Supplementary methods

1. Background to temporal resolution task design

When a series of perceptual decisions occur in a fixed predictable order, perceptual reports are subject to strong biases, due to expectations, range effects and anchoring effects (1, 2). These well established effects are mitigated by randomising the order of presentation, making decisions less vulnerable to systematic variations in criterion. For example in the standard staircase method, the threshold is taken as the first of three consecutive trials at which an observer reports that two stimuli were felt (retrospectively identified). Throughout the experiment, over trials, the interval between stimuli gradually increases (figure 1A). The subject being tested has some awareness of the protocol design due to the instruction. Such an approach allows prior belief (e.g. based on previous trials) to influence decision behaviour about upcoming trials:

Prior belief (e.g. "Last few trials have been single, certain of this, can't be close to threshold yet") ➔ Stimulus ➔ Judgement ➔ Response ➔ Updating of belief (e.g. " I'm not sure about that trial, other trials have been single, I'll wait for next trial before considering changing my response')

Use of a forced choice randomised design minimises the use such priors, since the order of 1- and 2-stimulus cannot be predicted. It therefore provides a more accurate measure of the quality of sensory information available for that trial, minimising the confounding influences:

Stimulus ➔ Judgement ➔ Response

In order to make the correct answer unpredictable, a mixture of both 1- and 2- stimulus trials are needed. Previous studies have used a single 200µs pulse with no change in stimulus strength as catch trials and we had intended to use this as our 1- stimulus trial. However in pilot testing we found that this was easily discernible from a 2- stimulus trial, not due to the absence of a gap, but due to their subjectively weaker *intensity* by virtue of the fact that the quanta of charge delivered is half, and also because the total *duration* of the two pulses delivered during 2 stimuli trials is 400µs i.e. previous paradigms have had a difference of length of 200µs between 1- stimulus and below threshold 2 stimuli trials.

We left the parameters of two-stimuli trials unchanged compared to previous paradigms (200µs pulse width). For 1- stimulus trials, a second stimulator in parallel was therefore configured to deliver an equivalent pulse quality to below threshold 2- stimulus trials. Firstly a longer pulse was used. As a pulse width of 400µs was not possible with available electrical stimulators we used 500µs for 1- stimulus trials; a difference of length of 100µs between 1- stimulus and below threshold 2- stimulus trials. Secondly, at the start of the experiment the intensity of the electrical stimulation was titrated such that 1- stimulus and 2- stimulus separated by 1ms were indistinguishable. Individual plots for each subject are shown in supp. figure 1. At small intervals all subjects now could *not* discern a gap (first data points close to floor of function) and participants subjectively reported that the 1- stimulus trials were perceived as identical to the 1ms interval 2-stimulus trials.

Previous paradigms set stimulation intensity at 2x or 3x the perceptual threshold. In pilot data we found that in certain subjects this resulted in stimulation strength was too painful to continue (it is unlikely that stimulation strength (mAmp) and intensity perception have a linear relationship in all subjects). As we were interested in the timing qualities of stimuli rather than strength we adjusted stimuli to a level that salient but not painful for all subjects.

2. Psychophysical analysis

Data from both tasks were modelled using the cumulative Gaussian (Φ), a mathematical function of sigmoid shape:

$$y = \Phi\left(\left(\log(x) - mu\right)/sigma\right)/2 + 0.5 \qquad \text{(Equation 1)}$$

In the temporal resolution task, y is the proportion of responses on which “two stimuli” were perceived, and x is interval duration. In addition the false positive rate (FP, the proportion of trials where only one stimulus was delivered in which subjects *incorrectly* identified an interval) defined the floor of the function.

$$y = \Phi\left(\left(\log(x) - mu\right)/sigma\right)/2 + 0.5 \times (1 - FP) + FP \qquad \text{(Equation 2)}$$

The temporal resolution threshold (mu) was defined as the interval at which the probability of either answer is equal (T_{50}). The slope of the function at T_{50} is equal to the inverse of the standard deviation (1/sigma) of the response distribution. Previous studies of timing in this patient group may only probe responses towards the right of the psychometric function, i.e. when the subject is more certain that there are two stimuli, or when there are a higher proportion of 'two stimuli' responses. Therefore in order to facilitate comparison to other paradigms, we also calculated interval thresholds for T_{75} and T_{98} at which points the probability of reporting "two stimuli" was 0.75 and 0.98 respectively (figure 3A).

The psychometric function fitted the responses of all 44 participants extremely well (supp. figure 1). Akaike's Information Criterion (AIC) was used evaluate the fit of the psychometric model for each subject. This takes into account both the statistical goodness of fit (log-likelihood (LL)) and penalises for an increasing number of parameters (k) estimated to achieve that degree of fit. AIC_{model} was compared to a model of guessing (with a mean AIC of 207.9) with lower values indicating the preferred model. The AIC_{model} was 101.5 indicating that model predicted the individual participants' choices extremely well (with no difference in fit values obtained for controls and patients t(42)=-1.32, p=0.191).

$$AIC_{model} = -2(LL - k) \quad \text{(Equation 3)}$$

Modelling response behaviour in this manner is similar to the non-parametric bootstrapping method which was recently described (point of subjective equality = T50, slope = 1/standard deviation)(3). However the bootstrapping analysis was applied to an ascending staircase methodology rather than a randomised paradigm and as such our study design is significantly different.

For the interval discrimination task the psychometric function was fitted to each subset of data corresponding to each set interval (50ms, 100ms, 200ms) each containing a third of the total trials (supp. figure 2). The point of subjective equivalence (response probability equal for either answer) was used as the threshold value (I_{50}) and the slope was also calculated at this point. In the absence of bias, I_{50} = fixed interval. Slope is a measure of sensitivity: a steep slope reflecting high resolution for the discrimination of interval length. A contrast index was calculated for each trial and was defined as the difference between intervals divided by their total length, (i1=interval one, i2 = interval two):

$$contrast = \frac{i1 - i2}{i1 + i2}$$ (Equation 4)

If the contrast index was negative this meant that interval 1 was longer than interval 2. If the contrast was zero there was no difference between the set and the variable interval length.

Drift diffusion model

Response accuracy and reaction times were fitted to the drift diffusion model of evidence integration using the Diffusion Model Analysis Toolbox. For both tasks, data were divided into seven conditions according to duration of the gap between stimuli (in the temporal resolution task) or contrast (for interval discrimination). These conditions thus varied the strength of evidence favouring a response. The diffusion starting point was fixed halfway between the boundaries, indicating that no information was available about the upcoming stimulus before each trial (randomised nature of both tasks). To confirm that the information accumulation rate explained the difference between conditions, four competing models were evaluated and the model fit was evaluated by total Akaike information criteria. Model 2, in which drift rate varied across conditions but decision threshold was fixed was the optimal model for both tasks. Temporal resolution analysis: 77% of subjects were adequately fitted by the model (as defined by AIC values < 3 SD from mean). This excluded four controls and six dystonic subjects from the subsequent analysis. Interval discrimination analysis: 92% of subjects were adequately fitted by the model which excluded three controls from the subsequent analysis.

	Model detail	**Temporal Resolution mean AIC**	**Interval Discrimination mean AIC**
Model 1	Null model. All parameters fixed across conditions	1414	1142
Model 2	Drift rate free. Decision boundary fixed.	909	841
Model 3	Decision boundary free. Drift rate fixed.	1536	3255
Model 4	Both drift rate and decision boundary free across conditions.	1388	3291

References

1. Wilson TD, Houston CE, Etling KM, Brekke N. A new look at anchoring effects: basic anchoring and its antecedents. J Exp Psychol Gen. 1996 Dec;125(4):387-402.
2. Hogarth RM, Einhorn HJ. Order effects in belief updating: The belief-adjustment model. Cognitive Psychology. 1992;24(1):55.
3. Butler JS, Molloy A, Williams L, Kimmich O, Quinlivan B, O'Riordan S, et al. Non-parametric bootstrapping method for measuring the temporal discrimination threshold for movement disorders. J Neural Eng. 2015 Aug;12(4):046026.