



Published in final edited form as:

*J Vestib Res.* 2016 November 03; 26(4): 375–385. doi:10.3233/VES-160587.

## Impact of artifacts on VOR gain measures by video-oculography in the acute vestibular syndrome

Georgios Mantokoudis<sup>a,b</sup>, Ali S. Saber Tehrani<sup>b</sup>, Amy Wozniak<sup>c</sup>, Karin Eibenberger<sup>d</sup>, Jorge C. Kattah<sup>e</sup>, Cynthia I. Guede<sup>e</sup>, David S. Zee<sup>b</sup>, and David E. Newman-Toker<sup>b,\*</sup>

<sup>a</sup>University Department of Otorhinolaryngology, Head and Neck Surgery, Inselspital Bern, Switzerland <sup>b</sup>Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA <sup>c</sup>Johns Hopkins Biostatistics Center, Baltimore, MD, USA <sup>d</sup>Department of Otolaryngology – Head and Neck Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, USA <sup>e</sup>Department of Neurology, Illinois Neurologic Institute, University of Illinois, College of Medicine, Peoria, IL, USA

### Abstract

**OBJECTIVE:** The video head impulse test (HIT) measures vestibular function (vestibulo-ocular reflex [VOR] gain – ratio of eye to head movement), and, in principle, could be used to make a distinction between central and peripheral causes of vertigo. However, VOG recordings contain artifacts, so using unfiltered device data might bias the final diagnosis, limiting application in frontline healthcare settings such as the emergency department (ED). We sought to assess whether unfiltered data (containing artifacts) from a video-oculography (VOG) device have an impact on VOR gain measures in acute vestibular syndrome (AVS).

**METHODS:** This cross-sectional study compared VOG HIT results ‘unfiltered’ (standard device output) versus ‘filtered’ (artifacts manually removed) and relative to a gold standard final diagnosis (neuroimaging plus clinical follow-up) in 23 ED patients with acute dizziness, nystagmus, gait disturbance and head motion intolerance.

---

\*Corresponding author: David E. Newman-Toker, MD PhD, Associate Professor, Department of Neurology, Johns Hopkins University School of Medicine, Johns Hopkins Hospital, CRB-II, Room 2M-03 North, 1550 Orleans Street, Baltimore, MD 21231, USA. Tel.: +1 443 287 9593; Fax: +1 410 502 7869; toker@jhu.edu.

#### Author Contributions

Georgios Mantokoudis: assisted in data collection and data analysis, drafted the manuscript, reviewed and critically edited the manuscript, and approved the final version.

Ali S. Saber-Tehrani: assisted in data collection and data analysis, reviewed and critically edited the manuscript, and approved the final version.

Amy Wozniak: conducted the statistical analysis, reviewed and critically edited the manuscript, and approved the final version.

Karin Eibenberger: Assisted in data processing, reviewed and critically edited the manuscript, and approved the final version.

Jorge C. Kattah: Assisted in data collection, reviewed and critically edited the manuscript, and approved the final version.

Cynthia I. Guede: Assisted in data collection, reviewed and critically edited the manuscript, and approved the final version.

David S. Zee: assisted in study design/conception, reviewed and critically edited the manuscript, and approved the final version.

David E. Newman-Toker: designed the study and analytic plan, reviewed and critically edited the manuscript, and approved the final version.

Potential conflicts of interest

None.

Statistical analysis

The statistical analysis was conducted by Amy Wozniak (MS) Johns Hopkins Biostatistics Center, Baltimore, Maryland, USA.

**RESULTS:** Mean VOR gain assessment alone (unfiltered device data) discriminated posterior inferior cerebellar artery (PICA) strokes from vestibular neuritis with 91% accuracy in AVS. Optimal stroke discrimination cut points were bilateral VOR gain  $>0.7099$  (unfiltered data) versus  $>0.7041$  (filtered data). For PICA stroke sensitivity and specificity, there was no clinically-relevant difference between unfiltered and filtered data—sensitivity for PICA stroke was 100% for both data sets and specificity was almost identical (87.5% unfiltered versus 91.7% filtered). More impulses increased gain precision.

**CONCLUSIONS:** The bedside HIT remains the single best method for discriminating between vestibular neuritis and PICA stroke in patients presenting AVS. Quantitative VOG HIT testing in the ED is associated with frequent artifacts that reduce precision but not accuracy. At least 10–20 properly-performed HIT trials per tested ear are recommended for a precise VOR gain estimate.

### Keywords

Eye movement measurements; vestibulo-ocular reflex; vertigo; vestibular neuritis; stroke; diagnosis

## 1. Introduction

Acute vestibular syndrome (AVS) is a well described syndrome consisting of continuous dizziness or vertigo, vegetative symptoms (nausea/vomiting), head motion intolerance, and imbalance with nystagmus lasting more than 24 hours [23]. An estimated 20–25% of patients with AVS may have a dangerous underlying central cause such as a stroke [23]. Careful bedside examination of three vestibular eye movements ('HINTS': Head Impulse, Nystagmus, Test of Skew) differentiates peripheral from central causes [3, 13, 17, 18] and even outperforms MRI neuroimaging in the acute phase [18, 21, 23].

The 'HINTS' clinical eye movement sign with the greatest combined sensitivity and specificity for stroke is the horizontal head impulse test (HIT) [23]. The horizontal HIT consists of a low-amplitude (10–20°), high-velocity (100–200°/s) passive horizontal head movement from lateral to center while the patient is stabilizing gaze using the VOR by fixing on a visual target. The relevance of this finding in posterior fossa stroke has been confirmed in quantitative studies using the laboratory gold standard for eye movement recording, magnetic scleral search coils [4]. Relative to quantitative testing, the sensitivity of clinical HIT for identifying vestibular hypofunction at the bedside ranges widely between 35–71% depending on the test technique and the extent of vestibular loss [2,9,10,22]. More importantly, examiner skill plays an important role in detecting an abnormal result [12], raising questions about whether inexperienced examiners should use the clinical HIT to make high-stakes triage decisions about stroke in acute dizziness in the emergency department (ED), as has been suggested [6].

Eye- and head movements can now be recorded at the bedside using light-weight portable video goggles with an integrated high-speed infrared camera [25]. Such video-oculography (VOG) devices are used to measure the vestibulo-ocular reflex (VOR) gain, which is a quantitative measure of vestibular function that corresponds to the clinical HIT. These VOG goggles, also referred as the video head impulse test (vHIT), assist physicians to correctly

perform a standardized HIT and also facilitate test interpretation. The goggles give an estimate of VOR gain in each semicircular canal plane tested, and have been validated for their measurement accuracy in laboratory settings against gold standard eye movement testing using magnetic scleral search coils [8, 14]. Their portability and ease of use promotes the application of such devices in daily practice, and such devices are now commercially available.

We recently demonstrated that portable VOG can be used in the ED in real time to help differentiate stroke from vestibular neuritis in patients with AVS [19]. In further studies, however, we also showed that artifacts in ED-based VOG testing are frequent [15]. Removing these artifacts manually through post-process filtering of the output of the device allows for accurate disease classification [16]. However, the impact, if any, of device artifacts on unfiltered gain measures and final diagnostic accuracy in AVS patients is unknown. In this analysis, we sought to characterize the impact of artifacts on portable VOG diagnosis in acute dizziness and vertigo.

## 2. Material and methods

### 2.1. Study design

Patients were recruited as part of a prospective observational study of AVS patients between August 2011 and December 2012 at two academic medical centers. We here report a cross-sectional comparison of VOG HIT results ‘unfiltered’ (standard device output) versus ‘filtered’ (artifacts manually removed) and relative to a gold standard final diagnosis (neuroimaging plus clinical follow-up). Research subjects gave written informed consent, and the study was approved by the institutional review board at both institutions.

We sought to assess the impact of artifacts on unfiltered VOR gain values. In particular, we examined whether artifacts biased VOR gain results (reducing accuracy) or if such measurement errors merely added random ‘noise’ without affecting the final mean VOR gain result (reducing precision). We also sought to calculate the minimum number of HIT trials for a stable mean VOR gain measure. This manuscript complements our previously published studies of artifacts and filtered gain values in these same subjects [15, 16].

### 2.2. Test subjects

From a series of 30 AVS patients presenting to the ED, we excluded 7 (3 for lack of confirmatory neuroimaging, 1 in whom calibration could not be properly performed, and 3 who had strokes in the territory of the anterior inferior cerebellar artery (AICA)). We report on VOR gain values in the remaining 23. From this pool of 23 patients, data from 21 patients were eligible for filtered ROC analysis (two patients had no remaining HIT traces after filtering) and 18 patients were eligible for within-subject paired comparison of filtered and unfiltered data (three additional patients had no traces with artifacts). We excluded AICA-territory strokes for this analysis, as in our prior manuscript [16], because AICA strokes often include labyrinthine involvement (producing a variable clinical phenotype that often has mixed central and peripheral features), making it difficult to identify an optimal ‘cut point’ for VOR gain that discriminates stroke from vestibular neuritis. AICA strokes are

important and discriminating them from neuritis in the ED currently requires additional bedside tests [11, 18] or analytic techniques not yet validated for vHIT [3]. Including such cases here would reduce our ability to assess the impact of using unfiltered data (as opposed to filtered data) on diagnostic discrimination, without adding new information to address the question of how best to diagnose those with AICA strokes closely mimicking vestibular neuritis.

Patients were classified into two diagnostic categories (central vs. peripheral) based on clinical features, neuro-otological investigations, and a diagnostic MRI with diffusion weighted images (MRI-DWI) obtained within the first 10 days after the onset of symptoms. Patients were followed for at least 90 days in order to reduce the risk of a potentially missed ‘pseudo-neuritis’ [5] at the initial visit. One patient with labyrinthitis (peripheral AVS with hearing loss) was classified with vestibular neuritis, and one patient with a small left middle cerebellar peduncle hemorrhage was classified with the PICA strokes for analysis. Additional recruitment and diagnostic methods have been described previously [15, 16].

### 2.3. VOG devices

We used a portable VOG device designed for quantitative HIT testing (ICS Impulse, GN Otometrics, Taastrup, Denmark; <http://www.icsimpulse.com/>). The device consists of a light-weight goggles frame which is firmly affixed to the head using a rubber strap. An integrated digital high-speed camera (250 frames/sec) sensitive to infrared light is connected through a firewire (IEEE 1394a) and an additional USB2 port to a laptop computer. The camera is mounted outside the patient’s line of sight to enable visual fixation and records monocular eye video via reflection using a transparent ‘hot’ mirror mounted on the frame. The ICS device has been validated to measure accurate HIT gains in the laboratory setting using simultaneous VOG and scleral search coils [14].

The device software (OTOSuite Vestibular Software Version 1–20 Build 310) uses a pupil-tracking algorithm to monitor eye position and combines this with data from an inertial sensor integrated in the goggles frame to measure head acceleration. Inadequate HIT trials are discarded automatically by a proprietary filtering algorithm. The software records and displays head and eye velocity traces from accepted HIT trials. It calculates VOR gain using area-under-the-curve methods [7]. It plots the VOR gain ratio for each HIT trial, and then calculates a side-specific mean across all trials for that patient.

### 2.4. HIT and analysis of VOR gains

Head impulses were collected at the bedside by two trained research fellows and one trained nurse using HIT physical examination methods described previously [15, 16]. Raw HIT trials extracted from the device were systematically classified as interpretable (no disruptive artifacts or fast phase eye movements) or uninterpretable traces using predefined quality standards by two masked, independent raters [15]. We here define ‘unfiltered’ device data as mixed interpretable and uninterpretable traces; ‘filtered’ device data as interpretable traces only; and ‘artifacts-only’ device data as uninterpretable traces only. Figure 1 shows an example of an unfiltered device output versus filtered (clean) traces.

We assessed VOR gain measures. We did not perform detailed re-fixation saccade analysis as done in a previous scleral search-coil-based study [4] because current VOG software does not provide for automated detailed saccade analysis (including removal of blink-related artifacts). This may be feasible in the future with new software modifications, given that small studies have found that VOG yields comparable results to search coils [24].

## 2.5. Statistics

Aggregate mean gain values of filtered HIT traces were compared to artifacts-only data using non-parametric, paired tests (Wilcoxon signed rank test). A receiver operating characteristics (ROC) analysis was conducted on unfiltered data and compared to our prior ROC analysis for filtered data [16]. To calculate VOG sensitivity and specificity for stroke diagnosis, we used the single-ear mean VOR gain from the lower-measured side (i.e., right or left), applying the optimal cut point derived from the ROC analysis, as in our prior analysis [16]. Single ear mean gains for 'filtered' results were not calculated if they contained fewer than five interpretable HIT results. To estimate the number of HITs required for a reliable result, we used simulation for variable numbers of HIT trials to calculate the within-subject variance of mean gain scores. 1000 samples containing 5, 10, 15, 20, 25, and 50 observations per person were generated from measurements on the ipsilesional side using unfiltered data and compared to filtered data. For each sample, the mean gain score per person and the area under the curve (AUC) were calculated. Summaries for the 1000 samples across the variable number of simulated observations are reported.

Analyses were performed using SAS version 9.2 (SAS Institute Inc, Cary, NC, USA), and a two-sided value of  $p$ -value $<0.05$  was considered statistically significant. For inter-rater agreement, Cohen's  $\kappa$  was calculated using SPSS software version 17 (SPSS Inc., Chicago, IL, USA).

## 3. Results

In the 23 AVS patients reported here, mean age was 60.2 (SD $\pm$  12.1, range 31–83), and 74% were men. Table 1 provides an overview of the clinical characteristics of the subjects. Overall, 1223 HITs were performed. All participants tolerated and completed the full exam protocol. Sixteen AVS patients were diagnosed with vestibular neuritis (or labyrinthitis) and seven patients had a posterior circulation stroke in the PICA territory.

Quality assessment of HIT traces (inter-rater agreement  $\kappa$  0.68) revealed 705 interpretable and 518 uninterpretable HITs. Table 2 shows mean VORgains of unfiltered (mixed filtered and artifacts) versus filtered and artifacts-only data. There were no statistical differences in mean VOR gain between filtered results and artifacts-only results using the Wilcoxon signed rank test ( $p = 0.24$ ). Figure 2 compares diseases showing histograms of VOR gains from all collected HITs for unfiltered (A) and filtered (B) data. Although unfiltered data showed a broader VOR gain distribution than filtered data, a clear segregation between stroke and peripheral was still possible, and the mean disease-specific VOR gains did not change (dotted lines, Fig. 2).

Figure 3 depicts individual mean VOR gains for each ear. There was no significant difference in unfiltered vs. filtered mean VOR gains with respect to discrimination between stroke and vestibular neuritis. Patients with vestibular neuritis had consistently low (<0.7) unilateral VOR gain values whereas PICA stroke had consistently bilateral normal (>0.7) VOR gain values. Figure 4 shows a scatterplot visualizing gain asymmetries. Patients with vestibular strokes had bilaterally normal VOR gains with data points clustering in the upper-right corner. Patients with vestibular neuritis had unilaterally low gains with data points clustering at the right inferior quadrant. Both filtered and unfiltered data are superimposed and show the same pattern.

The ROC curve presented for unfiltered ( $n = 23$ ) and filtered data ( $n = 21$ , two patients had fewer than five interpretable HITs after filtering) appear superimposed (Fig. 5). The area under the curve was 0.9821 for unfiltered data versus 0.9861 for filtered data. Optimal stroke discrimination cut points were bilateral VOR gain >0.7099 (unfiltered) versus >0.7041 (filtered). For sensitivity and specificity, there was no clinically-relevant difference between unfiltered and filtered data—the sensitivity for PICA stroke was 100% for both data sets and the specificity was almost identical (87.5% unfiltered versus 91.7% filtered). Based solely on a binary classifier using the lower-ear mean VOR gain (<0.7099 is neuritis, >0.7099 is stroke), the total diagnostic accuracy of unfiltered VOG HIT gains to differentiate PICA stroke from vestibular neuritis was 91% ( $n = 21$  of 23 correctly classified, 95% CI, 72–99%, Table 3).

The simulated within-subject variance of mean VOR gains was higher for unfiltered than filtered data, and decreased at a marginal rate as the simulated number of HIT trial observations per person increased (Fig. 6). For both unfiltered and filtered data, the mean VOR gain variance was lower after a total of 10 projected HITs and nearly reached an asymptote after 20 projected HITs. Thus, 10–20 HITs appear sufficient to obtain a stable VOR gain result from the device without any post-processing of data or manual filtering (although the variance on filtered data remained lower, regardless of the number of simulated observations). The variance in VOR gain units was 0.077 (unfiltered) versus 0.056 (filtered) for a sample of 20 HITs.

#### 4. Discussion

Our study shows that, although VOG artifacts are frequent, they do not meaningfully alter the diagnostic accuracy of VOR gain measurement in pursuit of discriminating central from peripheral causes of AVS. Artifacts do, however, affect the precision of results, indicating that testing should involve at least 10–20 impulses per tested canal. These results are important because they lend credibility to the concept that use of such devices in ‘real-world’ frontline care settings like the ED is practical, efficient, and can yield valid measures and accurate diagnoses [19] without specialized post-processing. Further, they provide a scientifically-rational basis for determining the minimum number of impulse test trials when testing is performed in less well controlled (non-lab), clinical settings such as the ED.

These results inform our understanding of VOG accuracy. There was neither a significant gain difference between filtered and unfiltered data, nor a significant difference in diagnostic

accuracy. In this analysis, mean VOR gain assessment alone (using unfiltered device data) discriminated PICA strokes from vestibular neuritis with 91% accuracy in AVS. This accuracy remains unchanged compared to filtered VOG exams after manually removing traces with artifacts [16]. Therefore, even when used in clinical care settings such as the ED, sensitivity of this device-based approach will likely exceed that of current diagnostic imaging (MRI DWI, sensitivity ~85% for all strokes and ~50% for strokes <1 cm in size) when performed in the first 48 hours after onset of symptoms [18, 21]. These results support the potential of such a device for rapid triage and early diagnosis of vestibular strokes in the ED, using an ‘eye ECG’ paradigm, in which the physiologic results guide downstream diagnostic decisions [19], even without secondary filtering of results by expert review.

These results inform not only accuracy but precision of VOG results. Our prior analyses suggested that individual HIT traces containing artifacts might need to be manually excluded from measures by direct inspection before relying on results for diagnostic classification [15]. Detailed, expert manual review of individual traces still likely offers the most accurate gain estimates. As such, expert vHIT operators should not rely solely on unfiltered VOR gain data to make important diagnostic decisions. Nevertheless, the present study results show that artifacts produce random noise in measurement, but do not significantly bias gain up or down. If artifacts are left in place, more trials are required to decrease the random noise; conversely, if artifacts are removed, fewer trials are required. We found robust mean VOR gain values without any filtering or manual interpretation as long as the number of correctly performed HITs is at least 10–20 per semicircular canal being tested. This number of impulses is practical and tolerable by even acutely vertiginous patients, because of the very low amplitude (10–20°) of the HIT maneuver. Most of our AVS patients had more difficulty shifting from a supine to sitting position than tolerating dozens of head impulses.

The nature of artifacts is of clinical relevance to those who might seek to use such devices, as they are common, occurring in almost half of traces [15]. Most of these artifacts were attributed to goggle slippage during head rotation, incorrect operation of the goggles (touching the goggle strap), eye blinks, pseudo-saccades (mini-blinks), or nystagmus [15]. Clinical technique is relevant, as we found that artifacts were slightly more prevalent using the ‘in front of the’ rather than ‘in back of’ the patient technique [15]. Overall, we favor the ‘in back of’ technique, which is recommended by the device manufacturer and generally preferred by vestibular testing laboratories. Although the ‘in front of’ technique replicates the clinical (non-quantitative) test, enabling device-based feedback to increase the examiner’s skill in clinically interpreting the HIT result, it is harder to avoid inadvertently touching the goggles or moving facial skin, which likely explains the small increase in artifacts when using this method.

Our results should be placed in context of an important recent study using dual magnetic scleral search coils to measure eye position [4]. Chen et al. examined patients with acute stroke and vestibular neuritis, demonstrating that both VOR gain and cumulative re-fixation saccade amplitude were strong predictors of stroke vs. neuritis, and that perhaps total saccade amplitude was even slightly better as a predictor than mean gain [4]. The search coil technique, which represents the gold standard in eye movement measurement, is less prone to artifacts and renders more precise VOR gain and re-fixation saccade measures. It is

unlikely, however, that the search coils approach will supersede VOG testing for quantitative measures in clinical practice for several reasons: (1) only a handful of vestibular research labs use the search coil technique; (2) on-site, portable bedside exam is not realistic for search coils; (3) search coils are invasive and somewhat uncomfortable for patients; (4) costs exceed those of vHIT; (5) specially trained personnel are needed to run the tests; and (6) testing time is longer than with vHIT. Nevertheless, search coils studies provide important reference data for future quantitative vHIT studies, and we hope that saccade analyses will soon be incorporated into VOG analytic software and subsequently validated against coil measures for their accuracy.

## 5. Limitations

We studied 1223 HIT traces, but our patient sample was small (especially for strokes), potentially limiting the generalizability of these findings to the larger population of AVS patients. We deliberately excluded patients with AICA-territory stroke (see Methods for rationale), so results related to diagnostic accuracy should not be extrapolated directly to undifferentiated patients presenting AVS. Accurate assessment of AICA stroke likely requires clinical evaluation of the full HINTS plus hearing battery [18] or quantitative gain and saccade analyses demonstrated previously using search coils [4]. Nevertheless, the current data strongly corroborate our prior work using clinical HIT and HINTS testing [13,17,18], and suggest that artifacts do not significantly impact overall diagnostic classification. Future studies should include a larger number of stroke patients, including those with AICA strokes, for the assessment of overall VOG-based diagnostic accuracy.

It is known that in acute vestibular neuritis ipsilesional gain, as measured by search coils, are rarely above 0.5 [1, 20], so the finding of gain between 0.5–0.7 could signal technical factors such as goggle (or facial skin) slippage, which can artificially increase gain. We applied a strong standard for diagnosing vestibular neuritis (clinical diagnosis by vestibular experts, MRI-DWI, and follow-up for stroke events), but there is always the theoretical risk of diagnostic misclassification. We used expert examiners trained in vestibular testing, and, although the device automatically discards improperly performed HITs, it remains unknown whether inexperienced examiners using VOG devices would yield the same results. We studied the impact of artifacts on VOR gain, but not refixation saccades (see Methods for rationale), so our results are silent on the impact of artifacts on saccade analyses. We did not perform a cumulative saccade amplitude analysis, as suggested by Chen et al. [3] because the manufacturer does not provide the necessary software tools. We tested only the horizontal semicircular canals in our patients, and it is unclear to what extent these results generalize to vertical canal testing. Finally, the device we used did not yet measure all three HINTS eye movements (recording skew and gaze evoked nystagmus in addition to HIT). Future studies using newer device software, now available, will need to determine whether a full battery of quantitative HINTS eye findings improves VOG accuracy.

## 6. Potential implications

This study suggests the use of VOG devices as a screening and triage tool for AVS patients could be a viable care option. VOG devices might assist neurologists or ED clinicians in



performing and assessing HITs correctly, avoiding the ‘covert saccade’ trap [25] in which corrective saccades are imbedded in the deficient VOR slow-phase response and, hence, not easily seen clinically. VOG devices simultaneously serve as an educational tool for psychomotor skills, providing immediate feedback on head impulse maneuver accuracy by rejecting improper impulses. From a research perspective, our results suggest next steps of assessing device use by non-expert examiners, extension of VOG to interpret the full HINTS test battery, incorporation of VOG test results into a diagnostic algorithm, and measuring cost-effectiveness. A multicenter, phase II, randomized clinical trial is currently underway to take these next steps (AVERT – Acute Video-oculography for Vertigo in Emergency Rooms for Rapid Triage, NIH/NIDCD U01DC013778; Clinical-Trials.gov #NCT02483429).

## 7. Conclusions

The bedside HIT remains the single best method for discriminating between stroke and vestibular neuritis in patients presenting AVS. Quantitative VOG HIT testing in the ED is associated with a high rate of artifacts which reduce precision but not diagnostic accuracy. Based on these results, for non-laboratory settings we recommend at least 10–20 properly-performed HIT trials per tested ear. Research is now underway that will address many practical questions about the use of VOG in a physiology-based ‘eye ECG’ approach to stroke diagnosis in acute dizziness and vertigo patients presenting to frontline care settings.

## Acknowledgments

This study was supported by the Swiss National Science Foundation PBBEP2 136573 (GM). Dr. David E. Newman-Toker’s effort was partially supported by a grant from the National Institutes of Health, National Institute of Deafness and Communication Disorders (U01DC013778). GN Otometrics and Interacoustics loaned VOG equipment for research.

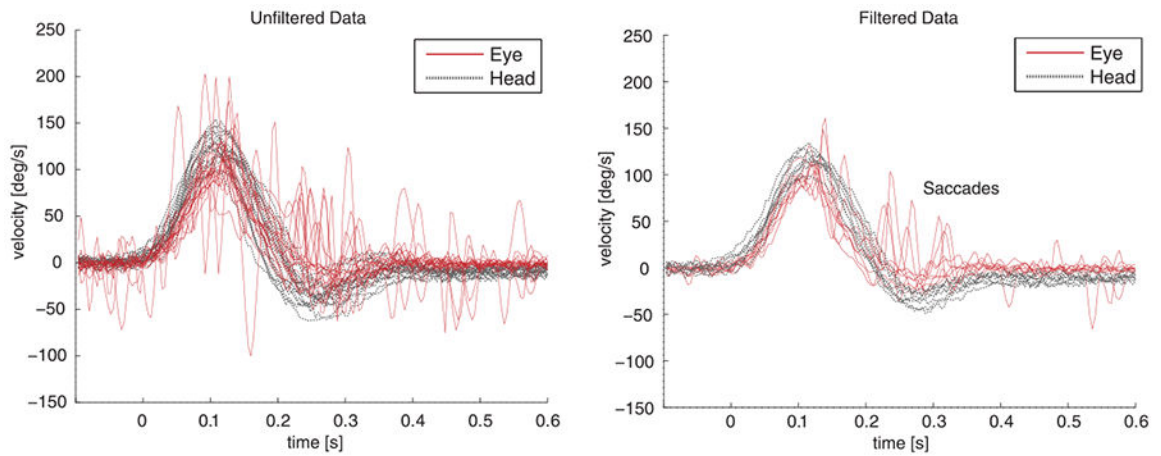
### Disclosures

Drs. Newman-Toker and Kattah have been loaned VOG goggles for research studies by GN Otometrics, Inc. (manufacturer). No authors have any financial interest in the company. Dr. Georgios Mantokoudis was supported by the Swiss National Science Foundation PBBEP2 136573. Dr. David E. Newman-Toker’s effort was partially supported by a grant from the National Institutes of Health, National Institute of Deafness and Communication Disorders (1U01DC013778–01A1).

## References

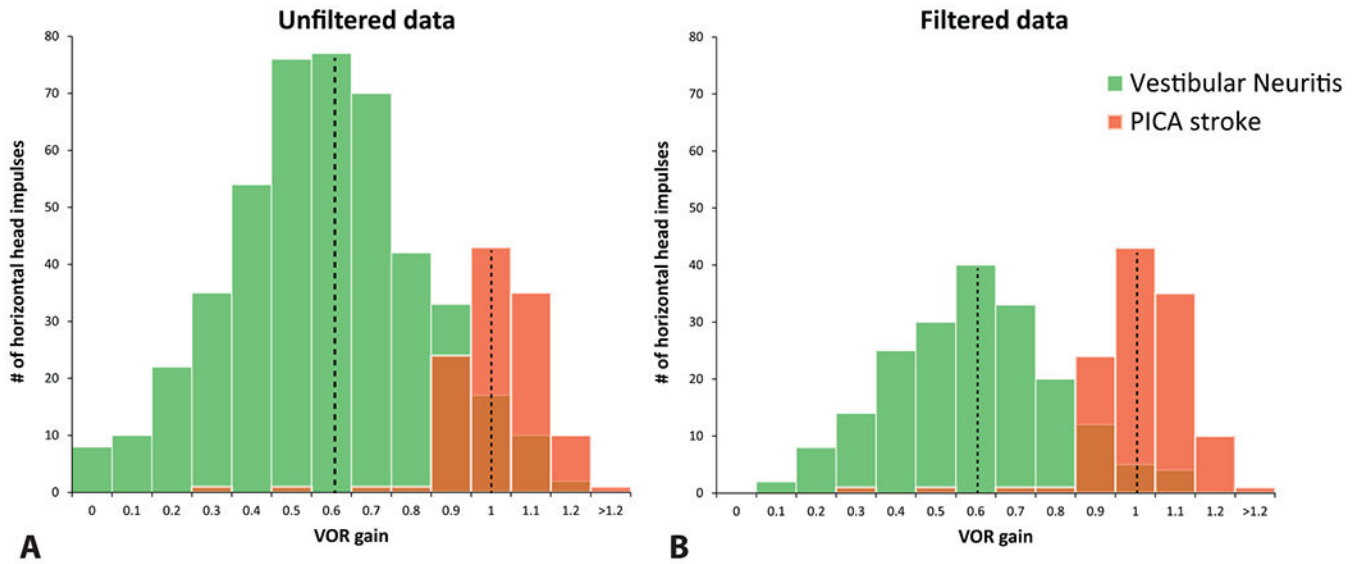
- [1]Aw ST, Fetter M, Cremer PD, Karlberg M, Halmagyi GM. Individual semicircular canal function in superior and inferior vestibular neuritis. *Neurology*. 2001; 57:768–774. [PubMed: 11552001]
- [2]Beynon GJ, Jani P, Baguley DM. A clinical evaluation of head impulse testing. *Clin Otolaryngol Allied Sci*. 1998; 23:117–122. [PubMed: 9597280]
- [3]Chen L, Lee W, Chambers BR, Dewey HM. Diagnostic accuracy of acute vestibular syndrome at the bedside in a stroke unit. *J Neurol*. 2011; 258:855–861. [PubMed: 21153732]
- [4]Chen L, Todd M, Halmagyi GM, Aw S. Head impulse gain and saccade analysis in pontine-cerebellar stroke and vestibular neuritis. *Neurology*. 2014; 83:1513–1522. [PubMed: 25253747]
- [5]Cnyrim CD, Newman-Toker D, Karch C, Brandt T, Strupp M. Bedside differentiation of vestibular neuritis from central vestibular pseudoneuritis. *J Neurol Neurosurg Psychiatry*. 2008; 79:458–460. [PubMed: 18344397]
- [6]Cohn B. Can bedside oculomotor (HINTS) testing differentiate central from peripheral causes of vertigo? *Ann Emerg Med*. 2014; 64:265–268. [PubMed: 24530107]
- [7]GN Otometrics. ICS Impulse FAQ. 2013; 2013

- [8]Halmagyi GM, Aw ST, Cremer PD, Curthoys IS, Todd MJ. Impulsive testing of individual semicircular canal function. *Ann NY Acad Sci.* 2001; 942:192–200. [PubMed: 11710461]
- [9]Harvey SA, Wood DJ. The oculocephalic response in the evaluation of the dizzy patient. *Laryngoscope.* 1996; 106:6–9. [PubMed: 8544630]
- [10]Harvey SA, Wood DJ, Feroah TR. Relationship of the head impulse test and head-shake nystagmus in reference to caloric testing. *Am J Otol.* 1997; 18:207–213. [PubMed: 9093678]
- [11]Huh YE, Koo JW, Lee H, Kim JS. Head-shaking aids in the diagnosis of acute audiovestibular loss due to anterior inferior cerebellar artery infarction. *Audiol Neurootol.* 2013; 18:114–124. [PubMed: 23296146]
- [12]Jorns-Haderli M, Straumann D, Palla A. Accuracy of the bedside head impulse test in detecting vestibular hypofunction. *J Neurol Neurosurg Psychiatry.* 2007; 78:1113–1118. [PubMed: 17220287]
- [13]Kattah JC, Talkad AV, Wang DZ, Hsieh YH, Newman-Toker DE. HINTS to diagnose stroke in the acute vestibular syndrome: Three-step bedside oculomotor examination more sensitive than early MRI diffusion-weighted imaging. *Stroke.* 2009; 40:3504–3510. [PubMed: 19762709]
- [14]MacDougall HG, Weber KP, McGarvie LA, Halmagyi GM, Curthoys IS. The video head impulse test: Diagnostic accuracy in peripheral vestibulopathy. *Neurology.* 2009; 73:1134–1141. [PubMed: 19805730]
- [15]Mantokoudis G, Saber Tehrani AS, Kattah JC, Eibenberger K, Guede CI, Zee DS, Newman-Toker DE. Quantifying the Vestibulo-Ocular Reflex with Video-Oculography: Nature and Frequency of Artifacts. *Audiol Neurootol.* 2014; 20:39–50. [PubMed: 25501133]
- [16]Mantokoudis G, Saber Tehrani AS, Wozniak A, Eibenberger K, Kattah JC, Guede CI, Zee DS, Newman-Toker DE. VOR Gain by Head Impulse Video-Oculography Differentiates Acute Vestibular Neuritis From Stroke. *Otol Neurotol.* 2014; 36:457–465.
- [17]Newman-Toker DE, Kattah JC, Alvernia JE, Wang DZ. Normal head impulse test differentiates acute cerebellar strokes from vestibular neuritis. *Neurology.* 2008; 70:2378–2385. [PubMed: 18541870]
- [18]Newman-Toker DE, Kerber KA, Hsieh YH, Pula JH, Omron R, Saber Tehrani AS, Mantokoudis G, Hanley DF, Zee DS, Kattah JC. HINTS outperforms ABCD2 to screen for stroke in acute continuous vertigo and dizziness. *Acad Emerg Med.* 2013; 20:986–996. [PubMed: 24127701]
- [19]Newman-Toker DE, Saber Tehrani AS, Mantokoudis G, Pula JH, Guede CI, Kerber KA, Blitz A, Ying SH, Hsieh YH, Rothman RE, Hanley DF, Zee DS, Kattah JC. Quantitative video-oculography to help diagnose stroke in acute vertigo and dizziness: Toward an ECG for the eyes. *Stroke.* 2013; 44:1158–1161. [PubMed: 23463752]
- [20]Palla A, Straumann D, Bronstein AM. Vestibular neuritis: Vertigo and the high-acceleration vestibulo-ocular reflex. *J Neurol.* 2008; 255:1479–1482. [PubMed: 18604466]
- [21]Saber Tehrani AS, Kattah JC, Mantokoudis G, Pula JH, Nair D, Blitz A, Ying S, Hanley DF, Zee DS, Newman-Toker DE. Small strokes causing severe vertigo: Frequency of false-negative MRIs and nonlacunar mechanisms. *Neurology.* 2014; 83:169–173. [PubMed: 24920847]
- [22]Schubert MC, Tusa RJ, Grine LE, Herdman SJ. Optimizing the sensitivity of the head thrust test for identifying vestibular hypofunction. *Phys Ther.* 2004; 84:151–158. [PubMed: 14744205]
- [23]Tarnutzer AA, Berkowitz AL, Robinson KA, Hsieh YH, Newman-Toker DE. Does my dizzy patient have a stroke? A systematic review of bedside diagnosis in acute vestibular syndrome. *CMAJ.* 2011; 183:E571–E592. [PubMed: 21576300]
- [24]van der Geest JN, Frens MA. Recording eye movements with video-oculography and scleral search coils: A direct comparison of two methods. *J neurosci Methods.* 2002; 114:185–195.
- [25]Weber KP, MacDougall HG, Halmagyi GM, Curthoys IS. Impulsive testing of semicircular-canal function using video-oculography. *Ann NY Acad Sci.* 2009; 1164:486–491. [PubMed: 19645955]

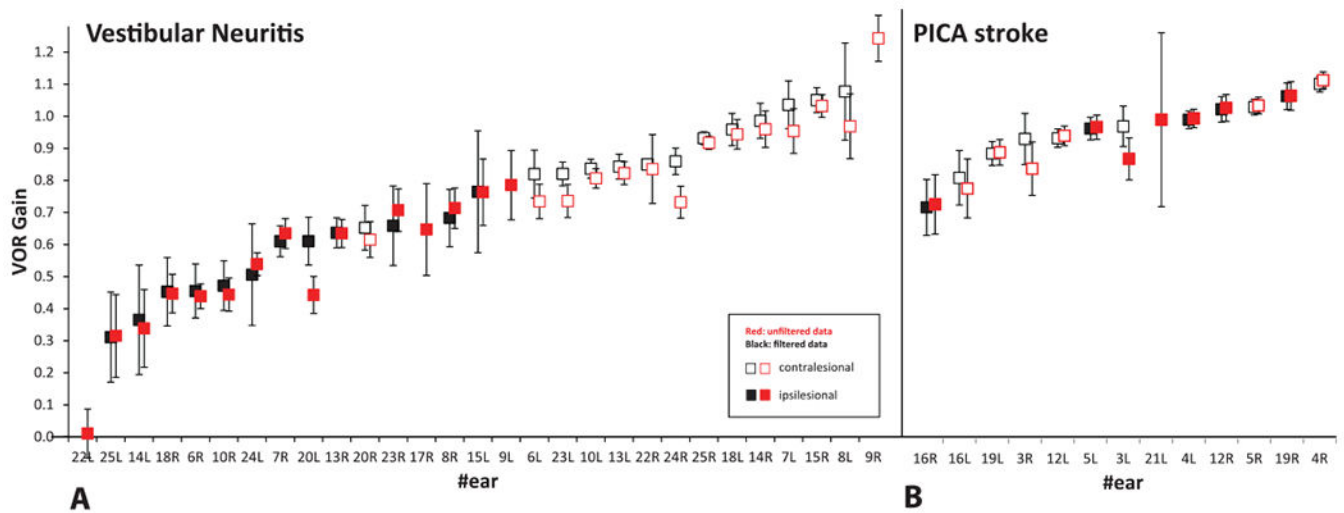


**Fig. 1.**

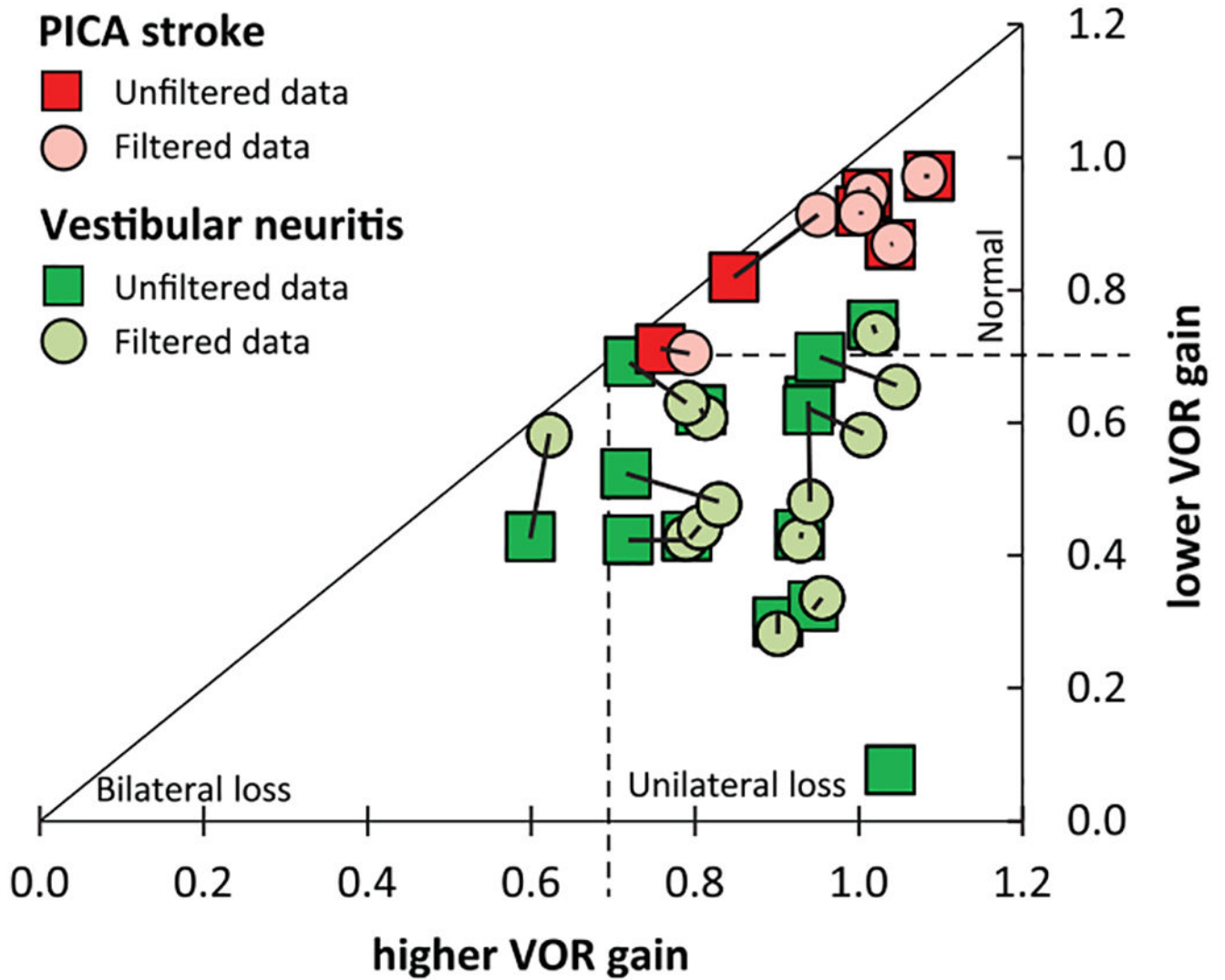
Unfiltered vHIT data versus filtered data. Velocity profiles from eye- and head VOG recordings derived from the contralesional (healthy) side in one patient with vestibular neuritis. The left panel shows unfiltered (raw device data, including artifacts) data, and the right panel filtered data (cleaned without artifacts). Both sets of data shown are from the same sequence of HIT trials in a single patient.

**Fig. 2.**

Population histograms depicting disease-specific VOR gain distributions, comparing unfiltered to filtered results. Population histograms of VOR gain values for patients with vestibular neuritis and PICA stroke. Histograms depict ipsilesional gain values for unfiltered vHIT data (A) and filtered vHIT data (B) in 23 patients with AVS. Data are not normalized to adjust for differences in the number of impulses per ear. Unfiltered (A) and filtered data (B) show an approximately normal distribution for each disease-specific population group with identical peaks at 0.6 VOR gain for vestibular neuritis and VOR gain of 1.0 for PICA strokes. Unfiltered data contained more HIT data from the larger neuritis population (A), however, filtering data (B) led to a disproportionately higher data removal from the neuritis group because abnormal HITs (not seen in PICA strokes) contained significantly more artifacts than normal HITs [15].

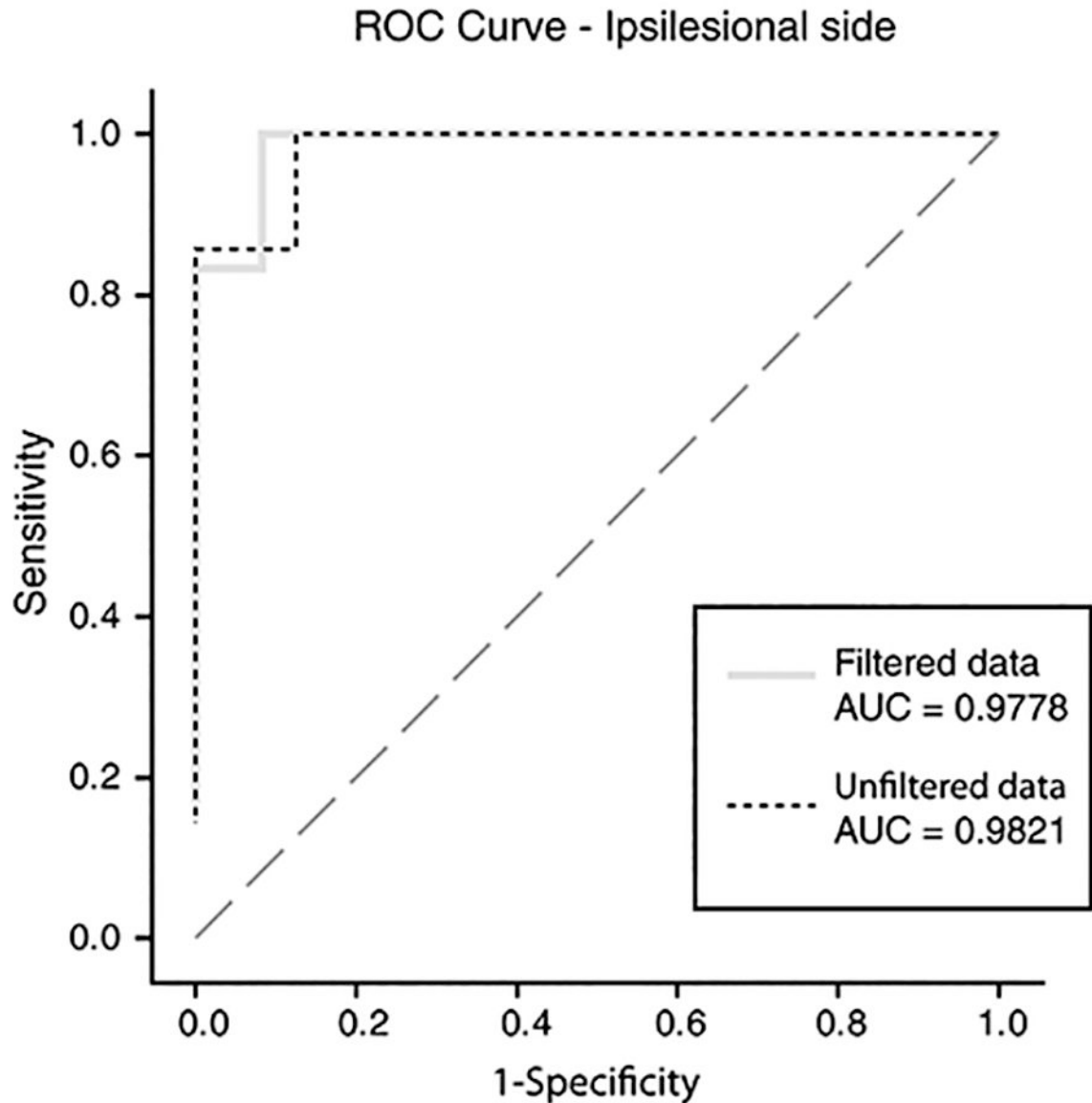
**Fig. 3.**

Mean VOR gains by ear with 95% confidence interval bars are shown for patients with (A) vestibular neuritis ( $n=15$ ), and (B) stroke (PICA,  $n=7$ ). Data are presented by increasing mean VOR gain values to highlight significant right-left asymmetries in peripheral disease and relative symmetry in PICA strokes. Unfiltered data (with artifacts) are depicted in red color, filtered data in black. VOR gains from five ears show only unfiltered results because there were fewer than five valid impulses (per ear) after filtering.

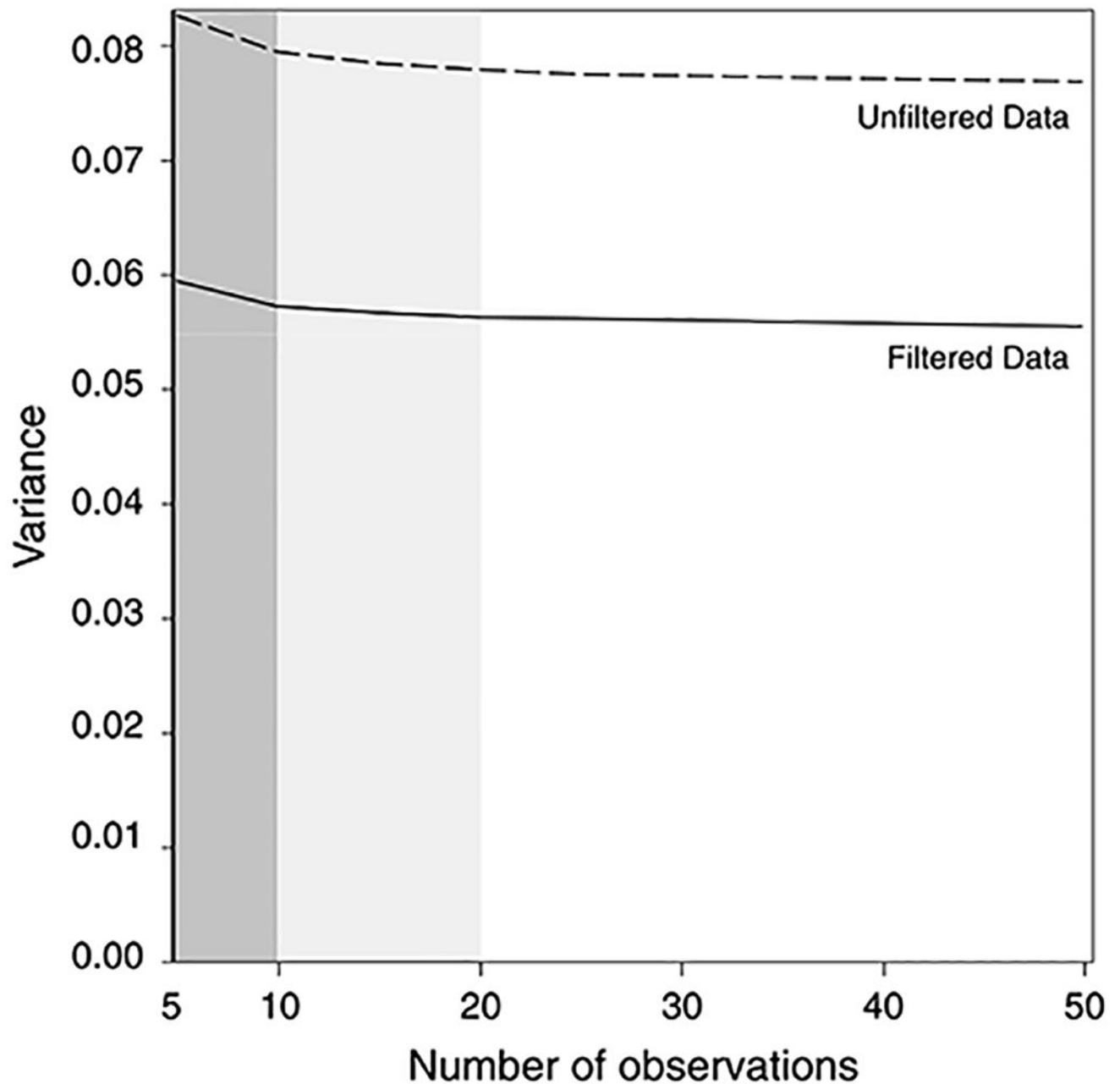


**Fig. 4.**

Scatterplot of patient-specific VOR gain asymmetries, comparing unfiltered to filtered results. Patient-specific VOR gain asymmetries are shown in a scatterplot, depicting lower VOR mean gain versus higher VOR mean gain for each patient. Only mean VOR gains based on five or more valid trials on both ears are shown (filtered data  $n=19$  [circles] and unfiltered data  $n=20$  [squares], corresponding data points connected). The triangle is divided into sections by the optimized cutoff of 0.70 for discriminating unilateral vestibular loss (neuritis, bottom-right quadrant), bilateral vestibulopathy (bottom-left corner) and strokes (bilateral normal VOR gain, upper-right corner). Note that strokes cluster in the upper-right corner independent of filtered vs. unfiltered data.



**Fig. 5.** ROC curves demonstrating sensitivity and specificity for stroke, comparing unfiltered to filtered results. This receiver operating characteristic (ROC) curve analysis plots the diagnostic accuracy (sensitivity and specificity) of quantitative VOG-derived VOR mean gains for identifying PICA stroke in AVS. We used ipsilesional vestibular neuritis and PICA stroke gains for ROC analysis and to determine an optimal VOR gain cut point for discrimination. The diagonal line indicates a hypothetical useless diagnostic test with a likelihood ratio of 1.0 at all diagnostic threshold cut points. The dashed line illustrates the ROC curve derived from unfiltered data from the VOG device, while the grey line shows the ROC curve from filtered data. Total diagnostic accuracy for stroke diagnosis at all thresholds, as measured by the area under the curve (AUC), is effectively identical for unfiltered vs. filtered results.



**Fig. 6.** Simulated, within subject gain variance (ipsilesional side) by number of anticipated HIT trial observations, comparing unfiltered to filtered results. Simulated, within-subject variance of mean VOR gains is shown on the y-axis, while the number of simulated HIT observations is shown on the x-axis. Filtered data are projected to have lower variance values than unfiltered data regardless of the number of HIT trials performed. Unfiltered data increase their precision substantially from 5 to 10 trials, and less beyond 20 trials.



**Table 1**AVS patient characteristics ( $n = 23$ )

Patient ID*	Age	Gender	Confirmed Diagnosis	Calorics (% Assym)	Clinical HIT	Vertical skew	Direction changing Nystagmus	Lesion side
1	66	f	Vestibular neuritis	38	Abnormal	No	No	left
3	62	m	PICA stroke	NA	Normal	No	No	left
4	65	m	Cerebellar peduncle hemorrhage	NA	Normal	Yes	Yes	left
5	31	m	PICA stroke	NA	Normal	No	No	left
6	72	m	Vestibular neuritis	NA	Abnormal	No	No	right
7	61	m	Vestibular neuritis	NA	Abnormal	No	No	right
8	62	f	Vestibular neuritis	70	Abnormal	No	No	right
9	59	f	vestibular neuritis	59	Abnormal	No	No	left
10	67	m	Vestibular neuritis	NA	Abnormal	No	No	right
12	63	m	PICA stroke	NA	Normal	No	Yes	right
13	46	m	Vestibular neuritis	51	Abnormal	No	No	right
14	83	m	Vestibular neuritis	NA	Abnormal	No	No	left
15	76	m	Labyrinthitis	NA	Abnormal	No	No	left
16	71	m	PICA stroke	NA	Normal	No	No	right
17	54	m	Vestibular neuritis	NA	Abnormal	Yes	No	right
18	72	m	Vestibular neuritis	NA	Abnormal	No	No	right
19	68	f	PICA Stroke	NA	Normal	No	No	right
20	43	m	Vestibular neuritis	43	Normal	No	No	left
21	64	m	PICA stroke	NA	Normal	No	No	left
22	48	m	Vestibular neuritis	72	Abnormal	No	No	left
23	40	f	Vestibular neuritis	9	Abnormal	No	No	right
24	55	f	Vestibular neuritis	NA	Abnormal	No	No	left
25	57	m	Vestibular neuritis	NA	Abnormal	No	No	left

m: male, f: female, NA: Not available.

\* Patient number indicates the order of ED admission. Missing patient numbers indicate three excluded AICA stroke patients from the series of 26 patients with complete test results.

**Table 2**

Paired analysis comparing unfiltered, filtered, and artifacts-only VOR gain data

Diagnosis	N*	Mean Gain (SD), unfiltered	Mean Gain (SD) filtered	Mean Gain (SD) artifacts-only	Paired Difference (SD) between filtered and artifacts- only
Neuritis	15	0.514 (0.194)	0.494 (0.188)	0.542(0.193)	-0.049 (0.107)
PICA stroke	3	0.866(0.165)	0.899 (0.175)	0.837 (0.159)	0.062 (0.094)
Neuritis and PICA	18	0.572 (0.229)	0.561 (0.239)	0.591 (0.216)	-0.030 (0.111) <sup>†</sup>

\* 5 patients were excluded from this analysis (2 patients had traces all with artifacts and 3 patients had completely clean data).

<sup>†</sup> no statistical differences were found in mean gain results between filtered and artifacts-only mean gain using the Wilcoxon signed rank test, *p*-value = 0.2379.

**Table 3**

Sensitivity and specificity with mean gain cut points of unfiltered data from ROC analysis ( $n = 23$ ) using the 'lower-gain side'

Mean Gain Cut-off	Sensitivity	Specificity	Number of Correctly Predicted Events	Number of Correctly Predicted Non-Events	Number of False Positive Events	Number of False Negative Events
$\geq 0.70992$	1.00000	0.875	7	14	2	0
$\geq 0.769737$	0.857142	0.9375	6	15	1	1
$\geq 0.818334$	0.857142	1.00000	6	16	0	1

\* Note that the lower-gain side (right or left) mean VOR gain was used as a predictive value for diagnosing stroke. The mean gain value cut points are depicted in the ROC curve (Fig. 5).