

# Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception

Waka Fujisaki · Shin'ya Nishida

Received: 1 October 2008 / Accepted: 16 May 2009 / Published online: 5 June 2009  
© Springer-Verlag 2009

**Abstract** To see whether there is a difference in temporal resolution of synchrony perception between audio–visual (AV), visuo–tactile (VT), and audio–tactile (AT) combinations, we compared synchrony–asynchrony discrimination thresholds of human participants. Visual and auditory stimuli were, respectively, a luminance-modulated Gaussian blob and an amplitude-modulated white noise. Tactile stimuli were mechanical vibrations presented to the index finger. All the stimuli were temporally modulated by either single pulses or repetitive-pulse trains. The results show that the temporal resolution of synchrony perception was similar for AV and VT (e.g., ~4 Hz for repetitive-pulse stimuli), but significantly higher for AT (~10 Hz). Apart from having a higher temporal resolution, however, AT synchrony perception was similar to AV synchrony perception in that participants could select matching features through attention, and a change in the matching-feature attribute had little effect on temporal resolution. The AT superiority in temporal resolution was indicated not only by synchrony–asynchrony discrimination but also by simultaneity judgments. Temporal order judgments were less affected by modality combination than the other two tasks.

**Keywords** Psychophysics · Temporal synchrony · Audio–visual–tactile · Synchrony–asynchrony discrimination · Simultaneity judgment · Temporal order judgment

W. Fujisaki (✉)  
National Institute of Advanced Industrial Science  
and Technology (AIST), Tsukuba Central 6,  
1-1-1, Higashi, Tsukuba, Ibaraki 305-8566, Japan  
e-mail: w-fujisaki@aist.go.jp

S. Nishida  
NTT Communication Science Laboratories,  
NTT Corporation, 3-1, Morinosato-Wakamiya,  
Atsugi-shi, Kanagawa 243-0198, Japan

## Introduction

Sensory information transmitted through our sensory systems can be divided into two categories. One is modality-specific information, such as color in vision and pitch in audition. The other is modality invariant or “supermodal” information, such as temporal and spatial locations, which can be conveyed through multiple modalities. Coincidences in the supermodal information, i.e., temporal synchrony and spatial alignment, serve as critical cues for signal binding across different sensory modalities.

Several interesting characteristics of multisensory temporal synchrony/simultaneity perception have been suggested in the audio–visual (AV) domain. For the purpose of understanding the underlying mechanisms from the viewpoint of system identification, we consider the following three characteristics of AV synchrony perception to be important. First, in comparison with within-modal synchrony perception, AV synchrony has a wide temporal window (Dixon and Spitz 1980; Fujisaki et al. 2004; Lewkowicz 1996; Zampini et al. 2005b), or, equivalently, a low temporal resolution (Arrighi et al. 2006; Fujisaki and Nishida 2005, 2007, 2008b; Recanzone 2003; Shipley 1964; Vatakis and Spence 2006c).<sup>1</sup> Second, whereas AV temporal judgments can be made between various combinations of visual features (e.g., changes in luminance, size, and position) and auditory features (e.g., changes in amplitude, frequency, and source location), the temporal resolution

<sup>1</sup> In this paper, we focus on the basic characteristics of multisensory temporal synchrony perception for simple, non-speech stimuli. The AV temporal window for complex stimuli is known to be wider than for simple stimuli (e.g., Dixon and Spitz 1980). Several recent studies also demonstrate that multimodal temporal synchrony perception for speech stimuli is somewhat special (Vatakis and Spence 2006a, b, 2007, 2008).

remains nearly the same regardless of the feature combination (Fujisaki and Nishida 2007). We will hereafter call this characteristic as “matching-feature invariance”. Third, the participants can select matching events of each modality through attention (Fujisaki et al. 2006; Fujisaki and Nishida 2007; van de Par and Kohlrausch 2004). For example, in a visual search task, a search of a sound-synchronized visual target among uncorrelated visual distracters requires a sequential shift of attention across space (at least when stimulus temporal density is sufficiently high; cf., Van de Burg et al. 2008). In contrast, in a selective attention task, interference by surrounding distractors on AV synchrony–asynchrony discrimination of a target can be completely excluded by pre-cuing of the target location (Fujisaki et al. 2006). These findings suggest that AV synchrony detection is preceded by nearly perfect attentive selection of the spatial position of a visual event. It is also possible to select AV matching events by feature-based attention (e.g., color, pitch), although, unlike in position-based attention, the selection is not perfect (Fujisaki and Nishida 2008b). We will hereafter call this characteristic as “attentive matching-feature selection”.

These characteristics are consistent with a hypothesis proposed by Fujisaki and Nishida (2005; see also Fujisaki et al. 2006; Fujisaki and Nishida 2007, 2008b; Nishida and Johnston 2002) that synchrony judgments across separate sensory channels, such as AV judgments, are based on the matching of temporally sparse, salient features extracted from each sensory signals. We consider this to be mid-level processing, slightly but significantly higher than peripheral low-level sensory processing. This is because the processing is under the control of attention, as suggested by attentive matching-feature selection, and because it does not directly compare raw sensory signals, but instead compares more abstract salient features extracted from within-modal signal streams. Salient features are like figures in figure-ground segregation, a general supra-modal representation common across attributes. We consider this is to be where matching-feature invariance arises.

In the present study, to understand the computational mechanisms underlying cross-modal synchrony perception in general, we broadened our scope of investigation to include the third modality, touch. Specifically, we compared the temporal resolution of synchrony judgments among audio–visual (AV), visuo–tactile (VT), and audio–tactile (AT) pairs to see how much similarity and difference there was in temporal resolution across different cross-modal combinations. We then tested attentive matching-feature selection and matching-feature invariance for AT pairs to investigate the applicability of our hypothetical model to non-AV cross-modal judgments.

To access the sensitivity limit of synchrony perception, we primarily used a synchrony–asynchrony discrimination

task (Fujisaki and Nishida 2005, 2007). In this task, participants are shown only two magnitudes of asynchrony, 0 ms (synchrony) and  $X$  ms (asynchrony), in a single block, and asked to make discrimination between the two alternatives. To avoid errors in assigning the responses of “synchrony” and “asynchrony” to the two magnitudes of asynchrony, it is crucial to provide feedback and to let the participants know which percept corresponds to which stimulus. This feedback allows the participants to properly perform the task even in an extreme situation where they feel a physical synchrony stimulus is less synchronous than the paired asynchronous stimulus. Given that the participants fully utilize the feedback information to optimize their performance, the resulting discrimination performance is expected to reveal limitations of the perceptual mechanisms.

There are two other popular methods for measuring the temporal accuracy of simultaneity/synchrony judgments. One is to estimate the range of apparent simultaneity by using simultaneity judgments (SJ), and the other is to estimate the just noticeable difference (JND) from the slope of psychometric function of temporal order judgments (TOJ).

In the SJ task, participants are shown a variety of asynchronies in a single block and judge whether the perceptual impression for each asynchrony falls in their subjective category of simultaneity (Dixon and Spitz 1980; Lewkowicz 1996). Depending on the criterion used by the participants, the width of the simultaneity window can change significantly. Two distinguishable asynchrony values may be classified into the same “simultaneous” category when the participants apply a generous criterion.<sup>2</sup> In contrast, in the synchrony–asynchrony discrimination, participants are not required to judge the perceived stimulus in relation to their subjective category of simultaneity. Instead they have to judge the similarity of the perceived stimulus to the memory of the two magnitudes of asynchrony established through feedback. To minimize

<sup>2</sup> Another problem with SJ arises when one attempts to measure the temporal acuity limit in terms of the upper temporal frequency for correct synchrony perception using repetitive stimuli. For the AV combination, and possibly for other modality combinations (Yau et al. 2009), when the stimulus temporal frequency is higher than the limit of synchrony perception, while lower than the limit of perception of visual flickers and auditory flutters, observers often report apparent synchrony even if the paired stimuli were different in temporal frequency. Further in AV, the visual flicker rate is apparently captured by the auditory flutter (auditory driving, see Recanzone 2003; Shipley 1964). These phenomena suggest the mechanism underlying SJ is complex, which makes the interpretation of the SJ data harder. In contrast, synchrony–asynchrony discrimination tasks can be used similarly for non-repetitive and repetitive stimuli because auditory driving is observed only at high frequencies where synchrony–asynchrony discrimination is impossible. This effect is negligible as far as observers’ performance is evaluated in terms of the threshold of discrimination.

potential effects of participants' criterion that we could not control, we primarily used synchrony–asynchrony discrimination, and checked the consistency with SJ data afterwards.

The participants' criterion also affects the performance of TOJ, but the expected effect is a horizontal shift of the psychometric function, not a change in its slope. Therefore, the JND of TOJ can be considered as a stable measure of temporal acuity. However, the processing required for judging the temporal order of two events may be significantly different from, and presumably more complex than, that required for judging synchrony (Pöppel 1988; van de Par et al. 2002). In fact, several studies have shown conflicting results between TOJ and SJ (Schneider and Bavelier 2003; van de Par et al. 2002; van Eijk 2008; Vatakis et al. 2008). A difference emphasized by Pöppel (1988), and one that most relevant to the present study, is that the JND of TOJ is relatively independent of the stimulus combination, while the window of subjective simultaneity is stimulus-dependent. The temporal acuity estimated by TOJ should be considered separately from the temporal acuity estimated from synchrony perception.

Our current interest is whether the temporal acuity of cross-modal synchrony perception is dependent on the sensory combination tested. The threshold time lag required to discriminate asynchrony from synchrony for AV pairs is about 80 ms (single pulse) or 120 ms (repetitive pulse) (Fujisaki and Nishida 2005). We examined whether the threshold lag was similarly large for VT and AT pairs. A few previous studies have measured the temporal resolutions of non-AV cross-modal pairs using TOJ or SJ (Harrar and Harris 2008; Hirsh and Sherrick 1961; Martens and Woszczyk 2004; Occelli et al. 2008; Zampini et al. 2005a). The classic study by Hirsh and Sherrick (1961) has shown that the temporal resolution estimated in terms of the JND of TOJ was similar across AV, VT, and AT pairs. More recently, Zampini et al (2005a) have shown that JND of TOJ obtained for AT was similar to, or slightly larger than, those obtained for VT (Spence et al. 2003) and AV (Zampini et al. 2003), although the participants were different in these studies. The data of Harrar and Harris (2008) also show a similar trend. However, some of these studies might not have optimized the stimulus parameters to compare the best performance across modality combinations. In addition, as noted above, it remains unknown how much JND of TOJ reflects the temporal resolution of cross-modal synchrony processing. As for SJ, as far as we know, no systematic comparison has been made among different cross-modal combinations. A conference proceeding (Martens and Woszczyk 2004) reported the range of simultaneity between sound and

whole-body vibration measured for the purpose of evaluating a bimodal display system. AT simultaneity window reported by this study appeared to be narrower than the AV windows reported by others (e.g., Zampini et al. 2005b; Fujisaki et al. 2004). However, this study neither exclude cross-talk artifacts nor measure comparable data for other modality combinations using common stimuli and participants.

This background motivated us to carry out a systematic study, as reported here, in which we measured synchrony–asynchrony discrimination thresholds for AV, VT, and AT pairs and compared them with the temporal resolutions obtained with SJ and TOJ. Since we had comparable within-modal data only for vision and audition (Fujisaki and Nishida 2005), we added tactile–tactile (TT) to the stimulus conditions of the present experiments.

In the first experiment, we measured synchrony–asynchrony discrimination performance for AV, VT, AT, and TT pairs to test whether the low temporal resolution was a common property of cross-modal synchrony regardless of the combination of modalities. The results showed that the temporal resolution was similar for AV and VT pairs but significantly higher for AT pairs.

The next two experiments were designed to test whether the higher temporal resolution of AT synchrony perception was mediated by a cross-modal mechanism similar to those responsible for AV and VT synchrony perceptions. We tested whether AT synchrony perception also showed attentive matching-feature selection and matching-feature invariance, which we hypothesize to be characteristic properties of mid-level cross-modal processing. These properties are unlikely to be obtained if AT superiority in temporal resolution is an artifact of the contribution of peripheral within-modal (purely auditory or tactile) mechanisms. The results suggest that despite having higher temporal resolution, the mechanism underlying AT synchrony perception is functionally similar to those for AV and VT synchrony perceptions, and thus captured by our salient-feature matching model.

In the last experiment, to assess task dependency, we measured temporal resolutions using SJ and TOJ and compared the data with those obtained with the synchrony–asynchrony discrimination by the same participants. The SJ showed AT superiority in temporal resolution very similar to that found in synchrony–asynchrony discrimination. A similar trend was also observed with TOJ, but the difference among different cross-modal combinations was not as evident as in synchrony–asynchrony discrimination and SJ. This might be a reason why AT resolution superiority has not been recognized in past TOJ studies.

## Experiment 1

### Methods

#### *Participants*

Participants were the two authors and five paid volunteers who were unaware of the purpose of the experiments. All had normal or corrected-to-normal vision and hearing, and were right handed. Informed consent was obtained after the nature and possible consequences of the studies were explained. The same participants took part in the subsequent experiments.

#### *Apparatus and stimuli*

Visual stimuli were presented with a VSG2/5 (Cambridge Research Systems). Auditory stimuli were presented with a TDT Basic Psychoacoustic Workstation (Tucker-Davis Technologies). Tactile stimuli were presented with a TDT RM1 Real-time Mini Processor (Tucker-Davis Technologies) and vibration generators (511-A, EMIC). The system was controlled by Matlab (The MathWorks) running on a PC (Dell Precision 360). Precise time control of audio-visual-tactile stimuli was accomplished by driving the TDTs by a sync signal from the VSG.

In a quiet dark room, the participant sat 57 cm from a monitor (Sony GDM-F500, frame rate 160 Hz). The visual stimulus was a luminance-modulated Gaussian blob (standard deviation  $2^\circ$ ) presented at the center of the monitor screen ( $21.5 \text{ cd m}^{-2}$  uniform field,  $38.7^\circ$  in width,  $29.5^\circ$  in height). The Gaussian blob was used, since the visual response is rapid to gradual luminance modulations (Kelly 1979). The luminance increment of the blob peak was temporally modulated between 0 and  $43 \text{ cd m}^{-2}$ . Nothing was visible during the off period. The fixation marker was a bullseye presented before stimulus presentation at the center of the monitor screen. Participants were instructed to view the visual stimulus while maintaining their fixation at this location.

The auditory stimulus was a 100% amplitude-modulated white noise [54 dB sound pressure level (SPL) at the peak of modulation] presented diotically via headphones (Sennheiser HDA 200) with a sampling frequency of 24,420 Hz. Faint white noise (about 40 dB SPL) was also presented to mask the sound produced by tactile stimulus generators.

The tactile stimulus was vertical movements of a tip of a vibration generator (511-A, EMIC). The participant touched the tip with the index finger, for which temporal discrimination is known to be good (Hoshiyama et al. 2004). Two vibration generators, each for the left or right hand, were placed on a chair seat with the tip-to-tip separation of 25 cm. The cushioned chair seat absorbed unwanted vibrations arising from the bodies of the vibration generators

and prevented the participant from sensing by a route other than through fingers. The chair was placed underneath a desk, so the participant could not see the vibration generators. A wrist rest was strung between the front two legs of the desk. The participants supported the weight of their arm with the wrist rest, and touched the vibration generator softly with the index finger. Although the strength of finger pressure was under control of the participants, we assume they adjusted the pressure to fall within the range where they could clearly feel the vibrations, and we informally observed that this range was relatively wide.

We used repetitive-pulse trains to measure the temporal resolution in terms of the upper temporal frequency and used single pulses to measure the temporal resolution in terms of the minimum time lag. For repetitive stimuli (Fig. 1a), the audio, visual, and/or tactile stimuli were modulated by the same periodic pulse train, resulting in a pair comprising a visual flicker and an auditory flutter (for AV condition), a visual flicker and a tactile flutter (VT), an auditory flutter and a tactile flutter (AT), or two tactile flutters presented to the two hands (TT). The pair was either in phase or  $180^\circ$  out of phase. Except for the TT condition, only the right index finger was used to sense the tactile stimulation. For non-repetitive stimuli (Fig. 1b), the AV, VT, AT, or TT stimuli were presented as a pair of single pulses. A single visual pulse was presented for one display frame (nominally  $1,000/160 = 6.25$  ms). An auditory pulse was presented for 6.25 ms. A single tactile pulse was an upward tip movement followed by a downward tip movement with an interval of 6.25 ms.

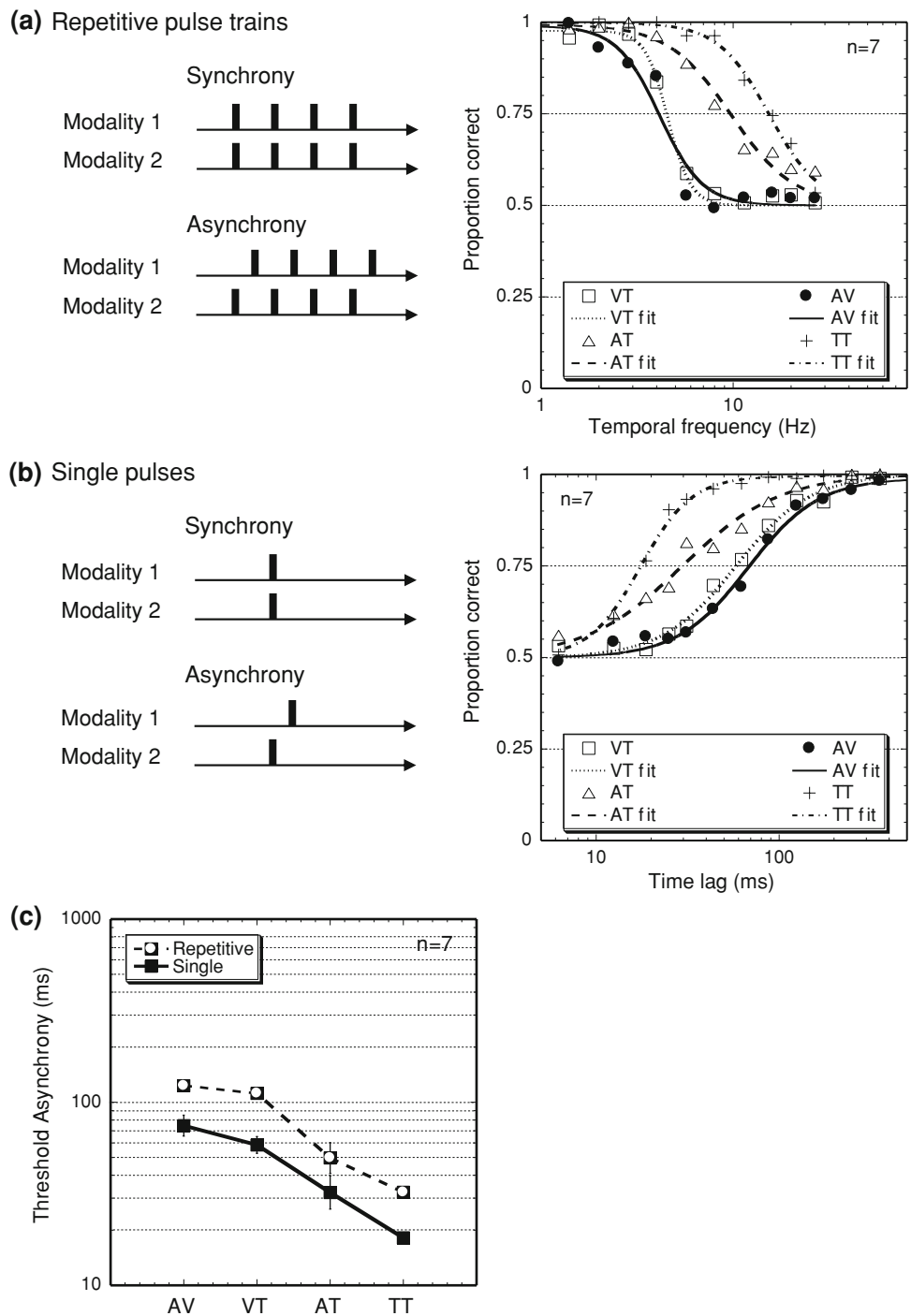
#### *Procedure*

For repetitive-pulse trains, the temporal frequency was changed from 1.4 to 26.7 Hz (10 steps). For single-pulse stimuli, the time lag was changed from 6.25 to 356.25 ms (12 steps).

The proportion correct for discriminating synchrony from asynchrony was measured in separate experimental blocks for different stimulus conditions. Each block consisted of ten trials, plus two initial practice trials, during which synchronous and asynchronous stimuli were presented in turn. In the next ten trials, five synchronous trials and five asynchronous trials were presented in a random order. Four blocks were run for each stimulus condition (e.g., AV, repetitive 1.4 Hz).<sup>3</sup>

<sup>3</sup> Since the numbers of synchronous and asynchronous trials were fixed, it was theoretically possible for the participants to use feedback to guess the next stimulus more accurately by chance. We consider, however, that the effect of this factor was negligible, since the obtained psychometric functions always dropped to the chance level for all the participants, including the authors.

**Fig. 1** Synchrony–asynchrony discrimination performance for audio–visual (AV), visuo–tactile (VT), audio–tactile (AT), and within-tactile (TT) conditions (experiment 1). **a** The results obtained with repetitive-pulse trains. For each stimulus combination, proportion correct is plotted as a function of the temporal frequency. Each data point is the average across seven participants. Smooth curves are the best-fit logistic functions. **b** The results obtained with single pulses. Proportion correct plotted as a function of the time lag of the asynchronous stimulus. **c** The threshold asynchrony, estimated for each participant from the 75% correct point of the logistic function, and averaged across participants (geometric mean). For repetitive stimuli, the threshold asynchrony is a half period of the threshold temporal frequency. *Error bars* indicate  $\pm 1$  SE. Threshold values for each participant can be found in Fig. 2d (repetitive-pulse trains) and in Fig. 3c (single pulses). The results show AT superiority over AV and VT conditions in synchrony–asynchrony discrimination



In a trial, 2 s after the last participant’s response, the fixation marker was removed. With no interval (repetitive pulse) or with an interval randomly chosen from the range between 800 and 1,200 ms (single pulse), a stimulus pair was presented, either synchronously or asynchronously. The participant had to make a two-alternative forced response (synchrony or asynchrony) by pressing a VSG response box key. A repetitive-pulse train lasted 6 s. For the purpose of preventing the participants from judging

synchrony from the onset and offset of the pulse train, a pulse train had 2-s cosine ramps both at the onset and offset of the stimulus. In addition, the stimulus delay started with a random phase, gradually shifted to the intended phase over the initial 2 s, and then kept that phase for the remaining 4 s. This procedure also excluded potential effects of task irrelevant prior events on discrimination performance, in particular the fixation marker offset. Feedback was given after each response by the color of the fixation marker. We

expected the feedback to exclude a type of error where the participants could discriminate the two lag conditions but not correctly label the physically “synchronous” pair as “synchronous.” In the single-pulse condition, delays were positive (modality 1 first) in half of the blocks and negative (modality 2 first) in the other half.

## Results

Figure 1a shows the proportion correct for repetitive-pulse trains as a function of temporal frequencies (in Hertz), and Fig. 1b shows the proportion correct for single pulses as a function of time lags (in milliseconds). Each psychometric function was fit by a logistic function to estimate the 75% correct point. The thresholds for all conditions are compared in Fig. 1c, where the data from repetitive-pulse trains are converted from temporal frequencies to time lags. Two-way repeated measure ANOVA of log threshold values indicates significant main effects of pulse type ( $F(1,6) = 39.99$ ,  $P = 0.0007$ ) and stimulus combination ( $F(3,18) = 51.07$ ,  $P < 0.0001$ ), but no interaction ( $F(3,18) = 0.9245$ ,  $P = 0.449$ ). For both pulse types, two-tailed paired  $t$  tests on log threshold values indicate significant differences for all comparisons ( $P < 0.036$ ) except that AT–TT differences were statistically marginal (repetitive,  $P = 0.051$ ; single  $P = 0.053$ ). These results indicate that for both repetitive and single pulses, temporal resolution was worst for AV, slightly better for VT, and much better for AT. However, the resolution of AT was not as high as that of within-modal TT.<sup>4</sup>

## Experiment 2

The first experiment demonstrated superiority of AT over the other cross-modal combinations (i.e., AV and VT) in temporal resolution of synchrony–asynchrony discrimination. The second experiment had two related purposes. One was to examine whether the AT perception, despite having higher temporal resolution, shares the property of attentive matching-feature selection with AV synchrony perception. The second purpose was to exclude the possibility of a within-modal artifact. Both are related to the question of whether a cross-modal mechanism mediates AT synchrony perception like it does AV and VT synchrony perception.

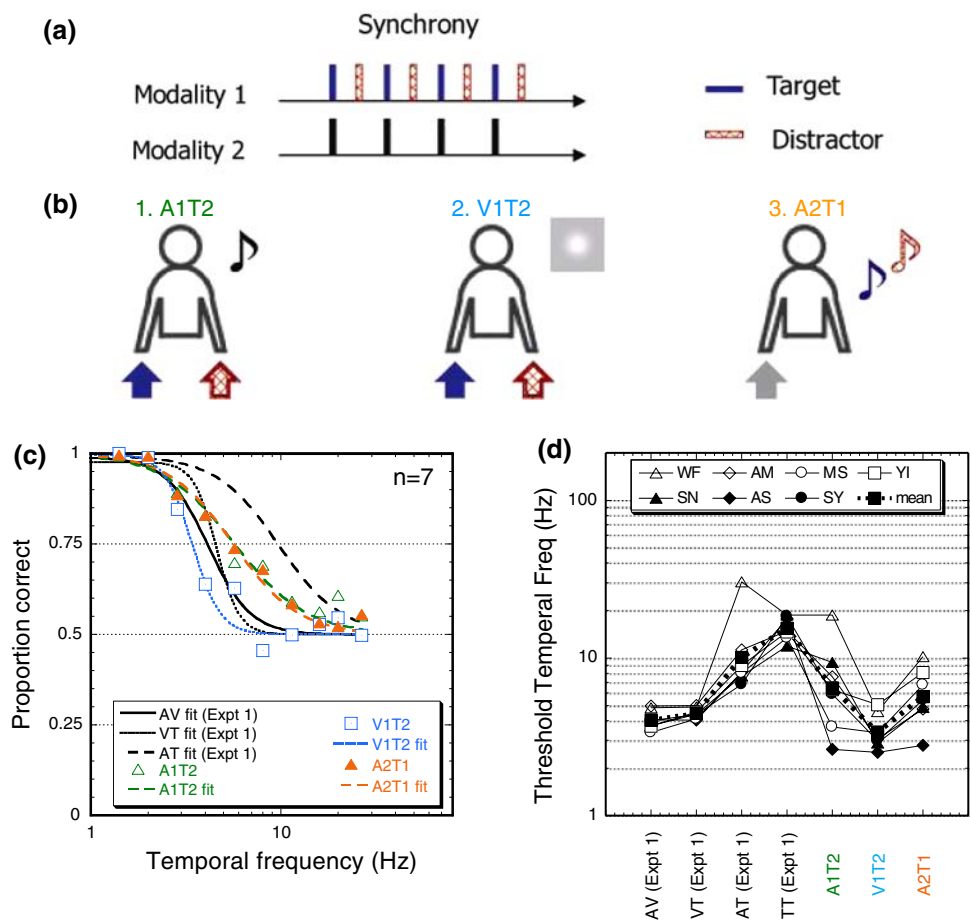
The design of the experiment is illustrated in Fig. 2a and b. We made three conditions from the repetitive-pulse train condition of experiment 1. For the first condition, A1T2,

the stimulus was a pair of repetitive auditory and tactile pulses. The difference from the original repetitive AT stimulus was that in addition to the target tactile sequence presented to the right index finger, a distractor sequence was presented in 180° anti-phase to the left index finger. In synchronous stimuli, the target tactile sequence was in synchrony with the auditory pulse sequence, whereas in asynchronous stimuli, the distractor tactile sequence was in synchrony. Discrimination of the two conditions will be impossible if the AT synchrony mechanism unselectively responds to the tactile signals from the two hands. In contrast, if the AT synchrony perception has the property of attentive matching-feature selection, the synchrony–asynchrony discrimination will be possible, although the performance may not be as good as the original AT condition due to the limitation of feature-based attentional selection (Fujisaki and Nishida 2008b). The second V1T2 condition was a variant of the A1T2 condition where the auditory sequence was replaced by a visual sequence. The third A2T1 condition was similar to A1T2 except that the roles of AT stimuli were swapped. Two auditory sequences, target and distractor, were presented in different pitches (622 and 1,480 Hz, triangular waves), and one of them was in synchrony with the tactile sequence presented to the right hand. If the AT synchrony judgment is based on a cross-modal process and the participant can attentively select one pitch, the synchrony–asynchrony discrimination will not be severely impaired. For the three stimulus conditions, we measured synchrony–asynchrony discrimination performance at various temporal frequencies using the procedure identical to the repetitive-pulse condition of experiment 1.

In these conditions, we could also examine whether the high temporal resolution of AT reflects a real cross-modal property, or instead is a product of a within-modal artifact. This is not only because attentive matching-feature selection is a characteristic property of cross-modal processing, but also because the synchrony–asynchrony task should be severely impaired under the conditions of the second experiment if it is based on within-auditory or within-tactile processing. The vibration generator itself produced small sounds when presenting tactile stimuli. We presented white mask noises that prevented participants from subjectively hearing any sound from the vibration generators. One might, however, think it hard to completely exclude the possibility that the tactile stimulus was detectable by the auditory system through, say, bone conduction. This artifact could change the AT task into a within-auditory (AA) synchrony judgment, which is known to be very rapid when peripheral sensors are involved. However, even if this is indeed the case, it would be difficult for the auditory system to assist synchrony–asynchrony discrimination under the A1T2 condition, since it would be hard to tell which finger is stimulated only by consciously inaudible sounds.

<sup>4</sup> It is suggested that the tactile system can sense much smaller inter-skin time differences than suggested by our TT data when touch source localization is tested using skin locations properly selected for this task (Von Békésy 1959).

**Fig. 2** A test of attentive matching-feature selection (experiment 2). **a** For one modality, two repetitive-pulse trains were presented 180° out of phase. The participants had to attentively select one of them (target) to make synchrony judgments with the pulse train of the other modality. **b** Three stimulus conditions. A1T2: a single auditory pulse and dual tactile pulses (to the left and right hands). V1T2: a single visual pulse train and dual tactile pulse trains. A2T1: dual auditory pulse trains (at different pitches) and a single tactile pulse train. **c** Proportion correct for synchrony–asynchrony discrimination plotted as a function of the temporal frequency. For the purpose of comparison, best-fit curves for AV, VT, and AT repetitive-pulse conditions (experiment 1) are also shown. **d** Threshold temporal frequency for each stimulus condition. Individual thresholds and their average (geometric mean) are shown. The results support attentive matching-feature selection in rapid AT synchrony–asynchrony discrimination



Therefore, successful discrimination in the A1T2 condition will be evidence against the within-auditory artifact. Another possible within-modal artifact is sound stimulation of tactile sensors around ears. In this case, the task could be changed into a within-tactile (TT) judgment. However, even in such a case, the tactile system would have difficulty assisting synchrony–asynchrony discrimination under the A2T1 condition, because it would be hard to discern the pitch only by tactile sensation. Therefore, successful discrimination in the A2T1 condition will be the evidence against the within-tactile artifact.

The results, shown in Fig. 2c and d, indicate that the participants could perform the discrimination task for both the A1T2 and A2T1 conditions. This would have been impossible if AT synchrony perception was mediated by a process that cannot discriminate target pulses from distractors. Such a process includes purely auditory or tactile mechanism (within-modal artifact). In comparison with the original AT condition of experiment 1, the discrimination performance deteriorated over a wide range of temporal frequencies, and the threshold frequency was reduced ( $P = 0.0343$  for A1T2 and  $P = 0.0128$  for A2T1 by two-tailed paired  $t$  tests). This can be ascribed to imperfect selection of the tactile target based on stimulated finger

(A1T2), and imperfect selection of the auditory target based on pitch (A2T1). Nevertheless, the discrimination performance for A1T2 or A2T1 did not drop below the performance for the AV condition in experiment 1. Temporal superiority of A1T2 or A2T1 to AV in terms of threshold frequency was statistically significant ( $P = 0.0368$  for A1T2 and  $P = 0.0436$  for A2T1, one-tailed paired  $t$  test).<sup>5</sup> These results suggest that the high resolution of AT synchrony perception is a property of a real AT process, and that this cross-modal process also has the property of attentive matching-feature selection.

The participants could also perform synchrony–asynchrony discrimination for the V1T2 condition at least at low temporal frequencies. This indicates that the mechanism

<sup>5</sup> Use of one-tailed tests is justified for testing whether the higher resolution for AT than for AV was preserved in the A2T1 and A1T2 conditions but a two-tailed  $t$  test does not indicate a significant difference in either case. Unfortunately, presumably because of the task complexity, A1T2 and A2T1 showed a large individual difference. In particular, the performance of one naïve participant was very low. For the other six participants, the threshold was consistently higher for A1T2 and A2T1 than for AV, and the differences were statistically significant even by two-tailed paired  $t$  tests ( $P = 0.0102$ ,  $P = 0.0104$ , respectively).

underlying VT synchrony perception also shows attentive matching-feature selection. The discrimination performance for the VIT2 condition was worse than for the original VT condition ( $P = 0.0187$ , two-tailed  $t$  test), which also suggests imperfect selection of the tactile target based on stimulated finger.

### Experiment 3

Matching-feature invariance is another important characteristic of AV synchrony perception (Fujisaki and Nishida 2007). In the third experiment, we examined whether the AT synchrony perception, which showed high temporal resolution, also shared this property. This is another test of whether AT synchrony perception is mediated by a standard cross-modal mechanism. In addition, if matching-feature invariance is observed for AT synchrony perception, then it will also provide another line of evidence against the possibility that the high temporal resolution of AT synchrony perception observed in experiment 1 was a product of within-modal artifacts.

The experimental procedure was the same as in experiment 1 except for the following. We changed the temporal structure of the auditory stimulus from a single pulse to a step onset while leaving the tactile stimulus as a pulse (Fig. 3a). The auditory change was defined either by amplitude modulation (AM), or by frequency modulation (FM). In the AM condition, participants were asked to judge whether the onset of a single tactile vibration pulse and a step onset of auditory white noise (about 50 dB SPL) were synchronous or asynchronous. In the FM condition, they were asked to judge whether the onset of a single tactile vibration pulse and an auditory pitch change were synchronous or asynchronous. The pitch change was an abrupt shift of a tone frequency from 440 to 2,200 Hz (7/3 octaves high) with its amplitude kept at about 60 dB SPL. If the high temporal resolution of AT synchrony perception is mediated by a cross-modal mechanism with matching-feature invariance, the stimulus modifications will not greatly affect the synchrony–asynchrony discrimination performance. However, this will not be the case if AT synchrony perception is mediated by such a mechanism, including within-modal one, which treats AM and FM signals in different ways.

Figure 3b and c shows the results. For both the AM and FM conditions, the threshold time lags were worse than the original single-pulse AT condition ( $P = 0.0002$  for AM and  $P = 0.0073$  for FM by two-tailed paired  $t$  tests on log thresholds). This could be because auditory signal was changed from a pulse to a step, with the tactile signal left as a pulse. However, the thresholds remained better than in original AV condition ( $P = 0.0055$  for AM and  $P = 0.0235$

for FM by two-tailed paired  $t$  tests on log thresholds). Most importantly, the thresholds were not significantly different between the AM and FM conditions ( $P > 0.10$ ), indicating that the defining features of auditory onset had little effect on AT synchrony perception.

Since the rapid AT synchrony perception shows matching-feature invariance, in addition to attentive matching-feature selection, we consider that it is likely to be supported by a cross-modal process qualitatively similar to that for AV synchrony perception. It should be noted that matching-feature invariance was also found for VT in one of our recent experiments (Fujisaki and Nishida 2008a).

### Experiment 4

Whereas we used synchrony–asynchrony discrimination threshold as a measure of temporal resolution, the tasks more often used in previous studies are temporal order judgments (TOJ) and simultaneity judgments (SJ). In apparent disagreement with our finding of AT superiority in temporal resolution, Hirsh and Sherrick (1961) concluded that the temporal resolution estimated by TOJ is similar for any stimulus combinations, including stimulus conditions we tested above (see also Harrar and Harris 2008; Occelli et al. 2008; Zampini et al. 2005a). In the fourth experiment, we, therefore, used TOJ and SJ to measure the temporal resolutions for our stimuli, and compared the estimated resolution variables with the synchrony–asynchrony discrimination thresholds.

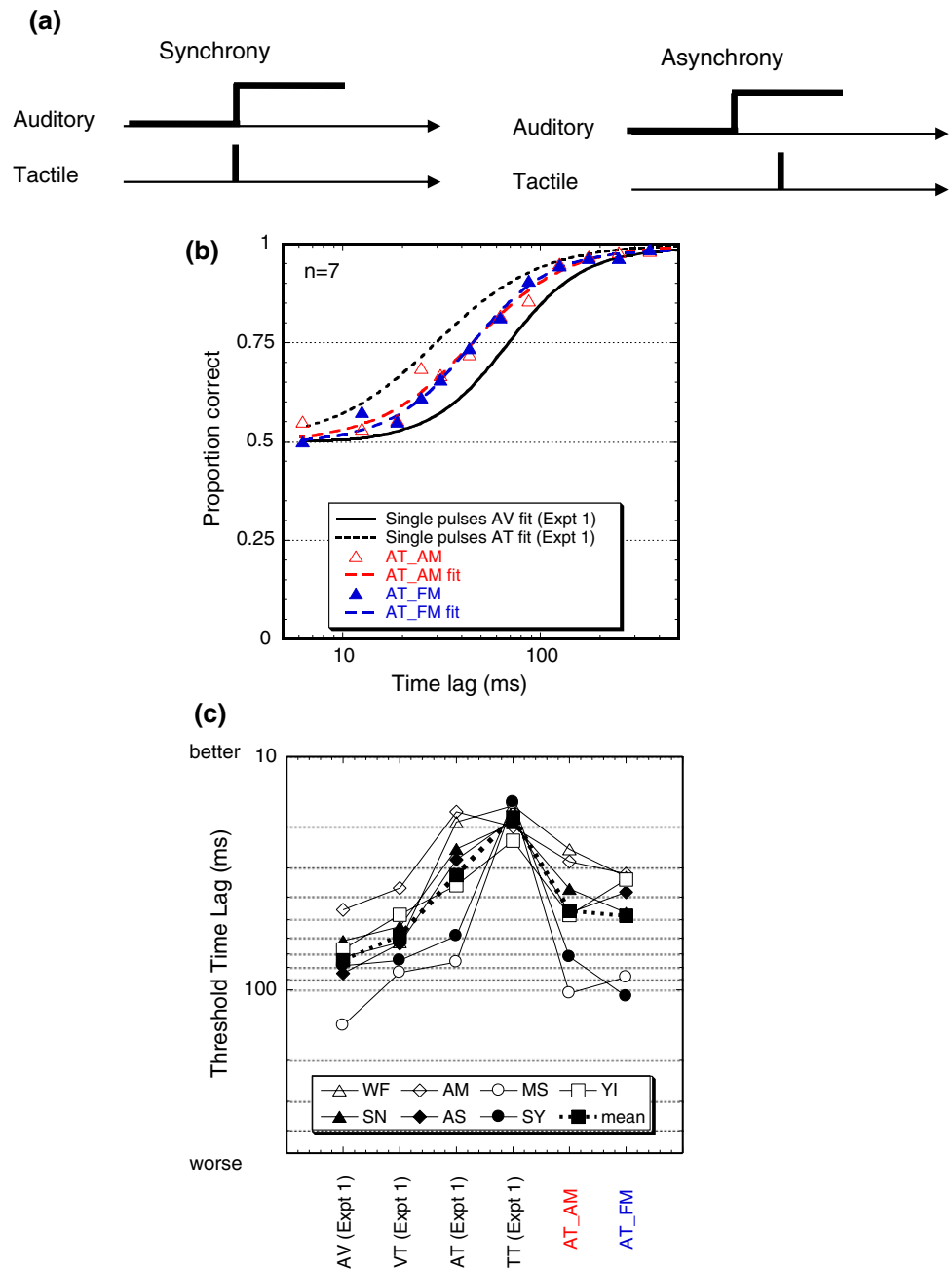
Apparatus and stimuli were the same as in experiment 1 except for the following. There were four conditions, one each for AV, VT, AT, and TT.<sup>6</sup> Two tasks (TOJ or SJ) were tested in separate blocks. The time lag varied from  $-300$  ms to  $+300$  ms (27 steps). Different time lags were shown in the same block in random order. At least 12 trials were carried out for each lag. For SJ, the participants made a binary response as to whether the two stimuli (which could be AV, VT, AT, or TT pair) were simultaneous or not. For TOJ, they made a binary response on the temporal order of the stimuli. No feedback was given in either task.

Figure 4 shows the response rate as a function of time lag. In comparison with AV, AT seems to have a narrower range of simultaneous response for SJ, and a steeper slope of psychometric function for TOJ. For comparison across stimulus conditions, we evaluated the performance of SJ in terms of the HWHH (half width at half height) of the best-fit Gaussian function (Fig. 5a), and that of TOJ in terms of the JND (Just Noticeable Difference, i.e., half of the difference between 75 and 25% points of the best-fit cumulative

<sup>6</sup> Additional data were also collected for the TT condition with hand crossed. See Discussion and footnote 10.



**Fig. 3** A test of matching-feature invariance of AT (experiment 3). **a** Auditory stimulus was a step onset of an amplitude or frequency change that was synchronous (*left*) or asynchronous (*right*) with a tactile pulse. **b** Proportion correct for synchrony–asynchrony discrimination plotted as a function of the time lag. Best-fit curves for AV and AT single-pulse conditions are also shown. **c** Threshold temporal frequency for each stimulus condition. Individual thresholds and their average are shown. The results support matching-feature invariance in rapid AT synchrony–asynchrony discrimination



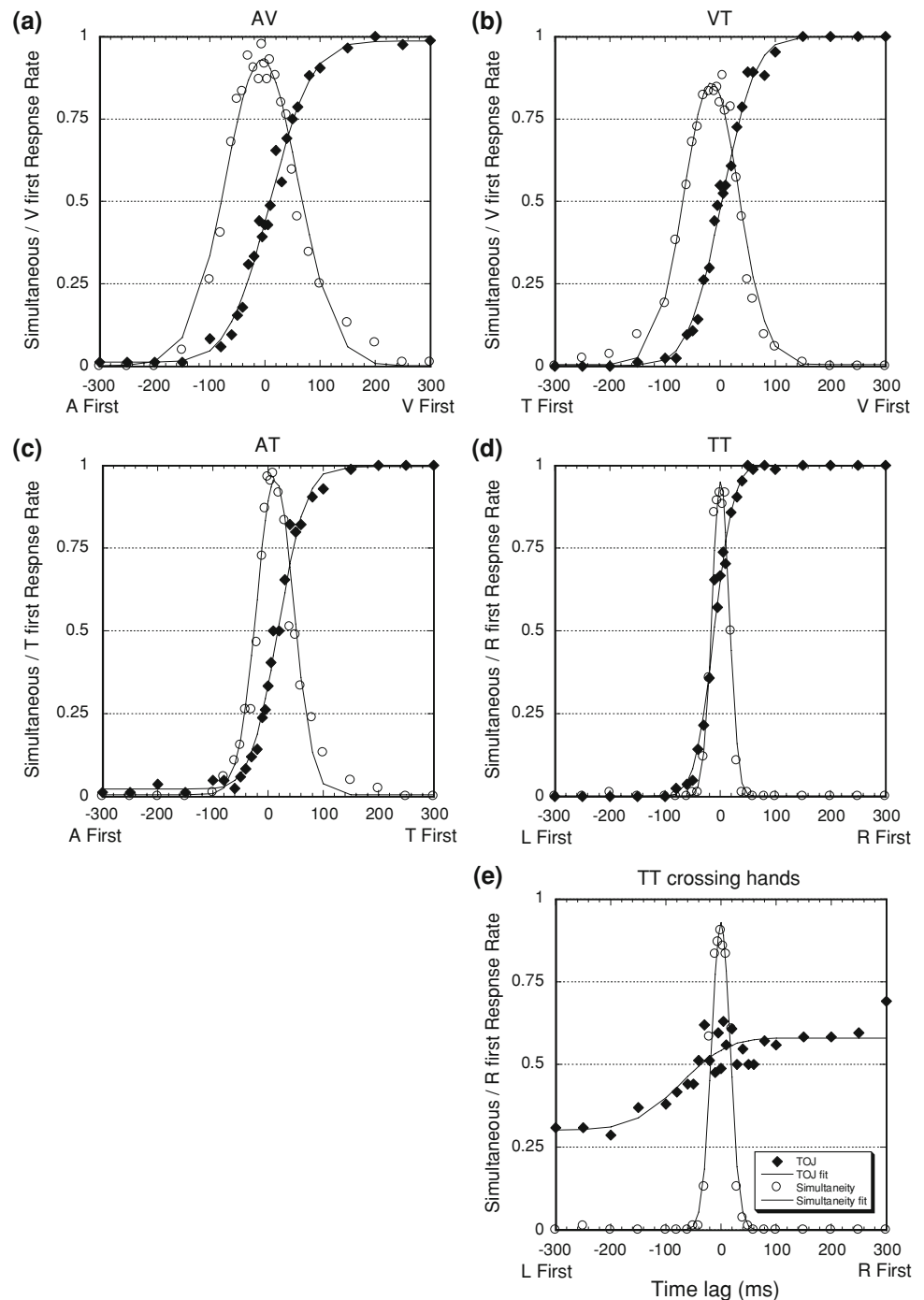
Gaussian function) (Fig. 5b). For both SJ and TOJ, the order of temporal resolution was the same: AV (worst), VT, AT, and TT (best). For SJ, two-tailed paired *t* tests on log HWHH values indicate significant differences for all comparisons ( $P < 0.017$ ). For TOJ, two-tailed paired *t* test on log JND values indicates significant differences between TT and AV ( $P = 0.0244$ ) and between TT and VT ( $P = 0.016$ ), while no significant differences among cross-modal conditions. Superiority of AT over AV was not clearly significant ( $P = 0.0577$  by a one-tailed *t* test).

In Fig. 5c, for further comparison across different tasks, the threshold time lags for synchrony–asynchrony discrimination, SJ, and TOJ are plotted together. The 75%

threshold of discrimination and the HWHH of SJ are directly comparable, since they are approximately the lag at which the asynchrony is correctly detected with 50% accuracy. To match JND for TOJ with HWHH of SJ in terms of the standard deviation of the underlying Gaussian function, we multiplied JND by 1.746.<sup>7</sup> Two-way repeated measure ANOVA indicates that the effect of task was not significant ( $F(2,12) = 1.661, P = 0.231$ ), while the

<sup>7</sup>  $1.746 = 1.178/0.674$ , where 1.178 is the *x* value at which the normal distribution function is one half of the peak (corresponding to HWHH of SJ), while 0.674 is that at which the cumulative normal distribution function is 75% (corresponding to JND of TOJ).

**Fig. 4** Simultaneity judgments (SJ) and temporal order judgments (TOJ) for AV (**a**), VT (**b**), AT (**c**) and TT (**d**), and control data for TT judgments with hands crossed (**e**) (experiment 4). Smooth curves are best-fit Gaussian function for SJ and best-fit cumulative Gaussian function for TOJ



effect of stimulus combination ( $F(3,18) = 29.649$ ,  $P < 0.001$ ) and the interaction of the two factors ( $F(6,36) = 3.615$ ,  $P = 0.007$ ) were significant. This interaction supports the notion that JND of TOJ was relatively stable against stimulus combination changes in comparison with the temporal resolutions measured by the other two synchrony tasks.

The effect of stimulus combination suggests a similarity between synchrony–asynchrony discrimination and SJ, and

a peculiarity of TOJ. In line with this idea, we found a high correlation of log threshold variations across stimulus conditions and participants between discrimination and SJ ( $r = 0.9441$ , 95% confidence interval: 0.8815–0.9741), while only moderate correlations between discrimination and TOJ ( $r = 0.6899$ , 0.4266–0.8454) or between SJ and TOJ ( $r = 0.7037$ , 0.4483–0.8529). There were significant differences between the first correlation coefficient with the latter two ( $P = 0.001051$  and 0.001464, respectively).

**Table 1** Comparison of the characteristics for three cross-modal synchrony perceptions and one within-modal asynchrony detection

	AV	VT	AT	Within-modal first-order visual motion
Temporal resolution	~4 Hz (Fujisaki and Nishida 2005)	~4 Hz	~10 Hz	20–30 Hz
Matching-feature invariance	Yes (Fujisaki and Nishida 2007)	Yes	Yes	No
Attentive matching-feature selection	Yes (Fujisaki et al. 2006; Fujisaki and Nishida 2007, 2008a, b)	Yes	Yes	No

## Discussion

The present study compared the temporal resolution of synchrony perception among three cross-modal conditions: audio–visual (AV), visual–tactile (VT), and audio–tactile (AT). The results of experiment 1 showed that the resolution was similar for AV and VT, but significantly higher for AT. The results of experiments 2 and 3 indicate that the AT synchrony perception is computationally similar to the AV synchrony perception in other aspects, and its higher temporal resolution is very unlikely to be a product of within-modal artifacts. The last experiment shows that higher resolution of AT synchrony perception is robustly found for the two types of synchrony task, but is less evident for TOJ. This may be the reason it has not been recognized in previous studies.

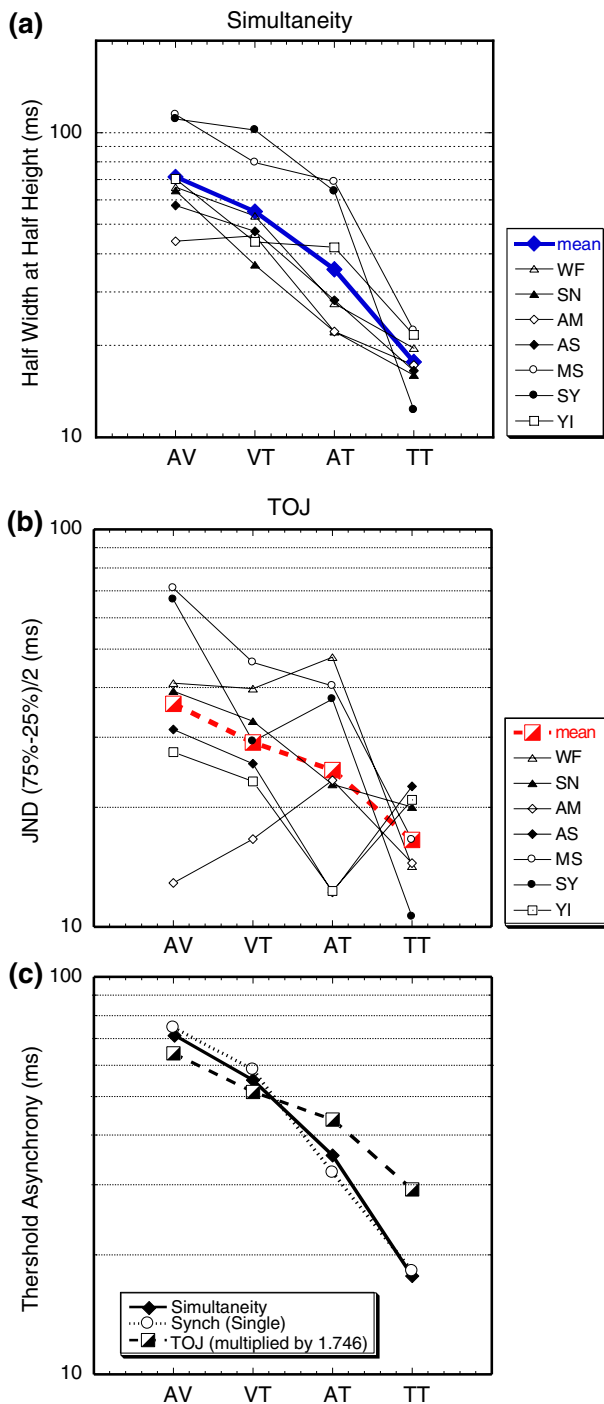
### AT superiority in temporal resolution over AV and VT

Although cross-modal temporal resolution can be affected by the type and strength of the stimuli, we would disagree with the idea that our finding of the AT resolution superiority can be ascribed to our specific choice of stimulus choice, rather than to the temporal characteristics of the underlying mechanism. We doubt that there are stimulus changes that could elevate the AV and VT resolutions to the same level as the resolution obtained with our AT stimuli. This is because we had already chosen visual stimulus parameters that would evoke rapid visual responses (high contrast, low-spatial-frequency luminance flicker) and because the obtained AV limit was indeed close to the best performance ever reported. We consider that the AT resolution superiority reflects a difference in the underlying processes. On the other hand, it is probably not difficult to lower the AT limit by manipulating audio and/or tactile parameters—for instance, by weakening stimulus amplitudes, blurring temporal waveform, or stimulating a body part other than the index finger. If there were some manipulations that could impair the AT resolution more than the AV and VT resolutions, one could apparently reduce the AT resolution superiority. We, therefore, do not expect that AT should always show higher resolution than AV and VT.

Table 1 summarizes the properties revealed by the present study. As an example of low-level within-modal temporal processing, it also shows the properties of first-order visual motion perception. The table indicates that AT processing has a high temporal resolution, but is functionally similar to other cross-modal processing (and different from within-modal low-level processing) in that it also has the properties of attentive matching-feature selection and matching-feature invariance. These findings are consistent with a hypothesis that cross-modal synchrony perception in general, including AT, is mediated by post-attentive mid-level mechanisms that match temporally sparse, salient features extracted from each modality (Fujisaki et al. 2006; Fujisaki and Nishida 2005, 2007, 2008b).

Given that the AT synchrony mechanism is functionally similar to the AV and VT synchrony mechanisms, the next question is why does AT have particularly high temporal resolution in synchrony judgments. From the viewpoint of system identification, there are two possibilities. One is that the difference reflects a variation in temporal resolution among different sensory modalities. The other is that it reflects the property of a comparator to match cross-modal signals. These possibilities are not mutually exclusive.

To put it simply, the first possibility ascribes the difference in cross-modal temporal resolution to the low resolution of the vision channel. Whenever vision is included, the synchrony–asynchrony discrimination threshold drops to around 4 Hz (repetitive pulse) or 80 ms (single pulse). This is consistent with the hypothesis that the temporal resolution of cross-modal synchrony perception is determined by the modality with the lower resolution, with the resolution being significantly lower for vision than for audition and touch. Indeed, in terms of the minimum inter-sensor time difference detectable by each modality, temporal accuracy seems to be worse for vision (e.g., Kandil and Fahle 2001) than for audition (Moore 2003) and touch (Von Békésy 1959, 1963). However, each modality has more than one temporal resolution. In general, temporal resolution decreases as the processing proceeds from the periphery to the center (Lee et al. 1993;



**Fig. 5** Task comparison. **a** Threshold time lag (half width at half height of the fit Gaussian function) of SJ. Individual thresholds and their average are shown. **b** JND (just noticeable difference) of TOJ. **c** Comparison of threshold time lag across three tasks, including the synchrony–asynchrony discrimination threshold (single-pulse condition of experiment 1). To match SJ and TOJ with regard to the standard deviation of the underlying Gaussian function, TOJ’s JND was multiplied by 1.746

Lu et al. 2001). It is highly unlikely that the AV and VT limits of the order of 4 Hz simply reflect a temporal limit of peripheral visual processing because peripheral resolution

in vision is much higher than 4 Hz. The achromatic flicker fusion limit is known to be higher than 30 Hz (Kelly 1979) and when two visual flickers are placed close together, the synchrony–asynchrony discrimination threshold is well beyond 20 Hz (Fujisaki and Nishida 2005). Therefore, even if it is true that cross-modal temporal limits reflect the signal temporal resolutions of the compared modalities, the resolutions are unlikely to be those of peripheral mechanisms, i.e., they would be those of more central processes. We consider that mid-level temporal resolution is determined by such factors as how rapidly the sensory system extracts salient features for cross-modal matching from low-level signal streams, and/or how precisely the temporal position of matching-feature is represented in the brain (Lavie and Tsal 1994). We are attempting to discover direct evidence for the expected difference in mid-level temporal resolution across modality.<sup>8</sup>

The second possibility is that AT resolution superiority may reflect rapid operation of the comparator for AT signals. Audition and touch are similar sensory modalities with regard to the point that both sense vibrations, whereas vision does not (Soto-Faraco and Deco 2009; Von Békésy 1959, 1963). Since AT signals are likely to be similar to each other in temporal profile, the comparison of the signals may be easier than it is for the other cross-modal combinations. It is also possible that the brain may regard AT as quasi-within-modal processing, and have particularly rapid comparison mechanisms for AT synchrony detection. In line with this idea, previous studies suggest a rapid perceptual interaction (Yau et al. 2009), and tight cortical connections between the two modalities—some show that tactile inputs activate the auditory cortex (Caetano and Jousmaki 2006; Levanen, Jousmaki and Hari 1998; Schroeder et al. 2001) and other reveal signs of AT integration within it (Foxe et al. 2002; Kayser et al. 2005). Even if these close affinities between the two modalities are related to the higher temporal resolution of AT, however, it is unlikely that AT synchrony detection is based on a fast comparison system specialized for a specific combination of low-level sensory signals, such as an inter-aural time difference detector, since matching-feature invariance was maintained for

<sup>8</sup> If mid-level temporal resolution is indeed found to be different across modalities, what kind of mechanism is responsible for the difference? Given that the mid-level signals are computed from low-level signals, although the temporal resolution significantly drops, the mid-level signals might still inherit a difference in temporal resolution from peripheral signals. Thus, vision is still worse than the other modalities even in terms of mid-level temporal resolution. Alternatively, the modality difference might come from a difference in the attentional load (Lavie and Tsal 1994) required to perform similar operations in different modalities.

AT synchrony perception (experiment 3)<sup>9</sup> (Fujisaki et al. 2004; Hanson et al. 2008; Harrar and Harris 2008; Navarra et al. 2007; Vroomen et al. 2004, Spence et al. 2003; Occelli et al. 2008; Zampini et al. 2005a).

#### Temporal limit of cross-modal processing

The threshold time lag to discriminate synchrony from asynchrony is lower for single-pulse trains than for repetitive-pulse trains. We first noticed this difference for AV judgments (Fujisaki and Nishida 2005), and now have found it to be a general tendency for VT, AT, and TT judgments (Fig. 1c). The present data further indicate that the threshold ratios between single and repetitive pulses were nearly constant regardless of the stimulus combination (1.54–1.92). This suggests that the single-pulse and repetitive-pulse thresholds may be limited by a common resolution-related factor for each stimulus combination. It would also mean that the single-repetitive difference is produced when the common resolution limit determines the thresholds measured for each pulse type. Intuitively, the risk of false matching could elevate the threshold lag for repetitive-pulse trains in comparison with single-pulse trains.

A simple interpretation of the lag threshold for single-pulse trains is that it indicates the width of a simultaneity window whose width reflects the temporal precision or tolerance of the temporal matching process. On the other hand, we have proposed that the temporal-frequency threshold for repetitive-pulse trains might reflect the temporal limit of the mid-level process in extracting salient features from sensory signals, beyond which temporal crowding renders correct feature matching impossible (Fujisaki and Nishida 2005). Despite the present finding of co-variation of the repetitive-pulse threshold with the single-pulse threshold, we still consider that the repetitive-pulse threshold reflects salient-feature extraction limit, in addition to the width of simultaneity window. This is because the width of simultaneity window alone cannot explain the severe deterioration of AV synchrony–asynchrony discrimination for high-density random-pulse trains

<sup>9</sup> A reviewer suggested that, because an event involving touch is likely to occur around the body, the superiority of AT over AV in temporal resolution might reflect the fact that asynchronies between A and T are usually smaller than those between A and V in a natural environment. Our sensory system might have developed the cross-modal difference in temporal resolution through adaption to the natural environment (e.g., Fujisaki et al. 2004; Hanson et al. 2008; Harrar and Harris 2008; Navarra et al. 2007; Vroomen et al. 2004). However, this idea cannot explain why VT does not have a high temporal resolution. The reviewer also suggested that the elevation of temporal acuity by spatial misalignment of stimulus pair (Spence et al. 2003) might be particularly strong for AT, but this is not supported by the finding that spatial coincidence has no effects on the temporal acuity of AT for sighted participants (Occelli et al. 2008; Zampini et al. 2005a).

(Fujisaki and Nishida 2007). We speculate that how rapidly the sensory system can extract salient features from low-level signal streams and how precisely the temporal position of a salient feature is represented in the mid-level process may be different sides of the same coin.

It is possible that the temporal-frequency limit of cross-modal synchrony–asynchrony discrimination, which we considered to reflect a mid-level processing limit, has some relevance to attention-related temporal resolutions suggested before (Aghdaee and Cavanagh 2007; Duncan et al. 1997; Holcombe and Cavanagh 2001; Lu and Sperling 2001; Shapiro et al. 1997; Verstraten et al. 2000). However, Fujisaki and Nishida (2008a) recently found that the temporal limit of cross-attribute feature binding, a representative measure of attentional resolution (Amano et al. 2007; Arnold 2005; Holcombe and Cavanagh 2001), significantly differs from synchrony–asynchrony discrimination in that it has temporal limits lower and much less variable against modality combinations (Fujisaki and Nishida 2008a). The present results only reveal a specific aspect of the temporal resolution of cross-modal processing.

#### Task difference

To estimate the sensory sensitivity of cross-modal synchrony perception, we measured the minimum time lag for discriminating a physically asynchronous stimulus from a physically synchronous one. A concern that the discrimination judgment might not necessarily reflect the subjective sense of synchrony is cleared by a finding of experiment 4 that the synchrony–asynchrony discrimination threshold agreed well with the threshold time lags for SJ (see also Fujisaki and Nishida 2007).

Compared to these two tasks, TOJ is slightly different. Although it also showed a tendency of AT superiority in temporal resolution over AV and VT, the difference across stimulus conditions was smaller. This explains, at least partially, why AT resolution superiority has not been recognized in the previous TOJ studies (Hirsh and Sherrick 1961; Zampini et al. 2005a; Occelli et al. 2008). In addition, the correlation of TOJ with the other synchrony tasks was moderate, and not as high as that between the two synchrony tasks. The present results confirm a partial dissociation of the sense of order from the sense of simultaneity (Pöppel 1988; van de Par et al. 2002; van Eijk et al. 2008; Vatakis et al. 2008).<sup>10</sup>

<sup>10</sup> A more striking difference between TOJ and SJ can be demonstrated with TT judgments with hand crossed (Fig. 4e). With our system and participants, we confirmed that TOJ was heavily disrupted by crossing the hands (Shore et al. 2002; Yamamoto and Kitazawa 2001), while SJ was unaffected (Axelrod et al. 1968). HWHH values are 18.2 and 19.8 ms in Figs. 4d and 5e, respectively.

Our data do not support the notion of strict stimulus independency of TOJ suggested by Hirsh and Sherrick (Hirsh and Sherrick 1961; Pöppel 1988). However, when we applied our analysis to the data for the AV, VT, AT, and TT conditions from Hirsh and Sherrick (1961), we found that their data also show a similar pattern. That is, JND was lower for AT (15.6 ms) and TT (18.8 ms) than for AV (25.4 ms) and VT (27.6 ms). There seems to be no real controversy between our study and Hirsh and Sherrick (1961). Both studies show that TOJ is affected by stimulus combination, but less than SJ and synchrony–asynchrony discrimination.

**Acknowledgment** Most of this work was carried out when WF was a JSPS Research Fellow at NTT Communication Science Laboratories.

## References

- Aghdaee SM, Cavanagh P (2007) Temporal limits of long-range phase discrimination across the visual field. *Vis Res* 47(16):2156–2163
- Amano K, Johnston A, Nishida S (2007) Two mechanisms underlying the effect of angle of motion direction change on colour-motion asynchrony. *Vis Res* 47(5):687–705
- Arnold DH (2005) Perceptual pairing of colour and motion. *Vis Res* 45(24):3015–3026
- Arrighi R, Alais D, Burr D (2006) Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *J Vis* 6(3):260–268
- Axelrod S, Thompson LW, Cohen LD (1968) Effects of senescence on the temporal resolution of somesthetic stimuli presented to one hand or both. *J Gerontol* 23(2):191–195
- Caetano G, Jousmaki V (2006) Evidence of vibrotactile input to human auditory cortex. *Neuroimage* 29(1):15–28
- Dixon NF, Spitz L (1980) The detection of auditory visual desynchrony. *Perception* 9(6):719–721
- Duncan J, Martens S, Ward R (1997) Restricted attentional capacity within but not between sensory modalities. *Nature* 387(6635):808–810
- Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D et al (2002) Auditory-somatosensory multisensory processing in auditory association cortex: an fmri study. *J Neurophysiol* 88(1):540–543
- Fujisaki W, Nishida S (2005) Temporal frequency characteristics of synchrony–asynchrony discrimination of audio–visual signals. *Exp Brain Res* 166(3–4):455–464
- Fujisaki W, Nishida S (2007) Feature-based processing of audio–visual synchrony perception revealed by random pulse trains. *Vis Res* 47(8):1075–1093
- Fujisaki W, Nishida S (2008a) Temporal limits of within- and cross-modal cross-attribute bindings. Paper presented at the 9th international multisensory research forum, Hamburg, Germany
- Fujisaki W, Nishida S (2008b) Top-down feature-based selection of matching features for audio–visual synchrony discrimination. *Neurosci Lett* 433(3):225–230
- Fujisaki W, Shimojo S, Kashino M, Nishida S (2004) Recalibration of audiovisual simultaneity. *Nat Neurosci* 7(7):773–778
- Fujisaki W, Koene A, Arnold D, Johnston A, Nishida S (2006) Visual search for a target changing in synchrony with an auditory signal. *Proc Biol Sci* 273(1588):865–874
- Hanson JV, Heron J, Whitaker D (2008) Recalibration of perceived time across sensory modalities. *Exp Brain Res* 185(2):347–352
- Harrar V, Harris LR (2008) The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. *Exp Brain Res* 186(4):517–524
- Hirsh IJ, Sherrick C E Jr (1961) Perceived order in different sense modalities. *J Exp Psychol* 62:423–432
- Holcombe AO, Cavanagh P (2001) Early binding of feature pairs for visual perception. *Nat Neurosci* 4(2):127–128
- Hoshiyama M, Kakigi R, Tamura Y (2004) Temporal discrimination threshold on various parts of the body. *Muscle Nerve* 29(2):243–247
- Kandil FI, Fahle M (2001) Purely temporal figure-ground segregation. *Eur J Neurosci* 13(10):2004–2008
- Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. *Neuron* 48(2):373–384
- Kelly DH (1979) Motion and vision. II. Stabilized spatio-temporal threshold surface. *J Opt Soc Am* 69(10):1340–1349
- Lavie N, Tsal Y (1994) Perceptual load as a major determinant of the locus of selection in visual attention. *Percept Psychophys* 56(2):183–197
- Lee BB, Martin PR, Valberg A, Kremers J (1993) Physiological mechanisms underlying psychophysical sensitivity to combined luminance and chromatic modulation. *J Opt Soc Am A* 10(6):1403–1412
- Levanen S, Jousmaki V, Hari R (1998) Vibration-induced auditory-cortex activation in a congenitally deaf adult. *Curr Biol* 8(15):869–872
- Lewkowicz DJ (1996) Perception of auditory–visual temporal synchrony in human infants. *J Exp Psychol Hum Percept Perform* 22(5):1094–1106
- Lu ZL, Sperling G (2001) Three-systems theory of human visual motion perception: Review and update. *J Opt Soc Am A Opt Image Sci Vis* 18(9):2331–2370
- Lu T, Liang L, Wang X (2001) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci* 4(11):1131–1138
- Martens W, Wozzyczek W (2004) Perceived synchrony in a bimodal display: optimal intermodal delay for coordinated auditory and haptic reproduction. Paper presented at the ICAD 04-tenth meeting of the international conference on auditory display, Sydney, Australia
- Moore B (2003) An introduction to the psychology of hearing, 5th edn. Academic Press, San Diego
- Navarra J, Soto-Faraco S, Spence C (2007) Adaptation to audiotactile asynchrony. *Neurosci Lett* 413(1):72–76
- Nishida S, Johnston A (2002) Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Curr Biol* 12(5):359–368
- Occelli V, Spence C, Zampini M (2008) Audiotactile temporal order judgments in sighted and blind individuals. *Neuropsychologia* 46(11):2845–2850
- Pöppel E (1988) *Mindworks: time and conscious experience*. New York, Harcourt Brace Jovanovich
- Recanzone GH (2003) Auditory influences on visual temporal rate perception. *J Neurophysiol* 89(2):1078–1093
- Schneider KA, Bavelier D (2003) Components of visual prior entry. *Cogn Psychol* 47(4):333–366
- Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC (2001) Somatosensory input to auditory association cortex in the macaque monkey. *J Neurophysiol* 85(3):1322–1327
- Shapiro KL, Arnell KM, Raymond JE (1997) The attentional blink. *Trends Cogn Sci* 1:291–296
- Shipley T (1964) Auditory flutter-driving of visual flicker. *Science* 145:1328–1330
- Shore DI, Spry E, Spence C (2002) Confusing the mind by crossing the hands. *Brain Res Cogn Brain Res* 14(1):153–163
- Soto-Faraco S, Deco G (2009) Multisensory contributions to the perception of vibrotactile events. *Behav Brain Res* 196(2):145–154

- Spence C, Baddeley R, Zampini M, James R, Shore DI (2003) Multi-sensory temporal order judgments: When two locations are better than one. *Percept Psychophys* 65(2):318–328
- Van de Burg E, Olivers CN, Bronkhorst AW, Theeuwes J (2008) Pip and pop: nonspatial auditory signals improve spatial visual search. *J Exp Psychol Hum Percept Perform* 34(5):1053–1065
- van de Par S, Kohlrausch A (2004) Visual and auditory object selection based on temporal correlations between auditory and visual cues. Paper presented at the 18th int. congress on acoustics, Kyoto, Japan, 4–9 April 2004
- van de Par S, Kohlrausch A, Juola F (2002) Some methodological aspects for measuring asynchrony detection in audio–visual stimuli. Paper presented at the Forum Acousticum Sevilla, Sevilla, Spain
- van Eijk RL, Kohlrausch A, Juola JF, van de Par S (2008) Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type. *Percept Psychophys* 70(6):955–968
- Vatakis A, Spence C (2006a) Audiovisual synchrony perception for music, speech, and object actions. *Brain Res* 1111(1):134–142
- Vatakis A, Spence C (2006b) Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neurosci Lett* 393(1):40–44
- Vatakis A, Spence C (2006c) Temporal order judgments for audiovisual targets embedded in unimodal and bimodal distractor streams. *Neurosci Lett* 408(1):5–9
- Vatakis A, Spence C (2007) How ‘special’ is the human face? Evidence from an audiovisual temporal order judgment task. *Neuroreport* 18(17):1807–1811
- Vatakis A, Spence C (2008) Investigating the effects of inversion on configural processing with an audiovisual temporal-order judgment task. *Perception* 37(1):143–160
- Vatakis A, Navarra J, Soto-Faraco S, Spence C (2008) Audiovisual temporal adaptation of speech: temporal order versus simultaneity judgments. *Exp Brain Res* 185(3):521–529
- Verstraten FA, Cavanagh P, Labianca AT (2000) Limits of attentive tracking reveal temporal properties of attention. *Vision Res* 40(26):3651–3664
- Von Békésy G (1959) Similarities between hearing and skin sensations. *Psychol Rev* 66(1):1–22
- Von Békésy G (1963) Interaction of paired sensory stimuli and conduction in peripheral nerves. *J Appl Physiol* 18:1276–1284
- Vroomen J, Keetels M, de Gelder B, Bertelson P (2004) Recalibration of temporal order perception by exposure to audio–visual asynchrony. *Brain Res Cogn Brain Res* 22(1):32–35
- Yamamoto S, Kitazawa S (2001) Reversal of subjective temporal order due to arm crossing. *Nat Neurosci* 4(7):759–765
- Yau JM, Olenczak JB, Dammann JF, Bensmaia SJ (2009) Temporal frequency channels are linked across audition and touch. *Curr Biol* 19(7):561–566
- Zampini M, Shore DI, Spence C (2003) Audiovisual temporal order judgments. *Exp Brain Res* 152(2):198–210
- Zampini M, Brown T, Shore DI, Maravita A, Roder B, Spence C (2005a) Audiotactile temporal order judgments. *Acta Psychol (Amst)* 118(3):277–291
- Zampini M, Guest S, Shore DI, Spence C (2005b) Audio–visual simultaneity judgments. *Percept Psychophys* 67(3):531–544