Exp Brain Res (2010) 200:277–282
DOI 10.1007/s00221-009-2134-5

RESEARCH NOTE

Combining multisensory temporal information for movement synchronisation

Alan M. Wing · Michail Doumas · Andrew E. Welchman

Received: 9 July 2009 / Accepted: 9 December 2009 / Published online: 29 December 2009

Abstract The ability to synchronise actions with environmental events is a fundamental skill supporting a variety of group activities. In such situations, multiple sensory cues are usually available for synchronisation, yet previous studies have suggested that auditory cues dominate those from other modalities. We examine the control of rhythmic action on the basis of auditory and haptic cues and show that performance is sensitive to both sources of information for synchronisation. Participants were required to tap the dominant hand index finger in synchrony with a metronome defined by periodic auditory tones, imposed movements of the non-dominant index finger, or both cues together. Synchronisation was least variable with the bimodal metronome as predicted by a maximum likelihood estimation (MLE) model. However, increases in timing variability of the auditory cue resulted in some departures from the MLE model. Our findings indicate the need for further investigation of the MLE account of the integration of multisensory signals in the temporal control of action.

Keywords Movement timing · Action · Synchronisation · Multisensory · Cue combination · Haptic auditory

A. M. Wing (✉) · M. Doumas · A. E. Welchman
Behavioural Brain Sciences Centre, School of Psychology,
University of Birmingham, Edgbaston,
Birmingham B15 2TT, UK
e-mail: A.M.Wing@bham.ac.uk

M. Doumas
Department of Psychology, K. U. Leuven,
Tiensestraat 102, 300 Leuven, Belgium

Introduction

Synchronisation is an ubiquitous aspect of human behaviour, important for skilled performance, social interaction and ensemble behaviour. When executing rhythmic tasks as diverse as dancing and rowing, a variety of cues to timing are available (Maduell and Wing 2007; Wing and Woodburn 1995). For instance, these cues may involve haptic input from the hands and body due to partner movements in dance or acceleration and deceleration surges of the boat in rowing, auditory information that comes from dance music or the call of the coxswain, or visual information that reflects movements of others engaged in the activity. Given a multiplicity of sensory timing cues, a key question is how does the central nervous system (CNS) utilise the information?

A problem faced by the CNS in synchronisation with a metronome is timing variance arising from variability in sensory registration of each event, in timekeeping, and in motor implementation (Wing 2002; Wing 1980; Wing and Kristofferson 1973). Each of these sources of variance affects participants' ability to match the times of their motor responses with the metronome events. Compensation for the asynchronies between motor responses and metronome events is required to ensure that the responses remain in phase and this has been modelled in terms of first-order linear phase correction (Vorberg and Schulze 2002; Vorberg and Wing 1996). In this model, the time to the next motor response is adjusted in proportion to the asynchrony between the previous motor response and corresponding metronome event. The mean and variance of asynchrony are then functions of afferent and motor variances and the constant of proportionality (correction gain). This model successfully accounts for action synchronisation in a variety of settings including synchronising with a periodic (Vorberg and Schulze 2002) or variable (Repp 2000, 2001) metronome.

The success of the linear phase correction model emphasises the importance of sensory cues to synchronisation in the control of timed behaviour. Synchronisation tasks often afford multiple cues to asynchrony which raises the important question of how are the various cues utilised? Previously it has been shown that, when visual and auditory pacing stimuli occur at different phases, synchronisation responses are drawn to the auditory events. This led to the suggestion that, in synchronisation, auditory signals dominate visual timing cues (Aschersleben and Bertelson 2003; Repp and Penel 2002). However, in the identification of spatial attributes of a stimulus, such as size or location, a model based on maximum likelihood estimation (MLE) suggests that, when multiple sensory cues are available, the CNS combines them by weighting them according to their relative reliability (van Beers et al. 1999; Ernst and Bülthoff 2004; Ernst and Banks 2002). The effect of such combination of cues is a reduction in variance of the underlying sensory representation, as evidenced through improved discrimination performance. The model also predicts a shift of the mean of the underlying distribution towards the more strongly weighted cue.

Could a cue combination model also apply to sensory cues for timing? Suppose that presenting both auditory and visual metronome cues gives rise to two perceived asynchronies. The shift in mean asynchrony, previously attributed to auditory dominance, might then reflect combination of asynchrony information from both cues, but weighted in favour of the auditory source. This account has some plausibility given that the auditory cue may be more reliable since the variance of tapping has previously been shown to be more variable with a visual metronome than with an auditory metronome (Kolers and Brewster 1985). However, more conclusive support for the cue combination model requires data on the variance of asynchrony, as well as its mean, in the paired metronome condition and this was the purpose of the present study. Because haptic stimulation (imposed movement) of one hand has previously been shown to have a pronounced effect on voluntary timing of the other hand (Ridderikhoff et al. 2005), we evaluated synchronisation with haptic and auditory metronomes. We wished to determine whether synchronisation variability would be less when haptic and auditory sensory cues were both available than when just one cue alone was provided. In addition we included conditions in which variance was added to the timing of the auditory cue to see if it would lead to greater dependence on the haptic cue.

Materials and methods

Participants

Five male and three female right-handed volunteers (mean age 33.7 years, standard deviation (SD) 10.7 years) took part in the study. Participants had no formal musical training and reported no auditory or neurological impairments. The experimental protocol conformed to the requirements of the School of Psychology human ethics committee.

Apparatus

Participants sat at a table and rested their left arm on an armrest. In the haptic condition, passive movements of the left index finger were produced by a lightweight robot (Phantom 1.5, SensAble Technologies, MA, USA) with thimble enclosing the finger tip. The robot was programmed to move the finger 20 mm vertically up and down, producing alternating extension and flexion movements at the metacarpophalangeal (MCP) joint. The trajectory of the passive movement (Fig. 1a) comprised brisk depress and release phases and approximated the form of finger-tapping trajectories recorded in previous studies (Doumas and Wing 2007; Semjen and Summers 2002). Auditory tones (frequency 1 kHz, duration 50 ms) were presented binaurally through headphones. Both auditory and touch stimuli were presented at an interstimulus interval of 600 ms. Participants tapped with the right index finger on a metal plate mounted on a force transducer (F241, Novatech Measurements, Sussex, UK). Force recordings made at 1 kHz yielded times of finger contact. The sound of tapping was not audible while wearing the headphones.

Procedure

Participants were instructed to be as accurate as possible in tapping their right index finger in synchrony with a metronome provided by auditory signals, haptic signals or both presented simultaneously. The auditory-alone metronome consisted of a series of auditory pulses with mean interstimulus interval (ISI) of 600 ms. The reliability of the auditory metronome was manipulated by adding temporal jitter (noise) drawn from a Gaussian distribution with mean zero (so that the mean phase was unchanged) and SD of 0, 50 or 100 ms. The haptic-alone metronome consisted of a series of regular passive flexion–extension movements of the left index finger (ISI 600 ms). The combined metronome consisted of a series of auditory pulses that coincided (in the mean) with peak flexion (downward) velocity of passive flexion–extension movements (ISI 600 ms) imposed on the left index finger (Fig. 1a). As with the

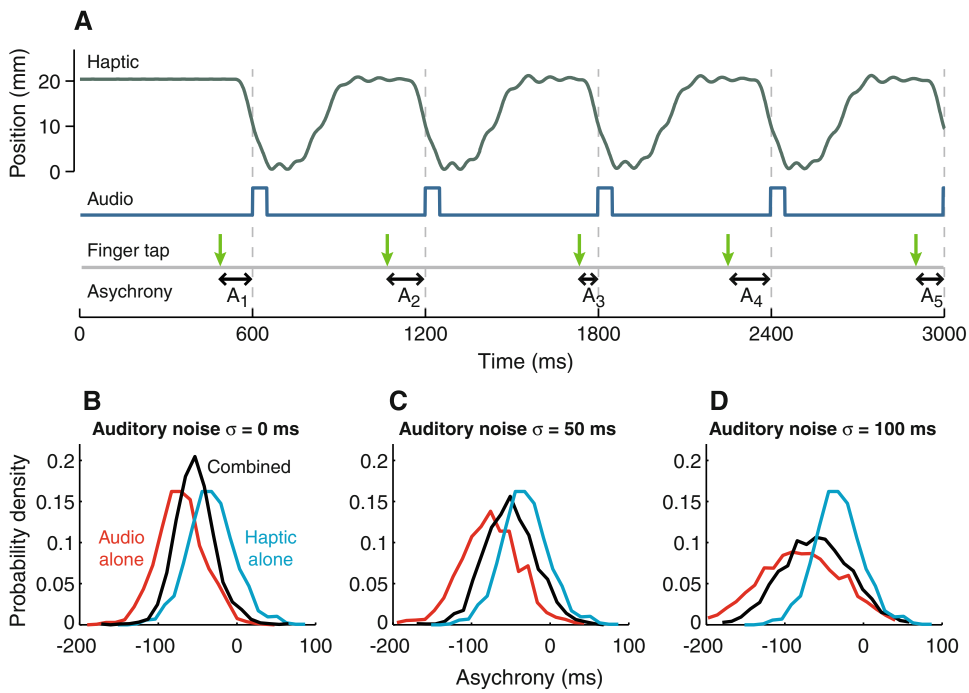

Fig. 1 **a** Synchronisation with the combined metronome. Passive movement trajectories presented to the left hand (*grey line*) and auditory tones (*blue line*). The finger-tapping responses are depicted by *green downward* arrows. We quantified tapping responses by the temporal asynchrony (A_n) between the tap and the stimulus at a given motor response (n). In the case of conditions involving the haptic and combined metronomes, asynchronies were defined as the temporal offset between the tap and the point of maximum velocity of passive movement in the flexion (downward) phase. **b** Probability density functions (mean of all subjects) of the asynchrony for the auditory metronome with no external noise (*red line*), haptic (*blue line*) and combined (*black line*) metronome tasks. **c**, **d** Asynchrony probability density functions for the corresponding conditions with noise introduced to the auditory metronome (**c**: SD = 50 ms; **d**: SD = 100 ms) (color figure online)

auditory-alone metronome, temporal jitter with 0, 50 or 100 ms SD was added to the auditory signal to produce different combined metronome conditions. Participants provided tapping responses in each of the seven experimental conditions (three auditory alone; one haptic alone; three auditory and haptic).

The experiment was carried out in one session of five blocks. Each block included seven trials with 50 metronome events spanning 30 s, one from each condition, presented in random order. Two practice blocks were performed to ensure that participants were familiar with the tasks and, in the case of the haptic metronome, the experimenter made sure that the movements were passive by checking that no appreciable resistive forces were registered by the robot.

Results

We quantified synchronisation behaviour using the stimulus–response asynchrony, i.e., the temporal offset between each finger tap and the auditory and/or haptic pulse defining the metronome. The distributions of asynchronies produced by participants differed according to the type of metronome (Fig. 1b–d). Large negative onset asynchronies (NOAs) were observed in the auditory-alone metronome conditions, i.e., finger taps tended to precede the metronome by a considerable margin.

In contrast, synchronisation with the haptic metronome resulted in smaller NOAs (Fig. 1b; $F_{1,7} = 8.829, p = 0.021$). This difference in NOA is consistent with afferent conduction delays that are shorter for auditory than haptic signals (Aschersleben 2002).

Following expectations based on the MLE model, under combined metronome conditions, the mean NOA lies between the values observed for the auditory and haptic conditions (Fig. 2b). Moreover, while the variability of the NOAs was not different for auditory and haptic conditions ($F_{1,7} < 1$, $p = 0.842$), we observed (Fig. 2a) lower variability in the combined metronome condition compared with either modality alone (haptic $F_{1,7} = 4.438$, $p = 0.037$; auditory $F_{1,7} = 3.464, p = 0.053$), consistent with the MLE model prediction of reduced variance through sensory combination.

These results for the mean and SD of the asynchronies suggest that, when temporal information is available from two modalities, the CNS combines the available signals rather than locking onto a dominant auditory channel.

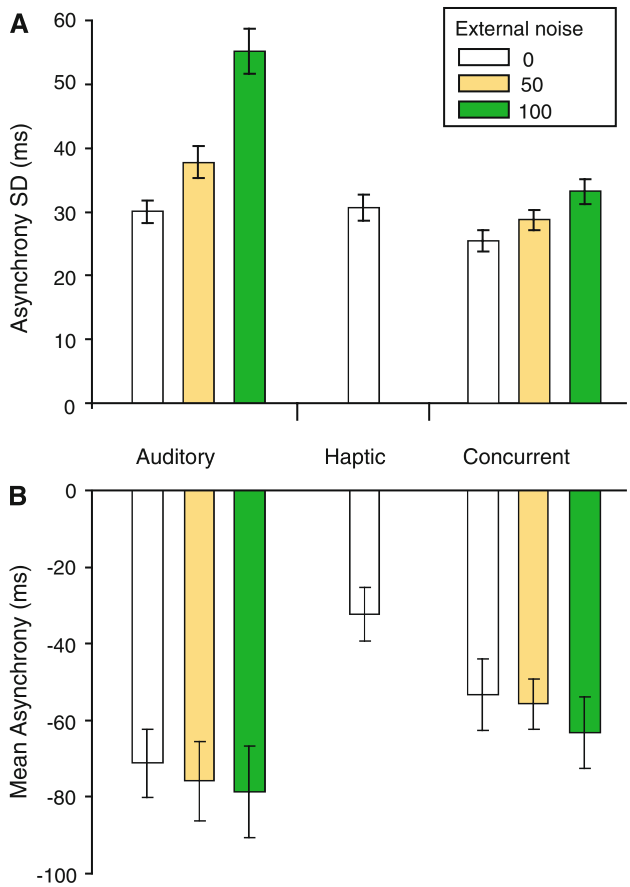

Fig. 2 SD (**a**) and mean (**b**) asynchrony for unimodal auditory, touch conditions and bimodal concurrent auditory and tactile conditions, where the audio cue was subject to 0, 50 or 100 ms jitter

We further assessed the MLE model by experimentally adding temporal noise to the auditory metronome to reduce its reliability (Fig. 1c, d). Figure 2b shows that, after adding noise, variability of synchronisation increased as a function of noise level ($F_{1,7} = 41.511$, $p < 0.01$). Performance in the auditory condition was more variable than in the combined condition ($F_{1,7} = 47.747$, $p < 0.01$) and the difference in variability between auditory and combined conditions increased with temporal noise level as shown by a significant noise by task interaction ($F_{2,14} = 12.857$, $p < 0.01$), reflecting greater variability reduction at higher noise levels in the combined condition, thus providing further evidence for improved timing in synchronisation with multiple cues. However, estimates of variability in the combined conditions when external noise was added to the auditory metronome were not reliably lower than in the haptic condition ($F_{1,7} < 1$, $p = 0.384$; $F_{1,7} = 1.309$ $p = 0.290$). Moreover, the temporal noise had no effect on mean asynchrony (Fig. 2b; $F_{1,7} = 2.307$, $p = 0.173$), contrasting with the expectation, based on the MLE model, that adding noise to the auditory metronome would increase the weight assigned to the haptic signal and result in a reduction of the mean asynchrony relative to the haptic modality.

In assuming weights are assigned in proportion to the reliability of the sensory channels being combined, the MLE model assumes that subjects are in possession of full information about channel reliability. However, in the present study, conditions were randomised. Thus, it might be thought that subjects would have required time to establish channel reliability. However, an analysis in which trials were analysed separately for first and second halves of each trial revealed the same effects as in the combined data. There was no tendency for the data later in the trial to be more strongly supportive of the MLE model.

Discussion

In the perception of spatial attributes, the accuracy of sensory estimation is improved by the combination of information from haptic and visual modalities. Thus, it has been shown that judgments of position (van Beers et al. 1999) and size (Ernst and Banks 2002) are more accurate when information is available from touch and vision, and positional judgments improve by combining auditory and visual information (Alais and Burr 2004). Such findings have led to the suggestion of a statistically optimal account of sensory integration, based on MLE, which assumes the sensory sources are independent and assigns them differential relative weights according to their reliability (i.e., inverse of the variability). If the variability of one information source increases, the relative weight shifts to the other. In such combination of sensory cues, the variability of the combined estimate is lower than that of either individual source. Moreover, if the estimates from the two sources differ in the mean, the average of the combined estimate lies between the two separate estimates, tending towards the more reliable one of the two sources.

In this paper, we have provided evidence for improved synchronisation resulting from the combination of different sensory cues to timing. Participants synchronised right index finger taps with auditory (brief tone pulses) and haptic (imposed movement of the left index finger) stimuli presented separately or concurrently. The mean asynchrony in the basic combined condition (haptic and audio without temporal jitter) lay between the asynchrony observed in each of the individual cue conditions, approximating predictions of the MLE model made from performance based on each cue alone. Also in agreement with the MLE model, the asynchrony variance was reduced in the basic combined condition compared to the variance in either of the single metronome conditions.

Although the results from the basic condition were consistent with the MLE model, a discrepancy was apparent when temporal jitter was added to the auditory stimulus. Under the MLE model, it was expected that the addition of jitter would result in the mean asynchrony converging on the haptic stimulus and the variance would move towards the value seen when synchronising with the haptic stimulus alone. Against this prediction, in the combined metronome conditions, the addition of jitter resulted in the mean asynchrony moving towards that for the auditory metronome, rather than that for the haptic metronome. Moreover, the variance of asynchrony tended to increase above (albeit not significantly), that for the haptic metronome, which is also not consistent with the MLE model. Taken together, the findings for the jittered auditory metronome suggest that the synchronisation was more sensitive to the manipulation of the auditory modality than would have been expected under the MLE model.

We suggest there are a number of possible reasons for the discrepant variance findings. The first relates to the assumption of independence of the timing cues being combined. In the MLE model, if the cues are correlated, the obtained reduction in variance is less than if they are independent (Oruc et al. 2003). Thus, at the limit, if the noise associated with each of two signals is identical (correlation of $+1$), there is no gain from using the combined signals compared to either signal alone. In the present case, there is reason to expect a correlation between the auditory and haptic asynchronies, because the two asynchronies have a single tap response as a common boundary. Thus, any variability in the associated proprioceptive afferent delay will result in common variance, and hence correlation, between the two asynchronies.

A second interpretation of the discrepancy between the effects of the jittered auditory cue and the prediction of the MLE model is that the weight given to audition was greater than under the model. Thus, the results could suggest that equal weight was given to auditory and haptic events, instead of weighting favouring the haptic modality. In effect, the subject may have been judging response asynchrony relative to a "virtual metronome" event defined by the midpoint of the interval between auditory and haptic events. This might indicate subjects have difficulty in rejecting the input from audition reflecting a bias for using timing signals from the auditory domain, established through long-term experience (Aschersleben and Bertelson 2003; Repp and Penel 2002).

A third account of the discrepancy from the predicted MLE effect relates to the costs and benefits associated with responding too early or too late. In everyday settings, movements are associated with a benefit (achieving the goal by moving within a time window centred on the synchronisation event) and a cost associated with moving too early or too late. (The facility to make corrective adjustments for early movements typically means that the cost associated with being too late is considerably greater.) This utility (cost–benefit) function should guide the participant in making their response; however, the movement chosen to maximise the gain will depend on the variability associated with that movement. In particular, more variability in movement production will prompt the participant to choose an earlier synchronisation point to avoid the possibility of moving into the cost region of the utility function. As variability increases, so the participant should programme their movement earlier in time (Mamassian 2008). In our case, the variability associated with the auditory metronome will cause synchronisation errors that may prompt observers to move earlier to avoid responding at time points normally associated with a cost.

While these various possibilities point to the need to investigate further the application of the MLE model to cue combination in synchronisation, our results do provide clear evidence for multisensory integration. In our study, the auditory stimulus involved a well-defined discrete event, but the haptic stimulus was relatively smooth and continuously varying in nature. In future research, it would be interesting to determine whether the weighting for the haptic stimulus might be increased if it included a more clearly marked event. For example, it has been shown in synchronising cyclic finger movements with an auditory stimulus, that asynchrony variance in tapping is less with contact compared to no contact during up–down finger movements (Elliott et al. 2009).

In summary, we have shown that the CNS uses sensory information from two modalities, auditory and touch, to improve synchronisation performance when both signals are available concurrently.

Acknowledgments This work was supported by the Medical Research Council (grant no. G9901257) and the Biotechnology and Biological Sciences Research Council (grant nos C520620 and E0095651). MD is a postdoctoral fellow of the Fund for Scientific Research (FWO) Flanders. We thank Nick Roach for assistance in setting up the haptic stimulus apparatus.

References

Alais D, Burr D (2004) The ventriloquist effect results from near-optimal bimodal integration. Curr Biol 14:257–262

Aschersleben G (2002) Temporal control of movements in sensorimotor synchronization. Brain Cogn 48:66–79

Aschersleben G, Bertelson P (2003) Temporal ventriloquism: cross-modal interaction on the time dimension. 2. Evidence from sensorimotor synchronization. Int J Psychophysiol 50:157–163

Doumas M, Wing AM (2007) Timing and trajectory in rhythm production. J Exp Psychol Hum Percept Perform 33:442–455

Elliott M, Welchman AE, Wing AM (2009) Being discrete helps keep to the beat. Exp Brain Res 192:731–737. doi:10.1007/s00221-008-1646-8

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415: 429–433

Ernst MO, Bülthoff HH (2004) Merging the senses into a robust percept. Trends Cogn Sci 8:162–169
Kolers PA, Brewster JM (1985) Rhythms and responses. J Exp Psychol Hum Percept Perform 11:150–167
Maduell M, Wing AM (2007) The dynamics of ensemble: the case for flamenco. Psychol Music 35:591–627
Mamassian P (2008) Overconfidence in an objective anticipatory motor task. Psychol Sci 19:601–606
Oruc I, Maloney LT, Landy MS (2003) Weighted linear cue combination with possibly correlated error. Vision Res 43:2451–2468
Repp BH (2000) Compensation for subliminal timing perturbations in perceptual-motor synchronization. Psychol Res Psychol Forschung 63:106–128
Repp BH (2001) Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. J Exp Psychol Human Percept Perform 27:600–621
Repp BH, Penel A (2002) Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. J Exp Psychol Human Percept Perform 28:1085–1099
Ridderikhoff A, Peper CL, Beek PJ (2005) Unraveling interlimb interactions underlying bimanual coordination. J Neurophysiol 94:3112–3125
Semjen A, Summers JJ (2002) Timing goals in bimanual coordination. Q J Exp Psychol A 55:155–171
van Beers RJ, Sittig AC, Gon JJ (1999) Integration of proprioceptive and visual position-information: an experimentally supported model. J Neurophysiol 81:1355–1364
Vorberg D, Schulze HH (2002) Linear phase-correction in synchronization: predictions, parameter estimation, and simulations. J Math Psychol 46:56–87
Vorberg D, Wing A (1996) Modelling variability and dependence in timing. In: Keele S, Heuer H (eds) Handbook of perception and action. Academic Press, New York, pp 181–262
Wing AM (1980) The long and short of timing. In: Stelmach GE, Requin J (eds) Tutorials in motor behavior. North Holland Publishing, Amsterdam
Wing AM (2002) Voluntary timing and brain function: an information processing approach. Brain Cogn 48:7–30
Wing AM, Kristofferson AB (1973) Response delays and timing of discrete motor responses. Percept Psychophys 14:5–12
Wing AM, Woodburn C (1995) The coordination and consistency of rowers in a racing eight. J Sports Sci 13:187–197