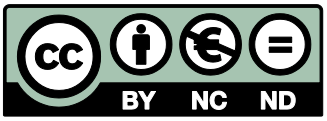
This is an author-final version. The final (press-formatted) version of this article will be published by Elsevier in *Cognitive Psychology* (DOI 10.1016/j.cogpsych.2015.10.002).



A model-based comparison of three theories of audiovisual temporal recalibration

Kielan Yarrow [1]*, Shora Minaei [1] & Derek H. Arnold [2]

[1] *Department of Psychology, City University London*

[2] *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

Observers change their audio-visual timing judgements after exposure to asynchronous audiovisual signals. The mechanism underlying this *temporal recalibration* is currently debated. Three broad explanations have been suggested. According to the first, the time it takes for sensory signals to propagate through the brain has changed. The second explanation suggests that decisional criteria used to interpret signal timing have changed, but not time perception itself. A final possibility is that a population of neurones collectively encode relative times, and that exposure to a repeated timing relationship alters the balance of responses in this population. Here, we simplified each of these explanations to its core features in order to produce three corresponding six-parameter models, which generate contrasting patterns of predictions about how simultaneity judgements should vary across four adaptation conditions: No adaptation, synchronous adaptation, and auditory leading/lagging adaptation. We tested model predictions by fitting data from all four conditions simultaneously, in order to assess which model/explanation best described the complete pattern of results. The latency-shift and criterion-change models were better able to explain results for our sample as a whole. The population-code model did, however, account for improved performance following adaptation to a synchronous adapter, and best described the results of a subset of observers who reported least instances of synchrony.

Keywords: Temporal recalibration; simultaneity judgement; adaptation; synchrony; audiovisual

Many sensory properties are subject to adaptation, whereby a period of exposure to a stimulus or set of stimuli affects subsequent perceptual reports. Perhaps the best-known example is the waterfall illusion, wherein prolonged exposure to one direction of motion makes a static test seem to move in an opposite direction (Addams, 1834). The list of stimuli that exhibit conceptually similar aftereffects after adaptation is substantial. Indeed, adaptation has been referred to as the psychophysicist's microelectrode (Frisby, 1979) in recognition of its wide adoption by experimenters and its potential to offer insights into the various processing steps that contribute to perception.

Audiovisual temporal recalibration (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Vroomen, Keetels, de Gelder, & Bertelson, 2004) is a relatively recent addition to the list of adaptation phenomena, and promises to help us understand how the brain determines the relative timings with which events occur. Although a seemingly simple mental operation, the perception of relative timing is likely to scaffold many aspects of higher-level cognition, such as making inferences about causality and integrating stimuli into coherent units. Temporal recalibration reveals a previously underappreciated malleability in the experience of temporal succession.

In a typical experiment participants are exposed to sequences of adaptors (e.g. flashes and beeps with a constant asynchrony) and then make judgements about either the simultaneity or order of subsequent tests. Such adaptors change peoples' reports about test timing, relative to baseline, with the typical finding being that the point of subjective simultaneity (PSS - a common summary measure in the literature on relative timing) has shifted toward the adapted asynchrony. It is as though, after having lived briefly in a world containing a constant audio or visual delay, people come to accept this timing relationship as synchronous. There are (at least) three accounts of the process underlying temporal recalibration.

Accounts of Temporal recalibration

The first explanation of temporal recalibration we will consider is the *latency-shift* account. This account suggests that the relative time at which audio and visual signals arrive (and are registered) at a hypothetical comparator has shifted, as though one modality were accelerated (or retarded) relative to the other (e.g. Di Luca, Machulla, & Ernst, 2009). Such a change might, for example, be implemented by adjusting one or more thresholds along the sensory pathways, yielding an effective change in overall transduction time.

The second explanation places the locus of adaptation at a higher level, positing that the effect is not really sensory in nature, but rather a consequence of changes in decisional criteria used to categorise stimuli as simultaneous or asynchronous (or indeed as occurring in a particular order). Under this *criterion-change* account (Yarrow, Jahn, Durant, & Arnold, 2011) simultaneity reports are malleable because people do not have a strong and persistent internal anchor regarding which physical timing relationship signifies synchrony. Hence they are prone to contextual biases when making this kind of judgement. This account is probably best understood with reference to detection-theoretic models (which we will outline shortly), but an intuitive description would be that when participants are repeatedly exposed to a small asynchrony, they become more willing to accept somewhat similar asynchronies as simultaneous because their frame of reference has changed.

The third explanation of temporal recalibration draws on a more established literature, relating adaptation to aftereffects in vision. This *population-code* account (Roach, Heron, Whitaker, & McGraw, 2011) suggests that the relative time between two crossmodal events is represented via the activity of a population of neurones, in a similar fashion to how spatial vision is supported by populations of orientation-tuned neurones in primary visual cortex (Hubel, 1988). In the case of relative time, multiple (hypothetical) neural units would each exhibit a preference for a particular asynchrony, but still respond (somewhat less vigorously) to similar asynchronies according to a "tuning curve." The complete population incorporates a wide range of preferences. A "labelled-line" readout would then decode the activity of this population to estimate audiovisual timing. Adaptation can be modelled as a reduction in the reactiveness of units maximally responsive to audiovisual offsets similar

to the adaptor. This tends to result in contrastive aftereffects, exaggerating differences between the adapted and similar audiovisual asynchronies.

Hence there are currently three quite different explanations of audiovisual temporal recalibration. In the burgeoning literature on this effect, different forms of (sometimes contradictory) evidence provide support for each. We will consider this evidence in more detail in the discussion, when placing our new results in a broader context. For now, we will explain the logic underlying our model-based comparison of predictions extracted from each of the three accounts of temporal recalibration.

Formalising and testing accounts of temporal recalibration via computational models

Each of the accounts outlined in the previous section can be formalised into a modelling framework. Formalising accounts in this manner has several benefits. It forces us to be precise and make explicit choices about the processes that might underlie theoretical descriptions. It also allows us to make quantitative predictions, which might not match the intuitions we initially held. The flip side of these strengths is that we are now testing a particular variant of a theory, not all conceivable instantiations. The approach we adopt here is to express each account as a bare-bones model that (we believe) captures its core features while minimising parametric flexibility. Having done so, we present an experiment for which each model predicts a distinct pattern of results.

Both the latency-shift and criterion-change accounts can be well captured using detection-theoretic latency models developed from the 1960s onwards (Baron, 1969; Gibbon & Rutschmann, 1969). A central feature of these models is that auditory and visual signals travel along separate pathways in the brain toward a decision centre (or comparator), and are subject to independent latency noise. This implies that the difference in their arrival times at the comparator (Δt, or the *subjective asynchrony*) is a random variable (i.e. it varies across trials even for repetitions of physically identical asynchronies) with a form that reflects the two contributing latency distributions, and a central tendency that reflects their mean difference (see Figure 1a).

These features led Sternberg and Knoll (1973) to describe such models as "independent channels" models. Different models within this class are distinguished by the decision processes applied at the comparator. For simultaneity judgements, the simplest conceivable process is to classify stimulus pairs as synchronous when Δt falls above a low criterion and below a high criterion (for example, when the subjective asynchrony at the comparator is above -100 ms and below +100 ms, with positive values denoting sound following light). A slightly expanded class of models, considered by Ulrich (1987) and labelled "general threshold" models, allow for decisional criteria to also be considered random variables reflecting, for example, an observer's inability to maintain a consistent decision criteria across multiple trials.

Here we adopt a specific variant of these models, previously found to be suitable for capturing key features of the psychometric function for simultaneity judgements (SJs) (Yarrow, Sverdrup-Stueland, Roseboom, & Arnold, 2013; Yarrow et al., 2011).[1] These key features include a potentially broad plateau, across which events are judged as synchronous, and a difference in slope for the two sides of the function (see Figure 4 for examples of these features). Our model assumes that the latency of both audio and visual signals is affected by independent Gaussian noise affecting their arrival times at the comparator, as a result of which Δt is considered a random variable with a Gaussian distribution and variance that is the sum of the variances of the individual signals. Decision criteria are also assumed to vary from trial to trial according to a Gaussian distribution, the spread of which might differ between the lower (sound before light) and upper (light before sound) criteria. These assumptions result in a psychometric function shaped as the difference of two cumulative Gaussians, and defined by four parameters (the means and standard deviations of the two contributing cumulative Gaussians). The meaning that can be ascribed to these parameters and the effects of changing them are more fully discussed in two previous publications (Yarrow et al., 2013; Yarrow et al., 2011) and in Appendix A of this paper, and the model is schematised in Figure 1a.

[1] Other variants have been described recently and shown to perform well (García-Pérez & Alcalá-Quintana, 2012a; García-Pérez & Alcalá-Quintana, 2012b) but are less suitable for characterising the criterion-change account we wish to test.

<INSERT FIGURE 1 AROUND HERE>

Previously, this four-parameter model has been fitted separately to data from each of several adaptation conditions to provide a descriptive analysis of the effects of adaptation. Here, we additionally attempt to test the latency-shift and criterion-change accounts by retaining some model parameters across all adaptation conditions, while allowing others to vary (according to the core features of the two accounts), thus fitting the model to all conditions simultaneously. We test four conditions: A baseline condition without any adaptation, a zero (synchronous) adaptor, a negative (AV) adaptor, and a positive (VA) adaptor. What changes would we expect across conditions under the two accounts outlined so far?

According to the latency-shift account, the mean latency of one or both signals can change following adaptation, resulting in a shift in the mean value of Δt. Tracing this effect through the machinery of our model leads to a yoked change in the means of the two cumulative Gaussians that contribute to the predicted psychometric function or, more simply, to a shift of the entire psychometric function. This prediction is schematised in Figure 2a. Although the mean of Δt might not be exactly zero for objectively synchronous events in baseline conditions (i.e. participants could have an initial bias in perceiving simultaneity, due for example to neural pathways of differing lengths), for simplicity we will assume this bias is small, such that adaptation to synchrony will have no effect. This means that all four conditions can be predicted using just six free parameters: four for the baseline condition, which are repeated exactly for the adapt-zero condition, plus an extra "shift" parameter in each of the negative and positive adaptation conditions.

<INSERT FIGURE 2 AROUND HERE>

The criterion-change model can also be captured under this scheme using six parameters. One could reasonably predict a tightening of decision criteria following exposure to a synchronous adaptor. Repeated exposures to an adapt-zero condition should result in more stable criteria than in unadapted baselines, as in the former condition all tests follow the same reference. However, here we will predict identical performance in baseline and adapt-zero conditions, as this simplifying assumption allows us to equate the number of free parameters in each of the three models under consideration.

The key assumption of the criterion-change model tested here is that, following negative and positive adaptation, only the decision criterion on the side of the adaptor should change. Thus after adapting to an audio lead, participants would regard more audio-lead scenarios as simultaneous, but make unchanged judgments concerning audio-lag scenarios. Adapting to a visual lead should result in the opposite contingency. These assumptions thus require two additional free parameters, one to adjust the criterion that determines the transition from judgements of audio lead to simultaneity following AV adaptation, and another to adjust the criterion that determines the transition from simultaneity to audio-lag judgements following VA adaptation. The end effect is that adaptation is predicted to expand the psychometric function for SJs outwards on the side of adaptation (see Figure 2b).

We now have formal models yielding predictions for the latency-shift and criterion-change accounts. We need to introduce a different class of model to describe the third, population-code, account of temporal recalibration (see Figure 1b). Given that only one population model has thus far been presented to describe audiovisual temporal recalibration (Roach et al., 2011) we have adopted that model here, with the addition of a decision rule suitable for simultaneity judgements. We provide some further exploration of this model and its parameters in Appendix A, but the key features of the model are as follows.

Relative time is assumed to be encoded via a population of neural units with a range of audiovisual offset preferences, starting with a zero-tuned neurone and expanding outwards in 50 ms steps. Each neurone has an identical Gaussian tuning curve, with a standard deviation describing how

its mean spiking rate falls off in response to stimuli that are increasingly distant from the unit's preferred audiovisual timing. Firing rates are also subject to Poisson noise. The population code is read out by a maximum-likelihood decoder (Jazayeri & Movshon, 2006) which is unaware of any adaptation. Adaptation itself is modelled using a proportional gain reduction parameter applied in full to neurones with preferences at the point of adaptation, and falling off from there with a Gaussian profile (requiring a further parameter to describe its spread). Thus far the model is exactly as described by Roach et al. (2011) for simulating absolute estimates of asynchrony, and it contains four free parameters. In order to make predictions about simultaneity judgements, we add a decision rule identical to that used in our earlier latency models, i.e. the model should give a simultaneous response if the decoded asynchrony falls between two decision criteria. This yields a model capable of predicting data in our four adaptation conditions with six free parameters.

We chose adaptation conditions in order to maximise differences between the predictions of the three models outlined above. In particular, we deliberately included both a baseline (no adaptation) condition and an adapt-zero condition, whereas previous investigations of audiovisual temporal recalibration have tended to use one or the other of these as if they were interchangeable. As Figure 2 and our previous discussion makes clear, this is approximately true for the latency-shift and criterion-change models, but the population-code account provides a clear prediction regarding exposure to synchronous adaptors.

Population codes with multiple neural units tend to generate "repulsive" after-effects (see Figure 2c). These arise due to an imbalance in potential responses. Prior to adaptation, it is assumed all units are equally responsive, so any physical input will generate a symmetrical pattern of response across the population of neural units. Post adaptation this symmetry is disturbed. The responsiveness of the adapted unit is maximally suppressed, and other units are affected as a function of proximity. This biases patterns of response in favour of units that are offset from the adaptor. In this context, for zero adaptation some positive tests end up being decoded as more positive and some negative tests

as more negative (relative to an unadapted code). For SJs, this shifts decoded asynchronies outside of the criteria for simultaneity, so the psychometric function becomes compressed.

Similar processes yield a distinct predicted pattern of results following AV and VA ordered adapters (illustrated for the AV case in Figure 2c). The difference in predictions, relative to latency-shift and criterion-change accounts, is particularly clear when adaptation occurs at points of transition from judgements of asynchrony to judgements of synchrony. While population-code models generally predict contrastive aftereffects, another important feature is that they don't tend to predict changes in perception for the adaptor itself, because there is no imbalance in the population relative to this point (Mitchell & Muir, 1976; Storrs & Arnold, 2012). Thus if one adapts at a category boundary, population coding predicts no change in perception for that particular timing relationship, but a contrastive aftereffect for more distal timings.

The population code predicts a combination of findings that contrast with other accounts of temporal recalibration. These are that perception of an adapted asynchrony will be unchanged but that the difference between this point and other timings will be exaggerated. As constant criteria for judging synchrony are assumed, the exaggerated difference between the adapted asynchrony and other timings will push some encoded audiovisual timings beyond the criterion on the opposite side of true synchrony from the adapted offset. So, rather than a uniform shift of the psychometric function (predicted by the latency-shift model) or an expansion on the side of adaptation (predicted by the criterion-change model) the population-code model predicts that the psychometric function should shrink in from the side *opposite* the point of adaptation (compared to an unadapted baseline condition) as long as adaptation is occurring very close to a decision boundary for judging events as synchronous. Note that these distinct predictions cannot be tested via the common practice of fitting data from baseline and adaptation conditions with an arbitrary function, like a Gaussian curve, and estimating a point of subjective simultaneity (PSS) from the fitted function's central tendency. All three models predict a shift in the resultant PSS, as is commonly observed.

The current Experiment

To summarise: We have formalised the core aspects of three accounts of audiovisual temporal recalibration into three models of equal parametric complexity (six parameters each). These models make distinct predictions about what should happen in zero-adaptation, negative-adaptation, and positive-adaptation conditions relative to an unadapted baseline. Our approach to testing the correctness of these predictions (and thus the correctness of these models) is to run the experiment, fit each model to *all data from each participant* separately, then compare the models according to their goodness of fit metrics. These will summarise the degree to which the different models are capable of describing the observed patterns of SJ data.

Methods

<INSERT FIGURE 3 AROUND HERE>

Participants

Initially, a convenience sample of 31 undergraduate psychology students was tested in a baseline block. They received course credit in exchange for participation. Of these, 22 (50% male, mean age = 26, range = 19-38) continued on to complete three subsequent adaptation blocks (see Figure 3). Selection was based on a preliminary fit to data from the baseline condition (see data analysis below). To continue, both of a participant's estimated decision boundaries for judging transitions from asynchrony to synchrony had to fall within 400 ms of objective synchrony. This ensured both that individually determined boundaries could be used to set the lags between light and sound for subsequent adaptation conditions, and that participants had understood the task and were performing competently. The experiment was approved following the procedures of the local research ethics committee.

Apparatus and stimuli

The experiment was controlled by a PC running a visual C++ executable instructing an A/D output card (National Instruments DAQCard-6715). Audiovisual stimuli, consisting of beeps and flashes, were presented from a speaker and a two-colour LED. The component unimodal signals were 10 milliseconds long and were generated at 44100 Hz, with onsets and offsets (the first and last milliseconds) smoothed using a Hanning window. The LED was placed centrally, with a chinrest used to maintain an observer distance of 57 cm, so that the visual stimulus subtended ~0.5 degrees visual angle. Computer outputs (+5v) caused green flashes for targets and red flashes for adaptors. Auditory signals (1000 Hz sinusoidal pure tones) were presented from a computer speaker, offset around 30° to the left, at a clearly audible level. Accurate stimulus timing was confirmed using a 20 MHz storage oscilloscope (Gould DSO 1604).

Design

The study had a within-subjects design with four conditions: Baseline (no-adaptation), synchronous (adapt zero), light leading (adapt VA) and sound leading (adapt AV). Each condition consisted of two blocks of 100 target stimuli. All participants received the baseline and adapt-zero conditions first, followed by counterbalanced presentations of AV and VA adaptation conditions (see Figure 3). In all conditions target stimuli could have stimulus onset asynchronies (SOAs) ranging from -450 to + 450ms, representing the delay between light and sound components (i.e. positive values represent sounds lagging lights). The SOA was randomly selected on each trial from a condition-specific distribution. Initially, distributions were uniform and ranged from -210 to +210ms in steps of 30ms, but they evolved over trials to reflect participant responses. If a participant responded simultaneous to a positive asynchrony the distribution was increased at higher positive SOAs than the one just tested, whereas if the response was non-simultaneous the distribution was increased at smaller SOAs. Opposite rules were applied following a negative SOA trial. Hence, the distribution was slowly adjusted to best sample the two transitions, from perceiving asynchrony to synchrony, for each

individual – an adaptive process based loosely on the generalised Pólya urn method of Rosenberger and Grill (1997).

Procedure

After providing consent, participants were seated in a darkened room and were given ten example "non-simultaneous" presentations, five with a beep preceding a flash by 300 ms, and five with a flash preceding a beep by 300 ms. Participants were asked to select simultaneous (right arrow key) only when they were sure that the tone and flash had happened at exactly the same time; otherwise they should select non-simultaneous (left arrow key). The delete key could be used to cancel a trial if there had been a lapse in concentration (in which case a replacement trial was added at the end of the block).

The baseline condition contained a sequence of target stimuli with random SOAs. There was a 500-1000ms (uniform random) delay from the participant's response until the midpoint of the next target stimulus pair. The baseline condition was followed immediately by the adapt-zero condition. All adaptation conditions started with 120 adaptors (red flashes and beeps with a constant SOA), before the first target pair was presented (distinguished by a green flash). After this initial exposure, 3-5 "top-up" adaptors appeared before each target. Adaptors were separated by 1100-1200ms. Since participants could not predict how many adaptors would appear before a target, they had to focus on the LED and speaker during the entire procedure.

When participants had completed the baseline and adapt-zero conditions, they received the choice of either a 15 minute break or to return another day. Most chose to complete the task on the same day. During the break individual boundaries for simultaneity perception were estimated using baseline data (see data analysis, below). The SOA of adaptors in the adapt-AV condition was then set to the estimated negative asynchrony/synchrony boundary, and vice versa for the adapt-VA condition.

Data analysis and modelling

In the baseline condition, the two transitions from judgements of asynchrony to judgements of synchrony were derived for each participant, in order to use these values as adaptors. Low (audio leading to simultaneity) and high (simultaneity to audio lagging) category boundaries, and the standard deviation associated with each, were estimated by performing a maximum-likelihood fit using Matlab (the MathWorks) using a function with the form:

(1) $P\ \text{“simultaneous”} = \Phi(B_{High}, \Delta t, \sigma_{High}) - \Phi(B_{Low}, \Delta t, \sigma_{Low})$

I.e. a difference of two cumulative Gaussians, with their means representing the high and low synchrony boundaries (Yarrow et al., 2011; Yarrow et al., 2013).

For our main analysis, the relative ability of the three different models of temporal recalibration to explain each participant's complete data set was assessed. We used Nelder-Mead simplex searches (Nelder & Mead, 1965; O'Neill, 1971) to find maximum-likelihood fits (with a binomial data model) for each participant across all four conditions simultaneously. To increase our chances of finding a global maximum, simplex searches were initiated from ten different starting parameter combinations, and then twice more, starting each time from the current best fitting parameter estimates. Because one of the parameters in the population-code model can take only discrete integer values (see below) and the simplex search is not optimised for this type of parameter (Lewandowsky & Farrell, 2011) we additionally completed a second set of 12 searches for this model, each of which combined a grid search on the integer parameter with repeated simplex searches on the remaining parameters.

For the latency-shift model, we used the same basic architecture outlined for our baseline fit above. To deal with the additional three conditions, we introduced two further parameters (for a total of six). These were "shift" parameters (S_{AV} and S_{VA}) that were added to the means of *both* cumulative Gaussians in Equation 1 (i.e. B_{Low} and B_{High}) in the adapt AV and adapt VA conditions respectively, thereby shifting the entire psychometric function. The two sides of the SJ psychometric function

retained the same slopes for all four conditions (i.e. sensory noise was not permitted to change).[2] For baseline and adapt-zero conditions, horizontal position was also held constant.

In the criterion-change model the slopes and boundaries were again held constant across baseline and adapt-zero conditions. Two "change" parameters (C_{AV} and C_{VA}) were introduced to capture adaptation. For the adapt-AV condition, C_{AV} was added to B_{Low}, to allow the low boundary to shift relative to baseline, while in the adapt-VA condition the high boundary was instead allowed to shift.

For the population-code model, we adapted the model described by Roach et al. (2011) in order to predict performance in an SJ task. The model assumes a population of 2N-1 neural units (i.e. 1,3,5 etc.) with stimulus preferences balanced about zero with a fixed spacing of 50 ms per unit. All units share the same standard deviation σ, so have tuning functions:

$$(2) \qquad f_i(SOA) = G_i e^{-(SOA-SOA_i)^2/2\sigma^2}$$

Where G_i is the gain of the *i*th neural unit, and SOA_i is its stimulus preference. Roach et al. did not specify the unadapted gain of units used in their simulations, so we arbitrary selected a value of G_0 = 100 (this sets the maximum mean spike rate for an unadapted neurone). Adaptation is assumed to affect neural gains depending on the distance between a neural unit's stimulus preference and the point of adaptation, SOA_a, with a maximal proportional reduction, α, at the point of adaptation. Adaptation then falls off with a Gaussian profile, with standard deviation σ_a:

$$(3) \qquad G_i = G_0(1-\propto e^{-(SOA_i-SOA_a)^2/2{\sigma_a}^2})$$

[2] A reviewer of this paper made the quite reasonable observation that noise affecting the latency of sensory signals might well scale with their latency, which implies that the variance of the difference (Δt) distribution could change following adaptation under a latency-shift account. We did investigate an eight-parameter model permitting changes of this kind, but do not report it here for reasons of brevity as the analysis was not very revealing; see footnote 6 in the discussion.

Each unit will produce R_i spikes in the critical interval after stimulus presentation, where R_i is a random variable reflecting the neural unit's mean firing rate f_i (based on its tuning function) and the presence of Poisson noise:

(4)[3] $$p(R_i = k|SOA) = \frac{f_i(SOA)^k e^{-f_i(SOA)}}{k!}$$

The population activity must now be decoded to infer the presented SOA, following the maximum-likelihood decoding scheme described by Jazayeri and Movshon (2006). Looking separately at each neural unit, the likelihood of each possible SOA is equal to the probability that the neural unit would fire R_i spikes given that SOA as an input. Working with logs (Jazayeri & Movshon, 2006, Equation 1) allows us to sum such likelihoods across neural units to generate a log-likelihood function for different possible SOAs:

(5)[4] $$LogL(SOA) = \sum_{i=1}^{I} R_i \log f_i(SOA) - \sum_{i=1}^{I} f_i(SOA) - \sum_{i=1}^{I} \log R_i!$$

The decoded SOA falls at the maximum of this function, which we estimated using the Newton-Raphson algorithm (applied iteratively to the differentiated log-likelihood function, from several starting points).

[3] Our equation here differs slightly from Roach et al.'s, as we have corrected what we believe to be a typo in their paper.

[4] Jazayeri and Movshon (2006) noted that the last term can be ignored as it is independent of the stimulus, and that the second term will sum to a constant for a homogenous representation, which allowed them to safely ignore that term as well. However, for the (unusual) bounded population code used here, the second term cannot be ignored. Our simulations produced different patterns compared to those reported by Roach et al., leading us to infer that they had decoded only after dropping this second term. In particular, Roach et al. argue that the compressive bias they observed in their absolute estimation data is predicted by a population code. Our simulations suggest that this is not true when using an optimal decoder (although it might be true for a decoder that fails to take account of the finite coverage of the encoding layer). As compressive/centring biases are often observed with absolute estimates (Gescheider, 1988) we suggest that this particular feature of their data probably arose from decisional rather than sensory processes.

In our experiment participants were presented with different physical SOAs and categorised them as either simultaneous or not, i.e. an SJ task. To achieve appropriate predictions, the four parameters of the Roach et al. neural population model were supplemented with two additional parameters, a low and high decision boundary. These permitted sorting of the population code model's output into a simultaneity function, with model predictions for each simulated trial either falling within or outside these boundaries. The final parameters for the population-code model were therefore: Number of neurons in population (N); standard deviation of each neuron's tuning curve (σ); depth (α, from 0-1) and bandwidth (σ_a) of adaptation; and a low and a high decision boundary (B_{Low} and B_{High}). When searching for the parameter values that maximised the model fit, predictions were simulated using 2000 simulated trials per SOA value tested.

Results

For each participant, we fitted their simultaneity judgement data (from all four conditions at once) to each of our three different models. Because the models all had six parameters, their goodness of fit can be compared directly using the log-likelihood of the MLE fit, where a higher score indicates a better fit. To illustrate our fitting procedures, we first present example data at the individual-participant level (Figure 4). The best-performing model varied from participant to participant, and it is useful to see how each of the different models are able to capture particular patterns of data. Hence Figure 4 illustrates the best-performing model fits for three different participants, selected because a different model performed best for each of them.

<INSERT FIGURE 4 AROUND HERE>

Figure 4a shows data from a participant whose behaviour was well captured by the latency-shift model. Their SJs show similar patterns in the baseline and adapt-zero conditions, but the distribution of simultaneous responses shifts uniformly to the left (i.e. toward sound leading

asynchronies) following AV adaptation, and uniformly to the right (i.e. toward sound lagging asynchronies) following VA adaptation.

The participant illustrated in Figure 4b fitted the predictions of the criterion-change model best. While there is little change from baseline to adapt-zero conditions, the distribution of simultaneous responses appears to expand outwards in the adapt-AV and adapt-VA conditions, selectively on the side of adaptation.

Finally, Figure 4c displays data from a participant who conformed well to the predictions of the population-code model. The distribution of simultaneous responses appears to contract inwards following adaptation to synchrony (compared to baseline). The distribution's centre of mass shifts left following AV adaptation, and right following VA adaptation, but this is achieved in each case by a shrinking of simultaneous responses in from the side *opposite* the adaptor, i.e. from the right for the adapt-AV condition and from the left for the adapt-VA condition.

Variability at the individual level might be considered as evidence of either individual differences or as simple measurement noise. What can we conclude at the group level? These data are shown in Figure 5. Panel A plots the mean log-likelihood of the MLE fit, averaged across all participants for each model. High values of log-likelihood indicate a better fit. Comparing the models, it is clear that the latency-shift and criterion-change models are performing better on average than the population-code model. This was confirmed by a repeated-measures ANOVA with the Greenhouse-Geisser correction for violations of sphericity ($F_{[1.19,24.89]} = 9.80$, $p = 0.003$) and Tukey's LSD corrected follow-up paired t-tests (shift vs. population $t_{[21]} = 3.77$, $p = 0.001$; change vs. population $t_{[21]} = 3.01$, $p = 0.007$; shift vs. change $t_{[21]} < 1$, N.S.).

<INSERT FIGURE 5 AROUND HERE>

We were concerned that the apparent success of the latency-shift and criterion-change models might be due to these models capturing patterns of data that the accounts they represented

did not *really* predict. Specifically, we noticed that some individuals yielded parameter estimates for the *shift* and *change* parameters that were opposite to predictions, e.g. an S_{AV} parameter with a positive value, indicating a shift to the right (toward sound-lagging asynchronies) following AV adaptation. To assuage this concern, we modified the models to constrain *shift* and *change* parameters to be negative-only or positive-only, in line with predictions of the underlying accounts. Although this modification led to slightly worse group mean fits for these models, the pattern of data (shown again in Figure 5a) changed little, and differences between both the latency-shift and criterion-change models and the population-code model remained significant ($F_{[1.12,23.55]} = 4.84$, $p = 0.034$; shift vs. population $t_{[21]} = 2.44$, $p = 0.024$; change vs. population $t_{[21]} = 2.18$, $p = 0.041$).

Mean parameters for best-fitting models are presented in Tables 1-3, for both the constrained and unconstrained models. Data are presented both for all participants, and for the subsets of participants for whom a model provided the best fit. Table 1 shows latency-shift model parameters, and suggests a shift that was more pronounced for AV adaptation than for VA adaptation (see the S, i.e. shift, parameters), with transitions from judgements of asynchrony to judgements of synchrony tending to occur when the stimuli were separated by around 200 ms (see the B, i.e. boundary, parameters). There is some evidence for a flatter SJ function on the light-leading side (i.e. $\sigma_{High} > \sigma_{Low}$). Table 2 shows criterion-change model fits, suggesting a pronounced effect of AV adaptation at the low boundary and VA adaptation at the high boundary (see the C parameters). Other trends are similar to those observed for the latency-shift model. Finally, Table 3 shows population-code model fits, suggesting a neural population with preferred SOA values ranging out to around +/-750 ms (based on the N parameter combined with 50 ms spacing). The tuning of each neurone (reflected in the σ parameter) is very broad (much wider than reported by Roach et. al. 2011). Adaptation leads to a roughly 20% spike-rate suppression within a quite localised region (see the α and σ_a parameters respectively).

<INSERT TABLES 1-3 AROUND HERE>

Given our theoretical interest in the mechanisms of temporal recalibration, we also wished to rule out a less interesting possibility as to why the population-code model might be underperforming. This model can only generate a near-symmetric psychometric function. This symmetry is primarily a product of the assumption of equally spaced channels across the encoded dimension: Different shaped functions might ensue if one were to assume an anisotropic distribution of channels across timing offsets, as posited to account for the oblique effect in tilt perception. In contrast to our population-code model, the particular modelling framework we selected as the basis for our latency-shift and criterion-change models allows psychometric functions to differ widely in slope on the AV and VA sides.

To address this issue we reduced the number of parameters in the better-performing models to five, by using a single σ parameter for both sides of the psychometric function (in place of a separate σ_{Low} and σ_{High}). As the models now contained different numbers of parameters, we adopted the Akaike Information Criterion (AIC) as a measure of goodness of fit, which includes a correction for such a difference.[5] The data are presented in Figure 5b, where a low value denotes a better fit. Despite being able to generate only symmetric psychometric functions, the latency-shift and criterion-change accounts still significantly outperformed the population-code model (unconstrained models: $F_{[1.20,25.10]} = 8.56$, $p = 0.005$; shift vs. population $t_{[21]} = 3.68$, $p = 0.001$; change vs. population $t_{[21]} = 2.83$, $p = 0.010$; constrained models: $F_{[1.20,25.10]} = 4.63$, $p = 0.035$; shift vs. population $t_{[21]} = 2.56$, $p = 0.018$; change vs. population $t_{[21]} = 2.12$, $p = 0.046$).

[5] We chose AIC primarily for simplicity. We note that the Bayesian Information Criterion (BIC) punishes extra parameters more severely, so would exacerbate the difference we obtained. However, it is difficult to assess whether the flexibility endowed by all the parameters in the population-code model is as great as that offered by those in the latency-shift and criterion-change models, and we suspect it may not be, particularly for N and σ, which both seem to mainly vary the magnitude of noise in the SOA estimates that the model produces. We ran one final ANOVA comparing the best-fitting log-likelihood values of the five-parameter latency-shift and criterion-change models (means -167.1 and -164.7 respectively) with that of the population-code model (mean -180.2) hence effectively assuming that the population-code model had only five parameters. We still found a significant difference between models ($F_{[1.20,25.10]} = 4.05$, $p = 0.048$) and specifically between latency-shift and population-code models ($t_{[21]} = 2.38$, $p = 0.027$) but the difference between criterion-change and population-code models became marginal ($t_{[21]} = 1.99$, $p = 0.059$)

To visualise the reasons some models were outperforming others, Figure 5c shows the raw data combined from all participants in each condition. There is a trend for simultaneous category judgements to fall off more quickly in the adapt-zero condition compared to the baseline condition, most consistent with the population-code model, but this model fails to predict the adapt-AV and adapt-VA conditions, where there is a substantial increase in synchrony judgements on the side of adaptation. This is the main change that seems to occur when these two adaptation conditions are compared to the adapt-zero condition (as predicted by the criterion-change model) but the situation is more complex when adapt-AV and adapt-VA conditions are compared to the non-adapted baseline, explaining why the criterion-change model is not significantly outperforming the latency-shift model in this data set. For comparison with previous reports, we also performed a more conventional analysis, in which data were fitted independently in each of the four conditions. This analysis is presented in Appendix B (and illustrates that our experiment yielded clear evidence of temporal recalibration when assessed in a typical fashion).

Finally, we also considered whether the subsets of participants best fit by each model varied systematically, in terms of best-fitting model parameters that could reflect how conservatively and/or precisely participants performed the SJ task. For simplicity, we performed this analysis on the five-parameter versions of the latency-shift and criterion-change models, as these provide a single estimate of sensory noise (σ) rather than two such values. This value did not differ across subgroups for either model (between-groups ANOVA; $p > 0.05$). It was not appropriate to assess sensitivity based on the parameters of the population-code model, because both σ, representing the tuning curve width of each neurone in the population, and N, representing the number of neurones in the population, contribute to precision (see Appendix A) and appeared to trade off with one another in different sub-groups. However, we did additionally consider the width of the SJ function (i.e. $B_{High} - B_{Low}$) which indicates how broadly participants place their criteria for judging events synchronous. Here trends in all three model fits were consistent. On average, the subgroup of participants best fit by the population-code model judged events to be simultaneous over a significantly reduced range of SOAs

(309 ms; $F_{[2, 19]} = 7.92$, $p = 0.003$) compared to those best fit by the latency-shift (527 ms; $p = 0.005$) and criterion-change (472 ms; $p < 0.003$) models.

Discussion

In this paper we have formalised three accounts of temporal recalibration (the latency-shift, criterion-change and population-code accounts) into computational models that each predict subtly different patterns of simultaneity-judgement data following exposure to audiovisual adaptors. Our participants completed four conditions (baseline, adapt zero, adapt AV and adapt VA) in a fairly typical audiovisual temporal adaptation experiment. Models were fitted to data across all four conditions at once, yielding metrics of goodness of fit for each model and participant. Although each model performed best for some subset of participants, at the group level the latency-shift and criterion-change models explained patterns of data better than the population-code model. This did not depend upon incidental model features that are not directly related to the different accounts of temporal recalibration, such as the freedom of some models to capture changes running opposite to predictions, or to predict an asymmetric psychometric function. Hence our data argue against a population-code model of temporal recalibration for participants in general.

This conclusion applies exclusively to the particular task and models we have compared. All of the models could have provided better fits if endowed with greater parametric flexibility. Indeed, by blending the models we could easily permit them all to predict each of the patterns of SJ data outlined here. For example, under the criterion-change model, decision criteria could be permitted to contract following adaptation to synchrony (at the cost of two further free parameters). This would still be in keeping with the spirit of the account.[6] A slightly more tenuous, but still reasonable, adjustment would be to permit movement of the decision criterion *opposite* the side of adaptation following AV or VA

[6] We did test a model of this kind, alongside an eight-parameter variant of the latency-shift account in which sensory noise was allowed to vary following AV or VA adaptation. However, the two models fitted equally well, and also showed similar AIC values to the simpler six-parameter models we have presented, so we chose not to complicate the paper further by including this analysis.

adaptation, as well as the one on the side of adaptation (for a ten-parameter model). This would be formally identical to a model combining latency shifts and criterion changes (see Yarrow et al., 2011 for further discussion of this model equivalence). Similarly, the population-code model could be combined with criterion changes at the boundaries where adaptors are presented, for a plausible eight-parameter model. We have not gone on to test all possible model variants in this way as it seems fairly clear that with a relatively modest increase in parameters, blended models will become indistinguishable in terms of their ability to predict the current data set.

Does this mean that our findings have no discriminative value? We would argue not. In particular, it seems to us that a simple population-code model, based on a loose analogy with the kind of representation known to exist for orientation in primary visual cortex and a simple feedforward projection to a decoding layer, is unlikely to ever fully capture the effects of audiovisual adaptation on simultaneity judgements *unless* supplemented with mechanisms operating beyond the sensory coding layer (such as a change in decision criteria). Many of our participants exhibited an increased tendency to report synchrony on the side of the adaptor, at SOAs *beyond* the point of adaptation. This is difficult to reconcile with a population code, even one that is elaborated substantially. There are probably other reasons that the population-code model fared worst in our comparisons, such as the fact that it yokes AV and VA adaptation via a single parameter, forcing a prediction of symmetrical adaptation effects for symmetrically positioned adaptors. However, we would still suggest that a fairly substantial proportion of naive participants change their SJs in ways that are highly suggestive of additional mechanisms beyond neural gain suppression.

In the remainder of the discussion, we provide some context for our current findings by presenting a brief review of other forms of evidence bearing on the question of which mechanism best accounts for audiovisual temporal recalibration. However, before doing so we make explicit some limitations in our study. First, our desire to adapt observers at their own individually determined synchrony-asynchrony boundaries led us to present the baseline condition first to all participants.

Although not uncommon in studies of temporal recalibration, this approach implies that some effects ascribed to adaptation might reflect practice.

Second, our model-fitting procedures differed slightly across models. In particular, the population-code model was simulated, not predicted from equations like the other two, which might imply a less smooth error surface, and thus greater difficulty locating the global maximum best fit. However, we do not think it is likely that this led to the population-code model's poorer performance, as we provided this model with extra opportunities via a second set of parameter searches (see methods).

Third, the fact that our first two models have been formalised to offer predictions without the need for simulation makes it much easier to be certain that all their parameters are offering some functional advantage (and to understand the particular kind of flexibility each parameter endows; see Appendix A). There is a real concern that the six parameters of the population-code model can trade off to a greater extent than those of the other models (for example N, the number of neurones in the population code, and σ, the standard deviation of their tuning functions, seem to have rather similar effects, both affecting the slope of the psychometric function). Hence comparisons like ours that simply adjust for the number of parameters should be treated with some caution.

Finally, it is worth noting that the population-code model provides a principled prediction for the exact degree of recalibration at *multiple* adaptation positions, not just the three tested here (i.e. it predicts how recalibration effects should scale for increasingly asynchronous adaptors). This would offer it greater parametric simplicity (compared to the latency-shift and criterion-change models) in future tests that could include a larger set of adaptation conditions.

Wider evidence bearing on accounts of temporal recalibration

Our analyses provide some support for both the latency-shift and criterion-change accounts of temporal recalibration, relative to the population-code account. Of course, these data are not the first evidence on this issue. We review other literature next. For reasons of brevity, we exclude

consideration of the related literatures on motor-sensory temporal recalibration (Stetson, Cui, Montague, & Eagleman, 2006) and Bayesian temporal calibration (Miyazaki, Yamamoto, Uchida, & Kitazawa, 2006).

Evidence in favour of the latency-shift account comes from experiments in which recalibration has been measured using both temporal judgements *and* simple reaction times. Navarra, Hartcher-O'Brien, Piazza, and Spence (2009) compared an adapt-zero baseline to asynchronous auditory-visual and visual-auditory adaptors (assessed in separate groups). They found that simple RTs to detect auditory stimuli changed in a manner consistent with a latency-shift account of recalibration. Auditory RTs increased following AV adaptation and decreased following VA adaptation. Visual RTs, meanwhile, were unchanged.

Around the same time, Di Luca, Machulla, and Ernst (2009) tested zero, AV and VA adaptors in a single group. They also examined an additional factor with two conditions, one where audiovisual events were co-localised, and one in which audio events were delivered via headphones. They found effects broadly consistent with latency shifts when comparing AV and VA adaptation (although trends relative to baseline did not follow recalibration predictions). The affected modality depended on the auditory stimulus. For co-localised stimuli, visual RTs were speeded by AV adaptation, whereas with headphone presentation auditory RTs were speeded following VA exposure. The former result is essentially opposite that of Navarra et al. (2009) in terms of the affected modality despite seemingly similar experimental conditions (although co-localisation was perhaps less precise, with auditory stimuli coming from two speakers on either side of a visual stimulus). Di Luca, Machulla, and Ernst 's (2009) second result is a better match to that obtained by Navarra et al., despite having a discrepant spatial layout.

Harrar and Harris (2008) also tested how VA adaptation affected simple reaction times (compared to an unadapted baseline) using co-localised stimuli. They found a slowing of visual RTs. An effect on visual RTs for co-localised stimuli echoes Di Luca, Machulla, and Ernst 's (2009) findings, but those authors found speeding for AV adaptation and no effects for VA adaptation. Harrar and

Harris (2008) also tested adaptation to other modality combinations (audiotactile and visuotactile) but found no significant RT changes. Sometime later, Yuan, Li, Bi, Yin, and Huang (2012; Experiment 2) revisited simple RT effects of AV and VA adaptation, this time using a more complex contingent adaptation paradigm (see later). They found no effects on the difference between visual and auditory RTs. Hence, summarising work with simple RTs, there is an overall tendency to find changes in response speed to one or the other modality in a manner that is broadly consistent with a latency-shift account of temporal recalibration, but inconsistent patterns of results suggests caution when interpreting these results.

Two papers described above included a second type of test pertinent to the latency account: Cross adaptation to different modality pairs. Di Luca, Machulla, and Ernst (2009) had participants make temporal order judgements about audiovisual, audiotactile and visuotactile pairs, following either AV or VA adaptation. They replicated classic audiovisual recalibration, but also observed effects on visuotactile judgements (when stimuli were co-localised) and audiotactile judgements (with headphone presentation of auditory stimuli), consistent with adaptors having induced latency shifts for visual and auditory stimuli respectively. Harrar and Harris (2008) also used temporal order judgements between these three modalities, with tests following VA, TV or TA adaptation regimes. In their experiment, VA adaptation yielded only audiovisual recalibration, with no significant transfer. In fact, the only significant cross adaptation they observed (from TV adaptation to audiovisual tests) ran opposite to that predicted if vision had been speeded.

Physiological measures

High temporal resolution neuroimaging, alongside behavioural measures, has also been used to investigate temporal recalibration. The simplest prediction of the latency-shift account is that there should be a shift in one or both unimodal event-related brain responses that matches the hypothesised change in neural latency. There is at least one report suggesting a relevant change in latency, specifically a speeding of the second response of bimodal neurones in superior colliculus to

staggered AV/VA stimulus presentations following adaptation (Yu, Stein, & Rowland, 2009). This shift, however, was dependent on the presence of an enhanced response to the first stimulus (i.e. it was absent when the second stimulus was presented alone), so the physiological mechanism suggested by these data is rather different to an independent acceleration of one modality.

Moving to human physiology, when Kösem, Gramfort, and van Wassenhove (2014) recorded magnetoencephalography during AV, zero, and VA adaptation, there were no shifts in latency of sensory event-related fields. However, over the course of adaptation phase shifts were observed in 1 Hz steady-state responses evoked by adaptors, in a direction consistent with the adaptation regime. An intriguing correlation between these phase shifts and behavioural responses was also reported.[7]

Summary of evidence supporting a latency-shift[8]

Some evidence supports a latency-shift account, but there are many inconsistencies. Evidence suggesting that changes in latency can generalise to decisions regarding a modality that had not been adapted runs contrary to the predictions of population-coding, which only predicts effects on the *relative* timing of the two adapted modalities, not on their individual components, and consequently not on the relationship between these components and a third modality. The criterion-change account would not predict changes in simple reaction time[9] or EEG components. This account could provide a reasonable decision-level explanation for transfer effects, for example by allowing that following VA adaptation, all punctate signals lagging vision are interpreted using a more liberal criterion.

[7] The result should probably be considered preliminary for a couple of reasons. First, the behavioural measure (post-adaptation PSS, rather than a change in PSS) was not completely matched to the neural one (a change in phase). Second, the behavioural measure would not be predicted to show strong recalibration, as there were no top-up phases during TOJ testing (and in fact recalibration went in the wrong direction for VA adaptation relative to baseline).

[8] Note that the effects of adaptation on implicit measures, which can be taken to support either the latency-shift or population-code accounts, are considered towards the end of the discussion.

[9] A different kind of criterion account could be considered, based on changes in the accumulation threshold used for detection in simple RT tasks, but this is really a particular possible implementation of the *latency-shift* account, and conceptually rather different to the criterion-change account tested here.

Evidence contradicting a latency-shift

Several reported results seem inconsistent with a latency-shift account of temporal recalibration. Perhaps the most striking example is the non-uniform biases in absolute estimates of SOA obtained following adaptation by Roach et al. (2011). Biases in that study were much reduced in the vicinity of the adaptor, as predicted by a population code (although these authors did not observe reliable reversals in bias, like those denoted in Figure 2c, perhaps because they did not sample SOAs widely enough). It was these data that led Roach et al. to propose a population-code account of temporal recalibration.

Roach et al's (2011) findings are at odds with those reported here for many of our participants (e.g. Figure 4a and 4b), who increased their reports of simultaneity beyond the point of adaptation, where repulsive biases generated by a population code should ensure that stimuli appear non-synchronous. Probably the most parsimonious way to explain these discrepancies would be to suggest a two-stage process, whereby a (non-uniform) sensory change must be interpreted by a bias-prone decisional process. Assuming this scheme, many of our participants showed profound changes in the decisional process. This may not have been the case for Roach et al. (2011), who used a task that might be prone to rather different kinds of decision bias, and also tested just a few highly practised observers, rather than a large naive sample. It is noteworthy that in our sample, the subgroup of participants who were best fitted by the population-code model were least likely to report synchrony (e.g. they had tight SJ functions) which might be expected of trained psychophysical observers.

Very recently, population-code predictions have been considered in some detail in a paper by Roseboom, Linares and Nishida (2015) making use of a three-alternative forced-choice procedure to assess changes in sensitivity following AV and VA adaptation. Their data revealed the hypothesised changes in sensitivity, which could be modelled as a combination of changes in the slope and lateral position of a transducer function relating objective time to perceived time. Slope changes in this function are predicted by a population code account, while lateral shifts are predicted by a latency-shift account (see our Figure 2 parts A and C). Interestingly, allowing a change in transducer slope was

particularly useful when modelling an adapt-zero condition, in line with the tendency towards contraction of the SJ psychometric function we observed here in our equivalent condition.

Evidence for and against the criterion-change account

Previous research also bears on the criterion-change account. The most targeted investigation in this context was that of Yarrow et al. (2011). They used logic similar to that presented here, alongside a ternary judgement task[10], arguing that a latency-shift account would most straightforwardly predict that both decision boundaries should change together following adaptation (i.e. a shift of the entire psychometric function) whereas a criterion-change account would predict a shift of just one decision boundary, on the side of adaptation. Based on separate fits to three conditions (adapt-zero, adapt AV and adapt VA), these investigators reported a significant change only at the boundary nearest the adaptor (relative to the adapt-zero baseline). These data were consistent with a criterion-change account, as is the data presented here if we consider only the same three conditions. A similar pattern can be seen in the group-averaged data presented by Fujisaki et al. (2004). However, Yuan et al. (2012, Experiment 1) reported data that was most consistent with a latency-shift (although failures to find any significant changes relative to baseline make this conclusion quite tentative).

This leads us to the recent literature demonstrating the existence of *opposite* contingent temporal recalibration effects, which are most easily reconciled with a criterion-change account. Such demonstrations include simultaneous and opposite temporal recalibrations to male/female faces presented with leading/lagging speech (Roseboom, Kawabe, & Nishida, 2013; Roseboom & Arnold, 2011), spatially lateralised left/right blobs paired with spatially congruent leading/lagging tones (Heron, Roach, Hanson, McGraw, & Whitaker, 2012), and horizontal/vertical gratings paired with low-

[10] Yarrow et al. actually used a combination SJ / TOJ task, but treated the data as ternary (before/same/after). Their final fitting procedure was slightly suboptimal (their models should really have been maximum-likelihood fitted using a multinomial rather than binomial data model) but the conclusions are unlikely to be much affected.

pitch leading / high-pitch lagging tones (Roseboom et al., 2013, but see Heron et al., 2012). Although it is possible to conceive of, for example, latency shifts applying in opposite directions for the auditory components of male and female voices, or multiple separate population codes representing SOAs for different spatial locations, it is rather easier to imagine participants learning to classify simultaneity in different ways for different feature combinations. In a similar vein, reports that AV recalibration can reflect the high-level grouping of the AV adaptors when temporal information is ambiguous (Yarrow, Roseboom, & Arnold, 2011) and can be modulated by paying attention not just to adaptors in general, but specifically to their temporal relationship (Heron, Roach, Whitaker, & Hanson, 2010), are easier to reconcile with a decision-level explanation.

We have, of course, already noted substantial evidence against the criterion-change account (in the form of evidence favouring other accounts). To this we should perhaps add some of the tasks tested by Fujisaki et al. (2004) in their seminal report showing, for example, that recalibration can be measured using an implicit task (the stream-bounce illusion). Arguably such tasks might be less affected by decisional criteria than explicit timing judgements, although one cannot discount the possibility that participants are assessing, and making decisions about, audio-visual timing relationships in such circumstances. More recently this finding has been supplemented by evidence of temporal recalibration when AV synchrony is assessed based on the peak timing of the McGurk illusion (Yuan, Bi, Yin, Li, & Huang, 2014). At present, there are few formal models of how relative timing might inform AV interactions (but see Colonius & Diederich, 2004). Such models would help researchers consider the potential neurocognitive loci of recalibration effects for these kinds of tasks.

Summary and Conclusions

In this paper, we have formalised three theories AV temporal recalibration into simplified models, and assessed their fit to SJ data from a fairly large sample of naive participants. We found that both latency-shift and criterion-change models describe the group's data better than a population-code model. However, individual differences suggest that some blending or development of these

models is required to fully account for temporal recalibration data. Reviewing the increasingly complex literature on this topic leads us to a similar conclusion: It is possible that some sensory level changes occur following exposure to audiovisual adaptors, but often act in combination with decision-level biases. The degree to which these different mechanisms manifest is likely to vary, depending on factors such as participant expertise, the experimental task, and the complexity/contingency of the stimuli. Hence we would caution against generalising too enthusiastically from one situation to another. We look forward to helping generate a more compelling unified account of audiovisual temporal recalibration, in order to better elucidate the underlying mechanisms by which humans perceive relative time.

Acknowledgements

DHA and KY's collaboration is supported by an Australian Research Council Discovery Grant. We thank Warrick Roseboom, Neil Roach and James Heron for helpful discussions on this topic.

References

Addams, R. (1834). An account of a peculiar optical phænomenon seen after having looked at a moving body. *The London and Edinburgh Philosophical Magazine and Journal of Science, 5*(29), 373-374.

Baron, J. (1969). Temporal ROC curves and the psychological moment. *Psychological Science, 15*, 299-300.

Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience, 16*(6), 1000-1009.

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), 7-16.

Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). *Recalibration of audiovisual simultaneity. Nature Neuroscience, 7*(7), 773-778.

Frisby, J.P. (1979). *Seeing: Illusion, brain and mind*. Oxford, UK: Oxford University Press.

García-Pérez, M. A., & Alcalá-Quintana, R. (2012a). On the discrepant results in synchrony judgment and temporal-order judgment tasks: A quantitative model. *Psychonomic Bulletin & Review, 19*(5), 820-846.

García-Pérez, M. A., & Alcalá-Quintana, R. (2012b). 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

Gescheider, G. A. (1988). Psychophysical scaling. *Annual Review of Psychology, 39*, 169-200.

Gibbon, J., & Rutschmann, R. (1969). Temporal order judgement and reaction time. *Science, 165*(891), 413-415.

Harrar, V., & Harris, L. R. (2008). The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Experimental Brain Research, 186*(4), 517-524.

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. doi:10.1007/s00221-012-3038-3

Heron, J., Roach, N. W., Whitaker, D., & Hanson, J. V. (2010). Attention modulates the plasticity of multisensory timing. *European Journal of Neuroscience, 31*, 1755-1762.

Hubel, D. (1988). *Eye, brain, and vision*. San Francisco, CA: W.H. Freeman.

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience, 9*(5), 690-696.

Kösem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *NeuroImage, 92*, 274-284.

Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Los Angeles, CA: Sage.

Mitchell, D. E., & Muir, D. W. (1976). Does the tilt after-effect occur in the oblique meridian? *Vision Research, 16*(6), 609-613.

Miyazaki, M., Yamamoto, S., Uchida, S., & Kitazawa, S. (2006). Bayesian calibration of simultaneity in tactile temporal order judgment. *Nature Neuroscience, 9*(7), 875-877.

Navarra, J., Hartcher-O'Brien, J., Piazza, E., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proceedings of the National Academy of Sciences of the United States of America, 106*(23), 9169-9173.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal, 7*(4), 308-313.

O'Neill, R. (1971). Algorithm AS 47: Function minimization using a simplex procedure. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 20(*3), 338-345.

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. doi:10.1098/rspb.2010.1737

Roseboom, W., Kawabe, T., & Nishida, S. (2013). Audio-visual temporal recalibration can be constrained by content cues regardless of spatial overlap. *Frontiers in Psychology, 4*, 189. doi:10.3389/fpsyg.2013.00189.

Roseboom, W., & Arnold, D. H. (2011). Twice upon a time: Multiple concurrent temporal recalibrations of audiovisual speech. *Psychological Science, 22*(7), 872-877. doi:10.1177/0956797611413293

Roseboom, W., Linares, D., & Nishida, S. (2015) Sensory adaptation for timing perception. *Proceedings of the Royal Society series B*, *282*: 20142833. doi:10.1098/rspb.2014.2833

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.

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.

Storrs, K. R., & Arnold, D. H. (2012). Not all face aftereffects are equal. *Vision Research, 64*, 7-16.

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.

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

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. doi:10.1016/j.concog.2011.07.003

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. doi:10.1037/a0024235

Yu, L., Stein, B. E., & Rowland, B. A. (2009). Adult plasticity in multisensory neurons: Short-term experience-dependent changes in the superior colliculus. *The Journal of Neuroscience, 29*(50), 15910-15922. doi:10.1523/JNEUROSCI.4041-09.2009 [doi]

Yuan, X., Bi, C., Yin, H., Li, B., & Huang, X. (2014). The recalibration patterns of perceptual synchrony and multisensory integration after exposure to asynchronous speech. *Neuroscience Letters, 569*, 148-152.

Yuan, X., Li, B., Bi, C., Yin, H., & Huang, X. (2012). Audiovisual temporal recalibration: Space-based versus context-based. *Perception, 41*(10), 1218.

Appendix A. *Further discussion of models and the effects of changing their noise parameters*

Latency model (used at the basis for latency-shift and criterion-change models)

Latency models assume that the latency with which both audio and visual signals reach a decision centre in the brain is a random variable. In our variant, these random variables follow a Gaussian distribution. Their difference (Δt) is then classified by a decision process, which sets lower and upper bounds (criteria) for values of Δt that will be classified as simultaneous. These bounds are themselves assumed to be random variables with Gaussian distributions.

Simulated from scratch as a series of steps using a computer, this model would have four quantities that contribute to the noise observed in a single psychometric function (i.e. variance in the two sensory signals, and variance in the two decision criteria). However, if the model were implemented in this way, these four parameters would be *non-identifiable* (a modelling term which indicates that they can trade off with one another to generate identical predictions). However, when formalised, the predictions of the model can be captured using just two noise parameters (σ_{Low} and σ_{High}) which describe the slope of the psychometric function at either side of the function. An increase in sensory noise (for either modality) decreases the slope on both sides of the function at once, whereas an increase in criterion noise decreases the slope on just one side of the function, specifically on the side where that criterion lies. Unfortunately, the relative contribution of each of these various underlying sources to the $\sigma_{Low/High}$ parameters can never be recovered using only SJs.

In the main text of this paper, we focus on the effects of changes in mean latency for one or both signals (the latency-shift account) or in the position of one or other decision criteria (the criterion-change account; see Figure 2) but for completeness, Figure A1 panels A-C show the effects of making changes in each source of noise.

<INSERT FIGURE A1 AROUND HERE>

Population-code model

Population-code models assume that neural units, each with a specific preference but collectively exhibiting a range of preferences, code the magnitude of an attribute unambiguously in their collated firing rates. In our variant, each neurone has a Gaussian tuning curve and shows Poisson variability in its spike rate. The preferences of different neural units vary by a fixed step size of 50 ms, but the number of neurones is a free parameter. Activity is interpreted by a sophisticated MLE decoder, yielding a single estimate that can then be compared with two decision criteria to reach a judgement about synchrony. Adaptation is modelled as the suppression of neurones with preferences close to the adaptor, with this suppression falling off with a Gaussian profile.

We do not have formally derived predictions for how this model behaves as parameters vary, so are somewhat limited in what we can say about how changes in the underlying architecture affect the psychometric function. However, the model can be simulated, and in Figure A1 panels D-F we illustrate the consequences of changing some parameters (specifically N and σ, which contribute mainly to observed precision) to provide a greater intuition about the workings of the model.

Appendix B. *Analysis based on separate fits to each condition*

In the temporal recalibration literature to date it has been common practice to fit data from each condition independently in order to assess the existence (and magnitude) of any recalibration effects. For completeness, we present such an analysis here. Data in each condition were fitted with a four-parameter latency SJ model identical to that used to establish synchrony boundaries in our baseline condition. We focus on the mean parameter estimates for the two boundaries where judgements changed from asynchronous (with audio coming first) to synchronous (the low boundary) and from synchronous back to (visual-first) asynchronous (the high boundary). Figure A2 shows the mean value of these boundaries in each of the four adaptation conditions. Averaging the low and high boundaries provides one method to estimate a point of subjective simultaneity. Hence an ANOVA incorporating both the two boundary estimates (as factor 1) and also the different adaptation conditions (factor 2) firstly tests for changes in the PSS (as a main effect of adaptation condition) and secondly reveals additional information about the relative effects of adaptation at each boundary (as an interaction).

<INSERT FIGURE A2 AROUND HERE>

Most previous temporal recalibration studies have used either a no-adapt baseline or a zero-adapt baseline, rather than both. We therefore ran three repeated-measures ANOVAs, the first considering adaptation relative to a no-adapt baseline, the second relative to a zero-adapt baseline, and the third comparing the two baseline conditions to each other. Note that we do not report the (trivial) main effects of boundary (which were always obtained).

The first ANOVA (baseline vs. adapt AV vs. adapt VA) revealed a main effect of adaptation ($F_{[1.30,27.26]} = 19.87$, $p < 0.001$) but the interaction narrowly missed significance($F_{[1.68,35.31]} = 3.44$, $p = 0.051$). This analysis would most straightforwardly support a latency-shift account, where both

boundaries move together during adaptation. Post-hoc t-testing suggested significant differences between baseline and adapt AV conditions and between adapt AV and adapt VA conditions (for data averaged across the two boundaries).

The second ANOVA (adapt zero vs. adapt AV vs. adapt VA) revealed a main effect of adaptation ($F_{[1.21,25.31]}$ = 20.91, $p < 0.001$) and an interaction ($F_{[1.46,30.74]}$ = 8.56, $p = 0.003$). Post-hoc t-testing suggested a significant low-boundary difference between adapt zero and adapt AV conditions and a significant high-boundary difference between adapt zero and adapt VA conditions. This analysis would most straightforwardly support a criterion-change account, as there is unequal change at the two boundaries following lag adaptation, and the change always occurs mainly at the boundary closest to the adaptors.

The third ANOVA (baseline vs. adapt zero) revealed a main effect of adaptation ($F_{[1,21]}$ = 21.11, $p < 0.001$) and an interaction ($F_{[1,21]}$ = 19.89, $p < 0.001$). Post-hoc t-testing suggested a significant difference only at the high boundary, which moved inwards, as predicted by the population-code model.

In sum, independently fitting data from each adaptation condition showed clear evidence of lag-adaptation effects in the sample as a whole, but did not differentiate between the three accounts of lag adaptation under consideration in this paper.

Tables

Table 1. *Means (SEMs) of best-fitting model parameters for the six-parameter latency-shift model (both unconstrained and constrained variants) for all participants (Ps), and for the subset of participants for whom this model provided the best fit*

	Mean best-fitting parameter (ms)					
	B_{Low}	B_{High}	σ_{Low}	σ_{High}	S_{AV}	S_{VA}
Unconstrained (All Ps)	**-199 (14)**	**215 (22)**	**109 (12)**	**144 (12)**	**-90 (10)**	**22 (21)**
Constrained (All Ps)	**-200 (14)**	**212 (22)**	**122 (13)**	**144 (12)**	**-76 (9)**	**53 (9)**
Unconstrained (Best-fitted Ps, N=6))	**-216 (33)**	**243 (32)**	**94 (16)**	**162 (12)**	**-116 (17)**	**4 (38)**
Constrained (Best-fitted Ps, N=3)	**-262 (52)**	**279 (34)**	**118 (26)**	**181 (16)**	**-119 (38)**	**80 (32)**

Table 2. *Means (SEMs) of best-fitting model parameters for the six-parameter criterion-change model (both unconstrained and constrained variants) for all participants (Ps), and for the subset of participants for whom this model provided the best fit*

	Mean best-fitting parameter (ms)					
	B_{Low}	B_{High}	σ_{Low}	σ_{High}	C_{AV}	C_{VA}
Unconstrained (All Ps)	**-193 (16)**	**175 (16)**	**115 (11)**	**137 (12)**	**-70 (24)**	**107 (28)**
Constrained (All Ps)	**-182 (14)**	**179 (16)**	**123 (14)**	**142 (11)**	**-91 (17)**	**129 (20)**
Unconstrained (Best-fitted Ps, N=10)	**-205 (23)**	**212 (22)**	**127 (22)**	**136 (13)**	**-110 (36)**	**158 (46)**
Constrained (Best-fitted Ps, N=11))	**-194 (16)**	**211 (20)**	**104 (7)**	**136 (13)**	**-133 (15)**	**169 (26)**

Table 3. *Means (SEMs) of best-fitting model parameters for the six-parameter population-code model for all participants (Ps), and for the subset of participants for whom this model provided the best fit (compared to constrained variants of other models)*

	Mean best-fitting parameter					
	N	σ (ms)	α	σ_a (ms)	B_{Low} (ms)	B_{High} (ms)
All Ps	**17 (1)**	**1914 (119)**	**0.21 (0.04)**	**86 (11)**	**-255 (14)**	**232 (20)**
Best-fitted Ps (N = 8)	**12 (1)**	**1380 (150)**	**0.2 (0.06)**	**101 (20)**	**-210 (25)**	**157 (22)**

Figure legends

Legend to Figure 1

Modelling schemes. **A.** In the latency models used here, two stimuli, such as an auditory beep and visual flash separated by 50 ms (bottom) accrue Gaussian latency noise as they pass through the brain. In this example, average latency is the same for each signal, so the mean relationship between objective asynchronies (measured at the sense organs; x-axis in the third graph down) and subjective asynchronies (at some intra-brain comparator; y-axis) is described by the line y = x. However, internal noise means that the subjective SOA actually falls somewhere in a Gaussian distribution centred (vertically) on this line. Hence, by slicing this graph vertically, we can focus in on the subjective SOA distribution predicted for multiple repetitions of a single (-50 ms) stimulus pair (second graph down). Stimuli are judged simultaneous when the subjective SOA falls between two decision criteria, but these criteria are not stable. They are modelled as Gaussian random variables (shown by graded shading in the figure, where dark indicates high probability density). In this example, the AV criterion is more stable than the VA criterion. Integrating the region between the (variable) decision criteria predicts the relevant point on the psychometric function for SJs (top graph). **B.** The same -50 ms SOA stimulus (bottom) yields a pattern of activity in a population code (third graph down) in which each neural unit has a Gaussian tuning curve that falls off around its preferred asynchrony. Firing rates reflect the position of the stimulus along a unit's tuning curve (plus Poisson noise) with adaptation suppressing the tuning curves of neurones with nearby preferences (inset). A decoding layer receives weighted inputs from all neural units and generates a maximum-likelihood estimate of the SOA. Because neural units are noisy, the same stimulus generates a distribution of decoded SOAs across multiple trials (second graph down). Stimuli are judged simultaneous if they fall between two decision criteria, so integrating the region between these criteria predicts the -50 ms SOA point on the SJ psychometric function (top graph).

Legend to Figure 2

Model predictions. The predicted biases in subjective SOAs (from baseline testing to post-adaptation conditions) are depicted (top) above the resultant changes in psychometric functions for simultaneity judgements (bottom). Predictions shown here were generated using the mean values of model parameters obtained experimentally. For clarity, the adapt-VA condition has been omitted. **A.** The latency-shift model predicts no shift in subjective SOAs following exposure to synchronous adaptors, but a uniform shift at all test SOAs following exposure to an auditory lead. The entire SJ psychometric function therefore moves a corresponding amount along the objective-SOA axis. **B.** The criterion-change model predicts no sensory changes at all, so no shift in subjective SOAs, but rather a movement of the decision boundary (dashed grey lines in upper figures) near the point of adaptation, specifically for non-zero adaptors. The SJ psychometric function therefore expands outwards on the side of adaptation. **C.** The population-code model predicts non-uniform repulsive changes in subjective SOA depending on the relative position of the test and the adaptor. The SJ psychometric function therefore contracts inwards, on both sides following synchronous adaptation, and from the side opposite adaptation when adaptors are presented at the decision boundary that divides judgements of an auditory lead from judgements of synchrony.

Legend to Figure 3

Schematic of methods. **A.** Flow-chart depiction of the experiment. Dotted grey lines indicate how values obtained in baseline fits were used as adaptors in subsequent conditions. **B.** Top-up and test procedure used in adaptation blocks, showing an example of VA adaptation. For black & white reproductions: LED flashes were red during adaptation and green during test.

Legend to Figure 4

Simultaneity judgement data (grey circles) from three different example participants (left to right) in all four conditions (top to bottom) alongside fits (solid black lines) from the best-fitting model

in each case. Positive SOA values indicate sound lagging light. Participant N (in part **A)** was best fitted by the latency-shift model, participant CC (in part **B)** by the criterion-change model, and participant X (in part **C)** by the population-code model. Circle size indicates the number of contributing judgements for each data point.

Legend to Figure 5

Group average data. **A.** Mean log-likelihood (indicating goodness of model fit) for all three six-parameter models, including unconstrained and constrained variants of latency-shift and criterion-change models (see main text for further details). Asterisks (*) denote a significant difference ($p < 0.05$) from all other conditions. Error bars denote standard errors of the mean. **B.** As part A, but showing mean Akaike Information Criteria (AIC) for five-parameter latency-shift and criterion-change models predicting only symmetric psychometric functions (plus their constrained variants) and the six-parameter population-code model. **C.** Simultaneity judgement data collated from all participants. Note that the adaptive method for stimulus selection meant that extreme asynchronies were less well sampled and predominantly reflect judgements from noisier participants. In this panel, error bars denote the Agresti-Coull 95% binomial confidence interval.

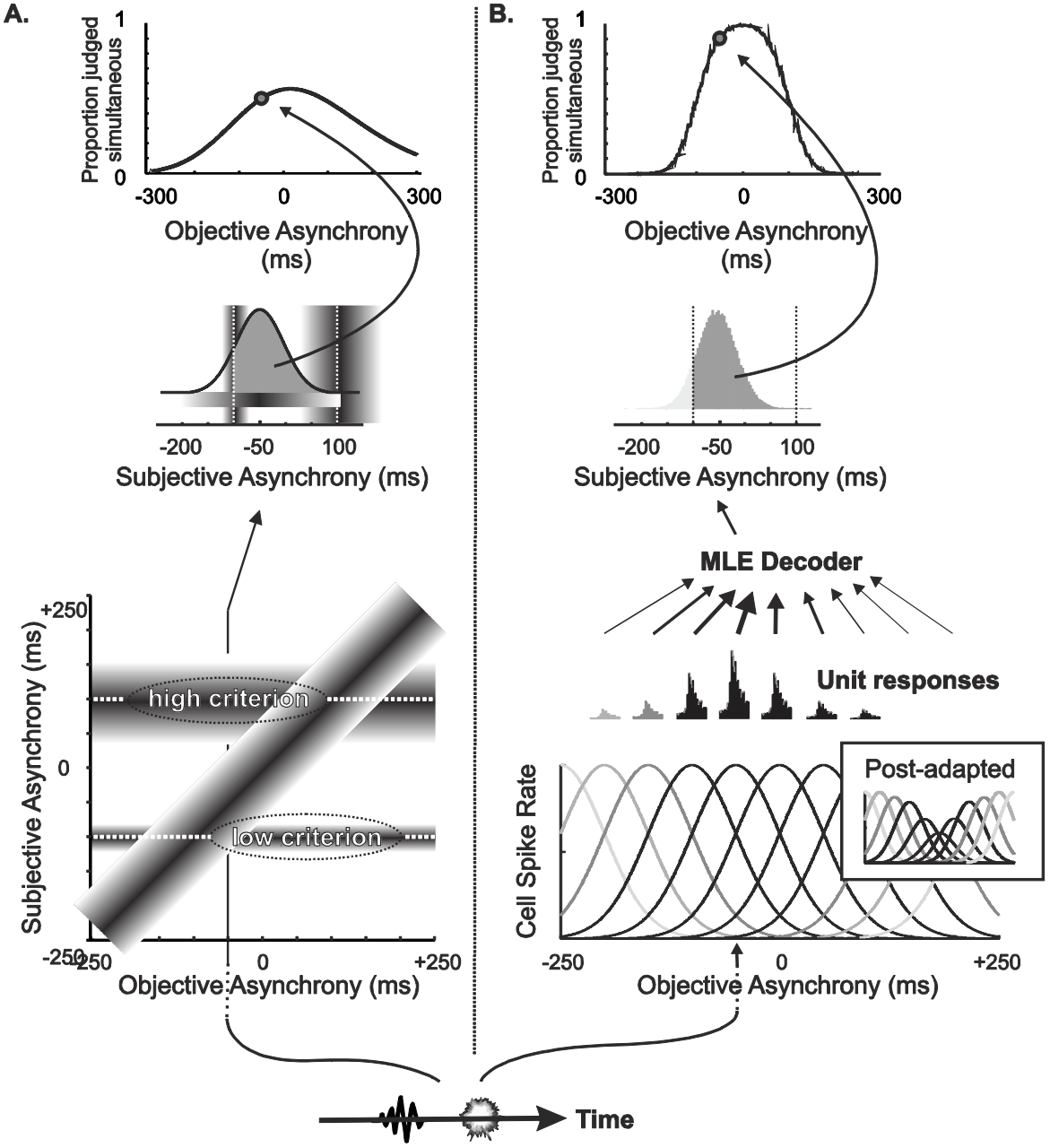
Legend to Figure A1

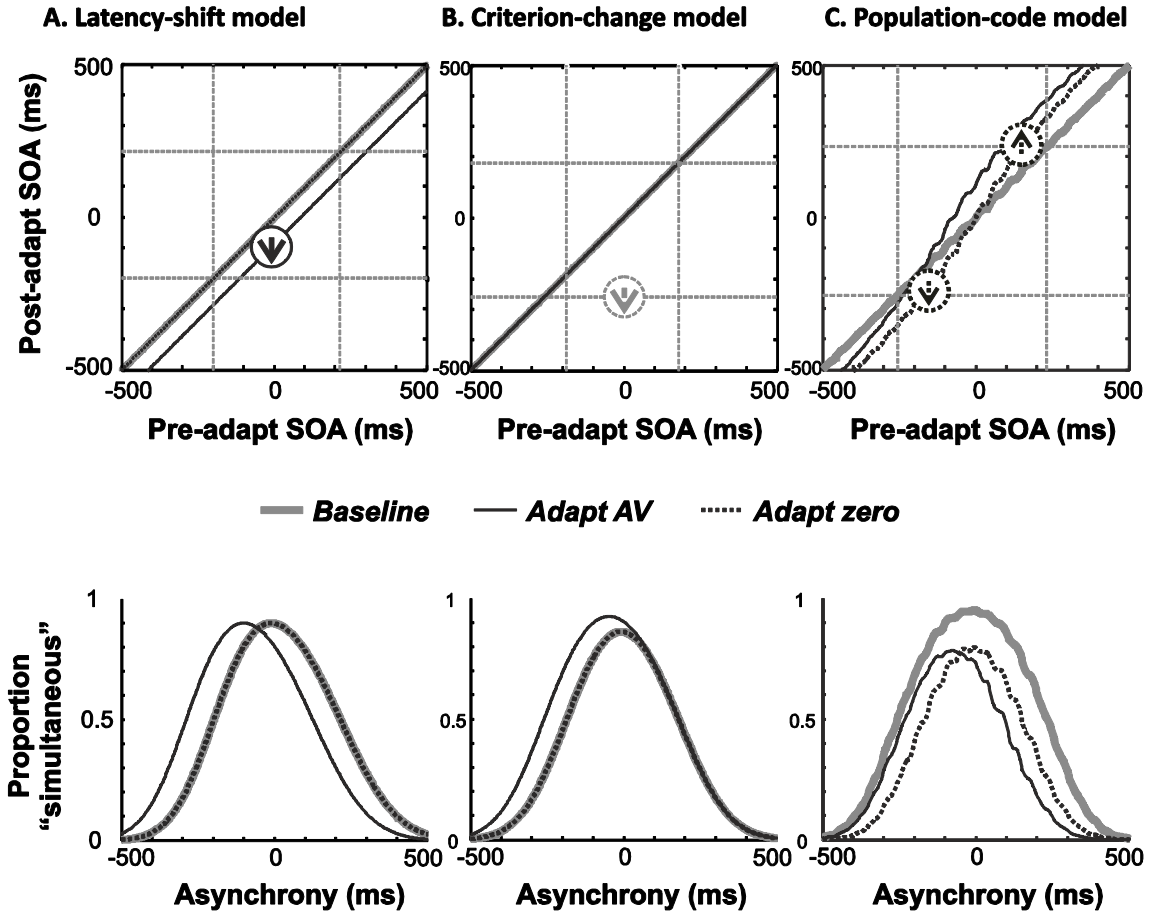
Examples of how changes to model parameters for a latency model (**A-C**) and a population-code model (**D-F**) affect the resultant psychometric function. **A.** Effects of changing the positions of decision criteria for judging events synchronous (B_{High} and B_{Low}) from +/-100 ms (thin black curve) to +/-200 ms (medium dark-grey curve) to +/-300 ms (thick light-grey curve) **B.** Effects of increasing variability in the Δt distribution of relative arrival times (affecting both σ_{Low} and σ_{High} parameters), with noise increasing from 10,000 ms^2 (thin black curve) to 20,000 ms^2 (medium dark-grey curve) to 30,000 ms^2 (thick light-grey curve) **C.** Effects of increasing variability in the placement of the B_{High} decision criterion (affecting only the σ_{High} parameter), with noise increasing from 0 ms^2 (thin black curve) to

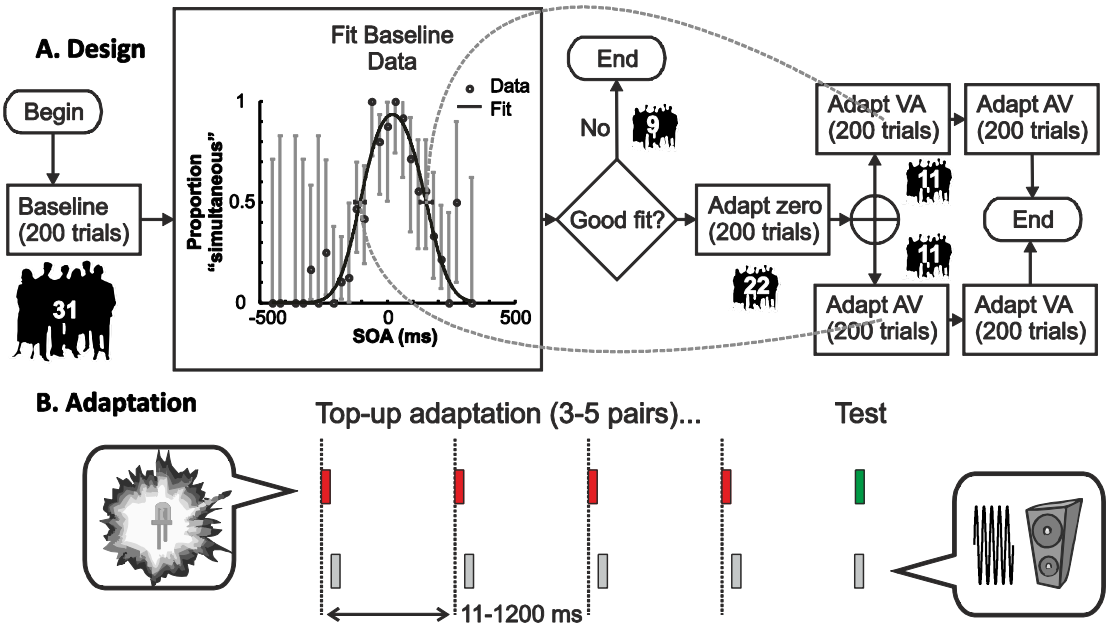
10,000 ms^2 (medium dark-grey curve) to 20,000 ms^2 (thick light-grey curve) **D.** Effects of changing tuning-curve widths (σ) for a population of 25 (N = 12) neural units. Increasing SD from 1200 ms (thin black curve) to 1500 ms (medium dark-grey curve) to 1800 ms (thick light-grey curve) flattens the psychometric function. **E.** Effects of changing number of neural units on either side of zero (N) for a population with tuning curve SD (σ) of 1500 ms. Increasing N from 9 (thin black curve) to 12 (medium dark-grey curve) to 15 (thick light-grey curve) sharpens the psychometric function. **F.** Although σ and N both modify the psychometric function in similar ways, varying them independently can lead to subtle differences. For a population with low N (N = 3, σ = 600 ms; thin black curve) a fairly similar psychometric function can be obtained from a larger population with broader tuning (N = 12, σ = 1400 ms; medium light-grey curve) but it has heavier tails. If a bias is present (B_{Low} increased from -100 to -50 ms and B_{High} increased from 100 to 150 ms) the low-N population produces a skewed psychometric function (thin dotted black curve) whereas the high-N population continues to produce a symmetric function (medium dotted light-grey curve).

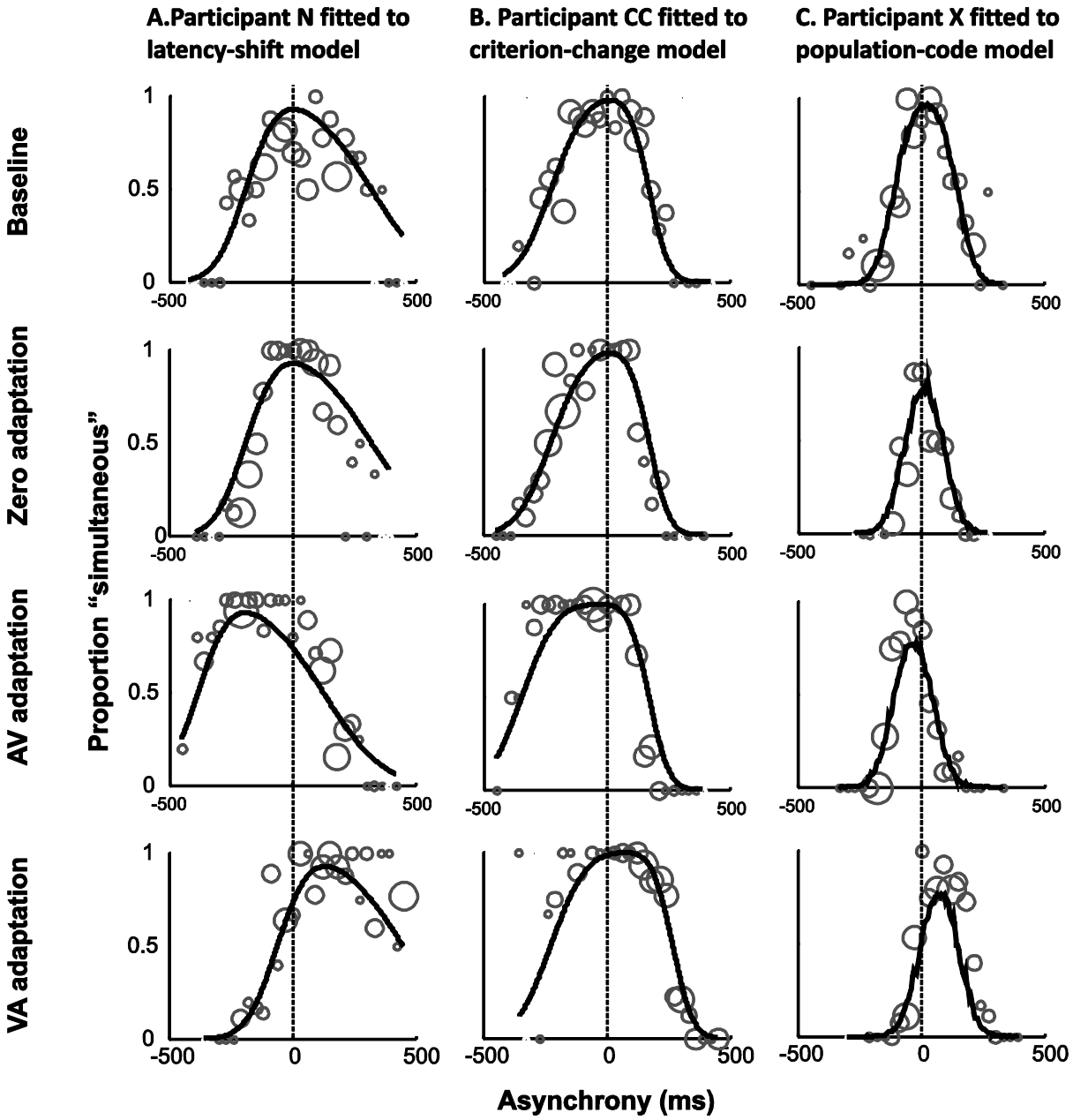
Legend to Figure A2

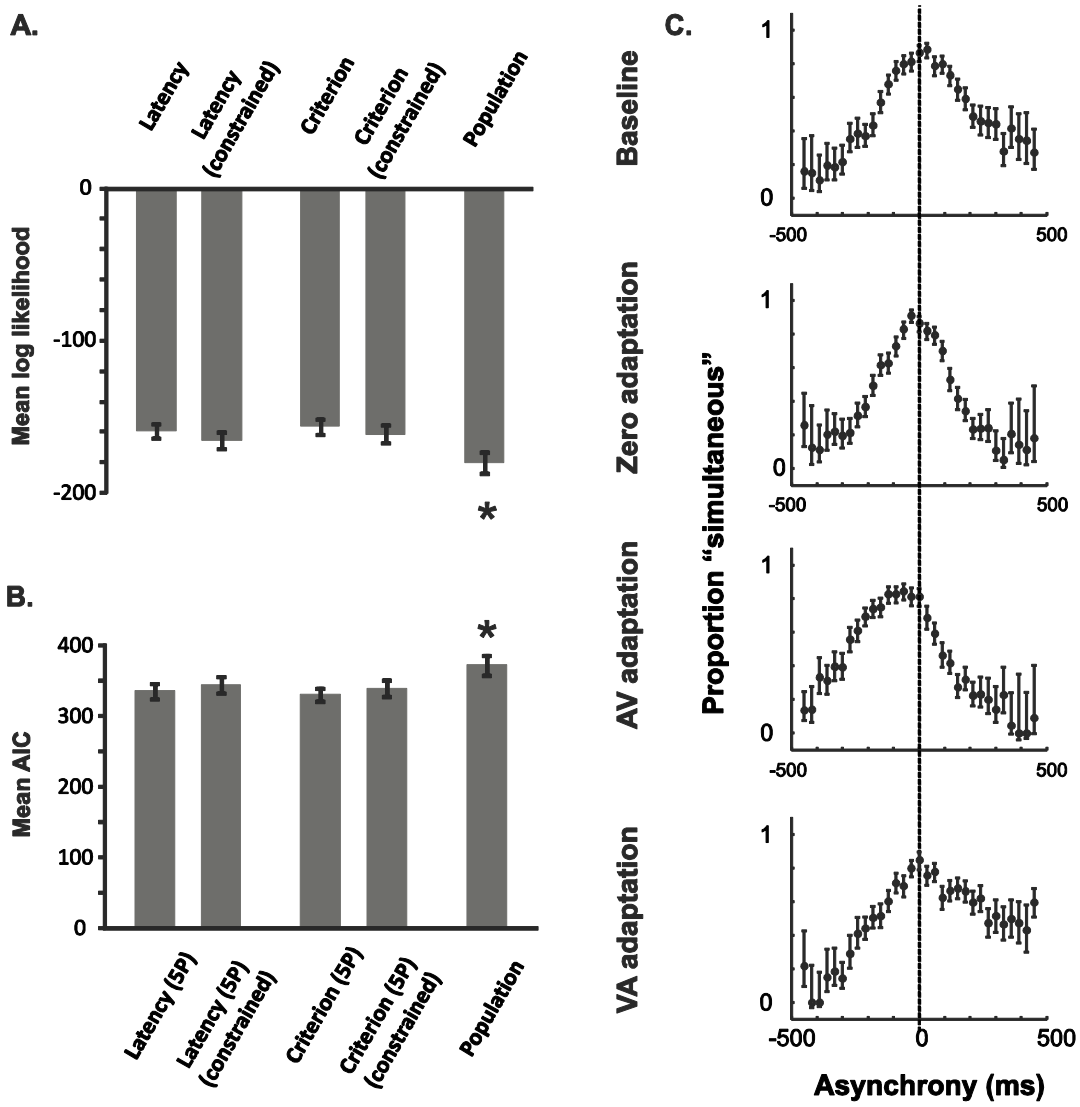
Summary of independent fits to each adaptation condition. In black: Mean parameter estimates for the low and high transition boundaries between judgements of synchrony and asynchrony (B_{Low} and B_{High}) in all four adaptation conditions (error bars denote standards errors). In grey: Psychometric functions for SJs generated using group-average parameters.

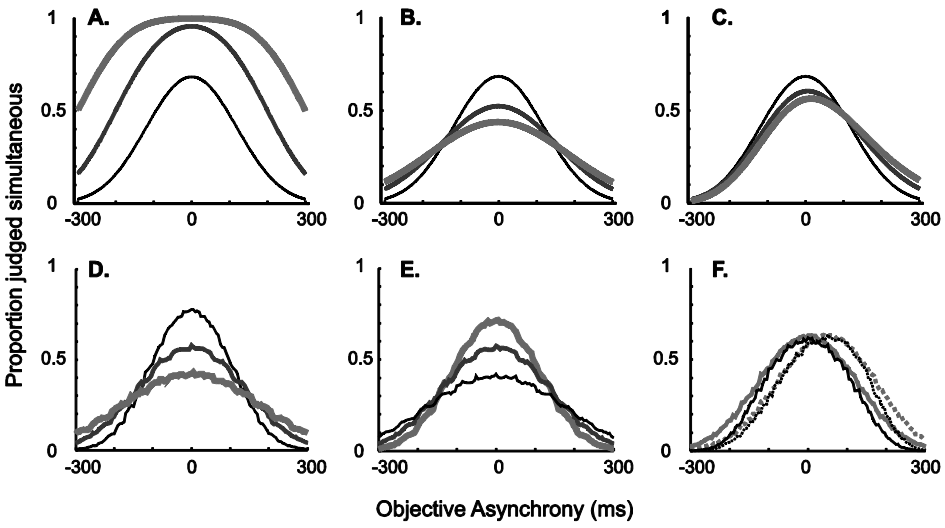

A.
B.
Proportion judged simultaneous
1
0
-300
0
300
Objective Asynchrony (ms)
-200
-50
100
Subjective Asynchrony (ms)
MLE Decoder
Unit responses
Post-adapted
Cell Spike Rate
Subjective Asynchrony (ms)
+250
0
-250
high criterion
low criterion
Objective Asynchrony (ms)
Time

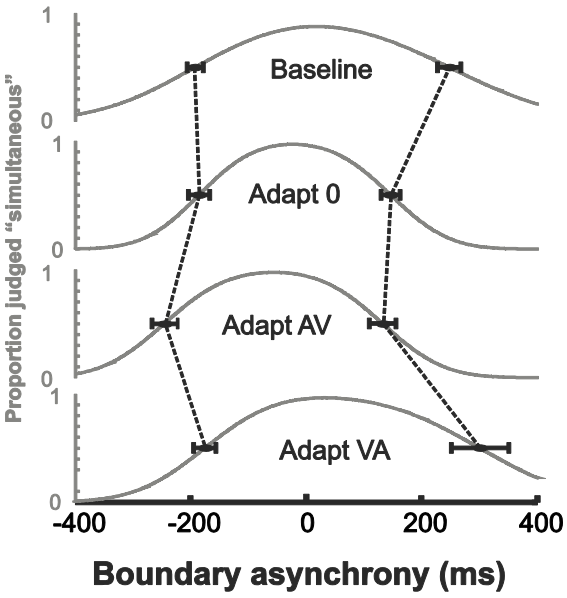

A. Latency-shift model
B. Criterion-change model
C. Population-code model
Post-adapt SOA (ms)
Pre-adapt SOA (ms)
Pre-adapt SOA (ms)
Pre-adapt SOA (ms)
Baseline
Adapt AV
Adapt zero
Proportion "simultaneous"
Asynchrony (ms)
Asynchrony (ms)
Asynchrony (ms)


A. Design
Begin
Baseline (200 trials)
31
Fit Baseline Data
Data
Fit
Proportion "simultaneous"
1
0.5
0
-500
0
500
SOA (ms)
End
No
9
Good fit?
Adapt zero (200 trials)
22
Adapt VA (200 trials)
Adapt AV (200 trials)
11
11
Adapt AV (200 trials)
Adapt VA (200 trials)
End
B. Adaptation
Top-up adaptation (3-5 pairs)...
Test
11-1200 ms


A.Participant N fitted to latency-shift model
B. Participant CC fitted to criterion-change model
C. Participant X fitted to population-code model
Baseline
Zero adaptation
AV adaptation
VA adaptation
Proportion "simultaneous"
Asynchrony (ms)
1
0.5
0
-500
0
500


A.
Latency
Latency (constrained)
Criterion
Criterion (constrained)
Population
Mean log likelihood
0
-100
-200
*
B.
Mean AIC
400
300
200
100
0
*
Latency (5P)
Latency (5P) (constrained)
Criterion (5P)
Criterion (5P) (constrained)
Population
C.
Baseline
Zero adaptation
AV adaptation
VA adaptation
Proportion "simultaneous"
1
0
-500
500
Asynchrony (ms)


A.
B.
C.
D.
E.
F.
1
0.5
0
-300
0
300
Proportion judged simultaneous
Objective Asynchrony (ms)


Proportion judged "simultaneous"
1
0
Baseline
Adapt 0
Adapt AV
Adapt VA
-400
-200
0
200
400
Boundary asynchrony (ms)