of-the-art motion tracker devices:

- Intersense ICube3, a sensor offering a low-profile, rugged aluminum enclosure, sourceless 3-DOF tracking with full 360° range, accuracy of 1° yaw, 0.25° pitch and roll with 180 Hz update rate and 4 ms of latency [26],
- Kinect for Windows and Microsoft Kinect API [27]. The Kinect face tracking API does not define accuracy, but appears to be heavily filtered.

Software was developed that can simultaneously collect data from the Isense ICube3 as well as from the Kinect for Windows (including the Kinect face tracker library). The ICube logging has been modified to also grab the associated raw accelerometer data. This can be thought of as a second sensor and is essentially the same as one might see from the accelerometer in a phone or other device with accelerometer. The Kinect log has both position of the head/face and orientation of the face.



Figure 1. The Intersense and the Kinect sensors.

Participants wear the Icube3 on headphones (just for mounting, no audio) on their head and were also tracked by the Kinect. Informal testing indicated no negative effects of the headphones on the user's head as tracked by the Kinect. The following four conditions were defined:

- sitting unsupported (SiU),
- sitting supported (SiS),
- standing unsupported (StU), and
- standing supported (StS).

When sitting supported, the user has backrest and head against the wall. The user also was asked to place their hands on the armrests. When standing supported, the user leans upright against the wall. We assumed this is a good proxy for a more elaborate setup such as adjustable headrest mounted to a chair or "standing stool".

However, in a real HRTF measurement a wall is not suited due to reflections. Also, we were concerned that the Kinect might not work with the participant so close to the wall, but we found that the Kinect was still able to identify the person.

We found five minutes for each condition to be rather boring and quickly uncomfortable for participants. Therefore, we reduced the trial duration to three minutes. Nevertheless, HRTF measurements even with impulse excitation can last longer [31].

Participants underwent four measurement conditions, mentioned previously, in randomized order and one minute breaks between conditions to relax.



Figure 2. Left photo shows the Kinect mounted on a stand aimed at the chair used for seated positions. The right photo shows headphones with the IntertiaCube3 mounted on top.

3. Results

A total of 22 participants completed the experiment. One participant restarted the session due to a software error. Another participant's data was discarded because one of

the conditions failed to get an estimate of face position using the Kinect and wasn't noticed in time to fix the problem. Five additional participants' data was excluded due to substantial failure of the Intersense tracking in the form of large azimuth drift. Azimuth estimates of the ICube must be corrected to the magnetometer measurements of the Earth's magnetic field and is susceptible to interference. We believe there was some intermittent interference or perhaps some occasional initialization failure in the sensor.

The demographic evaluation included height, gender, age and information about any kind of health issues, balancing problems etc. Of the 16 participants not excluded, 10 were male and 6 female. Reported heights ranged from 1.52 to 1.90 meters, with an average of 1.72 meters and a median of 1.74 meters. No health/balance issues were reported.

Before any statistics were calculated, measures were linearly interpolated to match each sensor's target measurement rate. This was done because measurement recording was sometimes delayed slightly due to various reasons including OS interruption, garbage collection, etc. While measurements were very close to uniformly spaced, they were not quite perfect. In the case of the Kinect, an occasional brief loss of a tracking lock was a source of measurement dropout, which could be much longer than the cases above. Kinect face tracking had 4 dropouts over 2 seconds long over all conditions and participants (durations of 14.5, 8.3, 8.0, and 2.2 seconds). Kinect head tracking had one dropout over 2 seconds long overall all conditions and participants of 7.9 seconds. The IntertiaCube3 had no dropouts over 2 seconds. The IntertiaCube3 has a sample rate of 180 Hz and the Kinect has a sample rate of 30 Hz.

For analysis, a simple means of determining how still the participants were for each posture was necessary. For linear measures (position, and acceleration), variance of position was calculated. For angle measures, circular variance was calculated. Circular variance is commonly used for polar coordinates and it can be adapted to 3DOF. Circular variance is defined as 1 minus the magnitude of the mean of the direction vector of an angle. To adapt for 3DOF we calculate the average of the circular variances of the three angles. The closer to 1.0, the tighter the grouping of pose measurements is. This seemed to be the best way to compare the angle measurements. The mean is not directly relevant for linear or angle measures as it is likely very different for each participant due to variations in body shape, sensor mounting, etc. The variance shows the measurements around that mean that indicate movement. For accelerometer values as well as position, the magnitude of the standard deviation of the measurement vector was used rather than presenting the full vector. The acceleration measurement reflects the effects of gravity. However, this influence is removed automatically in calculating the standard deviation.

Results of magnitude of standard deviation of Intersense acceleration (m/s^2) were 0.08351 (SiU); 0.06912 (SiS); 0.10440 (StU); and 0.09051 (StS) also indicating less movement in case of sitting and supported situations, furthermore, SD values less than $0.1 m/s^2$ are relatively low.

Table 2. Result of the Intersense. Circular Angle Variance (ICAV, left); Kinect Circular Angle Variance (KCAV, right)

	ICAV	KCAV
StU	0,0013	0,0031
StS	0,0010	0,0030
SiU	0,0004	0,0019
SiS	0,0001	0,0033

Table 3. Result of magnitude of standard deviance of position in meters for Kinect Head (left); and magnitude of standard deviance of position in meters for Kinect Face (right). The Kinect logs position of the head/face and orientation of the face used for automated reference.

	M.Stdv. Head	M.Stdv. Face
StU	0,0837	0,0440
StS	0,0208	0,0131
SiU	0,0411	0,0286
SiS	0,0189	0,0177

4. Discussion

One challenge in this sort of study assessing the suitability of a sensor is that ideally one will want to compare the sensor's measurements to a ground truth. This would likely consist of an additional measurement technique superior to the sensor(s) being evaluated. This superior measurement would serve as the ground truth and allow measuring the error of the test sensors. In the case of this study we did not have that ability. The Intersense does have published sensor measurement specifications and we can see that our measurement data fits within those tolerances. We can however compare the Intersense and Kinect to see if they appear to record similar movement across conditions, as well as also considering the practical issues of using the each sensor (e.g. setup, likelihood of tracking loss, etc.).

Results can be evaluated in comparison of the two systems as well as to compare the four conditions. Within the four conditions, there appears to be a general trend that shows that the supported postures allow the participants to be more still than the unsupported postures of the same type of posture. Furthermore, sitting appears to offer more support than standing. Though this has not been formally statistically tested.

Additionally the Intersense tracker seems to be more accurate than the Kinect, whether using angle estimates or raw accelerometer measures.

The Kinect Face Tracking appears to be the least accurate and does not agree with the other measures about stillness of participants across the different postures and support. We believe that this is probably due to the Kinect Face Tracking briefly losing a lock on the participants face and then reacquiring it with slightly different coordinates. This may be happening more often than our analysis of tracking dropout discussed previously, but

with very short durations such that the event doesn't trigger a tracking lost event in the Kinect API. We would otherwise expect the face tracking to at least be more accurate than Kinect head tracking. Also, the Kinect was experienced to be very sensitive to setup whereas the Intersense could be placed on the participant's head without much issue.

The subjective impression about the usability of the two systems let us conclude that the Intersense is much better in capturing posture information. The Kinect has serious problems with a chin rest in view as well (though a chin rest was not used in this study). Due to the nature of the face tracking algorithm, sometimes the lock is lost. We did notice that with visual tracking and debugging tools the Kinect face tracking is perhaps more stable than the Intersense when it has a good tracking lock but this appears to largely be related to very heavy filtering that does not pick up small movements. In fact, it appears that a tracked individual can move their face a bit before the Kinect face tracking updates the pose estimate.

Further products to be mentioned for the same tracking purpose could be the SmartTrack from Advanced Realtime Tracking (ART) [28]. This tracker is suited for tracking within around 2 meters from the camera with 6DOF and sub-millimeter accuracy, but is substantially more expensive. Another straightforward solution could be the Intersense IS900 6DOF tracker [29]. The author also has experience with FaceAPI from Seeing Machines [30] and believes it to be very similar to the performance of the Kinect for Windows with the Microsoft Kinect FaceTracking API.

Hirahara et al. measured spectral deviations of individual HRTFs of three subjects during a 95-minute measurement session [31]. Using the Fastrak sensor they observed excessive head movements in the pitch and yaw directions (up to 10°) but only small movements in the roll (less than 1°). No head fixation was applied, however, subjects were asked to gaze at a fixed point marked on the wall.

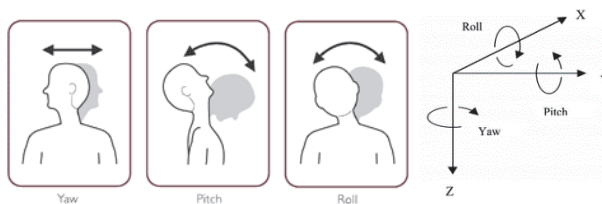


Figure 3. The yaw-pitch-roll coordinate system. Yaw corresponds to azimuth, pitch corresponds to elevation. Intersense and Kinect use a different coordinate system.

Yairi on the other hand reported large head movements in the roll and pitch directions, but small movements in the yaw direction [32]. Both studies reported large head movements after 5 minutes and suggested using head support during HRTF measurements. Table IV and Table V shows the results in signed degree values in the yaw-pitch-roll system (see Fig. 3.) for Intersense and Kinect Face.

Table 4. Result of the Intersense (mean, stdv) in signed degrees.

	mean			std.dev.		
	yaw	pitch	roll	yaw	pitch	roll
StU	2,00	2,02	-0,75	2,23	1,85	0,99
StS	-0,76	2,20	-0,83	1,65	1,59	0,94
SiU	1,32	0,96	-1,18	1,18	1,13	0,70
SiS	0,70	0,23	-0,02	0,80	0,43	0,25

Table 5. Result of Kinect Face Tracking (mean, stdv) in signed degrees.

	mean			std.dev.		
	yaw	pitch	roll	yaw	pitch	roll
StU	-1,15	-0,45	0,53	2,42	2,16	1,82
StS	-0,96	0,01	0,11	2,25	3,20	2,06
SiU	1,60	0,53	0,38	2,77	3,72	1,92
SiS	2,72	2,70	1,08	2,37	3,28	2,22

Our measurement also shows small errors in 3-minute sessions supporting the Japanese results however, standard deviation values are large. Mean measurement differences are defined from the first observed pose in the recorded session. A participant could have a mean measurement near zero and still have a large variance that indicates a lot of movement resulting in large STDV values.

If unsupported, greatest deviations in yaw directions were measured. If supported, differences in pitch can increase. Supported conditions reduce errors in the yaw directions, and partly in pitch and roll as well while seated. Based on the results of Intersense, supported sitting conditions can produce less than 1 degree error in all directions. This position is suggested for measurements however, actual HRTF measurements can be influenced by reflections coming from the legs while sitting. On the other hand, a standing position is uncomfortable and a reflecting wall behind the subject can also influence acoustic tests.

Tables IV and V also indicates large differences (accuracy) between sensors: the Kinect produces larger errors than the Intersense which is definitely the better sensor for this kind of measurements.

In HRTF measurements, 3-5 minute sessions are regarded to be very short. Even using impulse-like excitation, high spatial resolution requires longer recording times from 10-15 minutes up to 60-90 minutes. Our results indicate measurement inaccuracy already after some minutes and thus, breaking down longer measurement periods into shorter sessions is highly recommended if the head is not fixed and/or a rotating chair is used. These parameters introduce more problems to measurement accuracy and an exhaustive study incorporating all these parameters and long measurement sessions is suggested and put for further research.

5. Conclusion

16 subjects participated in a measurement to test different body positions (standing and sitting), different fixation methods (supported and unsupported) and to compare the Intersense and Kinect sensors for data capture. The goal was to determine the extent and variance of body movements if subjects are instructed to be still during individual measurements and recordings. Disagreement with Intersense versus Kinect on which postures allow the participant to be most still was measured. Intersense implies that the order of stillness is: SiS followed by StS or SiU and StU. Kinect head tracking has overlap between sitting and standing. Mean standard deviations of about 1.3-8.4 cm were measured for orientation and head position around the starting position corresponding to about 0-2.20 degrees in all directions. Furthermore, circular angle variance showed very little change between conditions and it was likely that the small unintended movements of the participants were beyond the capabilities of the Kinect to detect. Measurement sessions shorter than 3-5 minutes in a supported sitting situation can result in errors less than 1 degree and less than 2 degrees even in unsupported situations.

Acknowledgement

This research was realized in the frames of TÁMOP 4.2.4. A/2-11-1-2012-0001 „National Excellence Program – Elaborating and operating an inland student and researcher personal support system” The project was subsidized by the European Union and co-financed by the European Social Fund.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 643636 “Sound of Vision”.

References

- [1] Blauert J: Spatial Hearing: The Psychophysics of Human Sound Localization. The MIT Press, MA, 1983.
- [2] Begault DR: 3-D Sound for Virtual Reality and Multimedia, Academic Press, London, UK, 1994.
- [3] Cheng CI, Wakefield GH: Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space. *J. Audio Eng. Soc.*, Vol. 49, pp. 231-249, 2001.
- [4] Møller H, Sorensen MF, Hammershøi D, Jensen CB: Head-Related Transfer Functions of human subjects. *J. Audio Eng. Soc.*, Vol. 43, pp. 300-321, May 1995.
- [5] Hill PA, Nelson PA, Kirkeby O: Resolution of front-back confusion in virtual acoustic imaging systems. *J. Acoust. Soc. Am.*, Vol. 108, pp. 2901-2910, 2000.
- [6] Wightman FL, Kistler DJ: Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.*, Vol. 105, pp. 2841–2853, 1999.
- [7] Perrett S, Noble W: The effect of head rotations on vertical plane localization. *J. Acoust. Soc. Am.*, Vol. 102, pp. 2325–2332, 1997.
- [8] Wenzel EM: Localization in virtual acoustic displays. *Presence*, Vol. 1, pp. 80–107, 1991.

- [9] Sandvad J: Dynamic aspects of auditory virtual environments. 100th Convention of the Audio Engineering Society, Copenhagen, Denmark, Preprint 4226, 1996.
- [10] Wightman F, Kistler D: Measurement and validation of human HRTFs for use in hearing research. *Acta acustica united with Acustica*, Vol. 91, pp. 429-439, 2005.
- [11] Begault DR, Wenzel E, Anderson M: Direct Comparison of the Impact of Head Tracking Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. *J. Audio Eng. Soc.*, Vol. 49, pp. 904-917, 2001 Oct.
- [12] Wenzel EM, Arruda M, Kistler DJ, Wightman FL: Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, Vol. 94, pp. 111-123, 1993.
- [13] Møller H, Sorensen MF, Jensen CB, Hammershøi D: Binaural Technique: Do We Need Individual Recordings? *J. Audio Eng. Soc.*, Vol. 44, pp. 451-469, June 1996.
- [14] Wersényi Gy: Effect of Emulated Head-Tracking for Reducing Localization Errors in Virtual Audio Simulation. *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 17, pp. 247-252, Feb 2009.
DOI: [10.1109/TASL.2008.2006720](https://doi.org/10.1109/TASL.2008.2006720)
- [15] Van Dijk KR, Sabuncu MR, Buckner RL: The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*, Vol. 59, No. 1, pp. 431-438, January 2012.
DOI: [10.1016/j.neuroimage.2011.07.044](https://doi.org/10.1016/j.neuroimage.2011.07.044)
- [16] Thesen S, Heid O, Mueller E, Schad LR: Prospective acquisition correction for head motion with image-based tracking for real-time fMRI. *Magnetic Resonance in Medicine*, vol. 44, no. 3. pp. 457-465, September 2000.
DOI: [10.1002/1522-2594\(200009\)44:3<457::AID-MRM17>3.0.CO;2-R](https://doi.org/10.1002/1522-2594(200009)44:3<457::AID-MRM17>3.0.CO;2-R)
- [17] Forbes KPN, Pipe JG, Bird C, Heiserman JE: PROPELLER MRI: Clinical testing of a novel technique for quantification and compensation of head motion. *Journal of Magnetic Resonance Imaging*, Vol. 14, No.3. pp. 215-222, September 2001.
DOI: [10.1002/jmri.1176](https://doi.org/10.1002/jmri.1176)
- [18] Menke M, Atkins MS, Buckley KR: Compensation methods for head motion detected during PET imaging. *IEEE Transactions on Nuclear Science*, Vol. 43, No. 1, pp. 310 - 317, Feb. 1996.
DOI: [10.1109/23.485971](https://doi.org/10.1109/23.485971)
- [19] Green MV, Seidel J, Stein SD, Tedder TE, Kempner KM, Kertzman C, Zeffiro TA: Head movement in normal subjects during simulated PET brain imaging with and without head restraint. *J. Nuc. Med.*, Vol. 35, pp. 1538 -1546, 1994.
- [20] Basu S, Essa I, Pentland A: Motion regularization for model-based head tracking. *Pattern Recognition*, Vol. 3, pp. 611-616, Vienna, Austria, 25-29 Aug. 1996.
DOI: [10.1109/ICPR.1996.547019](https://doi.org/10.1109/ICPR.1996.547019)
- [21] Pentland A, Horowitz B: Recovery of nonrigid motion and structure. *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 13, No. 7, pp. 730-742, 1991.
DOI: [10.1109/34.85661](https://doi.org/10.1109/34.85661)

- [22] Azuma R, Bishop G: A frequency-domain analysis of head-motion prediction. Proc. of SIGGRAPH '95 - 22nd Conf. on Computer Graphics and Interactive Techniques, pp. 401-408, 1995.
DOI: [10.1145/218380.218496](https://doi.org/10.1145/218380.218496)
- [23] So RHY, Griffin MJ: Compensating Lags in Head-Coupled Displays Using Head Position Prediction and Image Deflection. Journal of Aircraft, Vol. 29, No. 6, pp. 1064-1068, November-December 1992.
DOI: [10.2514/3.46285](https://doi.org/10.2514/3.46285)
- [24] Blauert J: Untersuchungen zum Richtungshören in der Medianebene bei fixiertem Kopf, PhD dissertation, Techn. Hochschule Aachen, Germany, 1969.
- [25] Brungart DS, Simpson BD, McKinley RL, Kordik AJ, Dallman RC, Ovenshire DA: The interaction between head-tracker latency, source duration, and response time in the localization of virtual sound sources. Proc. of the 10th Int. Conf. on Auditory Display, Sydney, Australia, pp 1-7, July 6-9, 2004.
- [26] <http://www.intersense.com/pages/18/11/>
- [27] <http://www.microsoft.com/en-us/kinectforwindows/>
- [28] <http://www.ar-tracking.com/products/tracking-systems/smarttrack/>
- [29] <http://www.intersense.com/pages/20/14>
- [30] <http://www.seeingmachines.com/product/faceapi/>
- [31] Hirahara T, Sagara H, Toshima I, Otani M: Head movement during head-related transfer function measurements. Acoust. Sci. & Tech., Vol. 31, No. 2, pp. 165-171, 2010.
DOI: [10.1250/ast.31.165](https://doi.org/10.1250/ast.31.165)
- [32] Yairi S: A study on system latency of virtual auditory display responsive to head movement. PhD dissertation, Tohoku University, (in Japanese) 2006.