



Statistically robust measurement of evoked response onset latencies

Benjamin Letham, Tommi Raij*

MGH/MIT/HMS Athinoula A. Martinos Center for Biomedical Imaging, MA, USA

ARTICLE INFO

Article history:

Received 31 July 2010

Received in revised form 10 October 2010

Accepted 19 October 2010

Keywords:

Onset latency

Robust statistics

Evoked response

Brain

ABSTRACT

Onset latencies of evoked responses are useful for determining delays in sensory pathways and for indicating spread of activity between brain areas, therefore inferring causality. Previous studies have applied several different methods and parameters for detecting onsets, mainly utilizing thresholds based on the mean and standard deviation (SD) of the pre-stimulus “baseline” time window, or using *t*-tests of group data to determine when the response first differs significantly from the baseline. However, these methods are not statistically robust, have low power when the baseline data are not normally distributed, and are heavily influenced by outliers in the baseline. Here, we examine using a modified boxplot method known as the “median rule” for determining onset latencies. This rule makes no assumptions about the baseline distribution, is resistant to outliers, and can be applied to individual level data therefore allowing intersubject and interregional comparisons. We first show with simulations that the median rule is significantly less sensitive to outliers in the baseline than the SD method. We then use simulations to demonstrate the effect of skewness on onset latencies. Finally, we use magnetoencephalography (MEG) to show that the median rule can be easily applied to real data and gives reasonable results. In most situations the different methods give similar results, which enhances comparability across studies, but in data sets with a high noise level there is a clear advantage to using a statistically robust method. In conclusion, the median rule is an excellent method for estimating onset latencies in evoked responses.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Latency differences across brain areas reflect delays in sensory pathways, provide insight into how information flows through the brain for both bottom-up and top-down processing, and indicate causal relationships between areas (Fuxe and Simpson, 2002; Thorpe et al., 1996). In the cortex, latencies can be studied using evoked responses (ERs) recorded with electro- and magnetoencephalography (EEG/MEG). Since ERs in any given brain area consist of several waves of recurrent (re-entrant) activations, the clearest picture of how activations spread throughout the brain is achieved by studying the onsets of responses across areas.

Measuring onset latencies requires criteria to detect the presence of a response. This can be achieved by defining an amplitude threshold based on the pre-stimulus baseline typically consisting of all samples 100–200 ms before stimulus onset. A frequently

encountered method for defining the response threshold is a *t*-test, which is used to compare post-stimulus time points with the baseline distribution. In this method, the ERs of several subjects are used to create a distribution of response amplitudes at each time point. These distributions are then compared to the baseline and a significant departure from the baseline mean ($p < 0.05$) indicates the onset of the response (Allan and Rugg, 1997; Doniger et al., 2000; Fernandez et al., 1999; Fuxe and Simpson, 2002; Molholm et al., 2002; Rugg et al., 1993, 1995). However, this method can only be used to find the onset latency of an ER averaged across subjects. It does not inform about the distribution of onset latencies across subjects or the error in the measure of onset latency, making it difficult to accurately compare onset latencies across brain regions or subject populations (Fuxe and Simpson, 2002).

Another widely used method for defining the threshold, which can be applied to ERs of individual subjects, is to compute the mean and standard deviation (SD) of the baseline, and then to set a threshold at a certain level of SDs over the mean. The idea is that data points above the threshold are unlikely to come from the same distribution as the baseline (presumably consisting of noise) and thus are likely to be a part of the evoked response. However, noise levels vary across studies, and publications have used values as different as 2 SDs (Inui et al., 2006), 2.5 SDs (Fuxe and Simpson, 2002; Osman

Abbreviations: EEG, electroencephalography; EOG, electro-oculogram; ER, evoked response; MEG, magnetoencephalography; SD, standard deviation.

* Corresponding author at: Martinos Center, Bldg 149, 13th Street, Charlestown, MA 02129, USA. Tel.: +1 617 230 4341; fax: +1 617 726 7422.

E-mail address: raj@nmr.mgh.harvard.edu (T. Raij).

et al., 1992; Smulders et al., 1995), and 3 SDs (Martin et al., 2007; Raji et al., 2010) which makes comparisons across studies difficult.

In addition to defining a threshold, most studies have required that the response stays above the selected threshold for a minimum duration to be considered the real onset. This is useful because physiological signals often have brief peaks unassociated with the actual response which can easily exceed the 2–3 SD or $p < 0.05$ thresholds. Increasing the threshold safeguards against such outliers, but at the same time quickly decreases the sensitivity of the test, which is relevant especially for ERs with gradual onsets. Consequently, most studies balance sensitivity and specificity by combining a lower threshold with a minimum duration reflecting the number of consecutive outliers. The values for minimum response durations have varied widely from 10 ms (Foxe and Simpson, 2002), 16 ms (Rodriguez-Fornells et al., 2002; van Schie et al., 2004), 20 ms (Doniger et al., 2000; Molholm et al., 2002; Raji et al., 2010), 81 ms (Fernandez et al., 1999), 84 ms (Rugg et al., 1995), 100 ms (Osman et al., 1992; Osman and Moore, 1993; Smulders et al., 1995), even up to 112 ms (Allan and Rugg, 1997). When onsets are measured using t -tests, there are statistically rigorous methods for choosing a minimum duration (Achim, 1995; Guthrie and Buchwald, 1991); however, these methods are complex and most studies do not report which method, if any, was used to find the value. When using the SD method, statistically rigorous methods for choosing a minimum duration have to our knowledge not been described.

Clearly there is no consensus as to the proper threshold or minimum duration constraint. Moreover, the above methods are not statistically robust. Statistically robust methods are resilient to noisy data, particularly outliers, and do not assume anything about the form of distribution of the baseline data. A key assumption of the t -test is that the baseline is sampled from a normal distribution, which in general is not the case for MEG/EEG (Leonowicz et al., 2005). Moreover, the t -test has been shown to be very sensitive to outliers (Wilcox, 2005). Similarly, while an SD-based threshold does not explicitly assume a particular distribution, it has been shown to be overly sensitive to outliers and is a poor representation of data with a skewed distribution (Wilcox and Keselman, 2003).

When comparing different methods it is useful to examine how accurately they reflect the central tendency (value around which the data cluster) and dispersion (variability) of the data. For normally distributed data the mean and SD are efficient descriptions, but the presence of even just a few outliers in the pre-stimulus baseline can drastically influence the mean and SD and will lead to unrealistically high thresholds (Wilcox and Keselman, 2003). It is advantageous to use statistically robust measures instead, *i.e.*, measures that perform well for non-normal distributions, including those with outliers. Statistically robust measures of central tendency include the median, Winsorized means, and M-estimators, and of dispersion the interquartile range and the median absolute deviation (Hoaglin et al., 1983).

There are several statistically robust methods for outlier detection, an idea analogous to measuring onset latencies, one of which is the boxplot (Hoaglin et al., 1983). A particularly useful variant of the boxplot outlier method is called the median rule (Carling, 2000). This rule, like the SD method, is capable of estimating the onset latency from ERs of individual subjects, allowing the variance to be estimated. Here, we will use simulations to demonstrate that onset latencies measured using the median rule are in fact robust in the presence of outliers while those measured with the SD method are not, and we will examine the effect of skewness on measures of onset latency. We then use MEG data to demonstrate that the method can

easily be applied to real data and that it produces reasonable results.

2. Materials and methods

2.1. The median rule and properties

Boxplot rules for outlier detection are based on quartiles, which form a statistically robust description of the data. The second quartile (q_2) is the sample median, and the first and third quartiles (q_1 and q_3) define the interquartile range $q_3 - q_1$. The classical boxplot defines outliers as points that fall above $q_3 + 1.5(q_3 - q_1)$ and below $q_1 - 1.5(q_3 - q_1)$ (Tukey, 1977). The median rule is a modification of the boxplot outlier rule that performs better especially for small sample sizes (Carling, 2000). It uses the median together with the interquartile range and defines outliers as points that fall outside $q_2 \pm 2.3(q_3 - q_1)$. Boxplot rules have a breakdown point of 0.25, *i.e.*, up to 25% of the data can be outliers without problem. There are other rules with higher breakdown points (*e.g.*, using median absolute deviation), but for ERs a breakdown point of 0.25 should in practice be sufficient. The median rule performs well even with relatively small sample sizes (Carling, 2000), which are frequently used in neurophysiological studies.

2.2. Defining quartiles

The way in which quartiles are defined influences the performance of boxplot rules (Frigge et al., 1989). The several approaches differ mainly in how interpolation between data points is done. The best results with the median rule have been achieved using a method called ideal fourths (Carling, 2000; Hoaglin and Iglewicz, 1987). As described by Carling (2000), for an ordered data set x , each quartile is a linear interpolation between two data points in x : $(1 - g)x_{(j)} + gx_{(j+1)}$. For the first quartile, the indices of the data points used for interpolation (j) and the amount of interpolation (g) are the integer part (j) and remainder (g) of $n/4 + 5/12$. For the third quartile, j and g are the integer part and the remainder of $3n/4 + 7/12$. The second quartile is simply the sample median. Using ideal fourths reduces bias in the median rule compared with simple interpolation (Carling, 2000).

For neurophysiological time series data, the median rule can be applied by computing the quartiles of the baseline data and then finding the first post-stimulus point that falls outside $q_2 \pm 2.3(q_3 - q_1)$. Here we study waveforms with positive activation and therefore use the threshold $q_2 + 2.3(q_3 - q_1)$. For normally distributed noise, the median rule is equivalent to 3.1 SDs over the mean.

2.3. Simulation 1 – Robustness to baseline outliers

We first measured onset latency of a function with known onset to which normally distributed noise was added. Fig. 1A shows an example data set with the measured onset latencies. The base function extended from -200 ms to +200 ms, was sampled at 1 kHz, and had an onset at 30 ms to a stimulus at 0 ms. After onset, the response increased linearly to its maximum value of 100. Thus, the time window from -200 ms to 0 ms was the pre-stimulus baseline (200 samples) and the true onset latency 30 ms. Normally distributed noise with mean 0 and standard deviation 10 (*i.e.*, 10% of the maximum response amplitude) was added to the base function, and onset latencies were measured using the median rule and the SD method (3.1 SD threshold). The t -test method cannot be compared with the other methods that use individual simulated data sets and was thus not included in this comparison. Ten thousand simulated data sets were generated to achieve a stable estimate of onset latencies, each with the same base function but different noise sample.

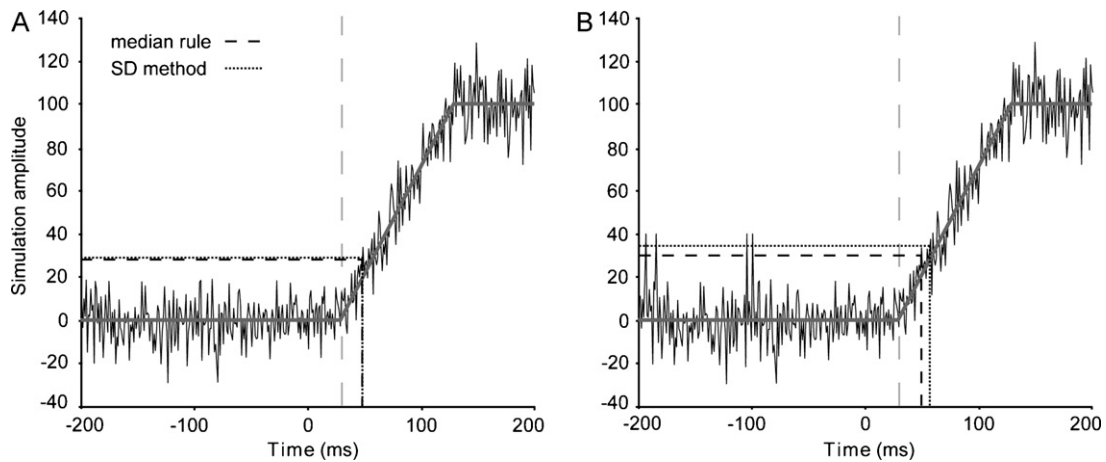


Fig. 1. Onsets for a simulated response with normally distributed noise. Both Panels A and B show a ramp function (thick solid grey line) with known onset at 30 ms (vertical dashed grey line) to which normally distributed noise (mean = 0, SD = 10) has been added. There is a “pre-stimulus baseline” from –200 ms to 0 ms. The black horizontal dashed line is the onset threshold for the median rule, $q_2 + 2.3(q_3 - q_1)$, and the black horizontal dotted line for the SD method (mean plus 3.1 SDs). The onset latency is the first post-stimulus point above threshold, and is marked with a corresponding vertical line. In Panel A no outliers are added to the pre-stimulus baseline, whereas in Panel B four randomly chosen baseline points are replaced with outliers at 4 SDs. Adding outliers significantly delays the onset latency measured with the SD method, whereas the onset measured with the median rule remains correct.

Fig. 1B shows data simulated in an identical way as above except that four randomly selected baseline points in each of the simulated data sets were replaced with outliers at 4 SDs. Onset latencies were re-measured for the simulated data sets with outliers. This measures robustness of the methods to outliers during the baseline.

2.4. Simulation 2 – Effect of skewness on onset latency

The second simulation studied the effect of skewness on measures of onset latency. Data sets were generated as in simulation 1 but with skewed instead of normally distributed noise. The noise was sampled from the *skew normal distribution*, a generalization of the normal distribution that is defined by a shape parameter in addition to the location and scale parameters of the standard normal distribution. When the shape parameter is set to 0, the skew normal distribution is identical to the normal distribution with the same location and scale parameters. A negative shape parameter corresponds to negative skewness and a positive shape parameter corresponds to positive skewness. With the skew normal distribution it is possible to vary skewness while holding mean and SD constant. For this simulation the mean was set to 0 and the SD was set to 15, while the skewness was varied from –8 (strongly negatively skewed) to +8 (strongly positively skewed). In order to use the *t*-test method for this simulation, we created “group averages” by averaging a small number ($N=4$) of simulated ERs. An example of such a group average can be seen in **Fig. 2A**. The average ER was used to measure onset latency with the median rule and the SD method, and the *t*-test method was applied to the group of 4 ERs. As before, this was repeated 10,000 times for each value of the skewness parameter to obtain a stable result. Since $N=4$ is unrealistically small for most applications, the simulation was finally repeated with $N=24$ in each group average.

2.5. Real MEG evoked responses

These data are a subset of those previously reported in **Raij et al. (2010)**. We presented 300-ms auditory, visual, and audiovisual (combined auditory and visual) stimuli to eight healthy right-handed human subjects (six females, age 22–30). The auditory stimuli were binaurally presented white noise bursts and the visual stimuli foveally presented static checkerboard patterns. The inter-

stimulus interval was pseudorandom, with a range of 1.15–3.45 s and mean of 1.5 s.

Whole-head 306-channel MEG (VectorView, Elekta-Neuromag, Finland) was recorded in a magnetically shielded room (**Cohen et al., 2002; Hämäläinen and Hari, 2002**) simultaneously with horizontal and vertical electro-oculogram (EOG). All signals were band-pass filtered to 0.03–200 Hz prior to sampling at 600 Hz. Responses were averaged offline time locked to the stimulus onsets with a time window of 250 ms pre-stimulus to 1150 ms post-stimulus. A total of 125 individual epochs for each stimulus type were acquired. Epochs exceeding 150 μV at any EOG channel or 3000 fT/cm at any MEG channel were automatically discarded from the averages. The averaged signals were digitally low-pass filtered at 40 Hz and amplitudes were measured with respect to a 200-ms pre-stimulus baseline. Gradient amplitude responses $\sqrt{x^2 + y^2}$ were calculated from the two planar gradiometers x and y at each sensor location and onsets were measured from these responses. Since such responses can only have positive values, we here examine onsets for upward deflections. However, the methods are equally applicable to normal EEG/MEG responses that have both positive and negative values.

Data are analyzed from two sensor locations, one over the supratemporal auditory cortex in each hemisphere, selected for each subject individually at the location that shows the strongest 100-ms response to the unimodal auditory stimuli. Here we report onset latencies to the audiovisual stimuli only, additionally constraining the onset to be no earlier than 15 ms. Onsets were computed for each individual’s data set using the median rule at $q_2 + 2.3(q_3 - q_1)$ threshold and the SD method at 3.1 SD threshold. The *t*-test method cannot be used for individual level responses and was thus applied only to the group data. Minimum duration constraints were not used with any method.

3. Results

P values below were computed using the Wilcoxon signed rank test, a non-parametric statistical test similar to a paired *t*-test. For onset latencies, a distribution free 95% confidence interval for the median was computed in R with the freely available program “sint” (<http://r-forge.r-project.org/projects/wrs/>) (**Wilcox, 2005**).

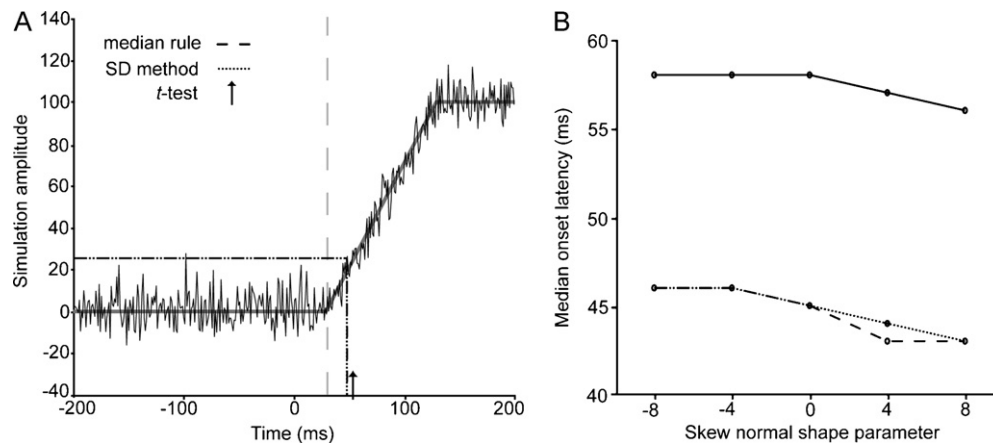


Fig. 2. Effect of skewness on onset latency. Simulated data were generated as in Fig. 1, but with varying skewness while holding mean and SD constant. Data were simulated in groups of four ($N=4$) which were then averaged to create a “group average” ER, an example of which is shown in Panel A. As in Fig. 1, the known onset is at 30 ms (vertical dashed grey line), and the black lines indicate the threshold (horizontal) and onset (vertical) measured with the median rule (dashed lines) and with the SD method (dotted lines). The onset of the group measured with the t -test is indicated with an arrow. The direction and degree of skewness is manipulated with a parameter, which can be negative, zero for unskewed, or positive. Panel B shows median onset latency for all three methods across 10,000 simulated “group averages” with varying amounts of skewness. For all methods, onset latencies decrease as skewness becomes more positive. The median rule and SD method (dashed and dotted lines) were identical except for moderate positive skewness. The onsets detected by the t -test (solid line) were clearly longer than for the other methods for all values of skewness. When the number of ERs in the average was increased ($N=24$) the effect of skewness became insignificant (not shown, see text for details).

3.1. Robustness to baseline outliers

Fig. 1 shows an example simulated data set with normally distributed noise, without added outliers in Fig. 1A and with four added outliers in Fig. 1B. The thresholds were computed using the median rule (dashed horizontal line) and the SD method (dotted horizontal line). Onset is the first post-stimulus time point above threshold, marked with vertical dashed and dotted lines for median rule and SD method, respectively. Across all 10,000 simulated data sets, with no added outliers (Fig. 1A) the median onset latencies were 49 ms (95% confidence interval 48–49 ms) with both the median rule and the SD method. These were not significantly different ($p=0.32$). With added outliers (Fig. 1B) the onset latency increased by 1 ms (2%) to 50 ms (50–50 ms) with the median rule, and by 5 ms (10%) to 54 ms (54–54 ms) with the SD method. These onset latencies were significantly different ($p < 2 \times 10^{-16}$).

3.2. Onset latencies with skewed noise

Fig. 2A shows an example “group average” of 4 ERs with positive skewness (shape parameter +8). The median rule and SD method onsets are labeled as in Fig. 1. The onset measured with the t -test is labeled with an arrow. The median onset latencies varied with skewness for all three methods as shown in Fig. 2B. Both the upper and lower bounds of the 95% confidence interval were equal to the median for all methods. Negative skewness delayed the measured onset while positive skewness resulted in earlier onsets. For negatively skewed noise the SD method and the median rule gave identical results, though for moderately skewed data the median rule onset was earlier by 1 ms ($p < 2 \times 10^{-16}$). Initially, median onset latency with the t -test method was less than 15 ms for all values of the shape parameter, which is earlier than the true onset and clearly a false positive. To receive any reasonable results using the t -test, we therefore had to apply the minimum duration requirement: four consecutive time points had to satisfy $p < 0.05$ for the onset (the first data point of the four below threshold) to be considered real (Gutherie and Buchwald, 1991). No such constraint was required with the median rule or the SD method. The resulting onset latency measured with the t -test was several milliseconds later for all amounts of skewness ($p < 2 \times 10^{-16}$).

With 24 ERs in the group average, noise levels were significantly reduced and the median onset latency was 37 ms (37–37 ms) with both the median rule and the SD method, and 38 ms (38–38 ms) with the t -test. These median onset latencies did not vary with skewness.

3.3. MEG evoked responses

Fig. 3 shows the MEG data collected across all subjects from sensors over the left (Panel A) and right (Panel B) hemisphere auditory cortices. The group level mean ER is indicated with the thick black line and the 95% confidence intervals calculated via percentile bootstrapping with the grey lines. The individual baselines for each of the ERs were checked for normality using the Lilliefors test, revealing that 12 of the 16 baselines (2 hemispheres in each of 8 subjects) were significantly different from the normal distribution ($p < 0.05$).

For the SD method, median onset latency was 37 ms (26–45 ms) over the left and 39 ms (28–47 ms) over the right auditory cortex. With the median rule, the median onset latency was 39 ms (28–43 ms) over the left and 39 ms (28–49 ms) over the right auditory cortex. The median rule did not differ significantly from the SD method in either hemisphere ($p=0.42$ for both). Using the t -test, onset latency of the group level mean data was 37 ms in the left and 33 ms in the right auditory cortex (error estimates or confidence intervals cannot be calculated using this method – thus, statistical comparisons between the t -test method and other techniques could not be done).

4. Discussion

The results show that the median rule is a useful tool for measuring onset latencies of ERs. Simulations indicate that this rule is not overly sensitive to outliers as are methods based on SD thresholding. The median rule can easily be applied to real MEG data, and at group level produces results comparable to those obtained with SD thresholding and t -tests.

Previously widely used methods for measuring onset latencies are suboptimal. The t -test assumes that the population from which the baseline is sampled is normally distributed. EEG/MEG data sets are non-stationary and heavy-tailed, which is inconsistent with an assumption of normality (Chau et al., 2004; Palus, 1996). Out-

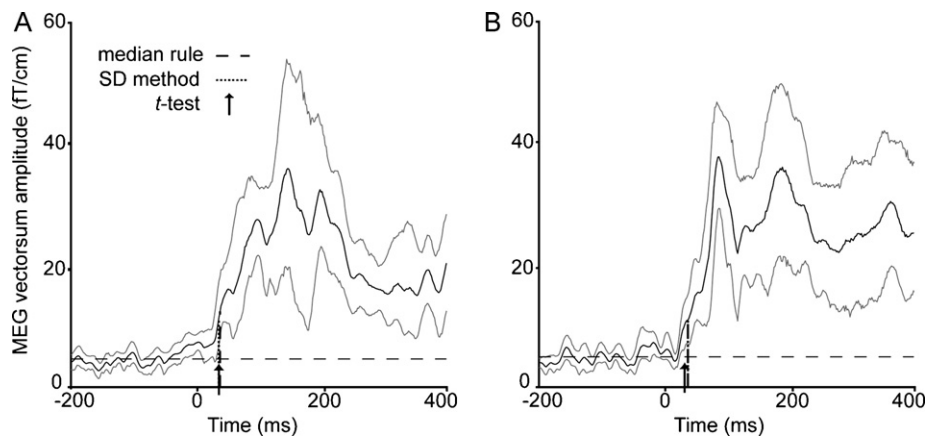


Fig. 3. Onsets for group level MEG responses. The median rule, SD method, and t -test method were applied to MEG ERs collected from 8 subjects. The group level mean ER (thick black line) is shown with 95% confidence intervals (grey lines) for selected sensor locations over the left (A) and right (B) auditory cortices. The t -test can only be used to find the onset latency at the group level (black arrows). For comparison, we also show the median values of the onsets across individual subjects for the median rule (dashed line) and SD method (dotted line). As expected for reasonable quality data, the results were similar across all three methods.

liers and heavy tails significantly reduce the statistical power of the t -test (Benjamini, 1983; Wilcox, 2005). In the context of onset latencies this translates to poor detection sensitivity. The t -test is also based on standard deviation, which is a non-robust measure of variability that is overly sensitive to outliers as demonstrated by the simulations. Most importantly, the t -test can only be used to measure onset latency from an ER averaged across subjects and gives no estimate of the variance in onset latencies across subjects. Without variances, it is impossible to make statistical comparisons. Some studies have adapted the t -test method to individual subjects by windowing the data to create the pointwise distribution rather than forming the distribution across subjects (Rodriguez-Fornells et al., 2002; van Schie et al., 2004). However, windowing further reduces sensitivity and requires an arbitrary selection of window size.

SD thresholds have been used extensively for outlier detection, despite having been shown to be ineffective due to their sensitivity to outliers (Wilcox and Keselman, 2003). Our simulations confirmed that the SD method is overly sensitive to outliers. With normally distributed noise, the onset latencies measured with the median rule and the 3.1 SD threshold were identical, as they theoretically should be. Replacing only two percent of the baseline data points with outliers delayed the SD onset such that it was significantly later than the median rule onset by 4 ms. The median rule produces a statistically more correct representation of the data, and is not affected as strongly by a relatively small number of outliers in the data.

All measures of onset latency are sensitive to skewness in data with high noise levels. When attempting to detect a positive deflection, positive skewness increases the probability of a random noise sample crossing threshold thereby decreasing the measured onset latency. Correspondingly, a negative skewness increases the observed onset. While in this study we analyzed responses that can only have positive values, typical evoked EEG/MEG responses have both positive and negative deflections. For negative deflections, the relations between direction of skewness and latency change are reversed: positive skewness increases and negative skewness decreases the observed onset latency. Skewness is often changed by elementary operations, for example if the noise in two planar gradiometers is normally distributed (unskewed) then the vector magnitude will follow a Rayleigh distribution (positively skewed). One should therefore exercise caution when comparing onset latencies between studies that have clearly different noise levels or use dissimilar computational methods that differentially

influence noise level and skewness. However, the results of the simulation with $N=4$ were a worst-case scenario with high noise and few averages. The effects of skewness on onset latency disappear with a large number of subjects.

Statistically robust methods would be particularly important for applications in which only a few trials (segments of raw data) are used for calculating ERs and noise levels are very high (Leonowicz et al., 2005). However, the median rule can be useful in all data sets, even in those with low noise levels. Using robust measures in future onset latency studies does not invalidate previously published results. In our real data from 8 subjects the difference between boxplot and SD onset latencies was not significant. The t -test latencies fell within the confidence intervals of both the other methods, although they could not be compared statistically because of the limitations of the t -test method. The lack of a consistent method for measuring onset latencies makes comparisons of different studies difficult. An ideal method should not make assumptions about the shape of the ER or the distribution of the baseline. For outlier detection, boxplot rules such as the median rule are widely used and have been proven versatile enough for a wide range of applications. The median rule is both statistically robust and easy to implement. Code for implementing the median rule is available in Wilcox (2005).

A requirement that the data stay above threshold for a certain period of time is arbitrary and therefore poorly justified unless done in a statistically rigorous way (Achim, 1995). A statistically rigorous requirement has not been developed for the median rule or SD method. Therefore we propose that when constraints are necessary, they should be based on the physiology of the response, which may vary across applications. For instance, in intracranial recordings it takes about 15 ms for activations to reach the auditory cortex (Celesia, 1976). Thus, when measuring the evoked response over auditory cortex, a physiologically relevant constraint would be to require that onset can occur only after 15 ms, as was done in this study. When interested in the onset of only the main (which is not necessarily the earliest) component of a response, a relevant constraint would be to choose the onset that stays above threshold longest (Maris and Oostenveld, 2007). Ideally, such constraints should be chosen when designing the experiment and not applied post hoc.

We have shown through simulations and with real data that the median rule, a statistically robust outlier detection method, is simple and effective for measuring onset latencies of evoked responses. As opposed to previously used methods, the median rule makes no assumptions about the distribution of the data, is not sensitive to

outliers in the pre-stimulus baseline, does not require additional constraints, and can be applied to finding the onset of individual ERs. Simulations show that the median rule performs well even for noisy and/or skewed data. The median rule can be easily applied to real data and produces results similar to other methods across subjects. This encourages comparison of results obtained with the median rule and previous studies using less robust methods, and may encourage use of the median rule in future studies.

Acknowledgements

We thank Drs. Jyrki Ahveninen, Matti Hämäläinen, Fa-Hsuan Lin, and the two anonymous Reviewers for insightful and valuable suggestions for the manuscript. This work was conducted with support from Harvard Catalyst|The Harvard Clinical and Translational Science Center (NIH Award #UL1 RR 025758 and financial contributions from Harvard University and its affiliated academic health care centers), NIH R01 NS048279, P41 RR14075, and the National Center for Research Resources. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, the National Center for Research Resources, or the National Institutes of Health.

References

- Achim A. Signal detection in averaged evoked potentials: Monte Carlo comparison of the sensitivity of different methods. *Electroencephalogr Clin Neurophysiol* 1995;96:574–84.
- Allan K, Rugg MD. An event-related potential study of explicit memory on tests of cued recall and recognition. *Neuropsychologia* 1997;35:387–97.
- Benjamini Y. Is the *t* test really conservative when the parent distribution is long-tailed? *J Am Stat Assoc* 1983;78:645–54.
- Carling K. Resistant outlier rules and the non-Gaussian case. *Comput Stat Data Anal* 2000;33:249–58.
- Celesia G. Organization of auditory cortical areas in man. *Brain* 1976;99:403–14.
- Chau W, McIntosh AR, Robinson SE, Schulz M, Pantev C. Improving permutation test power for group analysis of spatially filtered MEG data. *Neuroimage* 2004;23:983–96.
- Cohen D, Schlapfer U, Ahlfors S, Hamalainen M, Halgren E. New six-layer magnetically-shielded room for MEG. In: Nowak H, Haueisen J, Giessler F, Huonker R, editors. 13th international conference on biomagnetism. Jena, Germany: VDE Verlag; 2002. p. 919–21.
- Doniger G, Foxe J, Murray M, Higgins B, Snodgrass J, Schroeder C, et al. Activation timecourse of ventral visual stream object-recognition areas: high density electrical mapping of perceptual closure processes. *J Cogn Neurosci* 2000;12:615–21.
- Fernandez G, Efferen A, Grunwald T, Pezer NKL, Dimpelmann M, van Roost D, et al. Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science* 1999;285:1582–5.
- Foxe JJ, Simpson GV. Flow of activation from V1 to frontal cortex in humans. *Exp Brain Res* 2002;142:139–50.
- Frigge M, Hoaglin D, Iglewicz B. Some implementations of the boxplot. *Am Stat* 1989;43:50–4.
- Guthrie D, Buchwald J. Significance testing of difference potentials. *Psychophysiology* 1991;28:240–4.
- Hämäläinen M, Hari R. Magnetoencephalographic characterization of dynamic brain activation. Basic principles and methods of data collection and source analysis. In: Toga AW, editor. *Brain mapping: the methods*. New York: Academic Press; 2002. p. 227–53.
- Hoaglin D, Iglewicz B. Fine-tuning some resistant rules for outlier labeling. *J Am Stat Assoc* 1987;82:1147–9.
- Hoaglin D, Mosteller F, Tukey J, editors. *Understanding robust and exploratory data analysis*. John Wiley & Sons; 1983.
- Inui K, Sannan H, Miki K, Kaneoke Y, Kakigi R. Timing of early activity in the visual cortex as revealed by simultaneous MEG and ERG recordings. *Neuroimage* 2006;30:239–44.
- Leonowicz Z, Karvanen J, Shishkin S. Trimmed estimators for robust averaging of event-related potentials. *J Neurosci Methods* 2005;142:17–26.
- Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 2007;164:177–90.
- Martin T, McDaniel M, Guynn M, Houck J, Woodruff C, Bish J, et al. Brain regions and their dynamics in prospective memory retrieval: a MEG study. *Int J Psychophysiol* 2007;64:247–58.
- Molholm S, Ritter W, Murray M, Javitt D, Schroeder C, Foxe J. Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 2002;14:115–28.
- Osman A, Bashore TR, Coles MGH, Donchin E, Meyer DE. On the transmission of partial information: inferences from movement-related brain potentials. *J Exp Psychol* 1992;18:217–32.
- Osman A, Moore CM. The locus of dual-task interference: psychological refractory effects on movement-related brain potentials. *J Exp Psychol* 1993;19:1292–312.
- Palus M. Nonlinearity in normal human EEG: cycles, temporal asymmetry, nonstationarity and randomness, not chaos. *Biol Cybern* 1996;75:389–96.
- Raji T, Ahveninen J, Lin FH, Witzel T, Jaaskelainen IP, Letham B, et al. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *Eur J Neurosci* 2010;31:1772–82.
- Rodriguez-Fornells A, Kurzbuch AR, Munte TF. Time course of error detection and correction in humans: neurophysiological evidence. *J Neurosci* 2002;22:9990–6.
- Rugg M, Doyle M, Melan C. An event-related potential study of the effects of within- and across-modality word repetition. *Lang Cogn Process* 1993;8:357.
- Rugg MD, Doyle MC, Wells T. Word and nonword repetition within- and across-modality: an event-related potential study. *J Cogn Neurosci* 1995;7:209–27.
- Smulders FTY, Kok A, Kenemans JL, Bashore TR. The temporal selectivity of additive factor effects on the reaction process revealed in ERP component latencies. *Acta Psychol (Amst)* 1995;90:97–109.
- Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. *Nature* 1996;381:381.
- Tukey J. *Exploratory data analysis*. Reading, MA: Addison-Wesley Pub. Co; 1977.
- van Schie HT, Mars RB, Coles MGH, Bekkering H. Modulation of activity in medial frontal and motor cortices during error observation. *Nat Neurosci* 2004;7:549–54.
- Wilcox R. *Introduction to robust estimation and hypothesis testing*. 2nd ed. Amsterdam; Boston: Elsevier/Academic Press; 2005.
- Wilcox RR, Keselman HJ. Modern robust data analysis methods: measures of central tendency. *Psychol Methods* 2003;8:254–74.