

Yarrow, K., Sverdrup-Stueland, I., Roseboom, W. & Arnold, D. H. (2013). Sensorimotor temporal recalibration within and across limbs. *Journal of Experimental Psychology: Human Perception and Performance*, 39(6), pp. 1678-1689. doi: 10.1037/a0032534



**CITY UNIVERSITY
LONDON**

[City Research Online](http://www.city.ac.uk/researchonline)

Original citation: Yarrow, K., Sverdrup-Stueland, I., Roseboom, W. & Arnold, D. H. (2013). Sensorimotor temporal recalibration within and across limbs. *Journal of Experimental Psychology: Human Perception and Performance*, 39(6), pp. 1678-1689. doi: 10.1037/a0032534

Permanent City Research Online URL: <http://openaccess.city.ac.uk/3034/>

Copyright & reuse

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

Versions of research

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

Enquiries

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at publications@city.ac.uk.

The APA has copyright on this article. When published, it can be found at:

<http://www.apa.org/pubs/journals/xhp/index.aspx>

This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

Sensorimotor temporal recalibration within and across limbs

Kielan Yarrow ^{1*}, Ingvild Sverdrup-Stueland ¹, Warrick Roseboom ² & Derek H. Arnold ³

¹ *Department of Psychology, City University London*

² *NTT Communication Science Laboratories*

³ *School of Psychology, The University of Queensland*

* Author for correspondence:

Kielan Yarrow,
Social Science Building,
City University,
Northampton Square,
London EC1V 0HB

Tel: +44 (0)20 7040 8530

Fax: +44 (0)20 7040 8580

Email: kielan.yarrow.1@city.ac.uk

Abstract

Deciding precisely when we have acted is challenging, as actions involve a train of neural events spread across both space and time. Repeated delays between actions and consequent events can result in a shift, such that immediate feedback can seem to precede the causative act. Here we examined which neurocognitive representations are affected during such sensorimotor temporal recalibration, by testing if the effect generalises across limbs, and whether it might reflect altered decision criteria for temporal judgements. Hand or foot adaptation phases were interspersed with simultaneity judgements about actions involving the same or opposite limb. Shifts in the distribution of participants' simultaneity responses were quantified using a detection-theoretic model, where a shift of both boundaries together gives a stronger indication that the effect is not simply a result of decision bias. By demonstrating that temporal recalibration occurs in the foot as well as the hand, we confirmed that it is a robust motor phenomenon: Both low and high boundaries shifted reliably in the same-limb conditions. However, in cross-limb conditions only the high boundary shifted reliably. These two patterns are interpreted to reflect a genuine change in how the time of action is represented, and a timing criterion shift, respectively.

Keywords: Temporal recalibration, temporal judgement, adaptation, synchrony, action

Consider a falling apple. Our subjective experience of this well-known prompt for scientific insight is seamless and unified; the apple appears to fall as a coherent composite of colour, form and motion, before striking the ground with a dull thud. However, from the perspective of a homunculus (or indeed a neuroscientist) looking down upon the brain's activity in response to this sequence of events, the view is of a time-smeared and spatially distributed cacophony of electro-chemical signalling. This viewpoint leaves us wondering how the observer is able to decide with confidence that this particular auditory thud occurred at the same time as that particular visual collision.

Science has not yet provided a compelling answer to this question. There is, however, evidence to suggest that recent experience plays an important role, or, put another way, that the temporal relationships that constitute simultaneity can be learnt and relearnt. Several methods exist to try and establish the relative time at which two events appear maximally synchronous (known as the point of subjective simultaneity or PSS). The most common methods require that participants report either the order of two events (a temporal order judgement; TOJ) or whether two or more events occurred synchronously or asynchronously (a simultaneity judgement; SJ). Intriguingly, when brief auditory and visual stimuli are repeatedly presented slightly out of synch during a period of adaptation, participants subsequently change their judgements in these tasks, consistent with their having developed a new opinion about the most synchronous relationship (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004). This effect is known as temporal recalibration.

Assuming temporal recalibration represents a functional adaptation, it would make sense if the recalibration were a selective process. Humans exist in a cluttered environment, with many sensory signals originating in close spatial and temporal proximity. One would not want to adapt to spurious temporal correlations. Rather, one would want temporal recalibration to be selective for events that have a common

cause, like the sight and sound of the apple striking the ground. There are findings, both specific to temporal recalibration and more general to cross-modal timing perception, which point to this. For instance, auditory and visual events likely to be related, such as flashes of light and tonal beeps with a common spatial locus, can prove harder to discriminate in time relative to unrelated pairs (Aschersleben & Bertelson, 2003; Spence, Baddeley, Zampini, James, & Shore, 2003; Spence, Shore, & Klein, 2001). Moreover, rather than getting worse, the precision of timing judgments can actually be enhanced by additional events. An isolated pair of visual or audio-visual events, for instance, can seem synchronous at greater timing differences than the same events accompanied by additional audio or visual events, so long as the additional events promote separation of the timed events via perceptual grouping (Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Roseboom, Nishida, & Arnold, 2009; Scheier, Nijhawan, & Shimojo, 1999). In combination, these findings suggest that multi-sensory events adjudged likely to be related, either because of experience or because they are the only two events present, tend to be grouped perceptually across greater physical timing differences than other sensory signals.

In relation to temporal recalibration, it has been established that the probability of recalibration is modulated by the spatial and temporal proximity of the adapting events (Fujisaki et al., 2004; Heron, Roach, Hanson, McGraw, & Whitaker, 2012; Roseboom & Arnold, 2011; Yarrow, Roseboom, & Arnold, 2011, Yuan et al., 2012; but see Keetels & Vroomen 2007). This is presumably because proximity in both time and space suggests a common cause. Moreover, it has been shown that simultaneous, opposite, temporal recalibrations can be maintained at once (Heron et al., 2012; Roseboom & Arnold, 2011). For instance, people can adapt to a male actor whose lip movements lag a soundtrack and to a female actor whose lips precede the soundtrack, resulting in appropriate shifts of audio-visual timing perception for each actor (Roseboom & Arnold, 2011). These identity-contingent temporal recalibrations establish that temporal recalibration can be a selective process.

Another clear cue indicating that two events have a common cause arises when one's own actions are directly responsible for both events, such as when you turn on a light and see both the light turn on and hear the click of the switch. It is well established that the apparent timing of volitional actions and consequent sensory events, be they natural or augmented, are mutually attractive (e.g. Haggard, Clark, & Kalogeras, 2002; Hughes, Desantis, & Waszak, 2012; Yarrow, Haggard, Heal, Brown, & Rothwell, 2001; Yarrow, Haggard, & Rothwell, 2010; Yarrow & Rothwell, 2003), so one might reasonably predict that action should give rise to a particularly potent form of temporal recalibration. This intuition was confirmed in a study by Stetson, Cui, Montague, and Eagleman (2006; see also Cunningham, Billock, & Tsou, 2001), who described an effect we will refer to as *sensorimotor* temporal recalibration.

In Stetson et al.'s (2006) experiments, participants pressed a key, followed on the majority of trials by a flash of light with a predictable (either short or long) delay. The remaining trials had a variable delay, to permit the authors to draw inferences about how participants perceived the timing between an action and its sensory consequence. Temporal order judgements revealed that the point at which there was greatest uncertainty regarding the order of action and flash differed depending on context: The point of subjective simultaneity fell at a larger action-flash delay when the majority of trials had a long (rather than short) inserted delay. This effect was also greater in magnitude than that obtained in a control condition, which approximated traditional sensory-sensory temporal recalibration by having the key move to press the participant's finger, rather than the other way around.

Sensorimotor temporal recalibration of this kind has now been demonstrated in several labs. It can be obtained with either TOJs or SJs (Heron, Hanson, & Whitaker, 2009), and with delayed transient feedback signals encoded in various sensory modalities (Heron et al., 2009; Sugano, Keetels, & Vroomen, 2010) or indeed using delayed video feedback of the action itself (Keetels & Vroomen, 2012). Sensorimotor temporal

recalibration is marked by a similar temporal tuning relative to sensory-sensory temporal recalibration (Heron et al., 2009; Stetson et al., 2006), with the critical interval that determines whether it occurs being the time from movement termination until the feedback stimulus (Arnold, Nancarrow & Yarrow, 2012). It has been demonstrated mainly with manual actions, but can also be obtained for vocalizations (Yamamoto & Kawabata, 2011). Perhaps most interestingly, sensorimotor recalibration seems to generalise almost perfectly when the adapting feedback stimuli and subsequent test are different coloured flashes (Stetson et al., 2006) or even come from different sensory modalities (Heron et al., 2009; Sugano et al., 2010), at least when assessed using a temporal judgement task (Sugano, Keetels, & Vroomen, 2012).

This latter finding (i.e. generalisation across sensory modalities) is interesting because it begins to address the representational level at which sensorimotor temporal recalibration might be occurring. The most conceptually straightforward accounts of sensorimotor temporal recalibration would involve a shift in the perceived time of either the action (forwards), or of the sensory consequence of action (backwards), or some combination of both. As recalibration can be similarly pronounced even when the sensory consequence is changed from adaptation to test, this implies that it is probably the perceived time of the action that shifts (because such a shift should be equally measurable with any sensory reference; Sugano et al., 2010). However, actions and events can be represented at multiple neurocognitive levels. For example, actions have long been considered to depend in part upon high-level “generalised” motor programs, which can represent abstract features of a movement independent of specific lower-level parameters – such as the particular end effector that is being moved (Keele, 1968; Lashley, 1951). Hence another conceivable account for the generalisation of recalibration would be adaptation of a high-level multimodal representation of the time at which feedback was delivered. Finally, we must also consider recent and apparently contrary evidence, suggesting that generalisation across sensory modalities is asymmetrical when recalibration is assessed with a synchronised tapping task (recalibration occurred for visual adaptation and auditory testing, but not vice versa; Sugano et al., 2012). This finding might instead imply

shifts in low-level sensory representations, with visual events perhaps undergoing internal translation into an auditory code to yield generalised adaptation for auditory test stimuli. Additional tests of generalisation are therefore desirable, because they could help to further constrain the plausible locus of adaptation.

Another, rather different approach to investigating the mechanism of recalibration can be sought by modelling how observers carry out the task, and then fitting the model to data from different adaptation conditions. Because the parameters of models carry psychological meaning, changes in a particular parameter can make explicit the particular process(es) underlying the change, albeit with the important caveat that the interpretation is only as valid as the model and its underlying assumptions. Recently, Yarrow, Jahn, Durant, and Arnold (2011) applied this approach to audiovisual temporal recalibration by developing a detection-theoretic model of the simultaneity judgement task. In essence, the model assumes that the two signals being compared each arrive at a decision centre on any given trial with a mean delay plus some Gaussian latency noise. A relatively simple decision mechanism is then sufficient to generate realistic psychometric functions. The model is illustrated and described further in Figure 1.

<INSERT FIGURE 1 AROUND HERE>

Models of this kind make a conceptual distinction between the noisy internal representation upon which a decision is based (in this case the temporal difference between two events, which may be biased by unequal physical and neural delays) and the way in which the decision is actually reached (here by establishing low and high criteria, and judging values in-between as synchronous). Unfortunately, in practice, the parameters that quantify such a model can trade off against one another, which can make it difficult to determine what changes have generated the recalibration. However, Yarrow, Jahn, Durant, and Arnold (2011) developed an argument based on parsimony to conclude that in their data set, recalibration

was most likely to reflect a shift in one decision criterion in order to account for recent sensory experience, rather than a shift in the time at which a sensory event was represented to occur.

With these two methods for characterising the mechanism of recalibration in mind, we opted to extend the generalisation approach previously applied to sensorimotor temporal recalibration by Heron et al. (2009) and Sugano et al. (2010) in a new direction, whilst additionally using our own recent modelling work to further dissect how recalibration might come about. Hence we designed an experiment to induce recalibration for actions involving one of two different end effectors (the hand or the foot), and tested whether recalibration would generalise to the other effector (i.e. hand to foot and vice versa). By fitting a meaningful model of the simultaneity judgement task to data in all conditions, we were able to draw additional inferences about the psychological mechanisms that gave rise to within-limb and cross-limb recalibration. To anticipate our results, we found evidence for both within and cross-limb sensorimotor temporal recalibration, but whereas within-limb recalibration showed characteristics favouring a shift in the represented time of action, cross-limb recalibration might be better explained by a relaxed decision criterion for distinguishing simultaneity from succession.

Methods

Design

Participants were randomly allocated to one of two groups in a mixed 2x2x2 design. The first (between-subject) factor, *adapting effector*, compared hand and foot adaptation groups. The second (within-subjects) factor, *adaption-test correspondence*, compared same-limb and cross-limb test trials. The third (within-subjects) factor, *delay*, manipulated the time between the action and the consequent auditory feedback

during adaptation (0 vs. 150 ms), in order to induce recalibration. Both of the within-subject factors were blocked, with each participant completing four blocks of 50 useable trials. The order of the same-limb and cross-limb conditions was counterbalanced across participants, with a baseline (0 ms delay) block always preceding a recalibration (150 ms delay) block.

Participants

Twenty-four participants (18 females, mean age 22.5, $\sigma = 3.0$) completed the experiment, mainly undergraduates receiving course credit. They reported normal or corrected-to-normal vision, normal hearing and sense of touch. Seven participants had to be excluded (see data analysis, below) leaving a final sample of 17 (14 females, mean age 21.7, $\sigma = 3.3$). Counterbalancing was however maintained largely intact in both groups (with 4/8 and 4/9 participants completing the same-limb test condition first in the hand and foot groups respectively). Ethical approval was overseen by the City University Department of Psychology Research Ethics Committee.

Apparatus and Stimuli

A PC connected to a 20-inch CRT monitor controlled a combination of two National Instruments A/D input/output cards (DAQCard-6715 with a BNC2110 breakout box & DAQPad-6015). This was used to trigger auditory feedback (10 ms long 1000 Hz pure tonal beeps generated at 44100 Hz) following button presses during adaptation. On test trials, the hardware generated identical beeps and recorded copies of both the button press and the beep (sampled at 1000 Hz) so that the asynchrony for that trial could be determined using an edge-detection algorithm. Beeps were presented from a small speaker. The hand button and foot pedal were digital buttons that did not emit clicks when pressed, but to further mask any auditory cues associated with hand and foot actions, white noise was presented via a pair of computer speakers, with loudness set at 70 decibels dBA (measured at typical head position using a sound level meter;

Cirrus Research PLC CR152A). The timing of feedback during adaptation sections (0 or 150 ms delay from button contact) was verified using a 20 MHz storage oscilloscope (Gould 214 DSO 1604).

Procedure

Participants sat at a desk facing the computer screen at a distance of ~50 cm. The hand button was placed on the desk, and the foot pedal was placed on the floor underneath. The speakers generating the various auditory stimuli were all located to the front/left relative to a participant's head. Participants rested their dominant hand/foot over the button(s) and used their free hand to press the keyboard decision keys (←/→).

Each block consisted of an adaptation phase and a top-up/test phase. A two-tone metronome chirp (contiguous 25 ms 4000/3000 Hz pure tones) cycled at 1.25 Hz at the beginning of the adaptation phase to train participants to maintain a constant tap interval of ~800 ms. The metronome terminated after the first ten taps, and participants then continued to tap rhythmically for a total of 80 presses (~1 min). After each tap, auditory feedback was presented at a constant time (0 or 150 ms delay). Participants were instructed to try and maintain a roughly constant inter-tap interval (ITI). If a participant's ITI fell below 600 ms a warning message ("*Slow down!*") appeared on screen.

The top-up/test phase started immediately after adaption (signalled by a repeat of the distinctive metronome chirp). Each test trial consisted of seven top-up tap-feedback pairs, using the same lag as in the initial adaption phase, and then an eighth test pair, where the auditory beep came on at a random time relative to the button press (see Figure 2a and below). Note that this protocol allowed us to sample auditory beeps that both preceded and succeeded the button press on the critical 8th test presentation. Participants judged if their final tap on each trial, and its accompanying tone, were simultaneous or not by

making an unspeeded 'yes'/'no' responses to an on-screen prompt. After the response, the next top-up/test trial was initiated by the participant. Unusable trials (e.g. where an asynchrony could not be automatically determined) were discarded and repeated until 50 trials were logged, terminating the block.

To generate a suitable range of asynchronies at test, the median ITI from the last three top-up taps was used to predict the time of the participant's final (test) press. A random value drawn from a discrete probability distribution with steps of 30 ms was added to this prediction. The initial shape of this distribution was, for the majority of participants, set based on piloting derived from the first few participants. It was loaded towards central values (-90 to +90) with a smaller probability of selecting more extreme negative values (-120 to -180) because participants tended to tap slightly earlier than expected. The distribution was updated after each trial based on that trial's asynchrony and the participant's response. If the asynchrony was positive and the participant responded "yes" regarding simultaneity, the distribution expanded to cover more positive values. If the participant responded "no", however, the distribution was increased for smaller values than the last test. The opposite rule was applied when the asynchrony was negative. This adaptive procedure was designed to generate test values spanning the range necessary to capture both of the transitions from perceived asynchrony to synchrony. The approach is a modified version of the generalized Pólya urn model (Rosenberger & Grill, 1997) proposed for efficient sampling for temporal order judgements.

In the same-limb conditions, all eight taps (i.e. top-ups and test) were conducted with either the hand or the foot, whereas in the cross-limb conditions the first seven taps were conducted with one limb and the final tap with the other. Participants were instructed to maintain focus on the screen, and to avoid looking at their hand/foot (to minimise the use of visual timing cues), and also to be conservative when using the

simultaneous response. A five-minute break was imposed after the first two blocks, yielding an overall experimental duration of around 50 minutes.

Data analysis

For each participant and condition, proportional simultaneity responses were plotted as a function of test asynchrony. Data were trimmed to exclude trials where asynchronies exceeded ± 600 ms (2.3%) to remove cases where participants might have judged the incorrect pair of stimuli. Trimmed data were maximum-likelihood fitted with a “specific threshold” model of the SJ task (Ulrich, 1987; Yarrow, Jahn, Durant, & Arnold, 2011) using MatLab (The MathWorks, Inc.). This model reaches a decision based on whether the difference in latencies between two signals (which are each assumed to have Gaussian noise) falls between two criteria for simultaneity. It generates a difference of cumulative Gaussian curve, and has four parameters: The low boundary, the high boundary, and the standard deviations of the cumulative Gaussian functions associated with each. Note that by contrast, the more common approach to fitting SJ data, which uses a Gaussian or truncated Gaussian function, has, to our knowledge, not been underpinned with any theoretical foundation.

In this experiment, the low boundary relates to the boundary between judging the auditory stimulus as preceding the action and judging these events as simultaneous. The high boundary relates to the boundary between judging the two events as simultaneous and judging the action as having preceded the auditory stimulus. These were the primary dependent variables. Additionally, under this model the slopes of the functions describing the low and high category boundaries are estimated. These can vary to reflect three sources of noise: Latency noise associated with the timing of sensorimotor events (and contributing to both boundaries), and two further sources of noise, reflecting separate trial-to-trial variability in the placement of low and high criteria (See Figure 1). All three sources of noise cannot be uniquely estimated, but two

further psychologically meaningful measures can be estimated: An upper limit on the differential latency noise (equal to the smaller of the standard deviations associated with the low and high boundaries) and the difference in trial-to-trial variance between high and low criteria. The inclusion of separate measures of slope to capture noisy criteria has been justified previously based on a significant improvement in model fit for audio-visual simultaneity judgements (Yarrow, Jahn, Durant, & Arnold, 2011).

The deviance for each model fit was estimated for each participant in each condition. This value was compared with that arising from a fit to a simpler two-parameter model incorporating only a single (standard/inverse) cumulative Gaussian. This simpler model can capture cases where participants either fail to make meaningful discriminations (i.e. produce flat functions) or show a transition at just one or other boundary between judgements of simultaneity and successiveness (i.e. produce ascending or descending sigmoidal functions) but fails to capture the full inverse-U-shaped functions expected from the SJ task. It therefore represents an appropriate null model to exclude cases where sampling of test values was insufficient at *either* boundary to yield sensible parameter estimates. For nested models, deviance is expected to decrease as parameters increase even when the more complex model does not represent a real improvement. We therefore retained participants only when the improvement of fit in all four conditions fell above the 80th percentile of the appropriate null distribution (chi-squared, with 2 DF; Wichmann & Hill, 2001). As previously noted, seven participants were rejected on this basis.

For the remaining sample, standard two-tailed parametric tests were used to assess differences in model parameters across conditions at the population level. We also conducted analyses at the individual-participant level by running permutation tests (based on 999 random data shuffles and re-fits) to compare low/high boundary estimates between baseline (0 ms delay) and adapted (150 ms delay) conditions. For

these tests we applied a one-tailed alpha of 0.05 to categorise participants as either showing or not showing reliable recalibration.

Results

Analysis of low and high boundary data

Figure 2 shows data from one participant who took part in the *adapt-hand* conditions, selected because their data shows similar trends to the group as a whole. The x-axis shows asynchrony between the test button/pedal press and the test beep, with positive values denoting that beep followed press. For this participant, with same-limb testing, the whole distribution of simultaneous responses shifted rightwards following 150 ms delay adaptation relative to the 0 ms baseline (low boundary permutation test $p = 0.003$; high boundary = 0.005). This suggests a possible shift in the perceived time of either the action or the beep. However, for cross-limb testing only the high boundary was robustly shifted ($p = 0.006$) suggesting a selective relaxation of the action-then-sound criterion when judging simultaneity (see discussion).

<INSERT FIGURE 2 AROUND HERE>

Although it is common to report a single point of subjective simultaneity in studies using an SJ task, we retain the distinction between low and high boundary estimates here because they better reflect our model and can be individually informative about the psychological basis of any recalibration (Yarrow, Jahn, Durant, & Arnold, 2011). The model we use has an explicit theoretical basis that gives meaning to the parameters of the fit. Previous researchers have often fitted their data to estimate a single PSS, but the functions they use (e.g. Gaussians and truncated Gaussians) tend to have no explicit modelling basis. Thus, while it may *appear* that such fits provide a simpler or less theoretically laden interpretation of the data, in our opinion they do

so only by obfuscating the basis upon which a PSS has been estimated (and thus what this value might actually mean). However, for comparability with past reports, we provide an alternative analysis based on this parameter in our appendix.

Figure 3 shows average data for both adapt-hand and adapt-foot groups in all conditions. In the upper part of each panel, group mean low and high-boundary estimates are shown following adaptation to either a 0 ms delay or a 150 ms delay. The x-axis once again denotes asynchrony between the test button/pedal press and the test beep, with positive values denoting that beep followed press. The lower part of each panel summarises the mean magnitude of recalibration (i.e. the difference between the 0 and 150 ms adaptation conditions) again for both low and high boundaries.

<INSERT FIGURE 3 AROUND HERE>

In general, in baseline (0 ms delay) conditions, the distribution of simultaneous responses favoured positive asynchronies. This could reflect a baseline delay in the awareness of the action relative to the beep. Alternatively, under our model, the criterion for judging actions and beeps as simultaneous might have been more liberal when beeps followed button presses than vice versa. Both delayed and anticipatory awareness of action have been reported in the past, reflecting a range of methodologies applied to this question (see Yarrow & Obhi, in press, for a short review).

More critically, relative to this baseline condition, recalibration following adaptation to a 150 ms delay appeared substantial, affecting both low and high boundaries. This was confirmed with a 2 (adapting effector) x 2 (adaptor-test correspondence) x 2 (delay) mixed-measures ANOVA, applied separately to low and high boundary data, which revealed only main effects of delay (low boundary: $F = 21.5$, $df = 1, 15$, $p <$

0.001; high boundary: $F = 46.4$, $df = 1, 15$, $p < 0.001$) plus a main effect of adaptor-test correspondence specifically for the high boundary ($F = 18.1$, $df = 1, 15$, $p = 0.001$). This latter effect may suggest that the requirement to cross over from one limb to the other at test induced a more relaxed high criterion (both before and after adaptation to a 150 ms delay).

Although there were no two and three-factor interactions within the overall ANOVAs, given that three of the four combinations of adaptation and test actions investigated here have not been tested before, we also considered whether adaptation was individually robust in these new conditions. Paired t-tests revealed that recalibration was consistently significant at the high boundary (hand-hand $t_7 = 5.0$, $p = 0.002$; hand-foot $t_7 = 3.4$, $p = 0.011$; foot-foot $t_8 = 7.0$, $p < 0.001$; foot-hand $t_8 = 2.3$, $p = 0.048$) but *not* at the low boundary, where it was robust for the within-limb conditions (hand-hand $t_7 = 9.7$, $p < 0.001$; foot-foot $t_8 = 4.8$, $p = 0.001$) but failed to reach significance for either of the cross-limb conditions (hand-foot $t_7 = 1.8$, $p = 0.117$; foot-hand $t_8 = 0.8$, $p = 0.455$) or indeed when these conditions were pooled together ($t_{16} = 1.9$, $p = 0.081$). Hence low-boundary shifts seem to be less robust than high-boundary shifts, specifically for cross-limb conditions.

Although our experiment included only 50 trials per subject per condition, we also considered how often recalibration was statistically significant for each boundary at the individual participant level (using permutation tests). For the high boundary, recalibration was significant for the majority of participants in both same-limb (14/17) and cross-limb (11/17) conditions, with no evidence that recalibration was more common in one case than the other (sign test $p = 0.453$). By contrast, for the low boundary, recalibration was significant for most participants only in the same-limb conditions (13/17), not in the cross-limb conditions (2/17), a difference that was itself statistically reliable ($p = 0.007$) and thus reinforced the results of our group t tests.

Possible confounding effects of stimulus distributions

The stimulus levels (i.e. the range of asynchronies that were judged) were not fixed in our experiment, but rather varied for each participant and condition (see Figure 2 for an example). This partly reflected the adaptive method we used to try and capture a sufficient range of asynchronies to obtain a useable fit, but also the fact that we had to try and predict the likely timing of the final (test) press in each top-up / test trial, and present a stimulus accordingly. In fact, the asynchronies experienced by our participants differed systematically across conditions.¹ Overall, participants tended to press slightly earlier than expected (based on their immediately preceding top-up presses), but this bias was exaggerated (by around 80 ms) in cross-limb conditions compared to same-limb conditions, and mitigated (by around 35 ms) with adaptation to a 150 ms rather than a zero delay. Earlier button presses imply more test trials at positive asynchronies, which could have affected participants' responses. For example, they might have exhibited some bias towards using the two response categories equally, or taken account of the history of previous test asynchronies to form part of their current sensory estimate (as implied by Bayesian models of perception, e.g. Miyazaki, Yamamoto, Uchida & Kitazawa, 2006).

To address this issue, we re-ran our statistical analysis but included as a covariate the mean button bias (the mean actual time of the button press at test relative to the expected time predicted using the three preceding top-up taps) for each participant and in each condition. For the reanalysis of low-boundary data, this covariate was not a reliable predictor ($F = 1.0$, $df = 1, 59$, $p = 0.331$) and the ANCOVA yielded the same pattern of significance implied in our original ANOVA and t-tests (i.e. a main effect of delay, with pairwise comparisons finding significant effects for both of the within-limb conditions, but neither of the cross-limb conditions). However, for the reanalysis of high-boundary data, button bias was a reliable predictor ($F =$

¹ We thank Jean Vroomen for alerting us to this possibility, and for suggesting the supplementary analysis addressing the issue.

12.0, $df = 1, 59$, $p = 0.001$). Importantly, even when controlled statistically, the effect was to leave the main effect of delay intact (in fact, if anything, to enhance it, with all pairwise t-tests now showing p values ≤ 0.001) but to eliminate the less theoretically important main effect of adaptor-test correspondence that was found regardless of delay. This could suggest that swapping effectors only *appeared* to relax the high criterion via an indirect influence (i.e. by shifting the range of asynchronies that were consequently experienced) although an independent effect of our experimental manipulation on both of these variables is equally plausible.

Additional model parameters

The model fits provided two further measures derived from the slopes of cumulative Gaussians used to capture high and low boundaries: An upper limit on the differential latency noise ($\mu = 43$ ms, $\sigma = 21$) and a difference in trial-to-trial variance between high and low criteria ($\mu = 3441$ ms², $\sigma = 10717$). However, neither measure showed any significant main effects or interactions in a 2x2x2 ANOVA comparing values across conditions (all $p > 0.05$).

Discussion

Our participants were adapted to one of two temporal relationships between their own actions and the artificial auditory feedback triggered by those actions. Compared to a situation in which auditory feedback arrived immediately, adaptation to a 150 ms delay generated a change in the distribution of their judgements about simultaneity. This increased the likelihood that participants would judge test trials with a large positive asynchrony between action and beep as simultaneous, regardless of which limb was adapted and whether actions during adaptation and test were implemented by the same or different limbs (hand/foot). We also observed a corresponding *decreased* tendency to judge test trials with a negative

asynchrony between action and beeps as simultaneous (i.e. on trials when beeps preceded actions), but this was only true when the same limb implemented actions during adaptation and tests.

Our data are broadly consistent with previous reports of sensorimotor temporal recalibration (Cunningham et al., 2001; Heron et al., 2009; Keetels & Vroomen, 2012; Stetson et al., 2006; Sugano et al., 2010; Sugano et al., 2012; Yamamoto & Kawabata, 2011) which have shown how exposure to delayed feedback can shift the point of subjective simultaneity. Numerically, our effects are quite large (around 2/3rds of the additional 150 ms delay inserted between action and feedback), but sensorimotor temporal recalibration is known to be greater in magnitude relative to the sensory-sensory equivalent (Stetson et al., 2006) and experimental effects of comparable size to ours have been reported (c.f. Heron et al., 2009). To our knowledge we are the first to demonstrate recalibration for foot actions, or to assess generalisation of recalibration when testing with a different action to the one that underwent adaptation. Our data suggest that a form of recalibration does occur in these conditions, but it is of a somewhat different character to that obtained in standard conditions, as outlined in the next few paragraphs.

We fitted our SJ data with a specific threshold model (Ulrich, 1987; Yarrow, Jahn, Durant, & Arnold, 2011) based on detection-theoretic approaches to making temporal judgements (Allan, 1975; Gibbon & Rutschmann, 1969; Schneider & Bavelier, 2003; Sternberg & Knoll, 1973). This process suggested a reliable shift in the high boundary derived from the fitted model, regardless of whether the simultaneity judgements involved the same or opposite limb relative to adaptation, but a reliable shift in the low boundary only when adaptation and test actions were identical. Models of this kind imply that behavioural changes of the types observed here can happen for at least two reasons. First, the underlying representation of the time between two events can change, as if analyses of one signal had been given a head start. Second, the way the internal representation is interpreted could change. For example, where

once the observer chose to classify actions and feedback as simultaneous so long as the feedback came within 200 ms of the action (referred to as a high criterion *extent* of 200 ms), now they might make this classification using a relaxed criterion extent of, say, 300 ms.

Strictly speaking, it is very difficult to distinguish between these possibilities using a single SJ task, because changes to one or more decision criteria can trade off against shifts applied to the underlying representation.² However, some data patterns are more parsimoniously explained under one account than the other. When only one boundary shifts, we can explain it based on a change in just one decision criterion extent (e.g. a relaxation of the high criterion for simultaneity), whereas to explain it as a shift in the representation one must additionally invoke a shift/relaxation in the extent of the opposite decision criterion (to explain why the low boundary has not also moved). This second explanation is less parsimonious. Similarly, when both boundaries shift by a similar extent, it can be explained entirely by a shift in the underlying representation. To explain it with a change in decision criteria, one criterion extent must relax while the other contracts. Again, this is less parsimonious. Hence for our data relating to within-limb sensorimotor temporal recalibration, we favour an interpretation wherein recalibration is driven by a change in the underlying representation of the time between two events (expanded upon below) whereas we interpret cross-limb recalibration as largely resulting from a decision-level bias. As an aside, we note that the putative within-limb adjustment to the underlying representation of time for sensorimotor temporal recalibration contrasts with the decision bias account we reached when applying a similar approach to sensory-sensory (audio-visual) temporal recalibration (Yarrow, Jahn, Durant, and Arnold, 2011). In the sensory-sensory case, alternative experimental approaches have yielded data both consistent and inconsistent with a change in latency for one modality relative to the other (e.g. Di Luca, Machulla, & Ernst, 2009, and Roach, Heron, Whitaker, & McGraw, 2011, respectively).

² The two possibilities are impossible to distinguish with a TOJ task, and require extremely large shifts of the entire response distribution (relative to its width) under the SJ task; see Yarrow, Jahn, Durant, and Arnold (2011) for discussion.

Whilst we believe there is strong circumstantial evidence for our proposed dichotomy between processes of within-limb and between-limb sensorimotor temporal recalibration, our conclusions must remain cautious for at least four reasons. First, as outlined above, the argument rests on parsimony, which may end up being a poor guide. Second, inferences based on modelling are only as accurate as the model itself. However, detection-theoretic models have proved influential in Psychology and related fields, explaining numerous findings (Macmillan & Creelman, 2005), and specific threshold models like ours deal well with a range of temporal judgement data (Ulrich, 1987). Of course there may be data that our model cannot explain, and it may ultimately need to be expanded (for example by incorporating keying errors; Garcia-Perez & Alcalá-Quintana, 2012) or indeed supplanted entirely by a different type of model (e.g. Roach et al., 2011). At present, however, we consider our model to be as good a candidate as any.

The third reason for caution is statistical. Our data showed clear differences in terms of whether low-boundary shifts were reliable. However, our overall low-boundary ANOVA did not reveal an interaction between adaptation delay and adaptor-test correspondence. Hence we cannot state categorically that low-boundary recalibration differed reliably between our four conditions based on this analysis. Nonetheless, it is clear that a robust low-boundary effect was not obtained for cross-limb conditions, when considered separately from same-limb conditions. Furthermore, when we considered the frequencies with which individual participants showed reliable recalibration, this analysis did yield a significant difference between within-limb and between-limb recalibration conditions, thus mitigating the failure to find an interaction in the ANOVA.

Finally, we might also ask whether it is reasonable to see a decision bias arising in the cross-limb conditions, but not in the same-limb conditions. Where has it gone? We would argue that the decision bias is probably

a response to a perceived asynchrony. Hence if the underlying representation is adjusted by the adapting context, there is less reason for a decision bias to emerge. However, the pattern of means in Figure 3 does suggest somewhat greater effects of delay on the high boundary than the low boundary even for within-limb conditions. This suggests the existence of a residual decision bias here as well, alongside a change in the mapping from subjective to objective asynchrony.³

Caveats aside, what process(es) might underlie the compelling within-limb recalibration we observe? Recalibration of sensorimotor relationships has previously been shown to generalise across feedback sensory modalities when assessed using temporal judgement tasks (Heron et al., 2009; Sugano et al., 2010). This is consistent with either a shift in the perceived time of action, or with a shift in the perceived time of some high-level crossmodal representation of sensory event time. Here we found a recalibration that was suggestive of a true change in the representation of event time, but only when the adapted effector was used again at test. This argues against a shift in perceived time for a high-level representation of sensory event time, because such a shift would be measurable when comparing the affected sensory events with any action, not just the one used during adaptation. Hence, by elimination, data indicating the generalisation of sensorimotor recalibration to different modalities of sensory feedback favour a shift in the perceived time of action (as suggested by Sugano et al., 2010) but at the level of the individual effector. The neural representation of an action reflects a sequence of events, from abstract planning, to the generation of efferent signals, to the receipt of proprioceptive and tactile reafferents, with the perceived time of action likely to depend upon the combined internally represented timing of several of these events (Obhi, Planetta, & Scantlebury, 2009; Strother & Obhi, 2009; Yarrow & Obhi, in press). If the perceived time of action is indeed determined by some weighted average of the constituent events, we speculate that sensorimotor

³ This notion was supported by the results of a supplementary analysis in which boundary (low/high) was added into a four-factor ANOVA, along with delay (0/150 ms), adapting effector (hand/foot) and adaptation-test correspondence (same/cross). The relevant result was a significant delay x boundary interaction (with no higher-order interactions) which implied that the high boundary recalibrated to a greater extent than the low boundary in all conditions.

temporal recalibration might be implemented by changing the weights in favour of later (e.g. reafferent) cues, providing a delayed representation for action timing.

One complication for this account arises from a recent study which introduced an alternative method to assess sensorimotor temporal recalibration. Sugano et al. (2012) used a synchronised tapping task in place of an explicit temporal judgement in order to assess shifts in subjective time. When required to tap in time with a metronome, participants were consistently early, in line with classic findings (Dunlap, 1910; see Aschersleben, 2002, for review). However, following adaptation to a 150 ms visual or auditory delay, this bias was increased (i.e. participants tapped earlier than before), suggesting recalibration. Critically, in sensory crossover conditions, adaptation was observed for visual adaptation followed by auditory test stimuli, but not for auditory adaptation followed by visual tests, a clear contrast with previous findings.

The synchronisation task of Sugano et al. (2012) has some desirable properties, particularly in terms of its more implicit character relative to a typical timing judgement, that make this outlying finding difficult to ignore. However, the task is also somewhat distinct as a measure of the perceived time of action in that it appears to engender an unusually strong weighting of reafferent, as opposed to efferent, cues. In fact, when attempting to achieve synchronous tapping, participants act very much as though they are trying to align their reafferent feedback with the metronome (with both signals being subject to sensory delays; Aschersleben, 2002; Fraise, 1980; Paillard, 1949). Under the kind of “reweighting” account of sensorimotor temporal recalibration outlined above, this focus on afferent cues would provide very limited scope for additional forwards recalibration of the perceived time of action, implying that only a complementary backwards shift in sensory event time could be revealed. In fact, while statistically impeccable, the effects reported by Sugano et al. (2012) are numerically rather small (12-31 ms) when compared to findings from judgement tasks, suggesting that the authors may have isolated a purely sensory contribution to

sensorimotor temporal recalibration. Such a contribution might account for the non-significant trend toward a shift of the low boundary obtained in the cross-limb conditions of the experiment reported here.

In summary, we have demonstrated that people's judgements about synchrony between an action and a sensory event are modified following a period of adaptation to delayed feedback. Such behavioural changes are broadly similar for both hand and foot actions when the adapted limb is used again at test, implying some shift in the underlying representation of time. However, patterns of judgements recalibrate with a different character when testing with an alternate end effector relative to the one that was adapted, implying that an adjusted decision process might explain the apparent generalisation of sensorimotor recalibration across limbs. Overall, the compelling within-limb effect that we report suggests an altered sensory estimate for the time of action, which might involve a reweighting of the efferent and afferent cues that inform us about when our own actions occur.

References

- Allan, L. G. (1975). The relationship between judgments of successiveness and judgments of order. *Perception and Psychophysics*, *18*(1), 29-36.
- Arnold, D.H., Nancarrow, K. & Yarrow, K. (2012). The critical events for motor-sensory temporal recalibration. *Frontiers in Human Neuroscience*, *6*, 235:1-6.
- Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension. 2. evidence from sensorimotor synchronization. *International Journal of Psychophysiology*, *50*(1-2), 157-163.
- Aschersleben, G. E. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition*, *48*(1), 66-79.
- Cunningham, D. W., Billock, V. A., & Tsou, B. H. (2001). Sensorimotor adaptation to violations of temporal contiguity. *Psychological Science*, *12*, 532-535.
- Di Luca, M., Machulla, T. K. & Ernst, M. O. (2009). Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. *Journal of Vision*, *9*(12):7, 1–16.
- Dunlap, K. (1910). Reactions on rhythmic stimuli, with attempt to synchronize. *Psychological Review*, *17*, 399-416.
- Fraisse, P. (1980). Les synchronisations sensori-motrices aux rythmes [the sensorimotor synchronization of rhythms]. In J. Requin (Ed.), *Anticipation et comportement* (pp. 233-257). Paris: Centre National.

- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773-778.
- Garcia-Perez, M. A., & Alcala-Quintana, R. (2012). Response errors explain the failure of independent-channels models of perception of temporal order. *Frontiers in Psychology*, *3*, 94. doi: 10.3389/fpsyg.2012.00094
- Gibbon, J., & Rutschmann, R. (1969). Temporal order judgement and reaction time. *Science*, *165*(3891), 413-415.
- Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382-385.
- Heron, J., Hanson, J. V., & Whitaker, D. (2009). Effect before cause: Supramodal recalibration of sensorimotor timing. *PLoS ONE*, *4*(11), e7681.
- Heron, J., Roach, N. W., Hanson, J. V., McGraw, P. V., & Whitaker, D. (2012). Audiovisual time perception is spatially specific. *Experimental Brain Research*, *218*(3), 477-485.
- Hughes, G., Desantis, A., & Waszak, F. (2012). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychological Bulletin*, doi: 10.1037/a0028566
- Keele, S. W. (1968). Movement control in skilled motor performance. *Psychological Bulletin*, *70*(6), 387-403.
- Keetels, M. & Vroomen, J. (2007). No effect of auditory-visual spatial disparity on temporal recalibration. *Experimental Brain Research*, *182*, 559-565.

- Keetels, M., & Vroomen, J. (2012). Exposure to delayed visual feedback of the hand changes motor-sensory synchrony perception. *Experimental Brain Research*, 219(4), 431-440.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112-131). New York: Wiley.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). New York: Lawrence Erlbaum Associates.
- Miyazaki, M., Yamamoto, S., Uchida, S., and Kitazawa, S. (2006). Bayesian calibration of simultaneity in tactile temporal order judgment. *Nature Neuroscience*, 9, 875–877.
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Brain Research: Cognitive Brain Research*, 17(1), 154-163.
- Obhi, S. S., Planetta, P. J., & Scantlebury, J. (2009). On the signals underlying conscious awareness of action. *Cognition*, 110(1), 65-73.
- Paillard, J. (1949). Quelques données psychophysiologiques relatives au déclenchement de la commande motrice [some psychophysiological data relating to the triggering of motor commands]. *L'Année Psychologique*, 48, 28-47.
- Roach, N. W., Heron, J., Whitaker, D., & McGraw, P. V. (2011). Asynchrony adaptation reveals neural population code for audio-visual timing. *Proceedings of the Royal Society B: Biological Sciences*, 278(1710), 1314-1322.

- Roseboom, W., & Arnold, D. H. (2011). Twice upon a time: Multiple concurrent temporal recalibrations of audiovisual speech. *Psychological Science, 22*(7), 872-877.
- Roseboom, W., Nishida, S., & Arnold, D. H. (2009). The sliding window of audio-visual simultaneity. *J. Vis., 9*(12), 4.1-8.
- Rosenberger, W. F., & Grill, S. E. (1997). A sequential design for psychophysical experiments: An application to estimating timing of sensory events. *Statistics in Medicine, 16*(19), 2245-2260.
- Scheier, C. R., Nijhawan, R., & Shimojo, S. (1999). Sound alters visual temporal resolution. *Investigative Ophthalmology and Visual Science, 40*, 169-169.
- Schneider, K. A., & Bavelier, D. (2003). Components of visual prior entry. *Cognitive Psychology, 47*(4), 333-366.
- Spence, C., Baddeley, R., Zampini, M., James, R., & Shore, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception and Psychophysics., 65*(2), 318-328.
- Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General, 130*(4), 799-832.
- Sternberg, S., & Knoll, R. L. (1973). The perception of temporal order: Fundamental issues and a general model. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 629-686). London: Academic Press.

- Stetson, C., Cui, X., Montague, P. R., & Eagleman, D. M. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron*, *51*(5), 651-659.
- Strother, L., & Obhi, S. S. (2009). The conscious experience of action and intention. *Experimental Brain Research*, *198*(4), 535-539.
- Sugano, Y., Keetels, M., & Vroomen, J. (2010). Adaptation to motor-visual and motor-auditory temporal lags transfer across modalities. *Experimental Brain Research*, *201*(3), 393-399.
- Sugano, Y., Keetels, M., & Vroomen, J. (2012). The build-up and transfer of sensorimotor temporal recalibration measured via a synchronization task. *Frontiers in Psychology*, *3*, 246. doi: 10.3389/fpsyg.2012.00246
- Ulrich, R. (1987). Threshold models of temporal-order judgments evaluated by a ternary response task. *Perception and Psychophysics.*, *42*(3), 224-239.
- Vroomen, J., Keetels, M., de Gelder B., & Bertelson, P. (2004). Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Brain Research: Cognitive Brain Research*, *22*(1), 32-35.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Perception and Psychophysics.*, *63*(8), 1293-1313.
- Yamamoto, K., & Kawabata, H. (2011). Temporal recalibration in vocalization induced by adaptation of delayed auditory feedback. *PloS One*, *6*(12), e29414.

- Yarrow, K., Haggard, P., Heal, R., Brown, P., & Rothwell, J. C. (2001). Illusory perceptions of space and time preserve cross-saccadic perceptual continuity. *Nature*, *414*(6861), 302-305.
- Yarrow, K., Haggard, P., & Rothwell, J. C. (2010). Saccadic chronostasis and the continuity of subjective visual experience across eye movements. In R. Nijhawan, & B. Khurana (Eds.), *Space and time in perception and action* (pp. 149-163). Cambridge: Cambridge University Press.
- Yarrow, K., Jahn, N., Durant, S., & Arnold, D. H. (2011). Shifts of criteria or neural timing? the assumptions underlying timing perception studies. *Consciousness and Cognition*, *20*, 1518-1531.
- Yarrow, K., & Obhi, S.S. (In Press). Temporal perception in the context of action. In Arstila, V. & Lloyd, D. (Eds.), *Subjective Time*, MIT Press.
- Yarrow, K., Roseboom, W., & Arnold, D. H. (2011). Spatial grouping resolves ambiguity to drive temporal recalibration. *Journal of Experimental Psychology. Human Perception and Performance*, *37*, 1657-1661.
- Yarrow, K., & Rothwell, J. C. E. (2003). Manual chronostasis: Tactile perception precedes physical contact. *Current Biology*, *13*, 1134-1139.
- Yuan, X., Li, B., Bi, C., Yin, H., & Huang, X. (2012). Audiovisual temporal recalibration: Space-based versus context-based. *Perception*, *41*, 1218-1233.

Appendix: Comparison with traditional SJ parameters

Presenting two boundaries in place of a single PSS makes our presentation of results slightly more cumbersome, but, in our opinion, is more revealing about psychological processes that may give rise to the overall patterns of judgements that are observed. Our modelling makes clear that the PSS that is commonly derived in studies using the simultaneity judgement (based on the centre of mass of a fitted Gaussian or truncated Gaussian) conflates the positioning of the two transition boundaries from successiveness to simultaneity, and is thus less informative about the processes that could be generating a shift than the approach we have adopted here. Similarly, more typical measures of the “window of simultaneity” such as the full width half height or standard deviation of the fitted Gaussian, conflate sensory noise with the placement of the decision criteria. However, in order to facilitate comparison with other studies, we also calculated measures broadly equivalent to the PSS and window of simultaneity that are commonly reported (by averaging the two boundaries, and subtracting the high boundary from the low boundary, respectively). These data are reported in Table A1.

<INSERT TABLE A1 AROUND HERE>

Repeated-measures ANOVAs identical to those used in our main analysis of low and high criteria showed that the shift in the PSS generated by a 150 ms delay was significant as a main effect ($F = 66.0$, $df = 1, 15$, $p < 0.001$), as was the comparison between same-limb and cross-limb conditions ($F = 11.9$, $df = 1, 15$, $p = 0.004$) but that the effect of delay was larger for same-limb compared to cross-limb conditions (adaptor-test correspondence \times delay interaction $F = 6.0$, $df = 1, 15$, $p = 0.027$). The comparison between 0 and 150 ms delay conditions was individually statistically significant in all four cases (hand-hand $t_7 = 6.9$, $p < 0.001$; hand-foot $t_7 = 3.8$, $p = 0.007$; foot-foot $t_8 = 7.1$, $p < 0.001$; foot-hand $t_8 = 3.1$, $p = 0.015$). The window of simultaneity was larger for cross-limb than for same-limb conditions ($F = 9.5$, $df = 1, 15$, $p = 0.008$) and grew

slightly following adaptation to a 150 ms delay ($F = 5.3$, $df = 1, 15$, $p < 0.036$) but these factors did not interact. Hence a more traditional analysis would have suggested recalibration in all cases, with significantly greater recalibration in the same-limb conditions, but would not be diagnostic regarding the mechanisms of this recalibration.

Tables

Table A1. Means (SEMs) of PSS (Estimated as Average of Low and High Boundaries) and Width of the Window of Simultaneity (Estimated by Subtracting the Low Boundary from the High Boundary)

	PSS				Width			
	Adapt hand		Adapt foot		Adapt hand		Adapt foot	
	Adapt 0 ms	Adapt 150 ms	Adapt 0 ms	Adapt 150 ms	Adapt 0 ms	Adapt 150 ms	Adapt 0 ms	Adapt 150 ms
Same limb test	60 (10)	162 (7)	66 (14)	170 (19)	193 (27)	216 (28)	242 (54)	317 (57)
Cross limb test	144 (23)	229 (31)	117 (16)	174 (16)	295 (45)	326 (65)	312 (71)	376 (39)

Figure legends

Figure 1. Schematic overview of a detection-theoretic model of simultaneity judgements. The model is a variant of the general threshold models discussed by Ulrich (1987). (A): Graphical representation of the model parameters, illustrating how the objective asynchronies between an action and a sensory event are mapped internally to become an observer's subjective asynchronies, and then compared with the placement of two decision criteria to give rise to a psychometric function in an SJ task. On the left, the function that relates objective and subjective asynchrony is depicted. In this case, the relationship is veridical (i.e. there is no systematic tendency for information about the sensory event to register quicker than information about the time of movement, or vice versa). A temporal advantage for one or other event, were it present, would result in a function that was displaced along the x axis, no longer passing through the origin. The presence of latency noise is represented by the shading variations superimposed on the diagonal objective-to-subjective asynchrony function, with darker shading indicating higher probability density. This noise means that on any given trial, the subjective asynchrony may not exactly equal the objective one. The dashed vertical lines indicate the observer's criteria (in this case unbiased for the two possible arrival orders, so centred on a subjective asynchrony of zero, and indicating fairly conservative criterial extents, set at ± 50 ms). The observer only classifies the stimuli "simultaneous" when the subjective asynchrony falls between the low criterion and the high criterion. However, the criteria may not be perfectly stable across the trials of the experiment, but rather may vary somewhat from trial to trial around their mean position. This potential instability is illustrated by the shading variations, with darker shading again indicating higher probability density. Moving to the central column of the figure, two probability density functions are shown, each indicating how various subjective asynchronies will be generated over repeated trials at just one of two example objective asynchronies (-50 and 50 ms). These can be thought of as slices through the objective-to-subjective asynchrony function (from the left-hand graph). Given latency and decision noise, for each objective asynchrony, sampled over multiple trials, a proportion of trials will be classified simultaneous (shaded area, central column). We can predict this proportion by measuring the distance from the left-hand side of the probability density function to the high criterion (i.e. a first cumulative Gaussian) and subtracting

the distance from the left-hand side of the probability density function to the low criterion (a second cumulative Gaussian). This produces the psychometric function shown on the right. (B): Psychometric functions predicted by the two-noisy-criterion SJ model when model parameters change. The small schematic insets denote changes in model parameters in terms of the graphical representation developed in part A. Initially (thin black lines) there is no temporal advantage for one signal compared to the other, and the two signals generate $10,000 \text{ ms}^2$ of combined latency noise. Variations in model parameters are then denoted using grey lines. From top left to right, we show the effects of 1) a relative shift in the subjective timing of the two events; 2) a relaxation of one of the two criteria for judging simultaneity; and 3) a relaxation of both criteria. From bottom left to right, we illustrate 1) a further relaxation of one criterion for an observer who began with fairly liberal criterion extents; 2) an increase in the noise that derives from variability in the latency of registering the two events; and 3) an increase in the noise that derives from instability in the positioning of just one criterion. See Yarrow, Jahn, Durant & Arnold (2011) for further discussion. Matlab code for fitting the model is available from the corresponding author.

Figure 2. A. Schematic of experimental procedure (top-up/test phase). The effector could either remain the same or change between the final adaptation press and the test press, in separate blocks. B & C. Data and model fits for one representative participant from the adapt-hand group (B: Adapt-hand-test-hand trials; C: Adapt-hand-test-foot trials). Test asynchrony was subject to random variation reflecting the participant's final press time, hence many asynchronies are sampled but generally only once each (small data points represent single observations at a given asynchrony, while large data points represent double observations).

Figure 3. Average data for both the adapt hand (A, C) and the adapt foot (B, D) groups. Upper panels show same-limb test conditions, lower panels show cross-limb test conditions. In each panel, the upper graph shows mean low and high-boundary estimates in baseline (adapt 0) and recalibrated (adapt 150) conditions.

In these plots, error bars denote standard errors while the black curves denote model predictions based on group mean parameter estimates (solid line = adapt 0, dashed line = adapt 150). The lower graph in each panel denotes the difference in parameter estimates in the two adaptation conditions (i.e. adapt 150 minus adapt 0, which corresponds to the recalibration effect) for both low and high boundaries, with error bars denoting 95% confidence intervals on this value and asterisks () denoting significant differences at $p < 0.05$.*

Figure 1

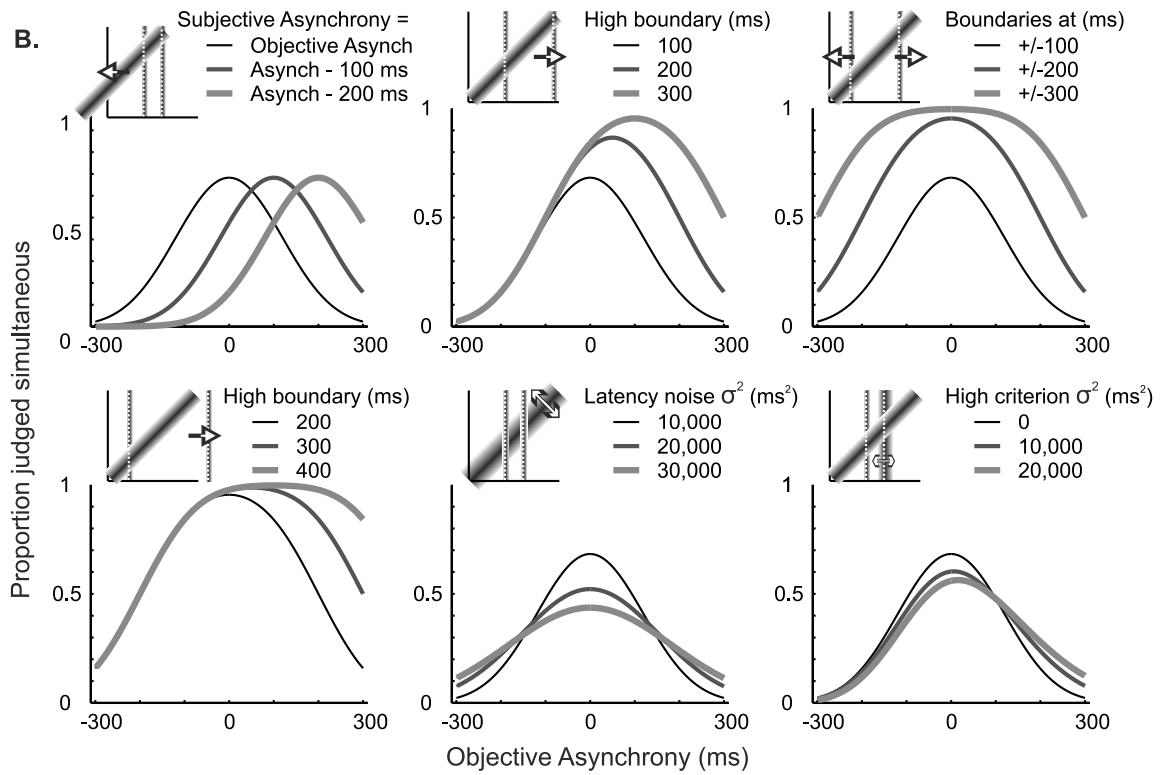
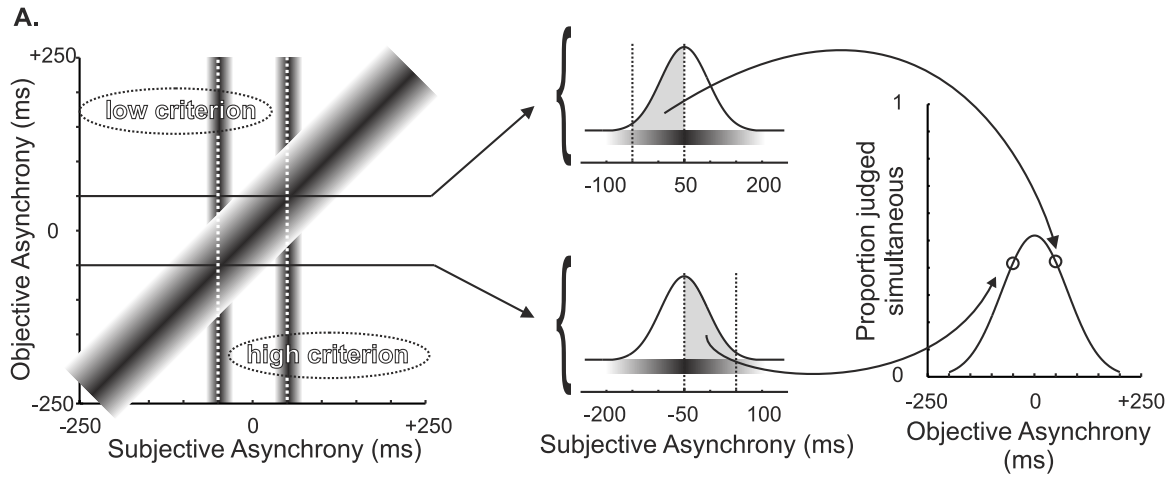
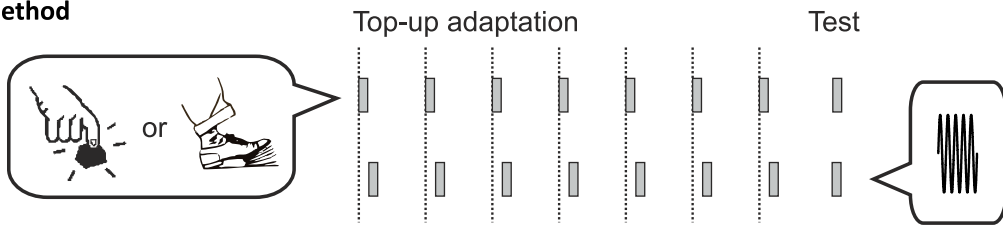
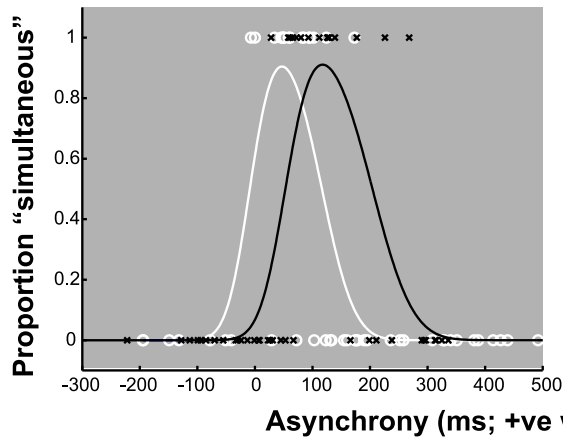


Figure 2

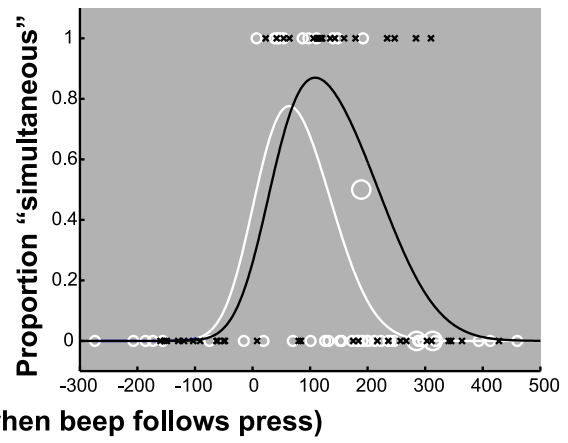
A. Method



B. Hand-Hand



C. Hand-Foot



○ Adapt 0 Data * Adapt 150 Data
— Adapt 0 Fit — Adapt 150 Fit

Figure 3

