Neural-latency noise places limits on human sensitivity to the timing of events

Abbreviated Title: Neural-latency noise limits timing sensitivity

Kielan Yarrow[1*], Carmen Kohl[2], Toby Segasby[1], Rachel Kaur Bansal[1], Paula Rowe[1] & Derek H. Arnold[3]

[1] Department of Psychology, City, University of London, London EC1V 0HB
[2] Department of Neuroscience, Brown University, Providence, RI 02912
[3] School of Psychology, The University of Queensland, Brisbane QLD 4072

* Author for correspondence:

Kielan Yarrow,
Rhind Building,
City, University of London,
Northampton Square,
London EC1V 0HB

Tel: +44 (0)20 7040 8530
Email: kielan.yarrow.1@city.ac.uk

Abstract

The brain-time account posits that the *physical* timing of sensory-evoked neural activity determines the *perceived* timing of corresponding sensory events. A canonical model formalises this account for tasks such as simultaneity and order judgements: Signals arrive at a decision centre in an order, and at a temporal offset, shaped by neural propagation times. This model assumes that the noise affecting people's temporal judgements is primarily neural-latency noise, i.e. variation in propagation times across trials, but this assumption has received little scrutiny. Here, we recorded EEG alongside simultaneity judgements from 50 participants in response to combinations of visual, auditory and tactile stimuli. Bootstrapping of ERP components was used to estimate neural-latency noise, and simultaneity judgements were modelled to estimate the precision of timing judgements. We obtained the predicted correlation between neural and behavioural measures of latency noise, supporting a fundamental feature of the canonical model of perceived timing.

Keywords

Time perception, timing, simultaneity, synchrony, order, intersensory.

1. Introduction

The temporal sequencing of events provides narrative structure for our experiences, and likely supports important cognitive operations such as inferring causal relationships (Michotte, 1954) and perceptually binding or segregating sensory representations (Fujisaki & Nishida, 2010; Holmes & Spence, 2005). However, we don't yet know how the brain determines synchrony and order. Indeed, even basic premises, such as the idea that the timing of the neural activity that represents an event is causal for the experience of subjective timing – which we refer to as the brain-time account – remain controversial (Dennett & Kinsbourne, 1992; Moutoussis & Zeki, 1997; Nishida & Johnston, 2002; Paillard, 1949; Whitney & Murakami, 1998; Yarrow & Arnold, 2016).

The brain-time account has inspired several formal models of temporal sequencing. The canonical model (Sternberg & Knoll, 1973) represents a special case of signal detection theory (Green & Swets, 1966). Behaviourally, tasks assessing perceived event timing, such as temporal order and synchrony judgements, reveal variation in judgements even across trials presenting the exact same physical stimuli, yielding gently sloped psychometric functions as responses gradually transition from predominance of one judgement category to another (e.g. asynchronous to synchronous). This implies that some kind of *internal noise* limits performance. A key assumption of the canonical model is that this internal noise reflects *latency* noise, i.e. trial-to-trial differences in the latencies with which the signals representing events propagate through the nervous system toward a central decision centre. Modern variants of the canonical model retain the notion that latency noise is a key determinant of the psychometric function (García-Pérez & Alcalá-Quintana, 2012a), even when they allow for other contributory sources, such as instability in decision criteria from trial to trial (Ulrich, 1987; Yarrow, Jahn, Durant, & Arnold, 2011).

Discussions of the brain-time account often focus on the average subjective ordering of events, which could reflect neural propagation latencies. For example, participants are biased to tap earlier when synchronising tap responses with an auditory metronome, and this bias is exacerbated for foot tapping compared to hand tapping (Fraisse, 1980). This is consistent with an attempt to synchronise reafferent tactile and exafferent auditory signals in the brain, given generally longer somatosensory relative to auditory latencies, with the resulting bias exaggerated by lengthened neural pathways from the foot relative to the hand. A similar focus on subjective order is evident in more direct assays of timing in the human brain. For example, studies have related the average timing of neural activity, in the form of event-related potential (ERP) components, to attention-dependent changes in average perceived temporal order, known as prior entry effects (McDonald, Teder-Salejarvi, Di Russo, & Hillyard, 2005; Vibell, Klinge, Zampini, Spence, & Nobre, 2007).

A focus on average subjective order, with less scrutiny applied to the predicted consequences of latency variation, is understandable, as estimating a bias seems conceptually more straightforward than measuring and making predictions about noise (but see Yarrow, et al., 2011). Yet the impact of latency noise on the precision of timing judgements is a key diagnostic for the canonical model of event timing, which has not been thoroughly tested. Furthermore, the primacy of latency noise is by no means a given. In addition to conceptual criticisms (Dennett & Kinsbourne, 1992; Nishida & Johnston, 2002), several models exist which could imply primacy for other forms of noise. These include population-code models, formulated to explain aftereffects of event timing (Roach, Heron, Whitaker, & McGraw, 2011; Yarrow, Minaei, & Arnold, 2015), where noise is thought to reflect variation in the spiking activity of units tuned to specific timing relationships, and models that imply a series of linear operations on temporally filtered inputs (Burr, Silva, Cicchini, Banks, & Morrone, 2009; Parise & Ernst, 2016), where noise has been modelled as an add on at a decision stage.

Here, we test whether neural-latency variation across trials predicts (and thus may limit) the precision of timing judgements. We present auditory, visual, and tactile stimulations, in order to estimate latency variation from inter-trial changes in ERP components. We then apply a variant of the canonical model (GLINC – Gaussian Latency Independent Noisy Criteria; Yarrow et al., 2011) to estimate the precision of synchrony judgements concerning audio-visual (AV), audio-tactile (AT), and visuo-tactile (VT) stimulus pairs. Our analytic approach is schematised in Figure 1. We find that the precision of subjective timing judgements can be predicted from formally near-equivalent measures of inter-trial latency variation – consistent with the hypothesis that temporally noisy brains promote temporal imprecision in perception.

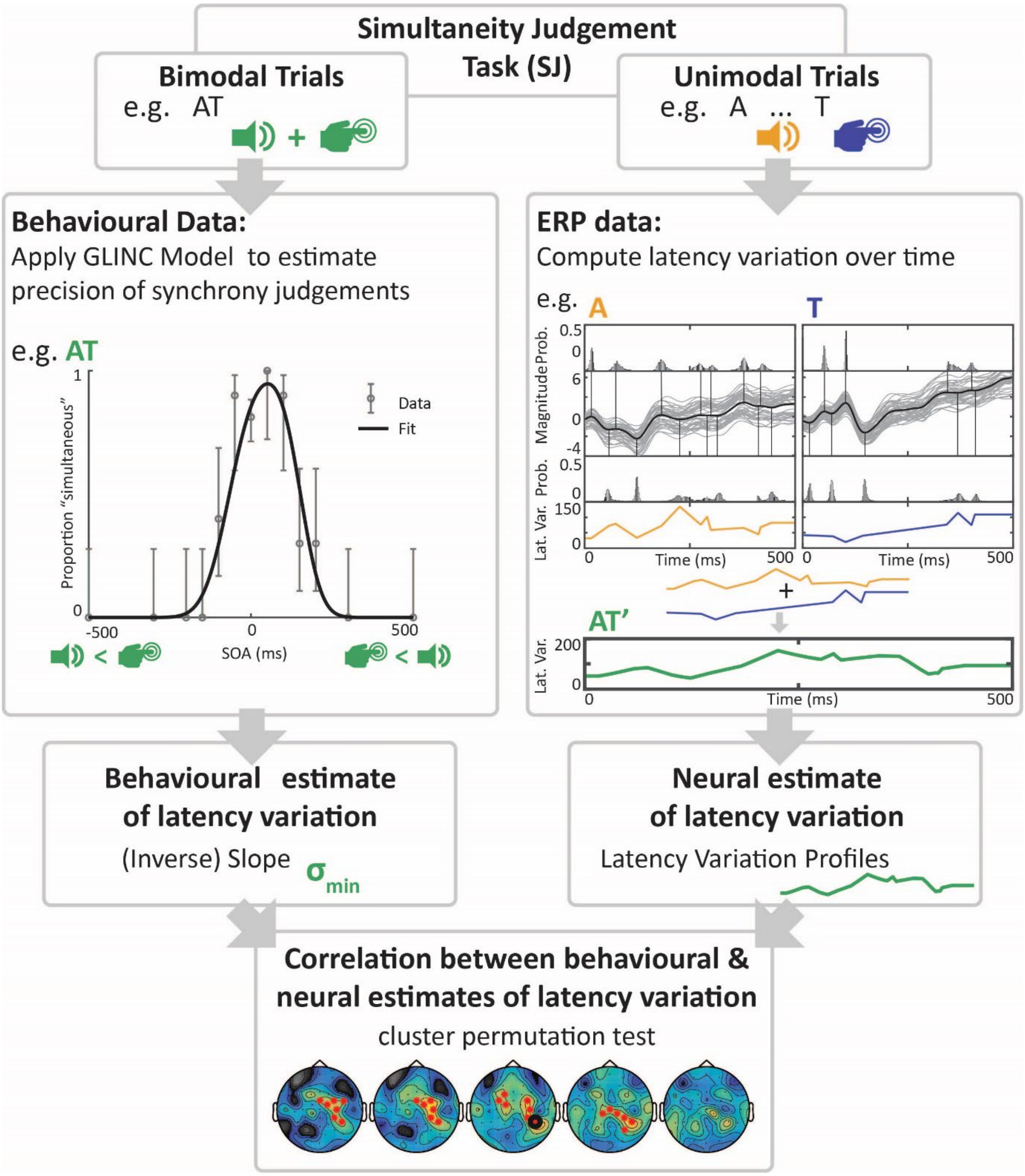

Fig. 1. Overview of the analysis workflow used to establish correlations between behavioural and neural estimates of latency variation. Top Panel: Participants completed unimodal and bimodal judgement trials. Left Panels: Bimodal-trial (here, AT) performance was estimated using the GLINC model (expanded in Figure 2) with the steeper of the two slopes from the resulting psychometric function inversely related to a behavioural estimate of latency variation (σ_{min}). Right Panels: EEG data associated with unimodal, i.e. single stimulus, trials (here, A trials and T trials) were used to compute event-related latency variation profiles (see Figure 3 for further details). These profiles were then combined to create an AT' profile, representing a neural estimate of latency variation. Bottom Panel: The behavioural and neural estimates of latency variation were then tested for correlation using cluster permutation tests (see Figure 5).

2. Materials & Methods

2.1 Participants

The combination of an unknown effect size and a complex familywise correction applied across a spatiotemporally correlated neural signal (via cluster tests; see below) made *a priori* power calculations challenging. We opted to target a sample size of 50. This provides >80% power to detect an (uncorrected) correlation of 0.35 (with $p < 0.05$ under our one-tailed hypothesis). Data were successfully collected from 57 predominantly female[1] participants, but for six, SJ data were insufficient to properly constrain behavioural model parameters in one or more modality pairings (see data analysis, below) and for one, poor EEG data quality led to rejection of >50% of trials. The final (convenience) sample therefore contained 50 participants (mean age 27.6, SD 9.4) who reported normal or corrected to normal vision and hearing, and were reimbursed, either with course

[1] A loss of data regarding the gender of the final 17 participants means we cannot provide an exact proportion, but we estimate that our sample was 80% female.

credits (for undergraduate psychology students) or at a rate of £8 per hour. They provided informed consent following procedures approved by the City, University of London Psychology Department ethics committee.

2.2 Apparatus & Stimuli

The experiment was controlled by a PC running Matlab (The MathWorks, Nattick, U.S.A.) under Windows OS, utilising the Cogent toolbox (Wellcome Department of Imaging Neuroscience) and communicating with both the stimulus peripherals and a second PC hosting the EEG recording software via a pair of parallel ports. These ports were accessed via the inpoutx64.dll freeware driver (http://www.highrez.co.uk/) made accessible in Matlab via the IO64 mex file (http://apps.usd.edu/coglab/psyc770/IO64.html). EEG was recorded using a BrainAmp amplifier (BrainProducts; sampling rate: 1000 Hz; filter pass band 0.1-500 Hz) with 64 active electrodes placed equidistantly on the scalp (EasyCap, M10 Montage) and referenced to the right mastoid. Stimuli were delivered as a 10 ms on-off pulse via either a yellow LED for visual stimuli (located centrally, just beneath instructions on an LCD flat-screen monitor) or solenoid stimulators (tactors; Dancer Design, St. Helen's, U.K.) for auditory and tactile stimuli. The tactile tactor was pinched gently between left forefinger and thumb. The auditory tactor struck a metal surface (a badge) pinned to the participant near their left ear in order to produce a sharp click. Throughout the experiment, a white-noise machine (Wellcare model SC1752) masked the subtle sounds associated with tactile stimuli.

2.3 Design & Procedure

Following EEG preparation, participants sat comfortably in a dark, electromagnetically shielded room to complete the experiment (which took around 90 minutes). Each trial of the experiment contained either one or two events. Events could be central LED flashes, taps to the left hand, or left-lateralised audible clicks. Initially, participants received 35 practice trials, in which they used their right hand to judge stimuli “not simultaneous” or “simultaneous” using left/right keyboard arrow keys, respectively. Participants were instructed to also use the non-simultaneous response if they detected only a single stimulus. During practice they received feedback about the correctness of each response. Trials could contain a single visual (V), auditory (A) or tactile (T) stimulus (each with probability 10/90), a bimodal AV, AT or VT pair with stimulus onset asynchrony (SOA) of 0 ms (each with probability 8/90), or an asynchronous bimodal AV, AT, or VT pair with SOAs of -500, -300, 300 or 500 ms (each of these 12 possible combinations presented with probability 3/90). The practice sequence was random with replacement.

Participants next completed an experimental block of 900 trials (with breaks offered every 35 trials). They now received no feedback. Trial types remained the same as during practice except that a wider range of bimodal asynchronous trials was presented, consisting of AV, AT, or VT pairs with 12 possible SOAs (+/-500 ms, +/-300 ms, +/-200 ms, +/-150 ms, +/-100 ms, +/-50 ms) and each of these 36 possible combinations occurred with a probability of 1/90. The sequence was now random without replacement and hence yielded exactly 100 unimodal, 80 bimodal synchronous, and 120 bimodal asynchronous trials per modality or modality pairing.

Each trial began with the on-screen instruction “Look down at LED”. After one second the LED flashed five times across a 500 ms period (with a 50% duty cycle) to ensure attention was directed correctly. A random (800-1200 ms) fore-period preceded the onset of the first stimulus (or both stimuli in synchronous trials). For non-synchronous bimodal trials, the SOA determined the further delay to the second stimulus. After another 500 ms, the on-screen instruction changed to display the response options. Once the response was registered, 500 ms of feedback (on practice trials only) and/or a 500 ms blank response-stimulus interval completed the trial.

2.4 Data Analysis

2.4.1 Observer Model

A variant of the canonical model for relative timing judgements was applied to behavioural data from bimodal trials, separately for each participant, and in the AV, AT and VT pairs (200 trials per modality pairing). This “GLINC” observer model is schematised in Figure 2.

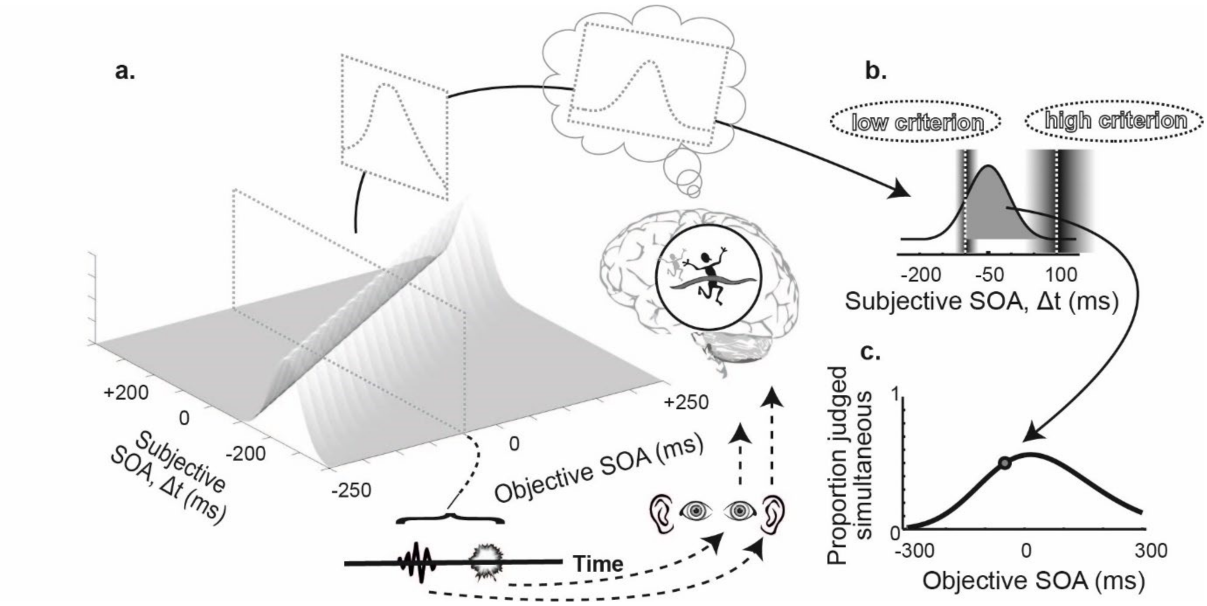

*Fig. 2. Schematic of GLINC observer model. Each signal must traverse a neural pathway to a decision centre, which receives both signals, and thus has access to their subjective difference in arrival times (Δt). (**a**) Each* stimulus onset asynchrony *(SOA) value (e.g. -50 ms) is presented many times during an experiment. Each presentation yields a noisy internal response (Δt). The relationship between objective and subjective asynchronies has unit slope and an intercept reflecting the average difference in transmission times between signals. However, the relationship is stochastic: Slicing for any given objective SOA yields the Gaussian distribution of resulting Δt values across trials, reflecting the signals’ combined latency noise. (**b**) This probability density function (PDF) is shown for a -50 ms SOA. Participants judge the trial synchronous when Δt falls between two decision criteria (solid greyed region). As the area under a PDF (to the left of any given point) is captured in the cumulative density function, the shaded region can be estimated as the difference of two cumulative Gaussians,*

one integrating all the way to the rightmost criterion, the other integrating only to the leftmost one. Variable shading around the criteria indicates additional criterion noise; each criterion is most likely to be placed where the shading is darkest, but varies across trials. ***(c)*** *Resulting psychometric function, with the point calculated in part b highlighted. Other points on the function are obtained in the same way. Precision is reflected in the slopes of the psychometric function. Under this observer model, both slopes combine latency noise and criterion noise, but the criterion noise is permitted to differ for each. Hence the steeper slope (σ_{min}) will align with the more stable of the two criteria, and thus better reflect (i.e. be more dominated by) latency noise (see main text for further details).*

Data were summarised as proportion judged simultaneous at each SOA. They were fitted with a four-parameter observer model which typically predicts a psychometric function representing the difference of two cumulative Gaussians:

(1) $$P(Simultaneous) \sim \Phi\left(\frac{SOA - C_{Low}}{\sigma_{Low}}\right) - \Phi\left(\frac{SOA - C_{High}}{\sigma_{High}}\right)$$

In Equation 1, ϕ is the standard normal cumulative distribution function. Under this model, the *c* parameters are the mean positions of two decision criteria (low and high) used to demarcate successive judgements from simultaneous judgements (i.e. the observer judges two stimuli simultaneous when the internal signals they generate arrive at a decision centre with a subjective SOA, Δt, that is both above the low criterion and below the high criterion). The associated *σ* values quantify (inversely) the slope on each side of the psychometric function. These are composite noise variables, used because they are formally identifiable in a model fit, whereas the various psychological constructs that feed into them are not. Each *σ*, when squared, represents the sum of two sources of variance. The first, the variance of Δt, is itself the sum of the (Gaussian) latency variance associated with each stimulus. This source contributes to the slope on both sides of the psychometric function (low and high). The second, the trial-by-trial (Gaussian) variance in a decision criterion, is unique on each side of the function, thus allowing the slopes to vary. Note that Equation 1 is an approximation, with simulation required in rare cases when the approximation breaks down.

Custom Matlab functions were used to find maximum-likelihood fits (assuming binomially distributed data). The Nelder-Mead simplex algorithm was used to find the best fit, with simplex searches initiated from the factorial combination of several positions per parameter (i.e. a grid search seeding a set of simplex searches). Observer models incorporated a fixed 1% keyboard error/lapse rate, to model occasional errors without increasing parametric complexity (and also simplify the calculation of log likelihood). In order to determine if participants had produced data of sufficient quality to incorporate into our main analysis, we assessed whether (for each bimodal condition) the four-parameter model provided a significantly better fit than a two-parameter cumulative Gaussian (deviance improvement, $\chi^2_{[2]}<0.01$, where deviance is -2 times the shortfall in log-likelihood relative to a saturated model). This represents the lab's standard approach to participant exclusion (Yarrow, 2018) with this null model used in place of a simpler guessing model, as it can capture both guessing, and cases where the range of stimuli is only sufficient to capture the decision boundary on one, but not both, sides of zero. For participants passing this test, we recorded their four best-fitting model parameters in each stimulus pairing, but in particular noted the smaller of the two σ values (i.e. the one associated with the steeper slope). This choice was guided by the particulars of the model – because both σ values contain the noise we are interested in (latency noise), but each overestimates it, as a results of also containing an additional nuisance source (criterion noise), the lower σ parameter should be the one less contaminated by this decision-level source.

2.4.2 EEG pre-processing

EEG data were pre-processed using custom Matlab scripts incorporating functions from EEGLAB (Delorme & Makeig, 2004). Data were initially band-pass filtered (0.1-45 Hz) before identifying bad channels (all channels were assessed via channel spectra, and electrode traces outlying from the norm or with extreme irregularities were removed). Next, data were re-referenced to an average reference, and data recorded during breaks were rejected by eye before running an independent component analysis (ICA) targeting blink components for removal. A second artefact rejection by eye was conducted to remove any remaining irregularities in the data, such as excessive muscular noise, electrode drifts and miscellaneous peaks. Finally, the missing (bad) channels were spherically interpolated from the new, clean dataset. Epochs (-200 to +800 ms relative to stimulus onset) were extracted for each unimodal (i.e. single-event) condition, with summary ERPs created following baseline correction to the mean of the first 200 ms. The artefact rejection steps left a median average of 91, 94 and 92 (range 56-99, 66-100, 58-99) unimodal trials for the auditory, visual, and tactile modalities respectively.

Note that by design our EEG analysis focussed on unimodal trials, which were included in the experiment specifically for the purpose of estimating neural latency variation. Bimodal trials were not utilised for our derived EEG measure, because they contain/conflate the brain’s response to two signals in a way that makes these responses difficult to separate, and we wanted to obtain an independent ERP for each individual modality, in order to properly equate neural and behavioural noise under the GLINC model (as described in the next section). Hence, because unisensory ERPs provide the bedrock for subsequent estimates of latency variation, we confirmed their information content via trial-by-trial decoding based on a 300 ms post-stimulus segment, using a nearest neighbour classifier with jack-knifed cross validation. Trials were classified as A, V or T based on similarities between measures of brain activity on a given trial, and average neural activation patterns elicited by each type of stimulus (individually for each participant) on training trials (all trials for that participant, bar the trial to be decoded on that iteration of the decoding process). On average, stimulus modality could be decoded correctly on 64.7% of trials (95% CI 61.9-67.3), i.e. around twice the chance expectation.

2.4.3 Event-related latency variation

In order to provide a time-varying measure of latency variation for the brain's response to isolated unimodal stimuli, we first calculated, for each participant and electrode, standard sensory ERPs, as the mean of all acceptable trials in a given condition, but with additional 20 Hz bi-directional (3^{rd} order Butterworth) low-pass filtering to minimise small oscillations and emphasize more substantial components. Within each ERP, local maxima and minima were identified out to 500 ms post stimulus, and their times recorded. Conceptually, the next step was to generate 1000 bootstrap resamples of the ERP (Efron & Tibshirani, 1994). A bootstrap resample is generated by resampling with replacement from the original sample, to create a new data set of equal size. The "with replacement" aspect of this procedure means that each resample is likely to contain some trials more than once, with some trials being entirely absent. Hence each resampled ERP was derived from a slightly different mixture of trials compared to the original ERP, and thus differed from it. For each such bootstrap ERP, we attempted to find the most sensible matches between its maxima/minima and those of the original signal, in order to build up bootstrap latency distributions for each turning point (see Figure 3). In practice, such a match is quite challenging, because a given bootstrap resample (calculated out to 600 ms to capture any delayed components) can generate more or less turning points than the original ERP, including some that are a poor match. Hence our bespoke Matlab function implemented a preliminary bootstrap (in order to identify likely time points where bootstrapping would generate spurious turning points) prior to the final bootstrap, where matching was achieved. Matching was based largely on correspondence of sign (i.e. being a maximum/minimum) and timing, but with some additional checks to try and ensure unique and sensible matches (specifically, a match was rejected where it better matched a spurious locus than an original ERP turning point, or where, despite being the closest match for a particular turning point, it was closer still to a different turning point). Where a convincing match could not be determined, none was recorded, such that the bootstrap latency distribution for any given turning point could contain fewer than 1000 values.

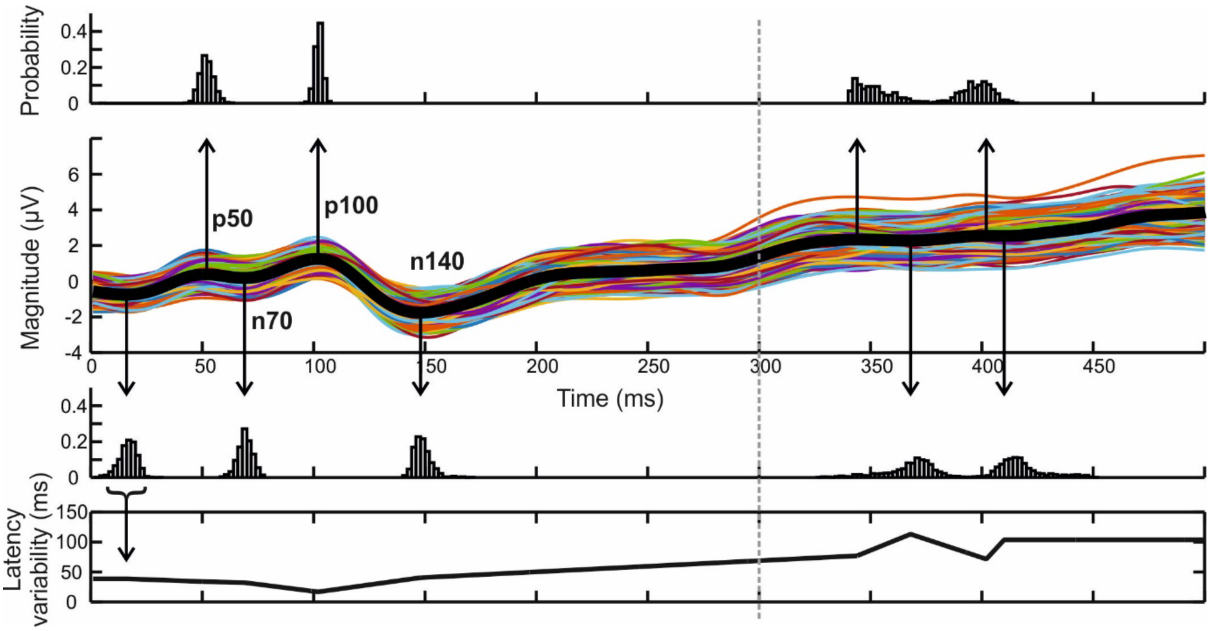

Figure 3. Process for determining the event-related latency variation profile of a given electrode. The central panel shows one illustrative participant's tactile ERP recorded by a contralateral centro-parietal sensor (bold black trace; EasyCap M10 electrode 12, selected because it contributes to a cluster emerging from our main analysis, presented in Figure 5). This ERP is presented alongside 1000 bootstrapped ERPs, derived from the same set of trials (coloured traces). Black vertical markers show the locations of turning points in the original ERP – the more prominent of which are typical for a posterior somatosensory potential evoked by a mechanical pulse – a P50, N70, P100, and N140, followed by a slow positive wave (Hämäläinen, Kekoni, Sams, Reinikainen, & Näätänen, 1990)*. For each bootstrap, turning points were determined, and an algorithm attempted to match these up with those present in the original signal, giving rise to bootstrapped latency histograms for each component (shown above/below the ERP). The width of these distributions was used to estimate latency variation at the time of each component. These values were linearly interpolated, to generate a complete event-related variation profile (bottom panel). The dashed vertical line at 300 ms indicates the upper time limit for signals exported to subsequent correlation analyses.*

The standard deviation of each resulting bootstrap distribution (i.e. the bootstrap standard error) was multiplied by the square root of the number of trials contributing to a condition in order to recover a value approximating the standard deviation of the latency of each ERP component across trials. In a final step, these scores (one for each component, and representing neural-latency variation at the time of that component) were linearly interpolated, so as to give a time-varying measure that respected the sampling rate of the original EEG signal. These event-related variation profiles were derived by interpolating between a median average of 11 turning points (minimum/maximum of 4 and 25 respectively across all electrodes/modalities/participants). We confirmed through simulation (using a tri-peaked difference of Gaussian to mimic underlying signal, and adding varying levels of latency, amplitude, and general 1/f noise) that our method generates estimates of latency noise that increase monotonically (albeit non-linearly) with simulated latency noise.

2.4.4 Cluster-based correlations

Visual, auditory, and tactile event-related latency variation profiles were combined in order to correlate them with behavioural measures that should (under the canonical model) also reflect latency variation. Because behavioural measures represent the standard deviation of Δt, which is formed from a combination of latency variation within each contributing modality, we created AV, AT and VT event-related latency variation profiles by squaring, summing, and square-rooting the two relevant profiles in each case (at each electrode). Correlating the resulting brain-based bimodal variation profiles (comprising 300 time points x 60 electrodes for each participant) with the relevant behavioural measure (e.g. σ_{min} from the relevant psychometric function; one per participant) presents a substantial multiple comparison problem, which we addressed via cluster-based permutation testing (Blair & Karniski, 1993; Groppe, Urbach, & Kutas, 2011) using functions from the

Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) (http://fieldtriptoolbox.org) to control familywise error (for each modality pairing) at a one-tailed alpha of 0.05, reflecting our *a priori* directional hypothesis. Tests were based on 9,999 permutations, with a minimum of two neighbours forming a cluster and a cluster threshold set to two-tailed $p < 0.05$.

3. Results

3.1 People differ in their ability to perform synchrony judgements

Fits of the GLINC model (schematised in Figure 2) to AV synchrony judgements are shown for two representative participants in Figure 4a. As expected, judgements of synchrony were more likely when events were physically synchronous, or separated by only a brief interval. However, slopes on either side of the psychometric function suggest the presence of judgement noise, with different decisions reached on repeated trials with the same physical stimulation. GLINC ascribes this noise to a combination of neural-latency variation across trials, and to criterion (i.e. decision-level) noise. For example, the less-precise observer illustrated in Figure 4a (on the right of the panel) has a steeper slope (and thus a lower σ parameter) for the low than for the high criterion. The interpretation of this based on GLINC would be that this observer is better able to maintain a consistent internal demarcation between auditory-leading and synchronous AV stimuli, compared to the demarcation of synchronous from visual-leading stimuli. This pattern has been observed before (e.g. Yarrow et al., 2011) and indeed was found for the majority of participants in the current sample (33/50, binomial $p = 0.033$).[2]

[2] A similar tendency was evident in AT data, with 37/50 participants having less noise at the low criterion associated with the categorisation of auditory-leading AT stimuli. No such tendency emerged for VT data (23/50 participants with less noise at the criterion associated with visual-leading VT stimuli).

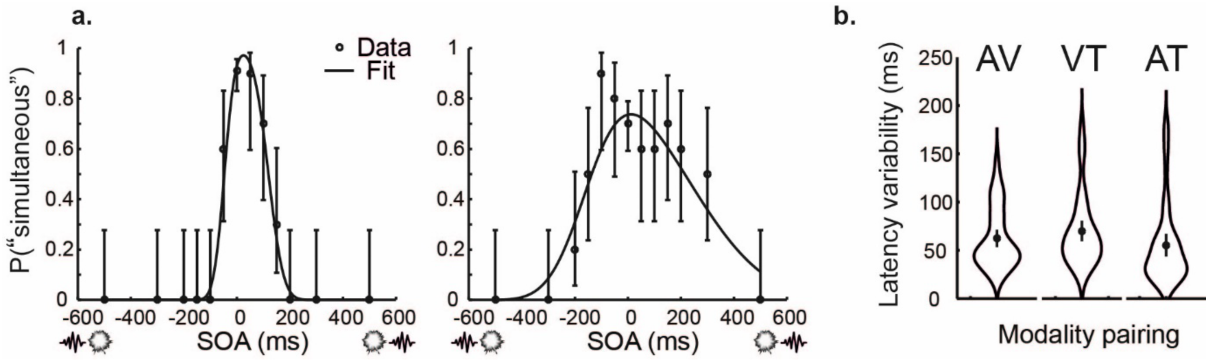

Figure 4. Behavioural data. Error bars show 95% confidence intervals. ***(a)*** *Example audio-visual SJ data for two participants (one relatively precise, one relatively imprecise).* ***(b)*** *Mean latency noise (σ_{min}) in each sensory combination from the complete sample of participants. Surrounding shape widths denote kernel probability density estimates. AV = audio-visual, VT = visuo-tactile, AT = audio-tactile.*

Under the GLINC model, the steeper of the two slopes (σ_{min}) will better isolate neural-latency variation, so this is used here to estimate this quantity (see Figure 2, especially legend to part c, and section 2.4.1). These behavioural estimates of latency noise are illustrated for the full sample of participants, and all three simultaneity-judgement (SJ) tasks, in Figure 4b. We also conducted split-half correlations on behavioural estimates of latency noise, with data split into odd and even-numbered subsets of trials for each stimulus onset asynchrony (SOA) category before fitting. These tests indicated reliable individual differences in behavioural noise for all three SJ tasks (r values of 0.534, 0.335 and 0.785; p values of <0.001, =0.0173, and <0.001; for AV, VT and AT SJ tasks respectively). This establishes that it is reasonable for us to investigate what neural processes might explain our reliable individual differences in the precision of behavioural timing judgements.

3.2 Behavioural differences are associated with changes in neural-latency variation

The canonical model predicts correlations between behavioural and neural estimates of latency variation. For our neural estimates, we used bootstrapping of ERPs recorded in response to isolated stimuli from each modality to estimate latency noise at each time point (out to 300 ms post stimulus) and each electrode (see Figure 3). Having estimated visual, auditory, and tactile event-related latency variation profiles at each time point and electrode, we combined estimates from each pair of modalities (see Figure 1 and sections 2.4.3 and 2.4.4) to form three bimodal neural-variation profiles. Like our behavioural measures, these composite neural-variation profiles provided evidence for reliable individual differences across participants (with mean split-half r values of 0.616, 0.611 and 0.627 for AV, VT and AT SJ tasks respectively, and r significant following permutation r_{max} correction (Blair & Karniski, 1993) at 82% of electrodes and time points). Given robust individual differences in both behavioural and neural estimates of inter-trial latency variation, we proceeded to perform correlations between them as a direct test of our hypothesis. Each composite neural-variation profile was correlated with the corresponding behavioural measure that should (under the canonical model) reflect the exact same latency variation (e.g. audio and tactile profiles were combined for correlation with the behavioural measure AT σ_{min}, see Figure 1). To achieve a non-parametric whole-brain control of familywise error, we used cluster permutation tests. Results are illustrated in Figure 5.

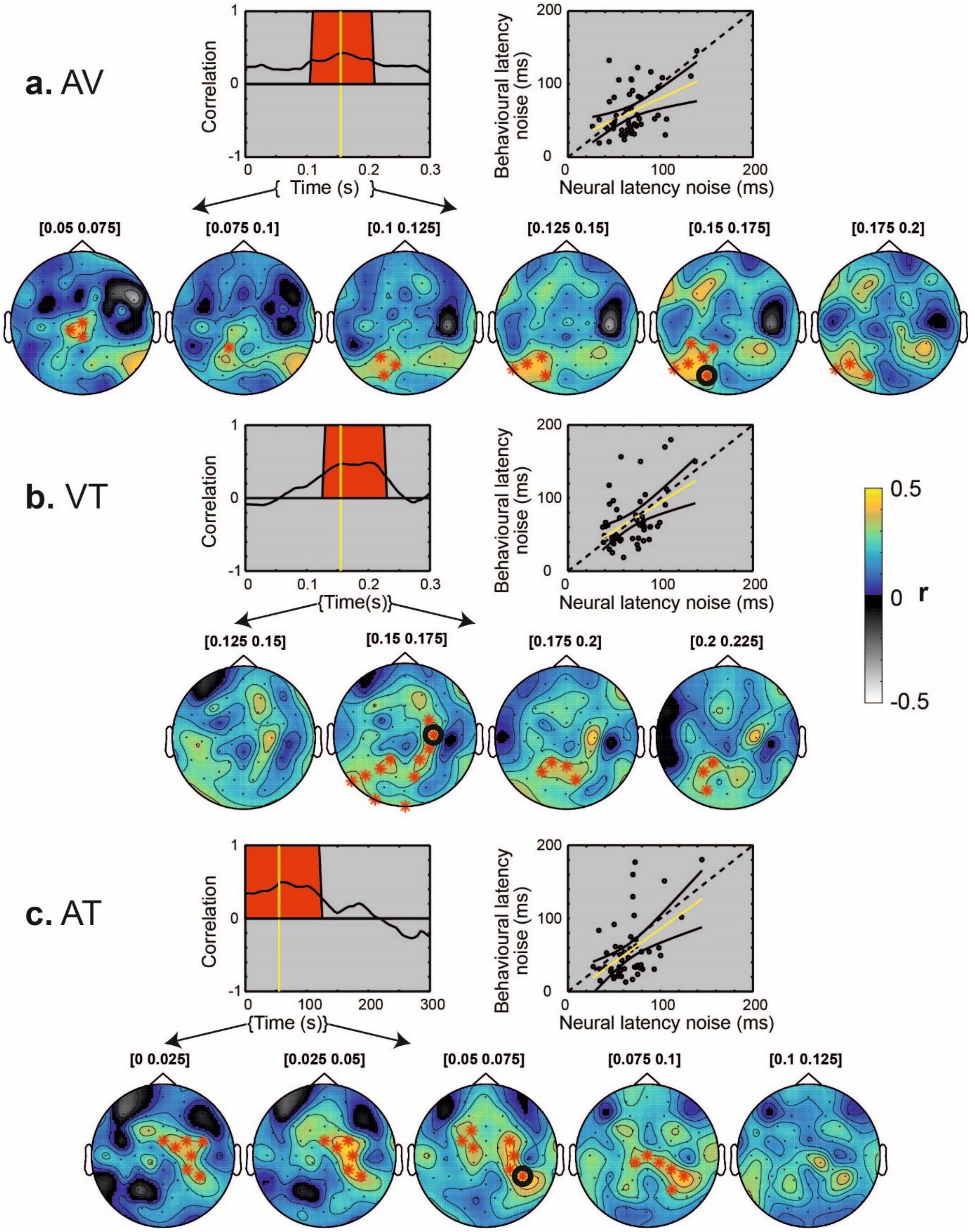

Figure 5. Summary of results from cluster-permutation tests of correlations between behavioural and neural estimates of latency variation, for ***(a)*** *audio-visual,* ***(b)*** *visuo-tactile and* ***(c)*** *audio-tactile synchrony judgement tasks. Within each panel, the lower row contains topoplots of average correlations, including all 25 ms epochs where a cluster remains significant throughout. Electrodes*

contributing continuously to the significant cluster are highlighted by red asterisks. One such electrode is further highlighted (black ring) for detailed illustration in the top row. Here, to the left, the correlation is plotted across time at this electrode. Correlations exceeding cluster thresholds are highlighted (red background region). One time point (yellow vertical line) is picked out for illustration in a scatterplot, shown on the right. Here, the line of equality is shown in dashed black, and the line of best fit in yellow (with 95% CIs in solid black).

Figure 5a shows correlations between audio-visual SJ precision and neural-latency noise, estimated from isolated audio and visual ERPs. The cluster permutation test revealed a single significant cluster ($p = 0.0245$). Topoplots illustrate the strength of correlation across the scalp at all epochs spanned by this cluster. This cluster seems to emerge at central electrodes (consistent with electrodes where strong auditory ERPs are observed) around 50 ms after stimulus onset, then spreads to occipital electrodes (suggestive of visual system involvement), and persists until around 200 ms post stimulation.[3]

Figure 5b shows correlations in the visuo-tactile case. The permutation test again revealed a single significant cluster ($p = 0.0328$), this time emerging at right-central electrodes from around 150 ms post stimulation, spreading to occipital electrodes, before disappearing around 225 ms post stimulation.

[3] We verified that the portions of contributing latency variation profiles which were coincident with this cluster contained many values estimated directly from ERP turning points (as opposed to being based entirely on interpolated values falling between ERP turning points). For the AV cluster, which spanned 17 channels, each for a duration ranging from 7 to 116 ms, across participants a median average 18 (minimum 10) turning points intersected coincident portions of visual variation profiles, while a median average 17 (minimum 12) turning points intersected coincident portions of auditory variation profiles. We went on to make similar calculations for the VT and AT clusters that are described next in the main text. For the VT cluster, which spanned 22 channels, each for a duration ranging from 1 to 80 ms, across participants a median average 17 (minimum 8) turning points intersected coincident portions of visual variation profiles, while a median average 14 (minimum 7) turning points intersected coincident portions of tactile variation profiles. For the AT cluster, which spanned 28 channels, each for a duration ranging from 1 to 107 ms, across participants a median average 26 (minimum 14) turning points intersected coincident portions of auditory variation profiles, while a median average 29 (minimum 20) turning points intersected coincident portions of tactile variation profiles.

In Figure 5c, correlations involving audio-tactile timing precision are plotted, once again highlighting a single significant cluster (p = 0.0242). In line with the contributing left-lateralised unisensory signals, this is largely right lateralised, and spreads from central to parietal electrodes. The cluster lasts until around 100 ms, and emerges very early at 0 ms, probably as an artefact of our interpolation process used to estimate latency variation, which assigns the noise estimate from the first ERP turning point to all earlier time points (Figure 3).

Based on a model-derived hypothesis, we have so far correlated neural-latency variability estimated from contributing unisensory stimulations with behavioural estimates of timing precision from bimodal stimulations (e.g. A and V variation profiles were combined and then correlated with estimates of the precision of AV synchrony judgements).[4] In principle, one might expect no such correlation between behavioural estimates and *non*-contributing unisensory signals (e.g. between AV behaviour and *tactile* ERPs). However, it is also plausible that some peoples' brains have a generally high temporal fidelity, and others a generally poor temporal fidelity, sharing this property across all sensory modalities, in which case correlations would still emerge. Testing for these relationships, we found no significant clusters for two of three tests (VT-A: smallest p = 0.0949; AT-V: no positive clusters to assess), but found a significant cluster for the final such test (AV-T; p = 0.0074) with an early (0-125 ms) occipito-parietal locus.

3.3 Average neural latency variability is higher for visual compared to tactile and auditory stimuli

Temporal acuity may vary between the senses. Our behavioural data are suggestive of greater variability for synchrony judgements involving visual stimuli (see Figure 4b – variability trend

[4] Our analysis followed, in a principled manner, from the model we have assumed as the basis for generating psychometric functions (i.e. the GLINC model). For this reason, we used the steeper slope of the psychometric function to estimate behavioural noise (see methods). However, in response to an anonymous reviewer request, we re-ran our three correlation analyses using the average of the two slopes to estimate behavioural noise instead. Headline results were very similar, with a single significant cluster emerging for all three modality pairs (AV p = 0.0461; VT p = 0.0299; AT, p = 0.0332).

suggests AV > VT > AT). Repeated-measures permutation tests with a tmax familywise correction for the three possible pairwise contrasts indicated that of these, just the outer contrast (AV>AT) was significant ($p = 0.013$). We sought a similar pattern in our neural data, calculating a crude measure of neural variability in each modality by averaging latency variability profiles across the full 300 ms x 60 electrodes included in our main correlation analysis. This measure showed a V > T > A pattern (with variability of 60, 55 and 54 ms respectively) that is somewhat consistent with our behavioural result. Tmax corrected permutation tests indicated significantly greater neural variability in response to visual stimuli, compared to both tactile and auditory stimuli ($p < 0.001$).

4. Discussion

The canonical model of multisensory timing perception formalises the brain-time account, i.e. the idea that the timing of particular operations in the human brain determines the perceived timing of sensory events.[5] Because the canonical model is a formal (if simple) process model, it makes clear predictions about the sources of noise that limit the precision of timing judgements. Specifically, the fidelity of timing judgements should be determined, to a substantial degree, by inter-trial differences in the speed at which contributing signals propagate through the central nervous system (measured as latency variation). Here, we tested this idea using synchrony judgements, completed alongside EEG recordings. Because it would be very difficult to estimate the latency noise affecting individual

[5] The brain-time account is usually invoked in discussions of event ordering. Event ordering can be seen as a prequel to other forms of time perception such as interval timing, although no clear consensus exists regarding the degree of neurocognitive interrelation between different forms of time perception (we use *timing* perception here to focus our discussion specifically on issues of relative order). In general, the brain-time account should probably be considered agnostic regarding the necessity of forming higher-order representations about time, such as of intervals, but specific formal accounts derived from it are required to be more specific. For example, our GLINC model implies that arrival order gives rise to a representation of intervening time (to which criteria can be applied to form judgements). Many formal models of interval timing go a step further, by acknowledging neural latency variability as a constant source of noise for interval judgements, but one that is typically dwarfed by interval-dependent “scalar” noise (Wearden & Lejeune, 2008). Such scalar noise is generally omitted in accounts of relative timing, because they focus on such tiny intervals.

trials (either behaviourally from SJs, or in the brain from the corresponding single-trial bimodal ERPs, somehow decomposed into their unimodal constituents) we have not attempted any within-participant, trial-by-trial correlations of neural and behavioural noise. Rather, we used responses across multiple trials to provide a model-based estimate of behavioural noise for each participant, and have correlated these with bootstrap-based estimates of neural noise derived from EEG. For all three modality pairs (AV, VT, and AT), we observed the predicted positive relationship between individual-difference measures, supporting a key assumption of the canonical model.

Inter-trial latency variation is likely to have a variety of physiological causes. Even operations as seemingly deterministic as propagations of action potentials show latency variance, at least for thin, unmyelinated axons (Faisal & Laughlin, 2007). Such latency noise is likely exaggerated greatly by stochasticity in the thresholding that occurs at synapses (e.g. Paraskevopouloua, Coon, Brunner, Miller & Schalk, 2021). The canonical model embraces such noise. However, several promising models of relative timing do not explicitly incorporate sensory latency noise (Parise & Ernst, 2016; Roach et al., 2011). Our data suggest that such noise may be an important feature that should be incorporated in modelling of time perception.

We estimated inter-trial latency variation based on a bootstrapping approach. Our overall approach is novel, although bootstrapping itself is well established, having become a textbook method for estimating standard errors. There are, of course, other ways to estimate neural latency noise from EEG data. Possibilities include attempting to clean the data sufficiently to enable estimations of ERP latencies on individual trials, which would also provide an estimate of latency variance across trials. However, the noise levels associated with EEG data makes this approach challenging. Another approach would be to use the variance of the EEG signal across trials at each time point (cf. Arazi, Yeshurun & Dinstein, 2019). Finally, for a non-time-varying estimate, one might select a temporal window of interest and compute cross correlations between all possible pairs of trials within that time window. The time delay that maximises each such correlation could then be calculated, with a summary statistic of these measures taken as an estimate of implied delays. We do not claim that our particular approach is a gold standard, but we do think it has some important strengths relative to these other possibilities. For example, the variance of EEG signals across trials, while straightforward to compute, would reflect both variability in the timing of ERP components, and variability in their magnitude, with each source contributing to an unknown extent. By contrast, our bootstrapping measure specifically targets latency noise (while still tracking changes in variability across time).

It is worth acknowledging that any method that derives an aggregate measure of (bimodal) neural noise by combining estimates based on unimodal trials is blind to possible early interactions between sensory channels. Such interactions might affect latency noise, or even act as an entirely separate cue for the detection of synchrony (Arnold, Hohaia, & Yarrow, 2020). Ignoring this putative issue is true to the assumptions of the canonical model, which Sternberg and Knoll (1973) explicitly labelled the “independent channels” model on this account, but the existence/importance of early interactions between bimodal signals is of course an empirical question that might be addressed in future work. For this investigation, we chose a particular variant of the canonical model (GLINC) to fit behavioural data and generate predictions – one we have outlined and used in previous publications (Yarrow et al., 2011; Yarrow, Sverdrup-Stueland, Roseboom, & Arnold, 2013; Yarrow et al., 2015; Yarrow, Martin, Di Costa, Solomon, & Arnold, 2016; Yarrow, 2018). Other variants exist, with some important differences (García-Pérez & Alcalá-Quintana, 2012a; García-Pérez & Alcalá-Quintana, 2012b; Sternberg & Knoll, 1973; Ulrich, 1987), but all assume latency noise is reflected in the slope of psychometric functions that describe subjective timing, and so all variants derive some support from our findings. We invite other authors to use our publicly available data (Yarrow, Kohl, Arnold & Rowe, 2021) to further test the predictions of different models.

We recognise that our focus on noise in sensory processes invites comparison with Bayesian models (e.g. Knill & Pouget, 2004), which have become popular when modelling various kinds of time perception (e.g. Jazayeri & Shadlen, 2010) including judgements of relative time (Ley, Haggard & Yarrow 2009; Miyazaki, Yamamoto, Uchida & Kitazawa, 2006; Roseboom, 2019). GLINC does not incorporate Bayesian information-processing stages, such as the integration of a current sensory estimate with a prior derived from past experience, but the model architecture could be elaborated to incorporate this. Bayesian model predictions are generally tested by estimating noise from behaviour, and such tests might usefully be supplemented by approaches like ours, which additionally estimate noise from brain recordings.

Our spatiotemporal illustrations should be considered, at best, suggestive. Caveats limit any inference regarding the spatial origins of neural signals from EEG scalp topography, and cluster significance does not imply significance for all contributing spatiotemporal points. Furthermore, ours are not classic contrasts, but rather correlations across participants. While some ERP components likely represent processing at regions critical for synchrony judgements, any ERP component correlated with these components would also emerge. For example, left-lateralised components evoked by a central visual stimulus might have a temporal fidelity limited by similar physiologically-imposed noise compared to right-lateralised components. The same applies to components preceding and following a critical component in time. Moreover, our method *sums* variance estimated from two contributing unimodal sensory signals, using spatiotemporal correspondence at the scalp, and it is not clear that this summation should accurately index a temporal comparator of different sensory modalities. Given these considerations, we feel the spatiotemporal loci of our correlations are surprisingly well matched to expectations, being most clearly right lateralised when both stimuli originated from the left (i.e. for audio-tactile stimuli), and broadly in line with regions of cortex relevant for each sensory pairing, and with expectations regarding processing latencies for different sensory modalities (e.g. central electrodes consistent with audio activations emerged early, right-central electrodes consistent with tactile activations slightly later, and occipital activations consistent with visual activation emerged last).

Although we observed the predicted correlations, they were modest, accounting at best for around 25% of the variance in behavioural performance. Several factors may be relevant. First, correlations are limited by the reliability of contributing measures. These reliabilities are unknown in the absence of a retest session, but split-half analyses of both behavioural and neural data generated r values of around 0.5, suggesting test-retest correlations would likely fall well short of a perfect correlation. Hence, we have imperfect but moderately reliable measures of behaviour and neural activity, reflecting practical trade-offs when determining the length of experimental sessions.

Second, only austere versions of the canonical model (e.g. Gibbon & Rutschmann, 1969) assume trial-by-trial latency variation is the *only* source of noise affecting timing judgements. Any additional sources of noise would suppress the correlations we have sought here. The GLINC model we have used, for instance, assumes criterion noise, i.e. an inability to make the same decisions about inputs, even if sensory coding and experiences are held constant across trials (Ulrich, 1987). Other variants assume participants cannot resolve relative timing when two signals arrive within some limited temporal window (Sternberg & Knoll, 1973). This refractory “moment” might be triggered by the arrival of the first stimulus (e.g. García-Pérez & Alcalá-Quintana, 2012a; Venables, 1960), but in this case it would not influence the *slope* of the SJ function, and thus should not act as an additional source of noise under our analysis. Indeed, this consideration informed our choice of task – we opted not to use temporal order judgements (TOJs), because TOJs seem more profoundly affected by additional sources of noise relative to SJs (Yarrow et al., 2016) perhaps including a flattening of the slope of the psychometric function resulting from something formally akin to a triggered moment (García-Pérez & Alcalá-Quintana, 2012a). We have previously concluded (via a very different kind of analysis) that variation in evoked responses does not have an easily detectable role in AV TOJ performance (Arnold, Hohaia, & Yarrow, 2020).

Another variant of the canonical model proposes a moving (i.e. non stimulus-locked) perceptual moment (Stroud, 1956). This has been linked to the alpha rhythm, for example when explaining individual differences in the double-flash illusion (Cecere, Rees, & Romei, 2015) and changes in visual-visual TOJ sensitivity across an entrained alpha cycle (Chota, Marque, & VanRullen, 2021). Perhaps of greatest relevance here, individual alpha frequencies have also been linked with the width of synchrony functions for visuo-tactile SJs (Migliorati et al., 2019), albeit without recourse to a formal observer model. A moving moment would increase noise in SJs much like criterion noise/variance under the GLINC model, because the time period within which the ordering of stimuli could not be resolved would vary from trial to trial, depending on where in the ongoing cycle the first stimulus happened to arrive. Hence, the modest degree of correlation in our data may provide some support for both criterion-noise and moving-moment variants of the canonical model.

In supporting the canonical model of relative timing, our data also support the broader brain-time account which it formalises. We recognise that our approach to testing the brain-time account is somewhat indirect, compared to the more common tactic of introducing experimental manipulations designed to vary mean transmission times while measuring corresponding changes in average timing perception and/or neural latencies (e.g. Fraisse, 1980; McDonald et al., 2005; Vibell et al., 2007). However, we believe our method makes a novel contribution to the wider debate. Of course, there are other findings that cast doubt on the brain-time account as a complete and sufficient theory. For example, the existence of contextual influences on perceived event timing (e.g. Bechlivanidis & Lagnado, 2016; Miyazaki et al., 2006; Yarrow, Whiteley, Haggard & Rothwell, 2006) suggests a softening of the brain-time account, to admit that some degree of (likely post-hoc) biasing or rationalisation can occur. However, as we have argued elsewhere (Yarrow & Arnold 2016), brain time remains viable as the fundamental basis for perceived temporal order, even if it is unlikely to be a complete account under all circumstances.

Other results may appear challenging to the brain-time account, but often bear closer examination. For example, the canonical model implies that neural latencies inform the point of subjective simultaneity (PSS), such that relative latency is one reasonable explanation on offer for non-zero or altered PSS values (e.g. Freeman et al., 2013; Grabot & van Wassenhove, 2017). However, most variants of this model also provide equally valid alternative explanations (e.g. differences in the positioning of decision criteria) such that PSS results that appear to refute the brain-time account (e.g. apparent dissociations between tasks; Love, Petrini, Chen & Pollick, 2013) may be less challenging when viewed through the lens of a formal model (Yarrow et al., 2016). Indeed, many such “dissociations” seem to result from comparing measures believed to be comparable on some intuitive basis (e.g. the width of an SJ function and the just noticeable difference derived from a TOJ function) but for which formal modelling reveals no such equivalence.

Returning to the current results: We have already noted limitations stemming from our correlational approach, and urge due caution when interpreting our findings. For example, the correlations we observe may be driven by an unmeasured third variable with putative effects on both our neural and behavioural measures, such as levels of arousal, focussed attention and so forth. We tried to make our measure of neural latency variability as specific as possible, but of course it is likely that this measure is itself related to more general forms of neural variability. However, although unmeasured variables might underlie the correlations observed here, we sought these correlations only because they are implied by the causal steps of a formal process model. This makes a causal attribution at least plausible.

As a final issue, we note that we have incorporated three tests of our one-tailed hypothesis into our design. Although each was subjected to appropriate statistical control of familywise alpha levels, one might argue that the experiment-wise alpha is higher. However, there is considerable overlap between measures informing the three cluster tests, so their independence is unclear. Furthermore, the average p value across the three tests still implies significance. Hence, while our inference is less robust than if we had *independently* verified the hypothesis in three separate data sets, we consider the degree of protection against false positives to be reasonable. We note, however, that a pilot for this project, with only AV stimuli and a less fully developed analysis, failed to detect the correlations we report here (Keane, 2019). As such, our findings would certainly bear replication.

To summarise: Our data suggest that better performers on cross-modal SJ tasks exhibit lower levels of neural-latency noise compared to worse performers, exactly as predicted by the canonical model of relative time perception. We therefore argue that viable models of relative timing should incorporate latency variability in neural transmission times as an explicit feature of human time perception.

Acknowledgements

We thank Lorenzo Guaita for assistance with data collection.

Author contributions

KY and DHA conceived and designed the experiment. CK, PR, RKB and TS collected the data. KY and CK performed the analysis. KY drafted the manuscript, which all authors edited and approved.

Data Availability

The datasets analysed during the current study are available in the City University of London figshare repository [https://doi.org/10.25383/city.11843274].

Competing Interests

The authors declare no competing interests.

References

Arazi, A., Yeshurun, Y., & Dinstein, I. (2019). Neural Variability Is Quenched by Attention. *Journal of Neuroscience*, *39* (30) 5975-5985.

Arnold, D. H., Hohaia, W., & Yarrow, K. (2020). Neural correlates of subjective timing precision and confidence. *Scientific Reports, 10*, 3098.

Bechlivanidis, C., & Lagnado, D. A. (2016). Time reordered: Causal perception guides the interpretation of temporal order. *Cognition, 146*, 58-66.

Blair, R. C., & Karniski, W. (1993). An alternative method for significance testing of waveform difference potentials. *Psychophysiology, 30*(5), 518-524.

Burr, D., Silva, O., Cicchini, G. M., Banks, M. S., & Morrone, M. C. (2009). Temporal mechanisms of multimodal binding. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 276*(1663), 1761-1769.

Cecere, R., Rees, G., & Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception. *Current Biology, 25*(2), 231-235.

Chota, S., Marque, P., VanRullen, R. (2021). Occipital alpha-TMS causally modulates temporal order judgements: Evidence for discrete temporal windows in vision. *Neuroimage, 237*, 118173.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21.

Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences., 15*(2), 183-247.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap.* CRC press.

Faisal, A. A., & Laughlin, S. B. (2007). Stochastic simulations on the reliability of action potential propagation in thin axons. *PLoS Computational Biology, 3*(5), e79.

Fraisse, P. (1980). Les synchronisations sensori-motrices aux rythmes [the sensorimotor synchronization of rhythms]. In J. Requin (Ed.), *Anticipation et comportement* (pp. 233-257). Paris: Centre National.

Freeman, E. D., Ipser, A., Palmbaha, A., Paunoiu, D., Brown, P., Lambert, C., ... & Driver, J. (2013). Sight and sound out of synch: Fragmentation and renormalisation of audiovisual integration and subjective timing. *Cortex, 49*(10), 2875-2887.

Fujisaki, W., & Nishida, S. (2010). A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities. *Proceedings of the Royal Society B: Biological Sciences, 277*(1692), 2281-2290.

García-Pérez, M. A., & Alcalá-Quintana, R. (2012a). Response errors explain the failure of independent-channels models of perception of temporal order. *Frontiers in Psychology, 3*, 94. doi:10.3389/fpsyg.2012.00094

García-Pérez, M. A., & Alcalá-Quintana, R. (2012b). On the discrepant results in synchrony judgment and temporal-order judgment tasks: A quantitative model. *Psychonomic Bulletin & Review, 19*(5), 820-846.

Gibbon, J., & Rutschmann, R. (1969). Temporal order judgement and reaction time. *Science, 165*(891), 413-415.

Grabot, L., & van Wassenhove, V. (2017). Time order as psychological bias. P*sychological science, 28*(5), 670-678.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology, 48*(12), 1711-1725.

Hämäläinen, H., Kekoni, J., Sams, M., Reinikainen, K., & Näätänen, R. (1990). Human somatosensory evoked potentials to mechanical pulses and vibration: Contributions of SI and SII somatosensory cortices to P50 and P100 components. *Electroencephalography and Clinical Neurophysiology, 75*(1-2), 13-21.

Holmes, N. P., & Spence, C. (2005). Multisensory integration: Space, time and superadditivity. *Current Biology, 15*(18), R762-R764.

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature neuroscience, 13*(8), 1020-1026.

Keane, B. (2019). *Neural correlates of human time perception.* Unpublished doctoral dissertation, University of Queensland, Brisbane.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences, 27*(12), 712-719.

Ley, I., Haggard, P., & Yarrow, K. (2009). Optimal integration of auditory and vibrotactile information for judgments of temporal order. *Journal of Experimental Psychology: Human Perception and Performance, 35*(4), 1005-1019.

Love, S. A., Petrini, K., Cheng, A., & Pollick, F. E. (2013). A psychophysical investigation of differences between synchrony and temporal order judgments. *PloS one, 8*(1), e54798.

McDonald, J. J., Teder-Salejarvi, W. A., Di Russo, F., & Hillyard, S. A. (2005). Neural basis of auditory-induced shifts in visual time-order perception. *Nature Neuroscience, 8*(9), 1197-1202.

Michotte, A. (1954). La perception de la causalité. Publications Universitaires De Louv.

Migliorati, D., Zappasodi, F., Perrucci, M. G., Donno, B., Northoff, G., Romei, V., & Costantini, M. (2019). Individual alpha frequency predicts perceived visuotactile simultaneity. *Journal of Cognitive Neuroscience, 32*(1), 1-11.

Miyazaki, M., Yamamoto, S., Uchida, S., & Kitazawa, S. (2006). Bayesian calibration of simultaneity in tactile temporal order judgment. *Nature neuroscience, 9*(7), 875-877.

Moutoussis, K., & Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 264*(1380), 393-399.

Nishida, S., & Johnston, A. (2002). Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Current Biology, 12*(5), 359-368.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 2011:156869.

Paillard, J. (1949). Quelques données psychophysiologiques relatives au déclenchement de la commande motrice [some psychophysiological data relating to the triggering of motor commands]. *L'Année Psychologique, 48*, 28-47.

Paraskevopouloua, S. E., Coon, W. G., Brunner, P., Miller, K. J., Schalk, G. (2021). Within-subject reaction time variability: Role of cortical networks and underlying neurophysiological mechanisms. *Neuroimage, 237*, 118127.

Parise, C. V., & Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration. *Nature Communications, 7*, 11543.

Roach, N. W., Heron, J., Whitaker, D., & McGraw, P. V. (2011). Asynchrony adaptation reveals neural population code for audio-visual timing. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 278*(1710), 1314-1322. doi:10.1098/rspb.2010.1737

Roseboom, W. (2019). Serial dependence in timing perception. *Journal of experimental psychology: human perception and performance, 45*(1), 100.

Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 629-686). London: Academic Press.

Stroud, J. M. (1956). The fine structure of psychological time. In H. Quastler (Ed.), *Information theory in psychology* (pp. 174-205). Glencoe, IL: Free Press.

Ulrich, R. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. *Perception and Psychophysics., 42*(3), 224-239.

Venables, P. H. (1960). Periodicity in reaction time. *British Journal of Psychology, 51*, 37-43.

Vibell, J., Klinge, C., Zampini, M., Spence, C., & Nobre, A. C. (2007). Temporal order is coded temporally in the brain: Early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. *Journal of Cognitive Neuroscience, 19*(1), 109-120.

Wearden, J. H. and Lejeune, H. (2008). Scalar properties in human timing: conformity and violations. *Quarterly Journal of Experimental Psychology, 61*(4), 569–587.

Whitney, D., & Murakami, I. (1998). Latency difference, not spatial extrapolation. *Nature Neuroscience, 1*(8), 656-657.

Yarrow, K., Jahn, N., Durant, S., & Arnold, D. H. (2011). Shifts of criteria or neural timing? the assumptions underlying timing perception studies. *Consciousness and Cognition, 20*, 1518-1531. doi:10.1016/j.concog.2011.07.003

Yarrow, K. (2018). Collecting and interpreting judgments about perceived simultaneity: A model-fitting tutorial. In A. Vatakis, F. Balci, M. Di Luca, Á. Correa (Eds.) *Timing and time perception: Procedures, measures, & applications* (pp. 295-325) Leiden: Brill.

Yarrow, K., & Arnold, D. H. (2016). The timing of experiences: How far can we get with simple brain time models? In B. Mölder, V. Arstila & P. Øhrstrøm (Eds.), *Philosophy and psychology of time* (pp. 187-201) Springer.

[dataset] Yarrow, K., Kohl, C., Arnold, D.H., & Rowe, P. (2021). Neural latency noise places limits on human sensitivity to the timing of events. City, University of London. https://doi.org/10.25383/city.11843274.

Yarrow, K., Martin, S. E., Di Costa, S., Solomon, J. A., & Arnold, D. H. (2016). A roving dual-presentation simultaneity-judgment task to estimate the point of subjective simultaneity. *Frontiers in Psychology, 7*:416.

Yarrow, K., Minaei, S., & Arnold, D. H. (2015). A model-based comparison of three theories of audiovisual temporal recalibration. *Cognitive Psychology, 83*, 54-76.

Yarrow, K., Sverdrup-Stueland, I., Roseboom, W., & Arnold, D. H. (2013). Sensorimotor temporal recalibration within and across limbs.

Yarrow, K., Whiteley, L., Haggard, P., & Rothwell, J. C. (2006). Biases in the perceived timing of perisaccadic perceptual and motor events. *Perception & psychophysics, 68*(7), 1217-1226.