Optimal encoding of interval timing in expert percussionists

Guido Marco Cicchini[1,2], Roberto Arrighi[1,2], Luca Cecchetti[2], Marco Giusti[2] & David C. Burr[1,2]

1. CNR Institute of Neuroscience, Via Moruzzi 1, Pisa 56100, Italy
2. Department of Psychology, University of Florence, Florence 50125, Italy

We measured temporal reproduction in expert drummers, string-musicians and non-musical subjects. While duration reproduction of the controls showed a characteristic *regression to the mean*, drummers responded veridically. This behavior is well explained by a model that combines optimally the sensory estimate for duration (more precise in drummers) with a *prior*, given by the average of the past few trials. These results highlight the efficiency and adaptability of sensori-motor mechanisms estimating temporal duration.

Jazayeri and Shadlen[1] recently reported that when subjects reproduce temporal intervals, their responses show a systematic regression towards the mean of the distribution from which the intervals are drawn (see Fig. 1). They explained their results with a performance-optimizing Bayesian model that considers the sample distribution as a *prior*, suggesting that the central nervous system incorporates knowledge of the temporal statistics of the environment in its estimate of duration.

We asked whether Jazayeri and Shadlen's results generalize to the entire population, particularly to expert percussionists whose profession requires accurate and reliable temporal production. We measured duration reproduction for 6 experienced drummers and 6 control subjects with no musical training. All were presented with brief visual stimuli of a given duration and asked to reproduce the duration by key-press (Supp. Mat. for details). The duration of each trial was drawn randomly from a uniform range of short, medium or long durations. The non-musical controls (red symbols of Fig. 1) behaved very much like Jazayeri and Shadlen's subjects, showing a strong tendency for regression towards the mean. This is evident both in the distributions of Fig 1a, showing three distinct reproduction patterns for 847 ms, depending on sample range, and on the PSE estimates of Fig. 1c which clearly regress towards the mean of each range. The performance of the drummers (blue symbols), however, was quite different: the reproduction distributions of Fig. 1b all align with each other, and the means are virtually veridical over the entire range tested, for all three temporal ranges. We also measured performance for string musicians (for whom precise meter is less imperative): their behavior was similar to controls (raw data in Sup. Mat.).

Following the convention of Jazayeri and Shadlen, we partitioned the errors into two components, the root-variance about a given mean (essentially precision) and bias away from true duration (essentially accuracy). Fig. 2a plots normalized root variance against normalized bias for each subject and for group means. The estimate of total error (considering precision and accuracy) is given by the Pythagorean sum of the two, the distance from the origin: 0.12 for drummers, 0.12 for string musicians and 0.14 for controls. Despite the completely different pattern of results for the three groups, the total error rates vary between the groups by only 17%.

In order to model these results, we obtained an estimate of individual temporal precision, using a separate temporal bisection task, where priors should not affect performance. Subjects reported whether the second flash of a triplet was closer in time to the first or the third, leading to a bias-free estimate of the *Weber fraction* (relative precision), which we plot on the abscissas of Figs 2b. The Weber fraction for the drummers (0.04) is much lower than the controls (0.11) and the string musicians (0.08), although this is not reflected in commensurate differences in total error (Fig 2a). This suggests that the regression-to-mean strategy is effective in minimizing total error. Fig. 1b shows the clear relationship between Weber fraction and regression to the mean, with higher Weber fractions associated with higher regression indexes.

The continuous curves of Fig. 2b show predictions of three models: the *no-prior* model that considers only the sensory data; the Bayesian Least Squares (BLS[1]) model; and our Bayesian *running average model*. The no-prior model clearly fails, as it predicts (by definition) that the regression index will always be zero. Jazayeri and Shadlen's Bayesian model captures much of the data, but falls short quantitatively in predicting the rate the regression index increases with Weber fraction. The running-average model (fully described in Sup. Mat.) fares better here. This differs from the BLS in that it does not consider all the information from previous trials, but calculates a running average of the last five trials, and maintains only the mean and standard deviation of this sample. Over time the distribution will approximate a Gaussian (central limit theorem), which combines with the sensory estimate in the standard "Bayesian" manner[2] (with weights inversely proportional to variance). If the sensory estimate is relatively good (low Weber fraction), this will dominate the average yielding veridical estimates; otherwise the running average will cut in, causing a regression towards the mean. A regression index of more than 0.5 implies that the prior is weighted more than the sensory estimate.

To test this model further, we asked whether the behavior of the subjects could also depend on stimulus type. We repeated all the measurements, both interval reproduction and bisection with auditory tones, as temporal discrimination is more precise in audition than in vision[3]. The results (open triangles of Fig. 2) show that both Weber fractions and the regression indexes are lower for audition than vision, again well predicted by the Bayesian running-average model.

This study extends Jazayeri and Shadlen's[1] important report in several ways. Firstly we show that the "regression to the mean" is not a universal property of time reproduction, but depends on circumstances: when the temporal judgment is imprecise – such as visual judgments with non-percussionists – then the mean can be beneficial; otherwise, there is no point in sacrificing accuracy. The results also suggest that training for a specific task – such as precision drumming – not only improves temporal resolution, but also changes the encoding strategies of those subjects. We also show that while the BLS model is qualitatively adequate, it predicts less regression than actually found; our running-average model predicts the pattern of data far better. While Jazayeri and Shadlen's is *ideal*, in the sense that it considers all information about the probability distribution from which the time samples are drawn, ours considers only the first two moments – mean and standard deviation – calculated on-line from the previous few trials. Keeping only two statistics is more biologically plausible, and makes the model more robust.

Much evidence[4-6] shows that humans can easily maintain a running average of a variety of sensory attributes, including size, color, shape and numerosity. Indeed this capacity is the basis of the psychophysical technique known as the *method of single stimuli*[4], where subjects report whether an individual trial is of higher or lower magnitude than the average of all seen to date. Subjects can keep simultaneously at least four separate averages[5], and the noise associated with the average seems to be less than that of the sensory judgments[6]. That subjects are so good at this task is consistent with the notion that continuous estimates are made of the mean, and perhaps the variance, of past sensory events.

The current study demonstrates the incredible plasticity and efficiency the processes leading to the internal sense of elapsed time. The system seems to have access to all available information, but use it only to confer a functional advantage. Although this and the previous[1] study were limited to time reproduction, the perceptual principles reported here almost certainly generalize to other sensory judgments. Pilot studies in our laboratory have shown similar results in other modalities, including size, position, saccade direction and numerosity judgment: indeed regression to the mean may be the basis for the supposed "logarithmic" representation of numbers often observed in children and uneducated adults[7].

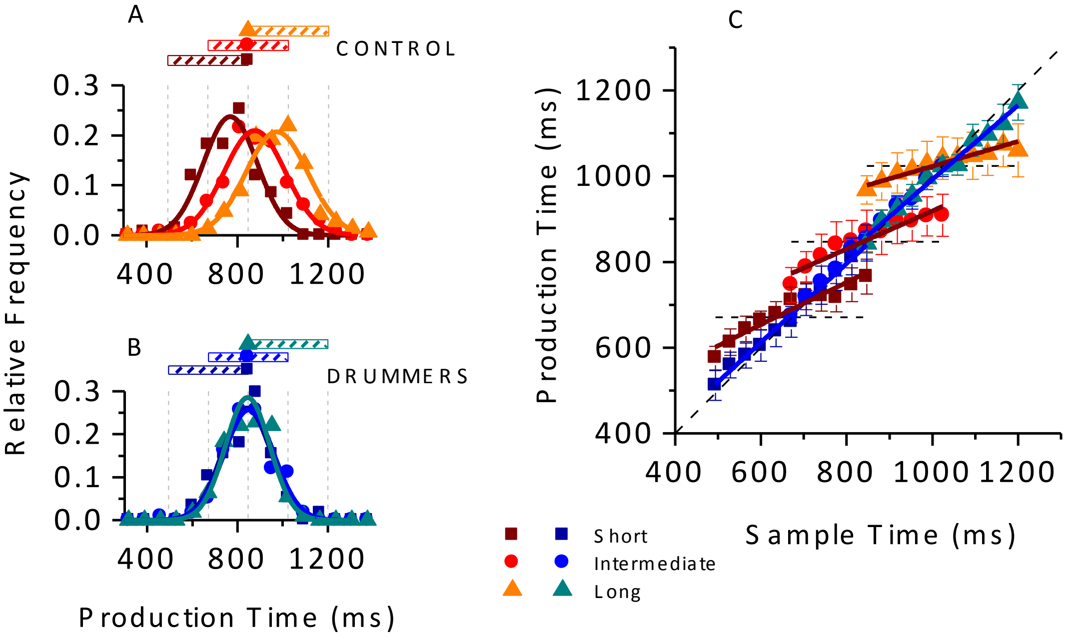

Figure 1

A. Reproduction distribution for non-musical controls for the duration 847 ms, during sessions where the intervals were drawn from short (494-847 ms: squares), intermediate (671-1024 ms: circles), or long intervals (847-1200 ms: triangles). **B.** Same as A for expert drummers. **C.** Average reproduction durations (calculated from distributions like those of A and B) for expert drummers (bluish symbols) and typical controls (yellow-reddish symbols) as a function of stimulus duration for the three stimulus ranges described above (same color code). The straight lines show best-fitting linear regressions. The *regression index* (of Fig. 2) is given the difference between the slope of these fits and the equality line (dashed). Whilst controls showed a strong regression towards the mean, performance for the drummers is virtually veridical.

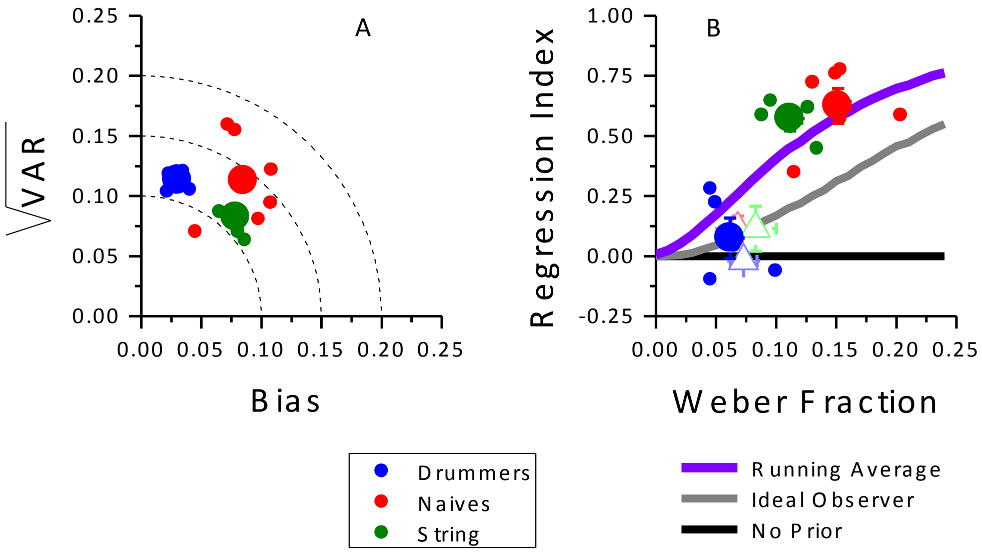

Figure 2

A. VAR$^{1/2}$ (the average root variance of the reproductions) plotted against BIAS (difference between average production time and physical sample interval), for the three subject groups (drummers red, controls blue, string musicians green). The total error (root-mean squared error) is given by the distance from the origin, similar for the three groups. **B.** *Regression index* plotted against *Weber fraction* for interval bisection. The regression index is the deviation from veridicality, given by 1- ρ, where ρ is the slope of the linear fit to reproduction data like those of Fig. 1c (0 for veridicality, 1 for total regression). Individual subjects are shown by small symbols and group means by large (drummers blue; string musicians green; controls red). Filled circles show results for the visual task, open triangle for audition. The three curves show the MLE model (black), BLS[1] model (grey) and the running average model (purple).

REFERENCES

1. Jazayeri, M. & Shadlen, M. N. Temporal context calibrates interval timing. Nat Neurosci 13, 1020-6 (2010).
2. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415, 429-33 (2002).
3. Burr, D., Banks, M. S. & Morrone, M. C. Auditory dominance over vision in the perception of interval duration. Exp Brain Res 198, 49-57 (2009).
4. Woodworth, R. S. Experimental psychology (Henry Holt & Co., New York, 1938).
5. Morgan, M. J. On the scaling of size judgements by orientational cues. Vision Res 32, 1433-45 (1992).
6. Morgan, M. J., Watamaniuk, S. N. & McKee, S. P. The use of an implicit standard for measuring discrimination thresholds. Vision Res 40, 2341-9 (2000).
7. Dehaene, S., Izard, V., Spelke, E. & Pica, P. Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. Science 320, 1217-20 (2008).