Matching auditory and visual signals: Is sensory modality just another feature?

Jeroen S Benjamins, Maarten J van der Smagt, Frans A J Verstraten
Experimental Psychology Division, Helmholtz Institute, Utrecht University, Heidelberglaan 2,
NL-3584 CS Utrecht, The Netherlands; e-mail: M.J.vanderSmagt@uu.nl
Received 2 March 2007, in revised form 29 June 2007; published online 1 May 2008

Abstract. In order to perceive the world coherently, we need to integrate features of objects and events that are presented to our senses. Here we investigated the temporal limit of integration in unimodal visual and auditory as well as crossmodal auditory – visual conditions. Participants were presented with alternating visual and auditory stimuli and were asked to match them either within or between modalities. At alternation rates of about 4 Hz and higher, participants were no longer able to match visual and auditory stimuli across modalities correctly, while matching within either modality showed higher temporal limits. Manipulating different temporal stimulus characteristics (stimulus offsets and/or auditory – visual SOAs) did not change performance. Interestingly, the difference in temporal limits between crossmodal and unimodal conditions appears strikingly similar to temporal limit differences between unimodal conditions when additional features have to be integrated. We suggest that adding a modality across which sensory input is integrated has the same effect as adding an extra feature to be integrated within a single modality.

1 Introduction
A coherent representation of objects and events in the world around us depends on integration of their multiple features across space and in time. Our brain appears to accomplish this task almost effortlessly, although integration of features, especially when they originate from different dimensions, has spatial as well as temporal limits.

In visual perception, temporal integration limits are usually explored by presenting rapidly alternating feature pairs. Typically, subjects are asked to match features that occur simultaneously. A study by Holcombe and Cavanagh (2001), for example, shows that correctly matching simultaneous alternations in orientation and colour of a stimulus is limited to alternations up to 18 Hz; this limit decreases to about 3 Hz when these features are spatially separated. Apparently, as features are separated over more than one dimension (space and time versus time alone), the temporal limit of integration changes dramatically.

In audition, models of temporal integration have been divided into two classes favouring different integration times. One line of research suggests that all auditory input that falls within a window of about 200 ms is integrated (5 Hz), while results based on gap detection and discrimination favour integration times of about 5 ms (200 Hz)—the so-called 'resolution – integration paradox' (de Boer 1985; Green 1985). Viemeister and Wakefield (1991) showed that increasing the noise level of a 100 ms interval between two sound pulses did not decrease detectability of the individual pulses, which argues against a large temporal window of integration. Moreover, even when the two pulses were separated by only 5 ms, they still appeared to be processed independently.

Clearly, auditory and visual perception have different temporal limits of integration. Yet, we often need to integrate across senses as well. Here we aim to determine the temporal limit of integration when we integrate across these two modalities. More specifically, we address the question whether the auditory – visual temporal integration limit differs from the limits set by either the visual or auditory modality.

So far, auditory–visual temporal *integration* research has typically focused on phenomena like auditory capture (Fendrich and Corballis 2001; Gebhard and Mowbray 1959; Recanzone 2003; Shipley 1964), synchrony judgments (Fujisaki et al 2004, 2006; Fujisaki and Nishida 2005; Noesselt et al 2005), and temporal-order judgments (Keetels and Vroomen 2005; Morein-Zamir et al 2003; Vroomen et al 2004). In these paradigms the auditory and visual signals are not presented synchronously. Subject's sensitivity to this auditory–visual *asynchrony* is measured, which is a measure of *segregation*. However, even though integration of specific features without segregation of others cannot exist, the temporal limit of *segregation* of auditory–visual signals is not necessarily the same as the temporal limit of *integration*. The same separate sensory events might be perceived asynchronously in one task, while they might be integrated into a single percept in another. Moreover, depending on whether the task is to integrate or segregate sensory input, spatial location, for example, can have opposite effects on temporal limits. As mentioned, matching in Holcombe and Cavanagh's (2001) unimodal task has a higher temporal limit when features alternate at the same spatial location, while in auditory–visual temporal-order judgments the just noticeable difference (JND) between auditory–visual signals is smaller (ie there is a higher temporal limit) when features are spatially separated (Keetels and Vroomen 2005). With this possible task-dependent effect on perceiving sensory events synchronously in mind, we specifically investigated the limits of temporal *integration* by using a matching paradigm in the visual and auditory domain as well as across these modalities.

2 Experiment 1: Crossmodal versus unimodal matching
In this experiment we aimed to find the temporal limit of matching auditory and visual input. The two sensory systems we tap into here both have separate limits of temporal integration (Holcombe and Cavanagh 2001; Viemeister and Wakefield 1991). The question arises whether one of these systems forms a bottleneck in auditory–visual integration, or whether the crossmodal integration limit differs from both. If the latter is the case it makes sense to expect that limit to lie below the integration limit of either modality.

2.1 *Methods*
2.1.1 *Participants.* Four subjects (two naive to the purpose of the experiment) participated in the experiment. All participants had normal or corrected-to-normal vision and no known hearing problems.

2.1.2 *Stimuli and procedure.* Stimuli were created and presented with Matlab 5.2.1 and the Psychophysics Toolbox extensions (Brainard 1997; Pelli 1997) on an Apple G4 computer. The visual stimuli were presented on a 19 inch Iiyama Pro Vision Master 454 monitor, set to a resolution of 640 by 480 pixels and a refresh rate of 200 Hz, yielding 5 ms timing precision. The auditory stimuli were presented through Sennheiser HD-250 linear II headphones, connected directly to the audio port of the computer. The participants were seated 86 cm from the monitor and their heads were supported by chin-and-forehead rests.

2.1.3 *Auditory – visual conditions.* The visual stimulus consisted of isoluminant (35.6 cd m^{-2}) red and green discs spanning 0.75 deg of visual angle. The discs alternated in colour with different frequencies. The rate of alternation was varied per trial. Alternation rate was defined as the inverse of the period from one onset of a coloured disc (eg red) to the onset of the next disc with the same colour (see figure 1a). Five alternation rates were used: 1, 1.9, 3.1, 4.2, and 5 Hz. For each trial, the coloured discs alternated for a total of 3 s. The discs were presented 3 deg above or below a 0.5 deg fixation spot (black/grey bull's eye) on a grey background (15.1 cd m^{-2}). The location of the visual stimulus was

varied to control for different temporal performance fields that have been demonstrated in visual perception (Carrasco et al 2004; Yeshurun and Levy 2003), and which might differentially affect auditory – visual integration as well. Results from the studies mentioned above show a temporal performance advantage in the lower visual field.

Two sounds [a high-pitched tone of 2 kHz, and a low-pitched tone of 1 kHz, 78.9 dB and 77.9 dB sound pressure level (SPL), respectively] were presented simultaneously with the onset of the red and the green disc. Each tone had a duration of 25 ms and contained no spatial information. Tone duration was kept constant, since changing duration of tones along with alternation rate yielded perceptually different sounds. When a high-pitched tone was presented at onset of the red disc, a low-pitched tone was presented at the onset of the green disc. Physical onset differences between sound and colour stimuli ranged from 1 to 1.88 ms, which coincided with the vertical build-up of the screen. Which tone pitch was presented simultaneously with which colour was randomly varied between trials.
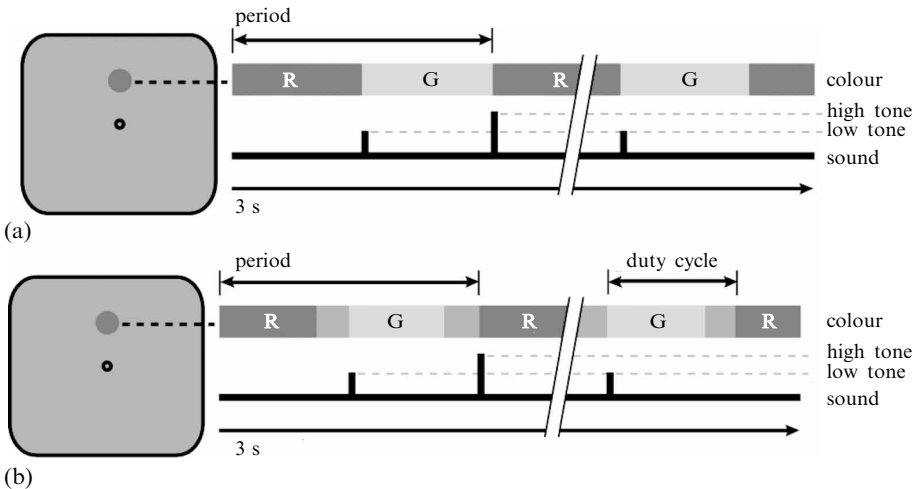
Figure 1. Schematic overview of a trial. Example of red (R)/high – green (G)/low presentation. (a) 100% duty-cycle trial, (b) 75% duty-cycle trial. The grey scale bar (dark grey = red, light grey = green, intermediate grey = coloured disc off) and pulse time line together indicate that a high tone is presented at onset of each red disc.

2.1.4 *Auditory – visual procedure.* The subject's task was to fixate the bull's eye, attend to both sound and colour of the stimulus, and, after the stimulus had ended, to indicate which sound accompanied the red disc. When a high-pitched tone was perceived simultaneously with the red disc, participants pressed the up-arrow key on the computer keyboard. When a low-pitched tone was perceived, they pressed the down-arrow key. To prevent participants from matching the correct sound to colour solely on the basis of their first or last simultaneous onsets, the first and last onsets of the coloured discs were not accompanied by sound. In each trial the number of first colour onsets without sound was randomly varied (either one or two colour onsets) as was the number of last colour onsets without sound (again either one or two colour onsets). Each participant was presented with a block of 150 trials (5 alternation rates, 30 times repeated in random order).

2.1.5 *Unimodal visual conditions.* In the unimodal visual conditions, the red/green alternating disc was identical to that in the auditory – visual conditions, but the auditory signal was replaced by a visual stimulus. A second set of smaller (0.25 deg) alternating black and white discs was superimposed on the coloured discs. As in Holcombe and Cavanagh's (2001) visual-matching paradigm, a second feature (contrast polarity of

the smaller disc) was superimposed on the coloured discs.[1] The participant's task was to indicate whether the white or the black disc coincided with the red disc. Otherwise, the procedure was identical to the auditory–visual procedure.

2.1.6 *Unimodal auditory conditions.* For the unimodal auditory conditions, the two sounds (a high-pitched tone of 2 kHz, and a low-pitched tone of 1 kHz) were presented monaurally for 25 ms. A second set of alternating tones with different pitch (3.3 and 4.3 kHz, 73.4 dB and 70.2 dB SPL, respectively) was presented to the other ear. The participant's task was to indicate whether the two higher-pitched tones were presented simultaneously (ie 2 kHz in one ear and 4.3 kHz in the other), or whether the high-pitched tone of one pair coincided with the low-pitched tone of the other. Otherwise, the procedure was identical to the auditory–visual procedure. It could be argued that by presenting to-be-matched auditory stimuli diotically, we introduce spatial separation of signals, thus making it harder to match simultaneously presented sounds. A second auditory matching paradigm was designed to test this. The high and low (2 kHz and 1 kHz) tones were presented in both ears and were presented simultaneously with two alternating types of bandpassed noise ('high' bandpass: 4.4–4.5 kHz, 'low' bandpass: 0.4–0.5 kHz). Pitch : noise ratio was set to 9 : 1. This prevented the tones from being drowned out by the bandpassed noise.

Again, the participant's task was to indicate whether the high-pitched tone was presented simultaneously with the 'high' noise, or whether the high-pitched tone coincided with the 'low' noise. The SPL of the pitch–pitch combinations was 79.35 dB (SD = 0.64 dB) and 76.02 dB (SD = 0.69 dB) in the pitch–noise paradigm. Thus, participants could not determine a match on the basis of SPL differences.

2.2 *Results and discussion*
Figure 2 shows the results collapsed across visuo–spatial conditions for both the unimodal and crossmodal conditions, since the effect of spatial location (upper versus lower visual field) as reported by Carrasco et al (2004) and Yeshurun and Levy (2003) was not apparent from our data. Both in crossmodal and unimodal visual conditions there was no significant difference between presenting the visual stimuli above or below fixation ($F_{1\,8;\,\mathrm{cross}}$ = 0.004, p = 0.951; and $F_{1\,8;\,\mathrm{vis}}$ = 5.16, p = 0.053—see additional online material at http://dx.doi.org/10.1068/p5783). It is immediately clear from the figure that the temporal limit for the unimodal visual matching lies well above 5 Hz (the highest alternation frequency used) as, on average, performance is still above 85% correct at this frequency. The unimodal auditory matching performance is generally worse than that for the visual conditions, demonstrating that this is a more difficult task. This is reflected in the larger variance in performance at the lower frequencies; two participants started out with a performance of 80% correct, while the other two showed 98% correct at 1 Hz for pitch–pitch matches. For pitch–noise matches these percentages were 78% and 93%. This auditory matching performance declines to 58% correct for pitch–pitch matches and 61% for pitch–noise matches for all participants at 5 Hz.

Mann–Whitney U tests (proportion correct is assumed not to be normally distributed) of the 5 Hz data show that unimodal matching performance still differs significantly (U = 0, p < 0.05 for visual; and U = 0, p < 0.05 for both auditory pitch–pitch and pitch–noise conditions) from chance performance, while crossmodal matching does not: as alternation rate increases, performance decreases from above 95% correct at 1 Hz alternation to chance level at alternation rates of 4.2 Hz and higher (dark bars).

[1] Matching one feature (colour) was tested in the pilot phase of the experiment as well. Another equiluminant set of smaller (0.25 deg) coloured discs (yellow–cyan; 95.5 cd m^{-2}) was superimposed on the red–green coloured discs. This yielded similar results as matching colour and contrast polarity.

At these frequencies observers were no longer able to match sound and colour correctly. This is consistent with a pilot study using more alternation rates (1 to 5.5 Hz with steps of approximately 0.5 Hz), which resulted in the same limit (see online material).

From these results it is clear that neither temporal integration limit of the visual modality, nor that of the auditory modality forms a bottleneck for the auditory–visual matching performance. Note that the same visual and auditory stimuli were used, where possible, in crossmodal and unimodal conditions. Crossmodal matching shows a limit around or below 4 Hz, while the limit of unimodal matching lies above 5 Hz. The fact that the unimodal matching limit lies above 5 Hz is in line with the results mentioned above (Holcombe and Cavanagh 2001; Viemeister and Wakefield 1991).
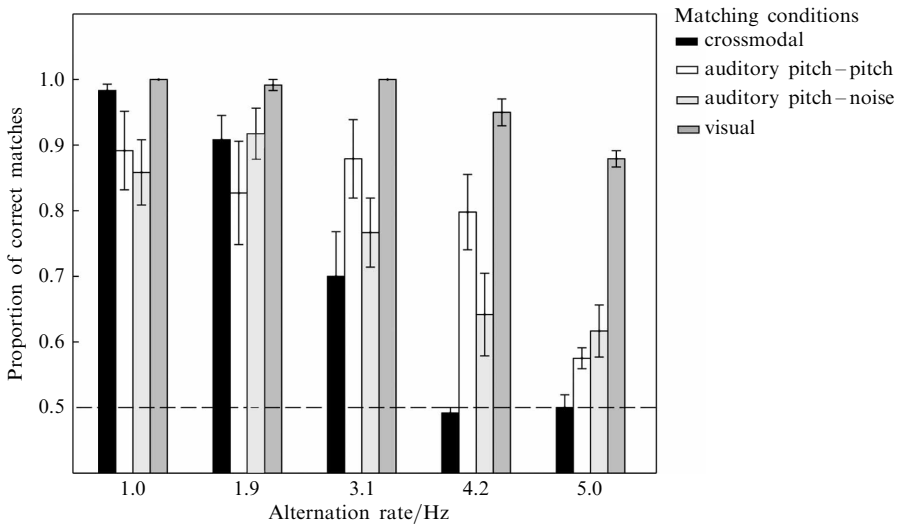
Figure 2. Average proportion of correct matches for the different alternation rates in crossmodal (solid black bars), visual (dark-grey bars), and auditory (white and light-grey bars) matching conditions. Error bars are ±1 SEM. The dashed line indicates chance performance. 30 trials per point per subject. $n = 4$.

3 Experiment 2: Transient onsets and offsets

In the first experiment, the auditory stimulus was a short transient stimulus with clear onsets as well as offsets. The coloured discs, however, were isoluminant (even though they were not isoluminant with the background). The transition of one colour to the other (onset and offset at the same time) coincided with the onset of the auditory stimulus. A possible explanation for the low performance in the crossmodal conditions is that the onset of the auditory stimulus sometimes becomes associated with the offset instead of onset of one particular colour. Therefore, colour (and luminance) offsets were introduced into the visual stimulus in the second experiment, which only tested crossmodal conditions.

3.1 Methods

Participants and procedure were identical to those in the auditory–visual conditions of experiment 1. The auditory stimuli were only 5 ms in duration. The visual stimulus differed in the following way: interleaved with the red–green alternating disc, a blank interval was presented to investigate the effect of colour offset. The duration of this blank interval was varied as a function of the duty cycle of a coloured disc in five possible conditions: 100% duty cycle (no blank interval, red changed directly to green; similar to the auditory–visual condition of experiment 1), 75%, 50%, and 25% duty cycle (see figure 1b for an example of a 75% duty-cycle trial) and a fixed 10 ms disc

presentation (each coloured disc was presented for a fixed duration of 10 ms). Each participant was presented with three blocks of 250 trials (5 alternation rates times 5 offset timing conditions, 10 times repeated per block in random order).

3.2 Results and discussion

Figure 3 shows that, as alternation rate increases, performance of auditory – visual matching decreases, as has already been shown in experiment 1. The shorter duration of the auditory stimulus did not influence the results, as is apparent from the solid black bars (100% duty-cycle condition, which is a replication of the crossmodal results from experiment 1). For all duty-cycle conditions, the proportion of correct matches decreases from about 90% at the lowest alternation rate to chance level at alternation rates of 4.2 Hz and higher. This is independent of the duty cycle of the visual stimulus. The one exception is the condition where the visual stimulus had a fixed presentation duration of 10 ms (most similar in duration to the 5 ms auditory stimulus across all alternation rates used). For this condition performance was even worse. At 1.9 Hz alternation, the performance is already at chance level. An additional experiment at a fixed alternation rate of 2 Hz, with varying presentation duration of the visual stimulus (10 ms, and 25 – 250 ms in 25 ms steps), but otherwise identical to the experiment above, showed that only at the lowest (10 ms) presentation duration this strong decrease in performance is apparent (see online material). At 25 ms presentation duration, performance is still above 70% on average correct. It seems plausible that at 10 ms presentation duration the discriminability (or visibility) of the visual stimulus is somewhat impaired, making the crossmodal matching task at 1.9 Hz already more difficult.
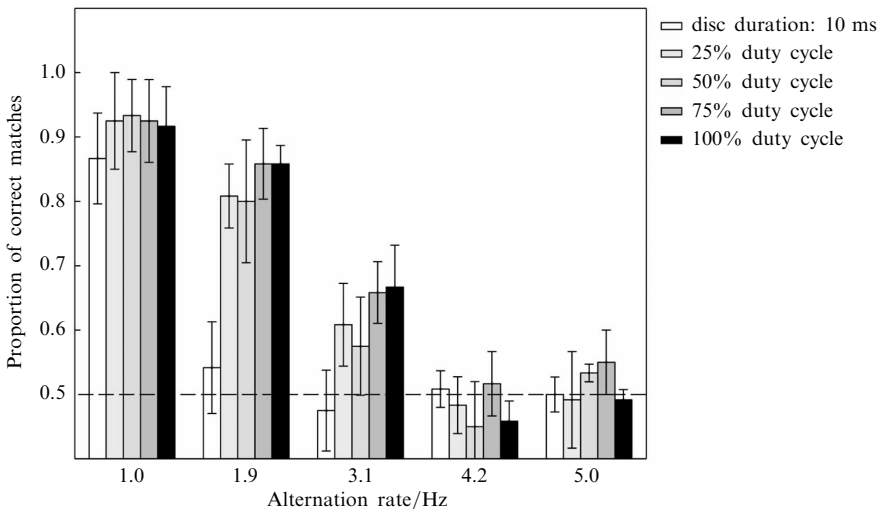
Figure 3. Average proportion of correct sound – colour matches for the different alternation rates. Error bars are ±1 SEM. The dashed line indicates chance performance. 30 trials per point per subject. $n = 4$.

Nonetheless, the overall results of this experiment show that adding a transient visual offset signal (and thus additional temporal markers to which the auditory signal can be matched or mismatched) neither facilitates nor impedes auditory – visual matching.

4 Experiment 3: Perceptual delay

Several studies have demonstrated a perceptual delay between auditory and visual signals (Fujisaki et al 2004; Keetels and Vroomen 2005; Lewald and Guski 2003; Vroomen et al 2004; Zampini et al 2005). The size of this perceptual delay ranges from 7 – 10 ms (Fujisaki et al 2004; Keetels and Vroomen 2005) to 50 – 100 ms (Lewald and

Guski 2003), which is an order of magnitude greater. In the previous experiments, any perceptual delay between sound and colour might account for decreasing matching performance when alternation rate increases. Therefore, in this final experiment we investigated whether low performance in the previous experiments is confounded by perceived asynchrony between visual and auditory stimuli. To this end physical sound and colour onset synchrony was systematically varied.

4.1 Methods
Stimuli and task were the same as in the 100% duty-cycle condition of the second experiment. Only two alternation rates were used: 1.25 and 2.5 Hz. The temporal differences in sound and colour onset were systematically varied over a range of $-189°$ to $189°$ phase shifts in 29 equal steps. Negative onset asynchronies meant that sound onsets preceded colour onsets (see figure 4). Three subjects (two naive) were tested in six blocks of 290 trials (29 phase shifts times 2 alternation rates, repeated 5 times per block).
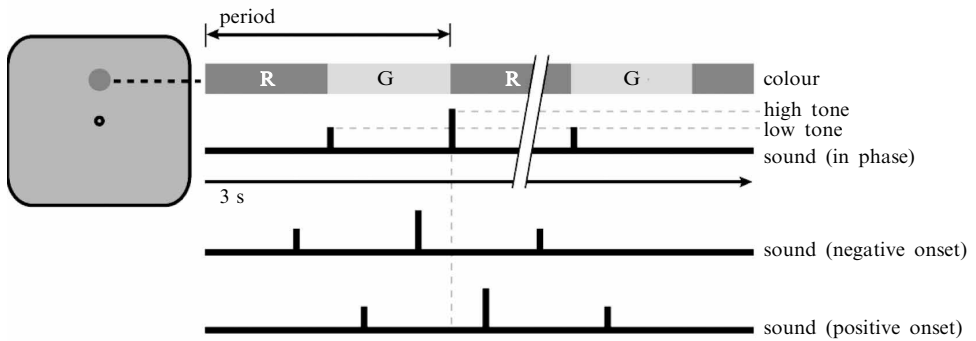

Figure 4. Schematic overview of a high-pitched tone co-occurring with red (R) either without sound-onset shift, with a negative sound-onset shift, or with a positive sound-onset shift.

4.2 Results and discussion
The proportion of the high-pitched tone being perceived with red in figure 5a is plotted as a function of the presented sound–colour onset asynchrony. Since the subjects are asked to report which (high-pitched or low-pitched) tone occurs simultaneously with the colour red, a proportion of 1 indicates that a high-pitched tone and red are indeed perceived to occur simultaneously. If this proportion is zero it means that the high-pitched tone is actually perceived with green.

　In the 1.25 Hz condition this proportion is expected to peak above 0.85, since that is the proportion of correct matches at 1.9 Hz with 100% duty cycle in experiment 2 (see figure 3, black bars). For a range of $-13.5°$ to $108°$ of sound–colour onset asynchrony (see figure 5a) the proportion of the high-pitched tone perceived with red is indeed above 0.85. This range, however, does not show a clear peak at any specific phase shift. Apparently, within a broad range of physical phase differences between sound and colour onset, subjects are able to match sound and colour correctly.

　For the 2.5 Hz condition, this proportion is expected to peak above 0.67, since that is the proportion of correct matches at 3.1 Hz with 100% duty cycle in experiment 2 (see figure 3, black bars). The maximum proportion is 0.75 in the 2.5 Hz condition at 94.5° phase shift and it remains above 0.67 for a range of 0° to 108° sound–colour onset asynchrony (see figure 5a). Again, a broad range of physical sound and colour onset asynchrony yields similar matching performance. This becomes even more apparent when the data are represented in a polar plot (figure 5b). Rather than showing a clear peak at a specific phase shift, matching performance in both conditions stays the same across a broad range of phasic delay between sound and colour onset.
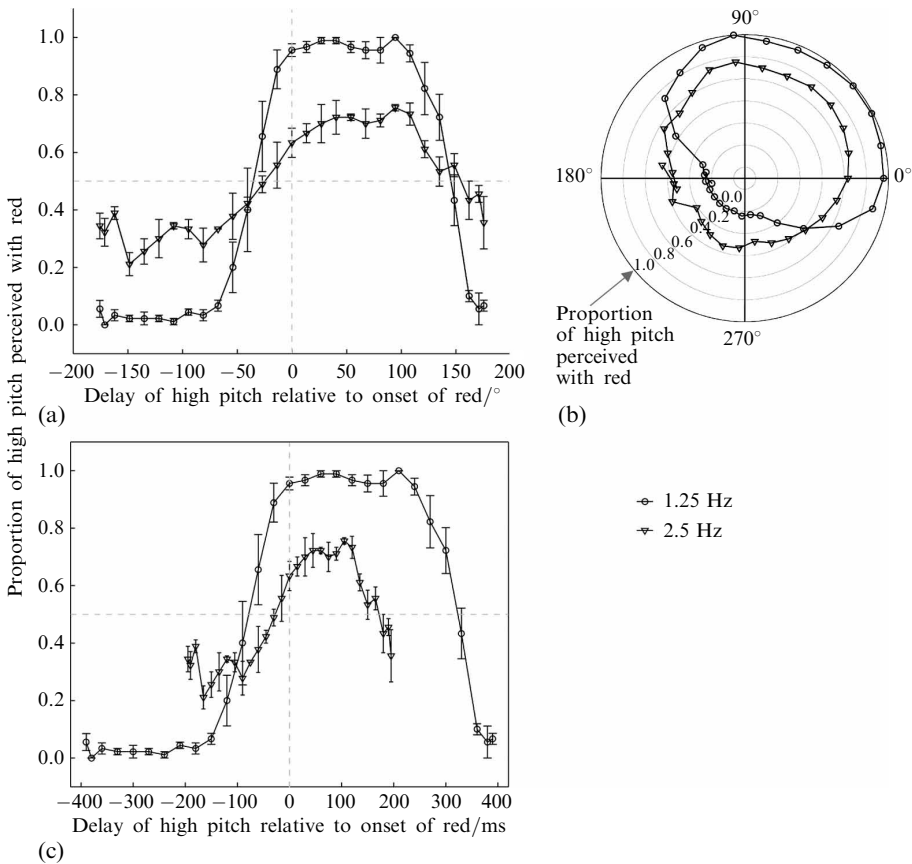
Figure 5. (a) Average proportion of the high-pitched tone being perceived with red as a function of the delay of the high-pitched sound relative to onset of red for both frequency conditions: (a) and (b) Cartesian and polar coordinate plots with delay in degrees; (c) Cartesian coordinate plot with delay in milliseconds. Error bars are ±1 SEM. 30 trials per point per subject. $n = 3$.

Apparently, even though a (slight) perceptual delay between sound and colour onset might be demonstrated in our setup, it does not appear to play a significant role in matching sound and colour when presented physically synchronous.

When the proportion of high-pitched tone perceived in synchrony with red is plotted as a function of the phase shift between sound and colour, the range of phase shifts that yield around maximum performance is similar for both alternation rates used: $-13.5°$ to $108°$ in the 1.25 Hz condition, $0°$ to $108°$ in the 2.5 Hz condition (figure 5a). However, when the proportion is plotted as a function of time delay (in milliseconds; see figure 5c) this range is clearly different: -30 ms to 270 ms in the 1.25 Hz condition, 0 to 135 ms in the 2.5 Hz condition. Interestingly, this means that the range of delays where sound and colour are perceived as occurring simultaneously differs depending on the alternation rate. We will return to this in the next section.

5 General discussion
Using an auditory – visual matching paradigm we have found a temporal integration limit for auditory – visual signals around or below 4 Hz. Furthermore, we have shown that this limit is not determined by a temporal integration limit in one of the two sensory systems, which in our experiment demonstrated temporal limits well beyond 4 Hz. This concurs with findings concerning limits of unimodal visual (Holcombe and Cavanagh 2001) and auditory temporal integration (Viemeister and Wakefield 1991), although the

auditory matching limit found in our experiment is notably lower than that reported by Viemeister and Wakefield. However, this difference can be explained by the fact that they merely measured detectability, while we measured matching performance.

The temporal limit of integration in the current experiment is not affected by timing of colour offsets, which shows that transient signals created by these offsets are not used for matching sound and colour. When each colour is presented for 10 ms, performance drops to chance level at presentation rates below the 4 Hz limit. How this very short duration of colour affects matching performance needs further investigation.

We also investigated whether perceptual delay between auditory and visual signals affected the temporal limit we found in experiments 1 and 2. The results from previous auditory–visual studies on perceptual delay have shown that this delay can vary from about 7 to 100 ms (Fujisaki et al 2004; Keetels and Vroomen 2005; Lewald and Guski 2003; Vroomen et al 2004; Zampini et al 2005). At two alternation rates below the integration limit found in experiments 1 and 2 (1.25 Hz and 2.5 Hz), matching performance appeared not to be affected by perceptual delay.

Interestingly, the range of auditory–visual onset asynchronies, where matching performance was above 80%, was similar at both the 1.25 Hz and 2.5 Hz alternation rate in *phasic delay*, which means that in *time delay* this range scales with alternation rate. Such a frequency-dependent scaling of perceptual synchrony has also been demonstrated recently for the perception of natural and artificial auditory–visual motion sequences (Arrighi et al 2006). Moreover, other reports on auditory–visual perceptual delay show that recalibration takes place after adapting to auditory–visual signals that are presented physically asynchronous (Fujisaki et al 2004; Vroomen et al 2004); the point of subjective auditory–visual synchrony due to the perceptual delay shifts with the physical asynchrony. Moreover, other research has shown that even the (simulated) difference between light and sound arrival time can be accounted for as well, when sound and vision (appear to) come from the same spatial location (Alais and Carlile 2005; Sugita and Suzuki 2003). Taken together, we conclude that auditory–visual integration has a great capacity for compensating different physical lag times.

5.1 Temporal limits of integration

The crossmodal temporal limits presented in the current study are similar to those found for the temporal matching of unimodal visual features, but only when these visual features alternate in time and are spatially separated (Holcombe and Cavanagh 2001). By separating features spatially, the maximum alternation rate at which colour and orientation can be matched decreases from 18 Hz to 3 Hz. This effect on integration limits of separating visual signals spatially was also confirmed and extended by a recent study by Bartels and Zeki (2006). In addition to spatially separating visual signals, they show a different temporal integration limit for matching visual signals *within* one visual attribute (3.3 Hz), compared to matching spatially separated signals *across* visual attributes (2.6 Hz). Again, when sensory events differ on an extra dimension (attribute, time, and space versus time and space), the temporal integration limit of this input decreases further. In our experiment, we find a limit of auditory–visual temporal integration that is similar to matching spatially separate visual features, while the sensory input in the current experiment differs only in modality and alternates in time. It could be argued that sound and vision are separated in space in our experiment, since the visual stimuli were presented on a monitor 86 cm away from the subject, while auditory stimuli were presented through headphones. However, a recent study (Arnold et al 2005) has shown that auditory and visual signals of an event that reach an observer at the same point in time tend to become perceptually bound, even when the input comes from different sources such as a monitor and a set of headphones.

Therefore it is tempting to suggest that adding a modality across which sensory input is integrated has the same effect as adding an extra feature when integrating sensory input within one modality. In other words, sensory modality might be processed as just another feature of sensory events. Recent independent confirmation can be found in a study by Bodelón et al (2007). They showed that the perceptual temporal resolution is higher for individual features compared to the temporal resolution of perceiving conjunctions of features.

It has been recently proposed (Fujisaki and Nishida 2005) that attentional or 'mid-level' mechanisms are involved in matching simultaneously occurring sensory events. These mechanisms select the salient features of the sensory input, which are then passed onto a temporal correlator. All selected features that fall into a window spanning about 200 ms become correlated. All correlated input is integrated into one multisensory percept. This results in a temporal limit of (crossmodal) integration of around 4 – 5 Hz. Our current results are in agreement with this model, since our limit of around 4 Hz alternation rate is close to that of the model. Thus, a similar temporal limit is found when either a matching paradigm or a synchrony/asynchrony judgment paradigm is used (eg Fujisaki and Nishida 2005). Apparently, the temporal limit of auditory – visual (a)synchrony sensitivity is similar to the temporal limit of auditory – visual integration. Note, however, that Fujisaki and Nishida have constructed this model of auditory – visual integration without taking Holcombe and Cavanagh's result (of a higher unimodal matching limit when visual features are spatially superimposed) into account. To incorporate this (and our) results into their model, their temporal correlation window should vary in size, depending on the number of features (including modality) to be matched.

Both in our paradigm and in synchrony judgments, *integration* of sensory inputs will enable observers to make a judgment about the events presented to them (Fujisaki et al 2006; Fujisaki and Nishida 2005). In comparison, auditory – visual temporal-order judgments, which are also used to determine auditory – visual sensitivity, rely heavily on *segregation* of the auditory – visual signals (Keetels and Vroomen 2005; Vroomen et al 2004). Results for this kind of experiment typically show JNDs in the order of 20 – 40 ms [which is of a different order of magnitude than the integration results presented here, and by Fujisaki et al (2006) and Fujisaki and Nishida (2005)]. Interestingly, in this type of research, adding a feature on which sensory events differ, specifically spatial location, actually yields better temporal performance. The JND between auditory – visual signals is decreased when sources of auditory and visual input are separated in space (Keetels and Vroomen 2005; Zampini et al 2003). Shore et al (2001) did not find such a performance benefit of spatial location in unimodal visual temporal-order judgments (although their experiments were designed to answer different questions). Future research should thus focus on investigating whether adding a spatial dimension to auditory – visual matching would yield an increase in auditory – visual matching performance (comparable to auditory – visual *segregation* experiments) or a drop in performance due to the addition of yet another feature (comparable to unimodal *matching/integration* experiments—eg Holcombe and Cavanagh 2001).

References
Alais D, Carlile S, 2005 "Synchronizing to real events: subjective audiovisual alignment scales with perceived auditory depth and speed of sound" *Proceedings of the National Academy of Sciences of the USA* **102** 2244 – 2247
Arnold D H, Johnston A, Nishida S, 2005 "Timing sight and sound" *Vision Research* **45** 1275 – 1284
Arrighi R, Alais D, Burr D, 2006 "Perceptual synchrony of audiovisual streams for natural and artificial motion sequences" *Journal of Vision* **6** 260 – 268

Bartels A, Zeki S, 2006 "The temporal order of binding visual attributes" *Vision Research* **46** 2280 – 2286

Bodelón C, Fallah M, Reynolds J H, 2007 "Temporal resolution for the perception of features and conjunctions" *Journal of Neuroscience* **27** 725 – 730

Boer E de, 1985 "Auditory time constants: a paradox?", in *Time Resolutions in Auditory Systems* Ed. A Michelsen (Berlin: Springer) pp 141 – 158

Brainard D H, 1997 "The Psychophysics Toolbox" *Spatial Vision* **10** 433 – 436

Carrasco M, Giordano A M, McElree B, 2004 "Temporal performance fields: visual and attentional factors" *Vision Research* **44** 1351 – 1365

Fendrich R, Corballis P M, 2001 "The temporal cross-capture of audition and vision" *Perception & Psychophysics* **63** 719 – 725

Fujisaki W, Koene A, Arnold D H, Johnston A, Nishida S, 2006 "Visual search for a target changing in synchrony with an auditory signal" *Proceedings of the Royal Society of London, Series B* **273** 865 – 874

Fujisaki W, Nishida S, 2005 "Temporal frequency characteristics of synchrony – asynchrony discrimination of audiovisual signals" *Experimental Brain Research* **166** 455 – 464

Fujisaki W, Shimojo S, Kashino M, Nishida S, 2004 "Recalibration of audiovisual simultaneity" *Nature Neuroscience* **7** 773 – 778

Gebhard J W, Mowbray G H, 1959 "On discriminating the rate of visual flicker and auditory flutter" *American Journal of Psychology* **72** 521 – 529

Green D M, 1985 "Temporal factors in psychoacoustics", in *Time Resolution in Auditory Systems* Ed. A Michelsen (Berlin: Springer) pp 122 – 140

Holcombe A O, Cavanagh P, 2001 "Early binding of feature pairs for visual perception" *Nature Neuroscience* **4** 127 – 128

Keetels M, Vroomen J, 2005 "The role of spatial disparity and hemifields in audiovisual temporal order judgments" *Experimental Brain Research* **167** 635 – 640

Lewald J, Guski R, 2003 "Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli" *Cognitive Brain Research* **16** 468 – 478

Morein-Zamir S, Soto-Faraco S, Kingstone A, 2003 "Auditory capture of vision: examining temporal ventriloquism" *Cognitive Brain Research* **17** 154 – 163

Noesselt T, Fendrich R, Bonath B, Tyll S, Heinze H J, 2005 "Closer in time when farther in space— spatial factors in audiovisual temporal integration" *Cognitive Brain Research* **25** 443 – 458

Pelli D G, 1997 "The VideoToolbox software for visual psychophysics: transforming numbers into movies" *Spatial Vision* **10** 437 – 442

Recanzone G H, 2003 "Auditory influences on visual temporal rate perception" *Journal of Neurophysiology* **89** 1078 – 1093

Shipley T, 1964 "Auditory flutter-driving of visual flicker" *Science* **145** 1328 – 1330

Shore D I, Spence C, Klein R M, 2001 "Visual prior entry" *Psychological Science* **12** 205 – 212

Sugita Y, Suzuki Y, 2003 "Implicit estimation of sound-arrival time" *Nature* **421** 911

Viemeister N F, Wakefield G H, 1991 "Temporal integration and multiple looks" *Journal of the Acoustical Society of America* **90** 858 – 865

Vroomen J, Keetels M, Gelder B de, Bertelson P, 2004 "Recalibration of temporal order perception by exposure to audiovisual asynchrony" *Cognitive Brain Research* **22** 32 – 35

Yeshurun Y, Levy L, 2003 "Transient spatial attention degrades temporal resolution" *Psychological Science* **14** 225 – 231

Zampini M, Shore D I, Spence C, 2003 "Audiovisual temporal order judgments" *Experimental Brain Research* **152** 198 – 210

Zampini M, Shore D I, Spence C, 2005 "Audiovisual prior entry" *Neuroscience Letters* **381** 217 – 222