EJN EUROPEAN JOURNAL OF NEUROSCIENCE

European Journal of Neuroscience, Vol. 31, pp. 1828–1835, 2010 doi:10.1111/j.1460-9568.2010.07205.x

Multisensory cues improve sensorimotor synchronisation

M. T. Elliott, A. M. Wing and A. E. Welchman
School of Psychology, University of Birmingham, Edgbaston, B15 2TT, UK

Keywords: linear phase correction, maximum likelihood estimation, movement synchronisation, multisensory integration

Abstract

Synchronising movements with events in the surrounding environment is an ubiquitous aspect of everyday behaviour. Often, information about a stream of events is available across sensory modalities. While it is clear that we synchronise more accurately to auditory cues than other modalities, little is known about how the brain combines multisensory signals to produce accurately timed actions. Here, we investigate multisensory integration for sensorimotor synchronisation. We extend the prevailing linear phase correction model for movement synchronisation, describing asynchrony variance in terms of sensory, motor and timekeeper components. Then we assess multisensory cue integration, deriving predictions based on the optimal combination of event time, defined across different sensory modalities. Participants tapped in time with metronomes presented via auditory, visual and tactile modalities, under either unimodal or bimodal presentation conditions. Temporal regularity was manipulated between modalities by applying jitter to one of the metronomes. Results matched the model predictions closely for all except high jitter level conditions in audio–visual and audio–tactile combinations, where a bias for auditory signals was observed. We suggest that, in the production of repetitive timed actions, cues are optimally integrated in terms of both sensory and temporal reliability of events. However, when temporal discrepancy between cues is high they are treated independently, with movements timed to the cue with the highest sensory reliability.

Introduction

Maintaining synchrony with a periodic event requires that the central nervous system (CNS) compensate for timing variation arising from sensory, decision and motor processing noise. In particular, keeping in time with a pacing source (metronome) requires continual corrections based on the timing error (asynchrony) between the metronome and our actions (Hary & Moore, 1987; Mates, 1994a,b; Vorberg & Wing, 1996; Pressing, 1998; Vorberg & Schulze, 2002). This process is described by linear phase correction and accounts for the time series statistics of human synchronisation performance (Vorberg & Wing, 1996; Vorberg & Schulze, 2002). Previous work has addressed synchronisation to unimodal (typically auditory) metronomes (Hary & Moore, 1987; Vos *et al.*, 1995; Semjen *et al.*, 1998; Repp, 2001; Elliott *et al.*, 2009a). Here, we consider the integration of multimodal temporal cues and test for improvements in synchronisation accuracy.

Faced with multisensory information from the environment, the CNS could alternate between cues depending on the demands of the task (Welch & Warren, 1980) or combine information from different senses. Work considering judgments of temporal events under multisensory conditions points to the dominance of auditory signals over other (visual or tactile) modalities (Fendrich & Corballis, 2001; Morein-Zamir *et al.*, 2003; Vroomen *et al.*, 2004; Bresciani *et al.*, 2005). Moreover, if participants are asked to synchronise their movements to conflicting multisensory cues, a periodic auditory distracter disrupts visual synchronisation but the converse is not true (Repp & Penel, 2002, 2004; Kato & Konishi, 2006). In contrast to the auditory dominance effects reported in work on judgments of time, perceptual studies of spatial judgments suggest that multisensory signals are integrated by weighting cues according to their reliability (Alais & Burr, 2004; Heron *et al.*, 2004) to achieve the best estimate in accordance with maximum likelihood estimation (MLE; Alais & Burr, 2004; van Beers *et al.*, 1999; Ernst & Banks, 2002).

Here, we consider whether the CNS integrates multisensory timing cues and whether information is weighted to achieve an optimal estimate of temporal events. We have previously found evidence of sensory integration of auditory and haptic timing cues when producing synchronised finger taps (Wing *et al.*, 2010). However, this study was limited by applying the MLE model directly to the observed asynchrony variance. Here, we assume the observed timing variability contains sources of variance (e.g. motor) that would be unlikely to contribute to the weighting of sensory information. Hence, we extend the linear phase correction model to account for the optimal integration of multisensory timing signals independent of other sources of variance present when producing synchronised actions. We apply the model to experimental data obtained when participants tapped in time to auditory, visual and tactile metronomes presented under unimodal and bimodal conditions. We further use ‘jitter’ (Repp & Penel, 2004) to manipulate the regularity of the metronomes, and hence determine whether the CNS weights multisensory temporal cues according to both sensory reliability and the regularity of the metronome cues.

Correspondence: Dr M. T. Elliott, as above.
E-mail: m.t.elliott@bham.ac.uk

Received 1 October 2009, revised 24 February 2010, accepted 25 February 2010

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

Materials and methods

Experimental setup

Experimental protocols were approved by the University of Birmingham Ethical Review Committee and complied with the Declaration of Helsinki. Participants were recruited from staff and students at the University of Birmingham. They provided written informed consent and were screened for sensory and motor deficits. Ten participants (six male, three left-handed, mean age 25.7 years) took part in audio–visual and audio–tactile experiments. A subset of six participants (four male, two left-handed, mean age 31.7 years) took part in tactile–visual experiments.

Participants were presented with a metronome (inter-onset interval 500 ms) from different sensory modalities: visual (a red LED positioned in front of the participant), auditory (a beep emitted from a piezoelectric auditory buzzer) and/or tactile (a tap on the nondominant index finger delivered by a solenoid-based tactile actuator (MSTC3; M & E Solve, UK)). Participants were instructed to tap the index finger of their dominant hand on a force-sensitive resistor in time to the metronome. They were further instructed to pay attention to all available cues during the trials. Responses were registered using a data acquisition device (USB-6229; National Instruments Inc., USA). Metronome presentation was controlled using the MatTAP toolbox (Elliott *et al.*, 2009b). To suppress auditory feedback from the tactile actuator and their own finger tap responses, participants wore headphones playing white noise. The volume was set such that the surrounding background noise was suppressed but the auditory buzzer was still heard clearly. The experiment took place in a darkened room, to reduce visual distraction and improve contrast of the visual metronome.

Each trial consisted of 30 metronome beats in either a single sensory modality (auditory, visual or tactile only) or in combined pairs of audio–visual, audio–tactile or tactile–visual. Participants were instructed to tap in synchrony with the presented metronome and completed 15 trials per condition over two 60-minute sessions performed on different days. The order of condition presentation was randomised across participants.

Data analysis was performed using functions developed within MATLAB (2007b; The Mathworks, MA, USA). Repeated-measures ANOVAs were used for statistical analysis of the results.

Metronome reliability

We assumed participants had a goal of minimising the size of corrections while at the same time minimising variability in the asynchrony between their movements and the metronome events (Vorberg & Schulze, 2002). When the metronome is an isochronous stream of events, corrections are only required based on the participants' own variability. By adding jitter to the metronome, such that the timing becomes more uncertain, we created an additional source of noise against which participants had to correct. Given multiple sources of information (across modalities) about metronome events, it would be reasonable for the brain to integrate this (possibly discrepant) information to acquire the most likely estimate of the onset of the multimodal metronome event.

In order to create metronomes of varying reliability, we 'jittered' the metronome by adding a random temporal perturbation taken from a Gaussian distribution, N(0, σ), to the regular onset time of each metronome beat (Fig. 1). The perturbations applied to each metronome event were independent and referenced to the time at which the regular beat would have occurred, i.e. there was no cumulative phase shift effect of the perturbations (see Supporting information, Appendix S1, for a discussion of the lag 1 covariance from this manipulation). This ensured that the mean interval of the jittered metronome remained (approximately) equal to that of the regular metronome. Increasing the standard deviation of the distribution, σ, from which the perturbation value was sampled, resulted in a more variable metronome and hence reduced the temporal regularity. We used three jitter conditions, with standard deviations of 0, 20 and 50 ms. The zero jitter condition meant both metronomes had regular beats and hence allowed us to consider the inherent sensory reliability of the cues, in addition to the manipulated metronome regularity. During bimodal presentation, the regularity of one modality [that judged to have a higher sensory reliability based on previous work (e.g. Repp & Penel, 2002) and our own pilot testing] was varied by applying the jitter, while the other modality remained isochronous (standard deviation of 0 ms).

We assumed participants integrate the multimodal cues to get a best estimate of the timing of the metronome beats. We therefore measured the asynchrony between finger tap onsets and the onsets of the underlying regular metronome beats (i.e. before application of jitter), quantifying accuracy as the variability (standard deviation) of the measured asynchronies (Fig. 1).

Linear phase correction model

To test whether participants extracted the best estimate of the beats of a multisensory metronome, we extended the standard linear phase correction model (Vorberg & Wing, 1996; Vorberg & Schulze, 2002) to incorporate multisensory information. This allowed us to estimate the different variance components that contribute to timing errors when producing synchronised movements. In particular, we calculated the asynchrony variance due to any irregularity in the metronome, as well as variance due to the sensory registration of the metronome beats under unimodal conditions. This allowed us to predict the (reduced) variability of synchronised movements when a bimodal metronome was presented, based on the assumption that multisensory cues are integrated optimally using a maximum likelihood framework (Cochran, 1937).

The original linear phase correction model described by Vorberg and colleagues defined the variance in the observed timing errors (asynchronies, σ_A^2) as a function of noise originating from (i) the internal representation of the metronome (timekeeper, σ_T^2), (ii) the process of executing a motor action (σ_M^2) and (iii) variations in the periodicity of the (external) metronome beats (σ_C^2) (Vorberg & Wing, 1996; Vorberg & Schulze, 2002). Under this framework, noise in the sensory registration of events is combined with timekeeper noise. For our purposes, this is not adequate as sensory and timekeeper noise components need to be separated. Thus, we modified the linear phase correction model (see supporting Appendix S1) to account for variance resulting from the sensory registration of temporal events. Hence, we define asynchrony variance as

$$\sigma_A^2 = \frac{1}{1-(1-\alpha)^2}\left(\sigma_C^2 + \sigma_S^2 + \sigma_T^2 + 2\alpha\sigma_M^2\right), \qquad (1)$$

where σ_A^2 is the measured asynchrony variance and σ_T^2, σ_M^2, σ_C^2, σ_S^2 are the timekeeper, motor, metronome and sensory variances respectively. The correction parameter, α, represents the (average) proportion of the asynchrony which is corrected for on the subsequent movement ($0 < \alpha < 2$).

The overall asynchrony variance described by Equation 1 can be partitioned into two separate noise components: the modality-dependent variances (S) and the modality-independent variances (K). Here, we assume that the timekeeper is a central process, independent of sensory modality (Rao *et al.*, 2001; Spencer *et al.*, 2003; Grahn &

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

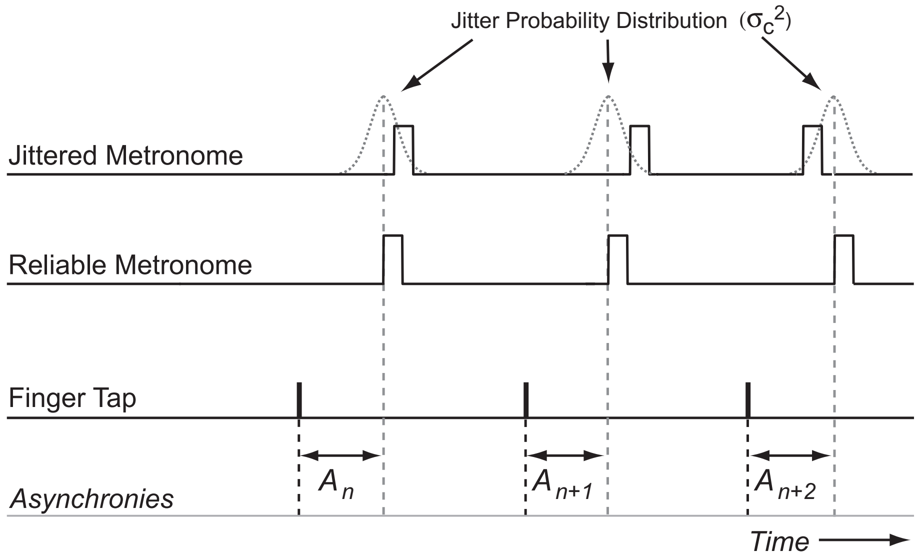

FIG. 1. Measurement of timing error (asynchrony) between metronome beats and corresponding finger taps. In bimodal conditions, participants were presented with two simultaneous metronomes with the same underlying period and phase. To create temporal uncertainty, jitter was added to one of the metronomes (Jittered Metronome). This involved adding a random value taken from a Gaussian distribution to the regular onset time of each beat. The standard deviation of the distribution was varied to set the level of uncertainty. The other metronome remained isochronous (Reliable Metronome). Asynchronies (*A*) were calculated between the tap onsets and the un-jittered onset of the corresponding metronome(s).

Brett, 2009) and that motor noise is also independent of the sensory modality which the metronome is presented in. As such, the modality-independent variance is defined as

$$K = \frac{1}{1-(1-\alpha)^2}\left(\sigma_T^2 + 2\alpha\sigma_M^2\right) \tag{2}$$

The variance associated with (i) the sensory registration of the events and (ii) the (manipulated) variance of the metronome will depend on the modality of the metronome cues, so the modality-dependent variance is

$$S = \frac{1}{1-(1-\alpha)^2}\left(\sigma_C^2 + \sigma_S^2\right) \tag{3}$$

For simplicity, when $\sigma_C^2 > 0$, the metronome beats are assumed to be independently and identically distributed (i.i.d). The application of jitter to the metronome (see 'Metronome Reliability' section) results in the metronome intervals exhibiting negative lag 1 covariance and hence are not truly i.i.d. However, as we show in the supporting Appendix S1, this has no direct effect on the predicted sensory variance calculations.

The sum of Equations 2 and 3 is equal to the asynchrony variance

$$\sigma_A^2 = K + S \tag{4}$$

Hence, the model assumes that when the metronome is presented in different sensory modalities, *S* changes according to the modality (i.e. different values of sensory noise), while *K* remains constant.

We used empirical data from self-paced tapping trials to estimate the modality-independent timekeeper and motor variance components (*K*) using the Wing–Kristofferson model (Wing & Kristofferson, 1973; see 'Estimation of modality-independent variance', below). Hence, from Equation 4, we were able to partition the asynchrony variance associated with modality-dependent variance using metronomes presented in different modalities by subtracting the estimate of modality-independent variance.

MLE of sensory events

To test the idea that multisensory temporal event signals are combined according to MLE, we used the values of *S* measured under unimodal conditions to predict performance under bimodal presentation. Specifically, we tested whether the observed asynchrony variance resulting from the integration of multiple sensory sources was lower than the variance associated with the individual component sources. This is statistically optimal if the sensory sources are uncorrelated and weighted according to their reliability (Cochran, 1937). Hence, if we know the individual sensory variances from two different modalities (S_1 and S_2), it is possible to predict the reduced variance if the signals are integrated optimally:

$$\sigma_{MLE}^2 = \frac{\sigma_{s_1}^2 \sigma_{s_2}^2}{\sigma_{s_1}^2 + \sigma_{s_2}^2} \tag{5}$$

If we assume that sensory estimation for movement synchronisation involves the integration of bimodal cues that accords to MLE, we can substitute the modality-dependent variances resulting from two unimodal presentations (S_1 and S_2) into Equation 5 to calculate the predicted asynchrony variance yielded by optimal integration of the two modalities:

$$\sigma_{A_{MLE}}^2 = K + \frac{S_1 S_2}{S_1 + S_2} \tag{6}$$

To account for individual differences, estimates of K, S_1 and S_2 were established using bootstrap methods on individual participant data. First, 10 000 random samples of *K* (modality-independent variance) were calculated from Gaussian distributions based on the statistics of the timekeeper and motor variances derived by the Wing–Kristofferson model (Wing & Kristofferson, 1973). Second, 10 000 pairs of unimodal asynchrony variances were sampled from the empirical data (using a sample and replace method). The *K* values were subtracted from the asynchrony pairs, leaving 10 000 pairs of unimodal estimates

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

of modality-dependent variance (S_1 and S_2). The K, S_1 and S_2 estimates were applied to Equation 6 and averaged to obtain the mean and 95% confidence intervals for the predicted bimodal asynchrony variances for each condition.

Estimation of modality-independent variance K

To estimate K, we asked participants to perform four additional trials of self-paced tapping. This allowed us to calculate an estimate of the variability associated with producing timed repetitive movements independent of sensory timing stimuli (specifically the timekeeper and motor variances).

At the start of each trial, participants were presented with five beats of an auditory metronome (inter-onset interval 500 ms). Participants started tapping during the metronome presentation and then continued to tap for 30 s after the metronome had been switched off. All other conditions remained as in the main part of the experiment to ensure any effects of sensory feedback (i.e. proprioception) were constant throughout.

We used measures of autocovariances (*acv*) of the intervals between taps to make estimates of the component variances associated with maintaining an internal timekeeper and producing the subsequent motor actions (Wing & Kristofferson, 1973). Specifically,

$$acv(0) = \sigma_T^2 + 2\sigma_M^2 \tag{7}$$

and

$$acv(1) = -\sigma_M^2 \tag{8}$$

where *acv(j)* is the lag *j* autocovariance of tap inter-response intervals.

Estimation of the correction parameter

We further calculated the level of correction made by participants to the jittered metronomes. The average correction parameter (α, Equation 1) was calculated using the lag 1 cross-covariance (*ccv(1)*) between the jittered metronome and tap intervals (Vorberg & Schulze, 2002) from synchronisation data:

$$ccv(1) = \alpha\sigma_C^2, \tag{9}$$

where σ_C^2 is the metronome variance.

For calculation of the correction parameter, we used data from the main experimental unimodal conditions where the metronome had been jittered. Additionally, participants completed extra trials in bimodal conditions where the two modalities were identically jittered. The matched jitter removed any relative weighting between modalities which would have affected our estimates of the correction parameter.

In analysis of these data we noted slight differences in the correction parameter across these conditions, but relatively large individual differences resulted in no significant effect. Hence, when applied to the model, the correction parameter was calculated as the overall mean value from the data collected.

Results

Using the modified linear phase correction model and the empirical data, we tested for optimal multisensory integration in sensorimotor synchronisation. We first report the estimates of the modality-independent parameters required for the modelling, followed by the main empirical results from the three bimodal conditions (audio–visual, audio–tactile and tactile–visual).

Modality-independent parameters

To utilise the model, our first task was to estimate a number of parameters (See Materials and Methods). First, we used data obtained during self-paced tapping to estimate the modality-independent variances associated with the central timekeeper and motor responses. For this purpose we measured the lag 0 and lag 1 autocovariances following Wing & Kristofferson (1973). As previously reported (O'Boyle *et al.*, 1996; Kampen & Snijders, 2002), this method can produce trials in which estimates of the component variances are invalid (motor variance is estimated to have a negative value due to a positive lag 1 autocovariance). Several methods have been employed to deal with trials that violate the assumption that intervals co-vary negatively at lag 1. These include discarding invalid trials (Turvey *et al.*, 1989; O'Boyle *et al.*, 1996), keeping all the trials, or setting motor variance to zero when it is estimated as negative (Ivry & Keele, 1989). As calculated negative motor variances are typically numerically small, these different methods have been shown to produce similar results (O'Boyle *et al.*, 1996). From a total of 24 trials (six participants, four trials each), we discarded ten invalid trials, resulting in 14 valid trials with a mean interval of 521.8 ms. [Due to the high rejection rate of trials, we ran the same task again but using bimanual tapping. This approach provided a higher success rate (2 out of 30 trials rejected) and also validated our original result with a combined motor and timekeeper variance of 489.47 ms^2 versus the original estimate of 471.21 ms^2.] These estimates were bootstrapped using sample-and-replace (10 000 repetitions) to determine the mean ± 1 standard deviation of the motor and timekeeper variances (Table 1).

Typically, the proportion of asynchrony corrected on the proceeding tap is between 0.5 and 1 when calculated from empirical data (Repp, 2008a,b; Repp *et al.*, 2008). We estimated the (average) correction parameter (α) calculated from the participants' empirical data as 0.64 (Table 1).

To test the assumption that the sensory signals were integrated optimally in bimodal conditions, we calculated the MLE based on the component sensory variance estimates. Using the parameters in Table 1, we calculated the modality-independent variance K and subtracted it from the empirical unimodal asynchrony variances to get the component sensory variance estimates.

Having described how we estimated the modality-independent contributions to asynchrony variance, we now present the main results that test the idea that multisensory signals are combined optimally to improve synchronisation performance.

TABLE 1. Estimated parameters of motor and timekeeper variance using the Wing–Kristofferson model applied to self-paced tapping data

Parameter	Mean	Standard deviation
Motor variance, σ_M^2	37.16	6.66
Timekeeper variance, σ_T^2	396.89	35.99
Correction factor, α	0.64	0.14

These values are consistent with previous studies using this approach (Wing, 2002). In addition α was calculated using measures of cross-covariance between tap and (jittered) metronome intervals.

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

Audio–visual cues

Participants tapped their index finger in time to metronomes presented as an auditory beep, a visual flash of an LED or both stimuli presented simultaneously (bimodal condition). The visual metronome remained reliable for all trials; however, jitter was applied (0, 20 and 50 ms) to the auditory metronome with the effect that asynchrony variability increased ($F_{2,18}$ = 39.396, P < 0.001; Fig. 2A). Comparing the unperturbed (zero jitter) conditions, we observed that variability in movement asynchrony was lower for auditory cues than visual ($F_{1,9}$ = 8.809, P = 0.019; Fig. 2A), in line with previous studies suggesting weaker synchronisation with a visual metronome. However, there was a clear benefit from having two timing signals available during bimodal presentations (Fig. 2A). In particular, asynchrony variability was lower than in single-modality auditory conditions ($F_{1,9}$ = 5.303, P = 0.047). Moreover, for low levels of jitter this reduction in variability fell within the range predicted by our model (Fig. 2A, grey shaded areas), suggesting statistically optimal estimation. However, for the high temporal jitter condition (50 ms) the combined audio–visual asynchrony was considerably larger than predicted. Thus, we observed only a slight reduction in asynchrony variability compared to the unimodal results, indicating suboptimal weightings.

Audio–tactile cues

Replicating the design of the audio–visual setup, we replaced the visual metronome with a tactile metronome presented as a tap to the finger of the non-dominant hand. The tactile metronome remained reliable for all conditions. We observed that asynchrony variability for unimodal tactile cues was lower than for the visual metronome ($F_{1,9}$ = 6.929, P = 0.027) and only slightly higher than that for unimodal auditory cues.

In bimodal conditions we observed a strong benefit associated with concurrent timing cues, with asynchrony variability reduced across all jitter levels, relative to unimodal auditory cues ($F_{1,9}$ = 13.235, P = 0.005; Fig. 2B). At zero or low (20 ms) jitter, we observed a good match to the model's predictions (Fig. 2B, grey shaded area). However, as was observed under audio–visual conditions, we observed a departure from optimal estimation at high levels of auditory jitter.

Tactile–visual cues

We further considered how participants performed in non-auditory settings. In particular, we tested synchronisation to visual and tactile timing signals presented alone or bimodally. As the visual metronome resulted in the highest asynchrony variance in the previous experiments we kept the visual metronome reliable for all trials and added temporal jitter to the tactile signal (0, 20 and 50 ms). We again noted a clear benefit associated with bimodal presentation, with asynchrony variability significantly reduced compared to single-modality tactile presentation at all jitter levels ($F_{1,5}$ = 7.620, P = 0.040). In contrast to the results presented for the previous two experiments, we observed a close match to the model's predictions (Fig. 2C, grey shaded areas) for all measured jitter conditions. Thus, performance was consistent with optimal use of sensory signals at all levels of metronome jitter considered.

Suboptimal weighting

To understand why there was a departure from optimal weighting at high levels of auditory jitter, we tested whether participants were

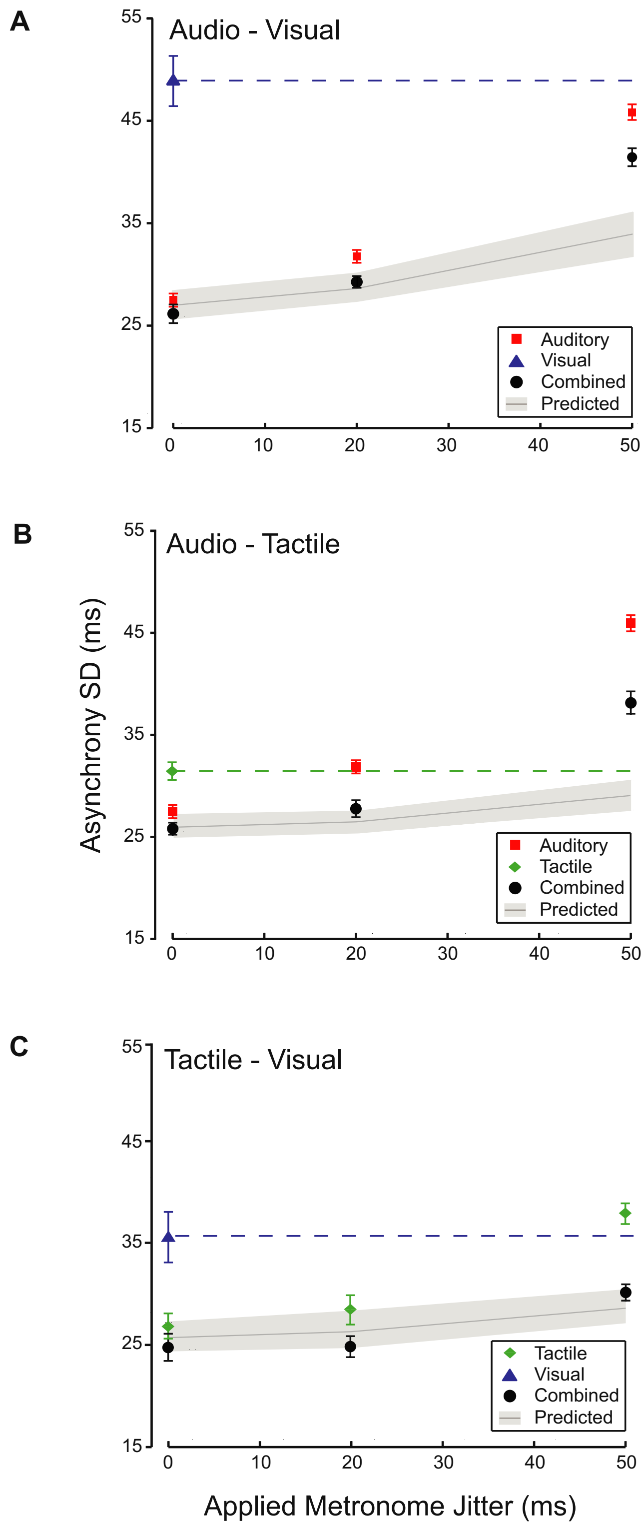

FIG. 2. Empirical results showing the asynchrony standard deviation against the level of applied jitter for (A) Audio–visual modalities, (B) Audio–tactile modalities and (C) Tactile–visual modalities. Unimodal results are shown for auditory (square)-, visual (triangle)- and tactile (diamond)-only conditions, whilst bimodal results are indicated by circles. Error bars represent standard errors. If the cues are weighted in proportion to their temporal reliability then the bimodal results should fall into the shaded region (bootstrapped 95% confidence intervals for the model's predictions). N.B. Individual differences in the visual-only conditions meant the subset of six participants running the visual-tactile experiment exhibited a reduced unimodal visual variability compared to the 10 participants in the audio–visual experiment.

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

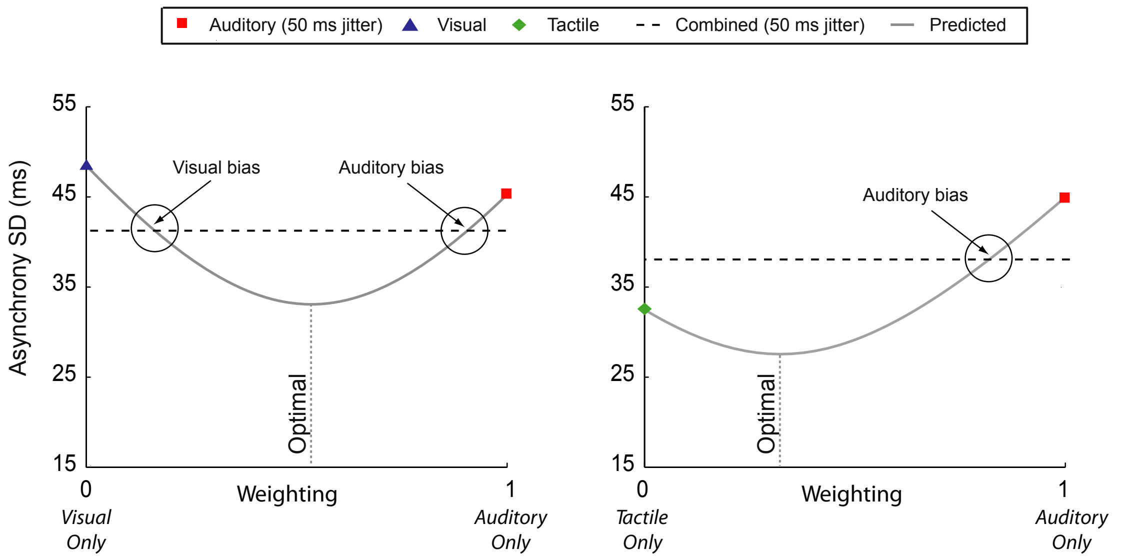

FIG. 3. Relative weighting between modalities in the high-jitter (50 ms) conditions. Using the unimodal asynchrony variances observed in (left) audio–visual and (right) audio–tactile conditions we plotted the predicted bimodal asynchrony standard deviations (SD) for all possible weightings between the two cues (solid line). The minimum of the curve indicates the optimal weighting to achieve the lowest variability in bimodal conditions. We overlaid the observed asynchrony variability (horizontal dashed line) to establish the weightings between modalities in the suboptimal conditions. For audio–visual conditions there are two possible weightings, one biasing towards auditory, the other visual. However, for audio–tactile conditions it is clear there is a strong weighting to the auditory metronome despite the large temporal irregularity of this modality.

biased towards one modality. We did this by simulating a range of weighting scenarios, generating different predictions against which we compared our empirical data. In particular, systematically varying the auditory weight between zero (exclusive reliance on the visual or tactile cue) and one (exclusive reliance on the auditory cue) changes the predicted outcome under bimodal conditions (solid grey curve, Fig. 3). We compared this range of predictions to the variance observed for audio–visual (Fig. 3A) and audio–tactile (Fig. 3B) conditions (auditory metronome with 50-ms jitter). For the audio–visual condition, the results were ambiguous as there were two valid intersections between the possible weights and the empirical data (i.e. there could have been a visual bias or an auditory bias, Fig. 3A). [Although the curves could (theoretically) be extrapolated below zero (i.e. negative cue weights, see Oruc *et al.*, 2003), this is possible only for highly correlated cues. Hence, we only consider a non-negative range of weights between zero and one.] However, in the audio–tactile case, it is clear that observers are using suboptimal weightings that are biased towards the auditory cue (Fig. 3B). It is likely that participants behaved similarly in the audio–visual condition; it thus appears that participants were relying more on the auditory signal, despite its high irregularity.

General discussion

We have considered the integration of multisensory cues to estimate temporal events in support of synchronisation tapping. By modifying the linear phase correction model we have been able to isolate the sensory noise associated with different sensory modalities, allowing us to test predictions based on the optimal weighting of event-timing signals. Under the majority of conditions tested, we found that our empirical observations matched closely the predictions of our model. However, when highly irregular auditory metronomes were presented, performance deviated from the model's predictions.

Weighting rhythmic cues

It is reported that temporal judgements are more reliable for auditory cues than visual or tactile modalities (Glenberg & Swanson, 1986; Glenberg & Jona, 1991) and it has been suggested that auditory metronome beats predominate over visual cues in synchronisation tasks (Repp & Penel, 2002, 2004; Kato & Konishi, 2006). However, an alternative possibility is that these results are explained by the CNS weighting multisensory cues according to their statistical reliability (van Beers *et al.*, 1999; Ernst & Banks, 2002; Alais & Burr, 2004; Heron *et al.*, 2004), with previous work using reliable auditory cues. In addition to the reliability of the sensory registration of events, here we have also considered the regularity in the timing of the event itself, as this is likely to be an important component of successful synchronisation behaviour. Hence, we tested whether the CNS would take account of both the statistical reliability of the sensory information and the beat regularity. Increasing jitter on the auditory metronome resulted in participants shifting their weighting more towards the isochronous visual or tactile modalities, despite them having a higher sensory uncertainty. This weighting of cues resulted in lower asynchrony variance under bimodal conditions. In particular, we found that in seven out of the nine conditions the reduced asynchrony variance corresponded to our model predictions, indicating an optimal weighting of modalities.

However, when high levels of temporal jitter were added to auditory signals we observed deviations from the optimal estimation model. In particular, there was an unexpected reliance on the auditory cue despite the high variability associated with this metronome signal. What might underlie this failure of optimal integration? We suggest its basis is in the temporal discrepancy between the two signals. In particular, in evaluating sensory signals that are temporally separated, the CNS should determine whether the signals have a common underlying cause and thus should be integrated or have separate causes and thus should be kept separate. The probability of integrating two signals depends on a temporal window of integration: if the temporal

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

discrepancy is within the window, signals are integrated (Meredith *et al.*, 1987; Yabe, 1998; Shams *et al.*, 2002; Colonius & Diederich, 2004). It is likely that the window of integration is not fixed, but rather depends on the likelihood of two signals originating from a common source based on prior knowledge, sensory reliability and the offset between the two signals (Roach *et al.*, 2006; Körding *et al.*, 2007; Sato *et al.*, 2007). Under our paradigm, the major determinant of whether the two metronome events should be integrated is their temporal offset caused by random jitter perturbations. If the temporal separation between different modality metronome events is sufficiently large, they should be judged independent. When this occurs, the CNS should choose one of the signals to compute asynchrony. Based on previous reports of the dominance of auditory cues in the context of producing timed actions (Jancke *et al.*, 2000; Repp & Penel, 2002, 2004; Repp, 2003; Kato & Konishi, 2006), it is likely that the CNS would rely on auditory signals when temporal offsets are large.

Our findings also suggest that the temporal window of integration depends on the sensory modalities involved. Specifically, while we observed large deviations from optimal weightings for high-jitter audio–tactile and audio–visual pairings, we found that tactile–visual cue pairs were integrated optimally at high jitter levels. This ordering is expected based on the unimodal measurements (auditory cues were more reliable than tactile or visual signals), as the temporal window within which signals should be accepted as originating from a common source will depend on the reliability with which the CNS can signal the temporal offset between signals. Thus, the CNS is more likely to accept a tactile–visual pair that differs in onset time as originating from a common source than an audio–tactile pair. As such, we would expect that introducing further temporal jitter for tactile–visual pairings would lead to the apparent dominance of the tactile cue.

We further suggest that future work should consider explicit manipulation of the phase offset of the metronomes between modalities (Repp & Penel, 2004; Kato & Konishi, 2006). This will provide better insight into the dynamics of the temporal integration window when integrating multisensory rhythmic cues and how this varies between modality pairs. In addition, individual tap analyses would help to establish whether the CNS integrates multisensory cues on a beat-by-beat basis (in accordance with the prevailing linear phase correction model which suggests corrections are made based on the previous asynchrony only: Vorberg & Schulze, 2002; Vorberg & Wing, 1996) or on a longer-term prior history (as suggested by the Bayesian framework of the causal inference models: Körding *et al.*, 2007; Sato *et al.*, 2007).

Modality-independence of the timekeeper

We extended the original linear phase correction model (Vorberg & Wing, 1996; Vorberg & Schulze, 2002) by partitioning components of the sensory noise. This was achieved by determining motor and timekeeper noise from self-paced tapping (Wing & Kristofferson, 1973) and then calculating the residual variance due to sensory input in a synchronisation task. Using maximum likelihood weighting, we predicted the optimal integration of bimodal sensory cues. This approach assumes that sensory integration occurs at an early stage, with a central timekeeper independent of modality. Both self-paced and synchronised finger tapping has been shown to activate the basal ganglia (Witt *et al.*, 2008), suggesting its use for centrally computing sub-second time intervals in the brain (Rao *et al.*, 2001; Grahn & Brett, 2009). However, other work has suggested that timing mechanisms are distributed across the brain, with different modalities invoking their own modality-dependent estimations (Jantzen *et al.*, 2005). We suggest future work could consider using a range of metronome intervals to establish the corresponding timekeeper gradients (Wing & Kristofferson, 1973) for different modalities: if the timekeeper is independent of modality then the gradients should be equal across sensory presentations.

Conclusion

In summary, our findings explain the role of multisensory information in the production of accurately timed actions. We have shown that in the context of rhythmic cues the brain will weight signals according to the relative reliability in the timing of the events across modalities, ensuring optimal movement production to the underlying event extracted from the signals. We further found this integration breaks down when there are large temporal discrepancies between cues, suggesting events judged to be independent are no longer integrated, and actions are then timed with respect to the cue with the highest sensory reliability.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. Derivation of model assuming correlated metronome noise.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset by Wiley-Blackwell. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Acknowledgements

This work was supported by Biotechnology and Biological Sciences Research Council Grants [EO09565, C520620]. The authors thank Dagmar Fraser for technical input.

Abbreviations

CNS, central nervous system; *K*, modality-independent variances; MLE, maximum likelihood estimation; *S*, modality-dependent variances.

References

Alais, D. & Burr, D. (2004) The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Curr. Biol.*, **14**, 257–262. doi: 10.1016/j.cub.2004.01.029.

van Beers, R.J., Sittig, A.C. & Gon, J.J.V.D. (1999) Integration of proprioceptive and visual position-information: an experimentally supported model. *J. Neurophysiol.*, **81**, 1355–1364.

Bresciani, J., Ernst, M.O., Drewing, K., Bouyer, G., Maury, V. & Kheddar, A. (2005) Feeling what you hear: auditory signals can modulate tactile tap perception. *Exp. Brain Res.*, **162**, 172–180. doi: 10.1007/s00221-004-2128-2.

Cochran, W.G. (1937) Problems Arising in the Analysis of a Series of Similar Experiments. *Suppl. J. R. Stat. Soc. Ser. A*, **4**, 102–118. doi: 10.2307/2984123.

Colonius, H. & Diederich, A. (2004) Multisensory interaction in saccadic reaction time: a time-window-of-integration model. *J. Cogn. Neurosci.*, **16**, 1000–1009. doi: 10.1162/0898929041502733.

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd

Elliott, M.T., Welchman, A.E. & Wing, A.M. (2009a) Being discrete helps keep to the beat. *Exp. Brain Res.*, **192**, 731–737. doi: 10.1007/s00221-008-1646-8.
Elliott, M.T., Welchman, A.E. & Wing, A.M. (2009b) MatTAP: A MATLAB toolbox for the control and analysis of movement synchronisation experiments. *J. Neurosci. Methods*, **177**, 250–257. doi: 10.1016/j.jneumeth.2008.10.002.
Ernst, M.O. & Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, **415**, 429–433. doi: 10.1038/415429a.
Fendrich, R. & Corballis, P.M. (2001) The temporal cross-capture of audition and vision. *Percept. Psychophys.*, **63**, 719–725.
Glenberg, A.M. & Jona, M. (1991) Temporal coding in rhythm tasks revealed by modality effects. *Mem. Cognit.*, **19**, 514–522.
Glenberg, A.M. & Swanson, N.G. (1986) A temporal distinctiveness theory of recency and modality effects. *J. Exp. Psychol. Learn. Mem. Cogn.*, **12**, 3–15.
Grahn, J.A. & Brett, M. (2009) Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex*, **45**, 54–61. doi: 10.1016/j.cortex.2008.01.005.
Hary, D. & Moore, G. (1987) Synchronizing human movement with an external clock source. *Biol. Cybern.*, **56**, 305–311. doi: 10.1007/BF00319511.
Heron, J., Whitaker, D. & McGraw, P. (2004) Sensory uncertainty governs the extent of audio-visual interaction. *Vision Res.*, **44**, 2875–2884. doi: 10.1016/j.visres.2004.07.001.
Ivry, R.B. & Keele, S.W. (1989) Timing Functions of The Cerebellum. *J. Cogn. Neurosci.*, **1**, 136–152. doi: 10.1162/jocn.1989.1.2.136.
Jancke, L., Loose, R., Lutz, K., Specht, K. & Shah, N.J. (2000) Cortical activations during paced finger-tapping applying visual and auditory pacing stimuli. *Cogn. Brain. Res.*, **10**, 51–66. doi: 10.1016/S0926-6410(00)00022-7.
Jantzen, K., Steinberg, F. & Kelso, J. (2005) Functional MRI reveals the existence of modality and coordination-dependent timing networks. *NeuroImage*, **25**, 1031–1042. doi: 10.1016/j.neuroimage.2004.12.029.
Kampen, J.K. & Snijders, T.A.B. (2002) Estimation of the Wing-Kristofferson model for discrete motor responses. *Br. J. Math. Stat. Psychol.*, **55** (Pt 1), 159–168.
Kato, M. & Konishi, Y. (2006) Auditory dominance in the error correction process: a synchronized tapping study. *Brain Res.*, **1084**, 115–122.
Körding, K.P., Beierholm, U., Ma, W.J., Quartz, S., Tenenbaum, J.B. & Shams, L. (2007) Causal Inference in Multisensory Perception. *PLoS ONE*, **2**, e943. doi: 10.1371%2Fjournal.pone.0000943.
Mates, J. (1994a) A model of synchronization of motor acts to a stimulus sequence. *Biol. Cybern.*, **70**, 463–473. doi: 10.1007/BF00203239.
Mates, J. (1994b) A model of synchronization of motor acts to a stimulus sequence. *Biol. Cybern.*, **70**, 475–484. doi: 10.1007/BF00203240.
Meredith, M.A., Nemitz, J.W. & Stein, B.E. (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J. Neurosci.*, **7**, 3215–3229.
Morein-Zamir, S., Soto-Faraco, S. & Kingstone, A. (2003) Auditory capture of vision: examining temporal ventriloquism. *Cogn. Brain. Res.*, **17**, 154–163. doi: 10.1016/S0926-6410(03)00089-2.
O'Boyle, D.J., Freeman, J.S. & Cody, F.W. (1996) The accuracy and precision of timing of self-paced, repetitive movements in subjects with Parkinson's disease. *Brain*, **119**(Pt 1), 51–70.
Oruç, I., Maloney, L.T. & Landy, M.S. (2003) Weighted linear cue combination with possibly correlated error. *Vision Res.*, **43**, 2451–2468. doi: 10.1016/S0042-6989(03)00435-8.
Pressing, J. (1998) Error correction processes in temporal pattern production. *J. Math. Psychol.*, **42**, 63–101.
Rao, S.M., Mayer, A.R. & Harrington, D.L. (2001) The evolution of brain activation during temporal processing. *Nat. Neurosci.*, **4**, 317–323. doi: 10.1038/85191.
Repp, B.H. (2001) Phase Correction, Phase Resetting, and Phase Shifts After Subliminal Timing Perturbations in Sensorimotor Synchronization. *J. Exp. Psychol. Hum. Percept. Perform.*, **27**, 600–621.
Repp, B.H. (2003) Rate limits in sensorimotor synchronization with auditory and visual sequences: the synchronization threshold and the benefits and costs of interval subdivision. *J. Mot. Behav.*, **35**, 355–370. doi: Article.
Repp, B. (2008a) Multiple temporal references in sensorimotor synchronization with metrical auditory sequences. *Psychol. Res.*, **72**, 79–98. doi: 10.1007/s00426-006-0067-1.
Repp, B.H. (2008b) Perfect phase correction in synchronization with slow auditory sequences. *J. Mot. Behav.*, **40**, 363–367. doi: 10.3200/JMBR.40.5.363-367.
Repp, B.H. & Penel, A. (2002) Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *J. Exp. Psychol. Hum. Percept. Perform.*, **28**, 1085–1099. doi: 12421057.
Repp, B.H. & Penel, A. (2004) Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychol. Res.*, **68**, 252–270. doi: 10.1007/s00426-003-0143-8.
Repp, B.H., London, J. & Keller, P.E. (2008) Phase Correction in Sensorimotor Synchronization with Nonisochronous Sequences. *Music Percept.*, **26**, 171–175. doi: 10.1525/mp.2008.26.2.171.
Roach, N.W., Heron, J. & McGraw, P.V. (2006) Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proc. R. Soc. Lond., B, Biol.*, **273**, 2159–2168. doi: 10.1098/rspb.2006.3578.
Sato, Y., Toyoizumi, T. & Aihara, K. (2007) Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Comput.*, **19**, 3335–3355.
Semjen, A., Vorberg, D. & Schulze, H.H. (1998) Getting synchronized with the metronome: comparisons between phase and period correction. *Psychol. Res.*, **61**, 44–55. doi: 10.1007/s004260050012.
Shams, L., Kamitani, Y. & Shimojo, S. (2002) Visual illusion induced by sound. *Cogn. Brain. Res.*, **14**, 147–152. doi: 10.1016/S0926-6410(02)00069-1.
Spencer, R.M., Zelaznik, H.N., Diedrichsen, J. & Ivry, R.B. (2003) Disrupted Timing of Discontinuous But Not Continuous Movements by Cerebellar Lesions. *Science*, **300**, 1437–1439.
Turvey, M.T., Schmidt, R.C. & Rosenblum, L.D. (1989) 'Clock' and 'motor' components in absolute coordination of rhythmic movements. *Neuroscience*, **33**, 1–10.
Vorberg, D. & Schulze, H.H. (2002) Linear phase-correction in synchronization: predictions, parameter estimation, and simulations. *J. Math. Psychol.*, **46**, 56–87.
Vorberg, D. & Wing, A. (1996) Modeling variability and dependence in timing. In *Handbook of perception and action*. Academic Perss, London, pp. 181–262.
Vos, P.G., Mates, J. & van Kruysbergen, N.W. (1995) The perceptual centre of a stimulus as the cue for synchronization to a metronome: evidence from asynchronies. *Q. J. Exp. Psychol. A*, **48**, 1024–1040.
Vroomen, J., de Gelder, B. & Vroomen, J. (2004) Temporal ventriloquism: sound modulates the flash-lag effect. *J. Exp. Psychol. Hum. Percept. Perform.*, **30**, 513–518. doi: 10.1037/0096-1523.30.3.513.
Welch, R.B. & Warren, D.H. (1980) Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.*, **88**, 638–667.
Wing, A.M. (2002) Voluntary Timing and Brain Function: An Information Processing Approach. *Brain and Cognition*, **48**, 7–30.
Wing, A. & Kristofferson, A. (1973) Response delays and the timing of discrete motor responses. *Perception and Psychophysics*, **14**, 5–12.
Wing, A.M., Doumas, M. & Welchman, A.E. (2010) Combining multisensory temporal information for movement synchronisation. *Exp. Brain Res.*, **200**, 277–282. doi: 10.1007/s00221-009-2134-5.
Witt, S.T., Laird, A.R. & Meyerand, M.E. (2008) Functional neuroimaging correlates of finger-tapping task variations: An ALE meta-analysis. *NeuroImage*, **42**, 343–356. doi: 10.1016/j.neuroimage.2008.04.025.
Yabe, H. (1998) Temporal window of integration of auditory information in the human brain. *Psychophysiology*, **35**, 615–619.

© The Authors (2010). Journal Compilation © Federation of European Neuroscience Societies and Blackwell Publishing Ltd