A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities

Waka Fujisaki^{1,*} and Shin'ya Nishida²

¹National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, 1-1-1, Higashi, Tsukuba, Ibaraki 305-8566, Japan

²NTT Communication Science Laboratories, NTT Corporation, 3-1, Morinosato-Wakamiya, Atsugi-shi, Kanagawa 243-0198, Japan

The human brain processes different aspects of the surrounding environment through multiple sensory modalities, and each modality can be subdivided into multiple attribute-specific channels. When the brain rebinds sensory content information ('what') across different channels, temporal coincidence ('when') along with spatial coincidence ('where') provides a critical clue. It however remains unknown whether neural mechanisms for binding synchronous attributes are specific to each attribute combination, or universal and central. In human psychophysical experiments, we examined how combinations of visual, auditory and tactile attributes affect the temporal frequency limit of synchrony-based binding. The results indicated that the upper limits of cross-attribute binding were lower than those of within-attribute binding, and surprisingly similar for any combination of visual, auditory and tactile attributes (2–3 Hz). They are unlikely to be the limits for judging synchrony, since the temporal limit of a cross-attribute synchrony judgement was higher and varied with the modality combination (4–9 Hz). These findings suggest that cross-attribute temporal binding is mediated by a slow central process that combines separately processed 'what' and 'when' properties of a single event. While the synchrony performance reflects temporal bottlenecks existing in 'when' processing, the binding performance reflects the central temporal limit of integrating 'when' and 'what' properties.

Keywords: vision; audition; touch; time; binding

1. INTRODUCTION

The human sensory system perceives different aspects of the surrounding environment through multiple channels. Multiple channels imply different sensory modalities (e.g. vision, audition and touch), and attribute-specific processing mechanisms within each modality (e.g. colour and shape processing pathways in vision) (Zeki 1993). For perception of a coherent world, the brain rebinds input signals through different sensory channels into representations of original multi-attribute objects and events.

Spatial coincidence and temporal coincidence are two representative cues for binding different attributes.¹ In past studies of cross-attribute binding, particularly those on vision, greater emphasis has been placed on spatial coincidence. In fact, the most influential theory of visual cross-attribute binding, the feature integration theory (Treisman & Gelade 1980; Treisman 1996, 1999) postulates that cross-attribute binding is accomplished by a serial scan of a map of spatial locations by a 'window of attention'. The theory's main hypothesis is that binding 'what' (an attribute content) to 'where' (a spatial location) through a serial scan of locations also binds the individual 'what'.

Just as input signals detected at the same location are likely to belong to the same *object*, input signals detected

synchronously are likely to belong to the same *event*. Although temporal synchrony is considered to be a relatively minor cue for within-attribute binding of spatially separate visual signals (Leonards *et al.* 1996; Alais *et al.* 1998; Lee & Blake 1999; Motoyoshi & Nishida 2001), it is a critical clue for binding signals across different attributes, in particular across different modalities (Munhall *et al.* 1996; Sekuler *et al.* 1997; Holcombe & Cavanagh 2001; Spence 2007). A good example is an audience binding audio–visual signals through temporal synchrony even when there is a mismatch in spatial location (ventriloquist effect) (e.g. Warren *et al.* 1981; Alais & Burr 2004). Compared with our understanding of binding from spatial coincidences, however, we know very little about the mechanisms underlying cross-attribute binding from temporal coincidences.

A fundamental question about cross-attribute/modality processing is whether it is mediated by peripheral neural mechanisms specific to each attribute combination, or by a central universal process common to all attribute combinations. In line with the notion of peripheral-specific processing, brain activities responsible for cross-modal/attribute interactions have recently been observed at very short latency and/or in peripheral sensory areas (Lebib *et al.* 2003; Schroeder & Foxe 2005; van Wassenhove *et al.* 2005; Mishra *et al.* 2007; Seymour *et al.* 2009). These findings suggest that cross-attribute temporal binding might also be processed separately for each combination of attributes/modalities in distributed networks involving attribute-specific cortical areas. On the

* Author for correspondence (w-fujisaki@aist.go.jp).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2010.0243> or via <http://rspb.royalsocietypublishing.org>.

other hand, just as the feature integration theory assumes for cross-attribute binding from spatial coincidences (Treisman & Gelade 1980; Treisman 1996, 1999), a centralized attention-driven process may play a critical role in cross-attribute binding from temporal coincidences.

A promising behavioural approach to the solution of this problem is to measure the processing speed of cross-attribute binding. If the underlying mechanisms are peripheral and specific, the binding speed is expected to be relatively high and/or variable among different attribute combinations. On the other hand, if the underlying mechanism is central and universal, the binding speed is expected to be relatively low and invariant against changes in attribute combination.

The processing speed of temporal binding can be psychophysically estimated from the upper temporal limit in judging the temporal relationship between two stimulus sequences. The highest temporal frequency for discriminating in-phase from reversed-phase stimulus pairs is considerably high for within-modal comparisons, in particular when peripheral phase sensors, such as visual motion detectors (Adelson & Bergen 1985), are likely to be involved. The limit reaches approximately 30 Hz for vision and even higher frequencies for audition (Victor & Conte 2002; Fujisaki & Nishida 2005). In comparison with these, the reported temporal limits for cross-attribute phase judgements are much lower (less than approx. 10 Hz) (Holcombe & Cavanagh 2001; Arnold 2005; Fujisaki & Nishida 2005, 2007, 2009; Bartels & Zeki 2006; Amano *et al.* 2007; Holcombe & Judson 2007). One could interpret these low temporal limits as evidence for the involvement of centralized attention-driven processes (He *et al.* 1997; Holcombe & Cavanagh 2001; Holcombe 2009).

However, from the viewpoint of peripheral-specific processing, the low temporal limits could be caused by sparse cortical connections between the attribute-specific areas, or by differences in neural processing latency between different attributes (Bartels & Zeki 2006; Seymour *et al.* 2009). Apparently in agreement with the peripheral-specific hypothesis, there is a significant variation, ranging between 2 and 10 Hz, in the reported temporal limits of cross-attribute temporal judgements.

A critical problem with previous data is that the temporal limits of cross-attribute binding were not collected using the same tasks but using two different temporal phase-discrimination tasks, i.e. a *binding* task (Moutoussis & Zeki 1997; Holcombe & Cavanagh 2001; Arnold 2005; Bartels & Zeki 2006; Amano *et al.* 2007; Holcombe & Judson 2007) and a *synchrony* task (Van de Par & Kohlrausch 2000; Fujisaki *et al.* 2006; Fujisaki & Nishida 2009). In both tasks, participants were asked to discriminate whether two regular repetitive sequences, presented at different locations or in different modalities, were in-phase or reversed-phase (180° anti-phase), but there were a slight change in the stimulus structure and a significant difference in the required task. Since different attribute combinations have been used for the two tasks, it remains obscure whether the temporal limit variation depends on the attribute combination or on other factors, including the task type.

The present study therefore compared the temporal limits of the binding and synchrony tasks for a wide

range of attribute and modality combinations. Figures 1 and 2 illustrate examples of the two tasks used in our experiment. In the binding task, each sequence was a repetitive alternation of two attribute values, such as red and green, and high- and low-pitch sound. The alternation was always synchronized between the two sequences, but the feature pairing was changed between the two phase conditions. The participants had to judge which features were presented simultaneously, e.g. whether the pitch was high or low when red was presented.² In the synchrony task (synchrony–asynchrony discrimination task), each stimulus sequence contained brief pulses at a given repetition rate. Participants had to judge whether the pulses of the two sequences, for instance visual flicker and auditory flutter, were synchronous or asynchronous. In addition to measuring the temporal limits of the two tasks for cross-attribute and cross-modality conditions, we also examined a few within-attribute binding and synchrony judgements in order to test the idea that the temporal judgements between the same visual attributes, but across long spatial separations, are mediated by an attention-driven process (Battelli *et al.* 2001; Aghdaee & Cavanagh 2007; Holcombe 2009).

Our main finding is that the temporal limit was very low and surprisingly similar (2–3 Hz) when and only when the task was binding and the comparison was across different attributes of the same or different modalities. This suggests that cross-attribute temporal binding is not mediated by a peripheral mechanism specific to each attribute combination but by a common slow central process that presumably binds ‘what’ (an attribute content) and ‘when’ (a temporal location) to find the binding of synchronous ‘what’.

2. MATERIAL AND METHODS

(a) General

See supplementary methods for further details about issues described in this section.

Participants were the two authors, and seven paid volunteers.

Visual stimuli were presented on a CRT monitor. One or two disc-shaped stimuli, each subtending 3.09° in diameter, were presented on a uniform yellow background. A white fixation bull’s eye was presented 1.6° above or below the centre of the monitor. The position of the fixation point was switched every trial to reduce retinotopic adaptation effects. Auditory stimuli were presented via headphones. Tactile stimuli were presented via two vibration generators.

One stimulus sequence lasted 6 s with 2-s cosine ramps both at the onset and offset of the stimulus. The relative phase of the two sequences 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.

(b) Stimuli for the binding task

For the binding task, we used five alternation sequences (figure 1b). V_{col} , a red and green colour change in a circular disc; V_{lum} , luminance increment and decrement in a circular disc; V_{ori} , 45° clockwise and anticlockwise tilts from the vertical of a grating shown in a circular disc; A , high-pitch (622.26 Hz) and low-pitch (261.63 Hz) of a complex (triangular wave) tone; T , vibrations (40 Hz) given to the right

Figure 1. Stimuli used for the temporal binding task. (a) Spatial configuration of a visual stimulus (colour (V_{col}) orientation (V_{ori}) combination). (b) Five attributes used for the binding experiment. V_{col} , visual colour change; V_{lum} , visual luminance change; V_{ori} , visual orientation change; A , auditory pitch change; T , tactile-stimulated hand (finger) change. (c) Thirteen attribute combinations tested. Three within-attribute, three cross-attribute and seven cross-modality combinations. A quartet of icons represents the 'in-phase' pairing of the attribute combination. (d) Schematic illustrations of stimulus time courses for the condition where the attribute combination comprised colour (V_{col}) and auditory pitch (A), and the temporal frequency was 1 Hz. Each sequence was modulated by a square wave. This resulted in two attribute values (i.e. red, green) alternated in turns at a given rate. Each sequence lasted 6 s with 2-s cosine ramps both at the onset and at the offset of the stimulus. The phase lag between two attributes started with a random value and gradually shifted to the intended phase over the initial 2 s. These manipulations were done to prevent the participants from making their judgement based on the onset and offset of the sequences. Top: 'in-phase' condition, where red appears when auditory pitch is low (shown by lime-green notes). Bottom: 'reversed-phase' condition, where red appears when pitch is high (shown by magenta notes).

and left index fingers. Combinations of these five resulted in 10 pairs: seven cross-modality pairs, and three visual cross-attribute pairs. In addition, we tested three visual within-attribute pairs (figure 1c). For within-vision comparisons, two visual sequences were presented in two discs on both sides of fixation with an inter-disc separation of 15.4° (figure 1a). It has been suggested that this separation is long enough to tap a high-level visual mechanism (Aghdaee & Cavanagh 2007).

We presented a pair of stimulus sequences while modulating each sequence by a square wave. This means two attribute values alternately appeared at a given rate. One wavelength

consisted of appearances of two attributes with two transitions in the attribute value. The time course of an example pair of stimulus sequences (colour–pitch combination) is shown in figure 1d. The temporal phase relationship between the two sequences was either in-phase or 180° reversed-phase. In both cases, the two sequences always changed synchronously, but the value combination was reversed between the two phase conditions. We arbitrarily determined one of the two value combinations as in-phase, regarding the other as reversed-phase. These two stimulus conditions were maximally separated in phase from each other, and neither had any physical-binding ambiguity.

Figure 2. Stimuli used for the synchrony task. (a) Four attributes used for the synchrony experiment. V_{col} , visual colour change; V_{lum} , visual luminance change; A , auditory pitch change; T , tactile-stimulated hand (finger) change. (b) Eight attribute combinations. Two within-attribute, one cross-attribute and five cross-modality combinations. (c) Schematic illustrations of stimulus time courses for the condition where the attribute combination comprised colour (V_{col}) and auditory pitch (A), and the temporal frequency was 1 Hz. Each sequence was a pulse train. Each pulse train kept a single attribute value, rather than alternating between the two values. Top: ‘in-phase’ condition, where red appears in synchrony with low pitch. Bottom: ‘reversed-phase’ condition, where red and low pitch appear asynchronously.

(c) Stimuli for the synchrony task

For the synchrony task, we used four pulse sequences defined by two visual attributes, one audio attribute and one tactile attribute (figure 2a). V_{col} , a colour change of a circular disc from yellow to equiluminant red; V_{lum} , luminance increment of a circular disc; A , an amplitude-modulated complex tone pip; T , a vibration given to the right index finger. Since an orientation change is inevitably accompanied by local luminance changes, we did not use orientation for the synchrony task. Combinations of the four attributes resulted in six pairs: five cross-modality pairs, and one visual cross-attribute pair. In addition, we used two visual within-attribute pairs (figure 2b).

We used different waveforms for binding and synchrony tasks to obtain the best temporal frequency limit for each task (see electronic supplementary material for a related control experiment). The temporal waveform for the synchrony task was a repetitive pulse train (figure 2c). One wavelength included only one pulse. Each pulse kept a single attribute value, rather than alternating between the two values. A pulse lasted three monitor frames for visual stimuli, and 18.75 ms for auditory and tactile stimuli. During inter-pulse intervals, only the background stimuli were presented. The pulse timing was in-phase (synchronous) or reversed-phase (asynchronous) between the two sequences.

(d) Procedures

We measured the temporal limits of the binding and synchrony tasks for various combinations of attributes. The

participant was instructed to judge the combination of attribute values that were presented at the same time (binding task), or to judge synchrony between the two sequences (synchrony task). Although the two tasks were subjectively different, in terms of physical stimulus, both tasks were discriminating of whether two repetitive sequences were in-phase or reversed-phase.

In each trial, a pair of stimulus sequences was presented. The participant had to make a two-alternative forced response about the relative phase. Feedback was given after each response by the colour of the fixation marker. For each of the 21 stimulus-task conditions (13 stimulus combinations for binding and eight for synchrony), the proportion correct was estimated at least at six temporal frequencies, which were chosen from the frequencies ranging from 1 to 16 Hz in a half-octave step. The proportion correct at a given frequency was computed from 20 to 40 judgements. The obtained psychometric function was fit by a logistic function by the maximum likelihood method for estimation of the 75 per cent correct point (see also electronic supplementary material).

3. RESULTS

For both binding and synchrony tasks, we measured the performance of binary phase discrimination as a function of temporal frequency and computed the 75 per cent correct point to estimate the threshold frequency. An example of the psychometric functions is shown in

Figure 3. The temporal limits of binding and synchrony tasks. (a, b) Examples of the obtained psychometric functions. The proportion correct (average over nine participants) of (a) binding and (b) synchrony judgement is plotted as a function of the temporal frequency (Hertz). The threshold temporal frequency was estimated from the 75 per cent point. Black circles and solid fitting curve, $V_{\text{col}}-A$ condition; white squares and dotted fitting curve, $A-T$ condition. (c) The threshold temporal frequency was estimated for each participant and averaged over the nine participants. The horizontal axis represents the combination of attributes. Blue indicates within-attribute comparisons, green cross-attribute ones and pink cross-modality ones. White squares indicate the results obtained in the binding experiment and black circles indicate the results obtained in the synchrony experiment. Error bar: ± 1 s.e.

figure 3a,b, and the estimated temporal limits are shown in figure 3c. The data in figure 3c are geometric averages over nine participants. Individual data were similar to the group average (electronic supplementary material, figure S1). Electronic supplementary material, tables S1–S3, summarizes the results of paired t -tests.

First, consider the results of cross-attribute conditions (right 10 conditions in figure 3c). Regardless of visual cross-attribute combinations ($V_{\text{col}}-V_{\text{lum}}$, $V_{\text{col}}-V_{\text{ori}}$, $V_{\text{lum}}-V_{\text{ori}}$) or cross-modal conditions ($V_{\text{col}}-A$, $V_{\text{col}}-T$, $V_{\text{lum}}-A$, $V_{\text{lum}}-T$, $V_{\text{ori}}-A$, $V_{\text{ori}}-T$, $A-T$), there was clear and systematic dissociation between the binding and synchrony tasks. When compared between the two tasks, the temporal limit was lower for the binding task than for the synchrony task for all the combinations where comparison was possible. When compared between the attribute combinations within each task, while the temporal limit of the synchrony task showed a large variation (specifically, the limit was higher for $V_{\text{col}}-V_{\text{lum}}$ and $A-T$ than for $V_{\text{col}}-A$, $V_{\text{col}}-T$, $V_{\text{lum}}-A$ and

$V_{\text{lum}}-T$), the temporal limit of the binding task was always 2–3 Hz and showed little variation among stimulus conditions. The present data are consistent with the previous data obtained under conditions similar to the current $V_{\text{lum}}-A$, $V_{\text{lum}}-T$ and $A-T$ synchrony conditions (Fujisaki & Nishida 2005, 2009) and the current $V_{\text{col}}-V_{\text{ori}}$ binding condition (Holcombe & Cavanagh 2001). The temporal limit of cross-attribute binding for another attribute combination (colour and motion) is also around 3 Hz (Arnold 2005; Bartels & Zeki 2006; Amano *et al.* 2007). The previous data however are too fragmentary, and the present data are the first to show that the temporal limit of cross-attribute binding is surprisingly similar for a wide range of attribute combinations, while that the synchrony limit is higher and more variable.

For visual within-attribute conditions ($V_{\text{lum}}-V_{\text{lum}}$, $V_{\text{col}}-V_{\text{col}}$, $V_{\text{ori}}-V_{\text{ori}}$), the temporal limit was 5–9 Hz, and there was little difference between the binding and synchrony tasks. These results are consistent with the previous data obtained under conditions similar to the

current $V_{\text{lum}} - V_{\text{lum}}$ synchrony condition (Battelli *et al.* 2001; Victor & Conte 2002; Aghdaee & Cavanagh 2007), and the current $V_{\text{ori}} - V_{\text{ori}}$ synchrony condition (Motoyoshi 2004). The present study further showed that within-attribute conditions were critically different from cross-attribute conditions in that the temporal limits of the within-attribute conditions were nearly the same for the binding and synchrony tasks, and that the temporal limits of the within-attribute binding were higher than those of cross-attribute binding.

To check the robustness of our main finding of constant temporal limit (2–3 Hz) for cross-attribute binding tasks, we carried out several additional experiments (see electronic supplementary materials for details). First, stimulus intensity had little effects on the cross-attribute binding limit (electronic supplementary material, figure S2). Second, stimulus waveform (pulse or square wave) had little effects on the cross-attribute binding limit (electronic supplementary material, figure S3). Third, the low temporal limit of cross-attribute binding could not be ascribed to potential shifts of apparent timing between different stimulus attributes (electronic supplementary material, figure S4). Fourth, matching the apparent spatial locations of the audio–visual signals had little effects on the temporal binding limit. Fifth, using different auditory attributes other than pitch (amplitude modulation, left/right ear alternation) had little effects on the temporal binding limit (electronic supplementary material, figure S5). Sixth, the temporal binding limit of cross-attribute binding remained to be 2–3 Hz even when it was measured under more natural binding situations with regard to spatial alignment (Spence *et al.* 2003) and attribute combination (e.g. Maeda *et al.* 2004; electronic supplementary material, figure S6). In addition, we obtained preliminary data suggesting that binding within the same attribute, but across different attribute values is considerably slow.

In conclusion, as far as we have tested, temporal limit of 2–3 Hz was a very robust property of cross-attribute binding.

4. DISCUSSION

The present data clarify that the limit of the binding task is always 2–3 Hz. This was the case not only for audio–visual and visuo–tactile combinations for which the synchrony limit was 4–5 Hz, but also for luminance–colour and audio–tactile combinations for which the synchrony limit was 7–9 Hz. A common binding limit suggests that the temporal bottleneck is a universal process, and not a stimulus-specific one. The present findings therefore provide clear psychophysical evidence that cross-attribute temporal binding of any combination, regardless of whether the combination is within-modal or cross-modal, is mediated by a universal slow central process.

In contrast, the temporal limit of cross-modal synchrony judgement is variable against a change in modality combination. This suggests that the mechanisms underlying cross-attribute synchrony judgements reside in more peripheral stages than those underlying cross-attribute temporal binding judgements. Nevertheless, the following three properties found in the previous studies led us to believe that cross-modal synchrony

perception is not a peripheral phenomenon either. First, the temporal limits of cross-modal synchrony perception are still much lower than the peripheral sensory limits, and second, it is affected little by the attribute combination as long as the modality combination does not change (Fujisaki & Nishida 2005, 2007; Fujisaki & Nishida 2009). For instance, the limit of audio–visual synchrony–asynchrony discrimination is 4–5 Hz regardless of which visual attribute is combined with which auditory attribute, while the visual and auditory limits are in the order of several tens of Hertz (Fujisaki & Nishida 2005, 2007; see also figure 3). Third, attention plays a critical role in the selection of matching features for cross-modal synchrony judgements (Fujisaki *et al.* 2006; Fujisaki & Nishida 2008). For example, in a visual search task, the search of a sound-synchronized visual target among uncorrelated visual distracters shows a steep increase in the search time with the set size. In contrast, in a selective attention task, interference by surrounding distractors on the audio–visual synchrony task can be completely excluded by pre-cuing of the target location (Fujisaki *et al.* 2006). It is also possible to select audio–visual matching pair by feature-based attention (Fujisaki & Nishida 2007). These findings suggest that audio–visual synchrony detection is preceded by nearly perfect attentive selection of the spatial position of a visual event. These three properties indicate that cross-modal synchrony perception is also mediated by central processing. Taken together synchrony and binding judgements indicate two central temporal bottlenecks of cross-attribute temporal processing, one in synchrony detection between each combination of modality, and the other in binding any combination of modality and attribute.

Recent studies indicate separate processing of ‘when’ and ‘what’ in the human brain (Nishida & Johnston 2002; Battelli *et al.* 2007, 2008). Both the synchrony task and binding task are judgements of relative temporal phase. However, the synchrony task primarily requires ‘when’ processing only, while the binding task requires ‘when’ processing and ‘what’ processing as well. As discussed below, the present findings may reflect this difference.

The cross-attribute synchrony task requires participants to match the timings of transient events between different attributes (e.g. between a visual pulse and an auditory pulse). We are proposing that those cross-attribute matching features should be temporally salient in such a way that they are distinctive from pre- and postsignals of the sequence (Fujisaki & Nishida 2005, 2007). Since they should be temporally sparse, cross-modal synchrony perception collapses for rapidly changing stimuli (Shipley 1964; Fujisaki & Nishida 2005, 2007; Vatakis *et al.* 2007). In a cluttered environment, selective attention plays a critical role in selecting matching signals for cross-modal synchrony judgements (Fujisaki *et al.* 2006; Fujisaki & Nishida 2008). In contrast, when an event is only a salient event within that processing channel, the event timing can be detected nearly automatically through bottom-up processing (Soto-Faraco *et al.* 2004, 2005). To compare the timing signals, participants do not have to access the information about stimulus contents (‘what’) that cause the timing signals. In fact, the temporal resolution is reduced

Figure 4. A hypothetical model for cross-modal temporal judgements. This model hypothesizes that ‘when’ and ‘what’ of a multi-attribute physical event are once processed separately and integrated at later stages.

significantly when participants have to access the stimulus contents to select matching features in cross-modal synchrony judgement (Fujisaki & Nishida 2008). Therefore, we conjecture that the temporal resolution of the synchrony judgement is determined by ‘when’ processing factors, such as the temporal accuracy to extract salient changes in each sensory channel and the temporal accuracy to compare the timing of salient changes.

The relationship between the timing signals of salient events (‘when’) is also important for the binding task. The occurrence of the illusory conjunction of colour and motion during perception of illusory asynchrony of the two attributes (i.e. apparent lead of colour oscillation relative to direction oscillation that is robustly observed at 1–2 Hz oscillations) (Moutoussis & Zeki 1997; Nishida & Johnston 2002) indicates that the sensory system uses the same timing signals for the cross-attribute binding and synchrony tasks. However, the binding task also requires participants to access stimulus contents (‘what’), i.e. to judge which combination of attribute values is presented at the same time. The present results indicate that this computation is common to any different attribute combinations, and is slow. This presumably implies that the specific combination of two attributes is not directly encoded, but is recognized through sequential identification of the attribute values that are linked to the same time point. Just as the feature integration theory (Treisman & Gelade 1980; Treisman 1996, 1999) postulates that cross-attribute binding from spatial synchrony is mediated by a slow central process that binds ‘where’ and ‘what’, we consider that cross-attribute binding from temporal synchrony may be mediated by a slow central process that binds ‘when’ and ‘what’.

Figure 4 shows a possible functional structure of cross-attribute temporal processing for synchrony and

synchrony-based binding consistent with the present and previous findings (see also Amano *et al.* 2007; Fujisaki & Nishida 2005, 2008; Nishida & Johnson 2002, 2010). ‘When’ and ‘what’ of a multi-attribute event are processed separately. A synchrony judgement is based on the timings of salient features extracted in ‘when’ processing, while a binding judgement waits for completion of ‘what’ processing and binds outputs from ‘when’ processing with those from ‘what’ processing. The scope of this model is limited to temporal judgements about the relationship between nearly simultaneous signals, not temporal perception in general including duration perception (e.g. van Wassenhove *et al.* 2008).

The binding of temporal limit might also depend on the speed of switching attention from one attribute to the other (Reeves & Sperling 1986), and/or the speed of binding attributes (Treisman & Gelade 1980), but it remains to be poorly understood whether these processing speeds are independent of attribute and modality. Why the cross-attribute temporal limit is always 2–3 Hz is a fundamental question, and the solution of this question will provide insights into the basic ‘clock’ of high-level sensory recognition.

A well-known exception of slow cross-attribute binding is the colour–orientation binding at the same location—the temporal limit of binding colour and orientation is much higher than 2–3 Hz when they are presented at the same location (Holcombe & Cavanagh 2001; Bodelon *et al.* 2007). The rapid colour–orientation binding is unlikely to reflect a general facilitation effect of spatial coincidence on cross-attribute binding, since the temporal limit of colour–motion binding is low even when they are presented at the same location (Arnold 2005; Bartels & Zeki 2006). The point is whether there is an

early neural mechanism that directly encodes the relationship between a specific attribute pair. In other words, without such a mechanism, a slow general central process mediates cross-attribute temporal binding. In addition, although dissociation between synchrony and binding judgement has been reported previously (Clifford *et al.* 2003), it is essentially different from the current dissociation, since that study examined pairs of superimposed colour and orientation with which rapid binding was possible, and the dissociation was concerned not with the upper temporal limit, but with apparent phase delay (see Nishida & Johnston (2010) for our view on possible mechanisms of colour–orientation asynchrony).

For temporal binding within the same attribute (and within the same attribute value, see the last additional experiment in electronic supplementary material), the temporal limit is relatively high and similar between the synchrony and the binding tasks. This suggests that both tasks are mediated by a mechanism that directly encodes the temporal relationship between spatially separated visual flickers. The mechanism responsible might be related to motion processing that encode spatio-temporal structure of the input pattern. Our finding does not reject the hypothesis that an attention-driven process determines the upper temporal limit of long-range within-attribute comparisons (Battelli *et al.* 2001; Aghdaee & Cavanagh 2007; Holcombe 2009), but that process is unlikely to be identical to the central cross-attribute binding mechanism we specified.

Some within-attribute bindings are rapid, even though they seem to include rather complex computations. This is presumably because they are also supported by low-level specialized binding mechanisms. Examples are global form perception from integration of local orientation signals (Clifford *et al.* 2004), and texture edge perception from local orientation difference (Motoyoshi & Nishida 2002). See also a recent relevant review on rapid and slow processing (Holcombe 2009).

Finally, a recent study suggests that a combination of different visual attributes (colour and motion) may be represented in a wide range of cortical areas including V1 (Seymour *et al.* 2009), whereas the present findings indicate that the main neural circuit for cross-attribute temporal binding should be localized more centrally. Potentially relevant loci include the parietal cortex, which has connections with multiple sensory areas (Lewis & Van Essen 2000) and plays a critical role in spatial cross-attribute binding (Friedman-Hill *et al.* 1995; Robertson 2003). Particularly of interest is the right inferior parietal lobe (IPL), which has been suggested to be a cortical area responsible for ‘when’ processing (Battelli *et al.* 2001, 2007, 2008). It has been reported, for instance, that damage to this area impairs synchrony judgement between spatially separated flickers (similar to the current $V_{lum} - V_{lum}$ condition; Battelli *et al.* 2001). The present findings, however, suggest that the mechanisms underlying this task are likely to be different from those mediating cross-attribute synchrony and binding tasks. In the future, it would be interesting to study how the right IPL activity correlates with the performance of synchrony and binding tasks for the various attribute combinations we have tested.

Despite rapid technological development in neuroscience, our knowledge about high-level sensory processing is still very limited because of the lack of a clear view of the global functional architecture. The present study reveals such a clear view, and suggests conceptual and methodological guidelines for studying the underlying cortical mechanisms by psychophysics, neuropsychology, physiology and neuro-imaging.

W.F. was partly supported by a Grant-in-Aid for Young Scientists (B) No. 21730606.

ENDNOTES

¹In this paper, we use the term ‘attribute’ to mean the basic sensory dimension (e.g. colour), and ‘attribute value’ to mean a specific instance of a given attribute (e.g. red). The term ‘feature’ is used mainly for referring to salient stimulus temporal features (e.g. red onset) used in cross-attribute temporal matching (see figure 4).

²When observers perceptually bind synchronous stimuli into a single cross-attribute event, they should be able to judge the combination of attribute values. However, observers might be able to perform the binding task without having the feeling of perceptual binding that the two signals originate from a common event. The ability to perform this task must be a necessary condition for binding perception, but may not be a sufficient condition under some situations.

REFERENCES

- Adelson, E. H. & Bergen, J. R. 1985 Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* **2**, 284–299. (doi:10.1364/JOSAA.2.000284)
- Aghdaee, S. M. & Cavanagh, P. 2007 Temporal limits of long-range phase discrimination across the visual field. *Vision Res.* **47**, 2156–2163. (doi:10.1016/j.visres.2007.04.016)
- Alais, D. & Burr, D. 2004 The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* **14**, 257–262.
- Alais, D., Blake, R. & Lee, S. H. 1998 Visual features that vary together over time group together over space. *Nat. Neurosci.* **1**, 160–164. (doi:10.1038/1151)
- Amano, K., Johnston, A. & Nishida, S. 2007 Two mechanisms underlying the effect of angle of motion direction change on colour–motion asynchrony. *Vision Res.* **47**, 687–705. (doi:10.1016/j.visres.2006.11.018)
- Arnold, D. H. 2005 Perceptual pairing of colour and motion. *Vision Res.* **45**, 3015–3026. (doi:10.1016/j.visres.2005.06.031)
- Bartels, A. & Zeki, S. 2006 The temporal order of binding visual attributes. *Vision Res.* **46**, 2280–2286. (doi:10.1016/j.visres.2005.11.017)
- Battelli, L., Cavanagh, P., Intriligator, J., Tramo, M. J., Henaff, M. A., Michel, F. & Barton, J. J. 2001 Unilateral right parietal damage leads to bilateral deficit for high-level motion. *Neuron* **32**, 985–995. (doi:10.1016/S0896-6273(01)00536-0)
- Battelli, L., Pascual-Leone, A. & Cavanagh, P. 2007 The ‘when’ pathway of the right parietal lobe. *Trends Cogn. Sci.* **11**, 204–210. (doi:10.1016/j.tics.2007.03.001)
- Battelli, L., Walsh, V., Pascual-Leone, A. & Cavanagh, P. 2008 The ‘when’ parietal pathway explored by lesion studies. *Curr. Opin. Neurobiol.* **18**, 120–126. (doi:10.1016/j.conb.2008.08.004)
- Bodelon, C., Fallah, M. & Reynolds, J. H. 2007 Temporal resolution for the perception of features and conjunctions. *J. Neurosci.* **27**, 725–730. (doi:10.1523/JNEUROSCI.3860-06.2007)

- Clifford, C. W., Arnold, D. H. & Pearson, J. 2003 A paradox of temporal perception revealed by a stimulus oscillating in colour and orientation. *Vision Res.* **43**, 2245–2253. (doi:10.1016/S0042-6989(03)00120-2)
- Clifford, C. W., Holcombe, A. O. & Pearson, J. 2004 Rapid global form binding with loss of associated colors. *J. Vision* **4**, 1090–1101. (doi:10.1167/4.12.8)
- Friedman-Hill, S. R., Robertson, L. C. & Treisman, A. 1995 Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science* **269**, 853–855. (doi:10.1126/science.7638604)
- Fujisaki, W. & Nishida, S. 2005 Temporal frequency characteristics of synchrony–asynchrony discrimination of audio–visual signals. *Exp. Brain Res.* **166**, 455–464. (doi:10.1007/s00221-005-2385-8)
- Fujisaki, W. & Nishida, S. 2007 Feature-based processing of audio–visual synchrony perception revealed by random pulse trains. *Vision Res.* **47**, 1075–1093. (doi:10.1016/j.visres.2007.01.021)
- Fujisaki, W. & Nishida, S. 2008 Top-down feature-based selection of matching features for audio–visual synchrony discrimination. *Neurosci. Lett.* **433**, 225–230. (doi:10.1016/j.neulet.2008.01.031)
- Fujisaki, W. & Nishida, S. 2009 Audio–tactile superiority over visuo–tactile and audio–visual combinations in the temporal resolution of synchrony perception. *Exp. Brain Res.* **198**, 245–259.
- Fujisaki, W., Koene, A., Arnold, D., Johnston, A. & Nishida, S. 2006 Visual search for a target changing in synchrony with an auditory signal. *Proc. R. Soc. B* **273**, 865–874. (doi:10.1098/rspb.2005.3327)
- He, S., Cavanagh, H. D. & Intriligator, J. 1997 Attentional resolution. *Trends Cogn. Sci.* **1**, 115–121. (doi:10.1016/S1364-6613(97)89058-4)
- Holcombe, A. O. 2009 Seeing slow and seeing fast: two limits on perception. *Trends Cogn. Sci.* **13**, 216–221. (doi:10.1016/j.tics.2009.02.005)
- Holcombe, A. O. & Cavanagh, P. 2001 Early binding of feature pairs for visual perception. *Nat. Neurosci.* **4**, 127–128. (doi:10.1038/83945)
- Holcombe, A. O. & Judson, J. 2007 Visual binding of English and Chinese word parts is limited to low temporal frequencies. *Perception* **36**, 49–74. (doi:10.1068/p5582)
- Lebib, R., Papo, D., de Bode, S. & Baudonniere, P. M. 2003 Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. *Neurosci. Lett.* **341**, 185–188. (doi:10.1016/S0304-3940(03)00131-9)
- Lee, S. H. & Blake, R. 1999 Visual form created solely from temporal structure. *Science* **284**, 1165–1168. (doi:10.1126/science.284.5417.1165)
- Leonards, U., Singer, W. & Fahle, M. 1996 The influence of temporal phase differences on texture segmentation. *Vision Res.* **36**, 2689–2697. (doi:10.1016/0042-6989(96)86829-5)
- Lewis, J. W. & Van Essen, D. C. 2000 Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* **428**, 112–137. (doi:10.1002/1096-9861(20001204)428:1<112::AID-CNE8>3.0.CO;2-9)
- Maeda, F., Kanai, R. & Shimojo, S. 2004 Changing pitch induced visual motion illusion. *Curr. Biol.* **14**, R990–R991. (doi:10.1016/j.cub.2004.11.018)
- Mishra, J., Martinez, A., Sejnowski, T. J. & Hillyard, S. A. 2007 Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J. Neurosci.* **27**, 4120–4131. (doi:10.1523/JNEUROSCI.4912-06.2007)
- Motoyoshi, I. 2004 The role of spatial interactions in perceptual synchrony. *J. Vision* **4**, 352–361. (doi:10.1167/4.5.1)
- Motoyoshi, I. & Nishida, S. 2001 Temporal resolution of orientation-based texture segregation. *Vision Res.* **41**, 2089–2105. (doi:10.1016/S0042-6989(01)00096-7)
- Motoyoshi, I. & Nishida, S. 2002 Spatiotemporal interactions in detection of texture orientation modulations. *Vision Res.* **42**, 2829–2841. (doi:10.1016/S0042-6989(02)00336-X)
- Moutoussis, K. & Zeki, S. 1997 A direct demonstration of perceptual asynchrony in vision. *Proc. R. Soc. Lond. B* **264**, 393–399. (doi:10.1098/rspb.1997.0056)
- Munhall, K. G., Gribble, P., Sacco, L. & Ward, M. 1996 Temporal constraints on the McGurk effect. *Percept. Psychophys.* **58**, 351–362.
- Nishida, S. & Johnston, A. 2002 Marker correspondence, not processing latency, determines temporal binding of visual attributes. *Curr. Biol.* **12**, 359–368. (doi:10.1016/S0960-9822(02)00698-X)
- Nishida, S. & Johnson, A. 2010 The time marker account of cross-channel temporal judgments. In *Space and time in perception and action* (eds R. Nijhawan & B. Khurana), pp. 278–300. Cambridge, UK: Cambridge University Press.
- Reeves, A. & Sperling, G. 1986 Attention gating in short-term visual memory. *Psychol. Rev.* **93**, 180–206. (doi:10.1037/0033-295X.93.2.180)
- Robertson, L. C. 2003 Binding, spatial attention and perceptual awareness. *Nat. Rev. Neurosci.* **4**, 93–102. (doi:10.1038/nrn1030)
- Schroeder, C. E. & Foxe, J. 2005 Multisensory contributions to low-level, ‘unisensory’ processing. *Curr. Opin. Neurobiol.* **15**, 454–458. (doi:10.1016/j.conb.2005.06.008)
- Sekuler, R., Sekuler, A. B. & Lau, R. 1997 Sound alters visual motion perception. *Nature* **385**, 308. (doi:10.1038/385308a0)
- Seymour, K., Clifford, C. W., Logothetis, N. K. & Bartels, A. 2009 The coding of color, motion, and their conjunction in the human visual cortex. *Curr. Biol.* **19**, 177–183. (doi:10.1016/j.cub.2008.12.050)
- Shipley, T. 1964 Auditory flutter-driving of visual flicker. *Science* **145**, 1328–1330. (doi:10.1126/science.145.3638.1328)
- Soto-Faraco, S., Navarra, J. & Alsius, A. 2004 Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition* **92**, B13–B23. (doi:10.1016/j.cognition.2003.10.005)
- Soto-Faraco, S., Spence, C. & Kingstone, A. 2005 Assessing automaticity in the audiovisual integration of motion. *Acta Psychol. (Amst.)* **118**, 71–92. (doi:10.1016/j.actpsy.2004.10.008)
- Spence, C. 2007 Audiovisual multisensory integration. *Acoust. Sci. Technol.* **28**, 61–70. (doi:10.1250/ast.28.61)
- Spence, C., Baddeley, R., Zampini, M., James, R. & Shore, D. I. 2003 Multisensory temporal order judgments: when two locations are better than one. *Percept. Psychophys.* **65**, 318–328.
- Treisman, A. 1996 The binding problem. *Curr. Opin. Neurobiol.* **6**, 171–178. (doi:10.1016/S0959-4388(96)80070-5)
- Treisman, A. 1999 Solutions to the binding problem: progress through controversy and convergence. *Neuron* **24**, 105–125. (doi:10.1016/S0896-6273(00)80826-0)
- Treisman, A. M. & Gelade, G. 1980 A feature-integration theory of attention. *Cogn. Psychol.* **12**, 97–136. (doi:10.1016/0010-0285(80)90005-5)
- Van de Par, S. & Kohlrausch, A. 2000 Sensitivity to auditory-visual asynchrony to jitter in auditory-visual timing. In *Proceedings of SPIE: human vision and electronic imaging V*, vol. 3959 (eds B. E. Rogowitz & T. N. Pappas), pp. 234–242. Bellingham, WA: SPIE.

- van Wassenhove, V., Grant, K. W. & Poeppel, D. 2005 Visual speech speeds up the neural processing of auditory speech. *Proc. Natl Acad. Sci. USA* **102**, 1181–1186. (doi:10.1073/pnas.0408949102)
- van Wassenhove, V., Buonomano, D. V., Shimojo, S. & Shams, L. 2008 Distortions of subjective time perception within and across senses. *PLoS ONE* **3**, e1437. (doi:10.1371/journal.pone.0001437)
- Vatakis, A., Bayliss, L., Zampini, M. & Spence, C. 2007 The influence of synchronous audiovisual distractors on audiovisual temporal order judgments. *Percept. Psychophys.* **69**, 298–309.
- Victor, J. D. & Conte, M. M. 2002 Temporal phase discrimination depends critically on separation. *Vision Res.* **42**, 2063–2071. (doi:10.1016/S0042-6989(02)00125-6)
- Warren, D. H., Welch, R. B. & McCarthy, T. J. 1981 The role of visual–auditory ‘compellingness’ in the ventriloquism effect: implications for transitivity among the spatial senses. *Percept. Psychophys.* **30**, 557–564.
- Zeki, S. 1993 *A vision of the brain*. Oxford, UK: Blackwell.