BRILL

Seeing and Perceiving 24 (2011) 623–638

brill.nl/sp

Synchronous Sounds Enhance Visual Sensitivity without Reducing Target Uncertainty *

Yi-Chuan Chen [1,**], **Pi-Chun Huang** [2], **Su-Ling Yeh** [3] **and Charles Spence** [1]

[1] Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD, UK

[2] Department of Ophthalmology, McGill University, Montreal, Canada

[3] Department of Psychology, National Taiwan University, Taipei, Taiwan

Received 12 January 2011; accepted 5 September 2011

Abstract

We examined the crossmodal effect of the presentation of a simultaneous sound on visual detection and discrimination sensitivity using the equivalent noise paradigm (Dosher and Lu, 1998). In each trial, a tilted Gabor patch was presented in either the first or second of two intervals embedded in dynamic 2D white noise with one of seven possible contrast levels. The results revealed that the sensitivity of participants' visual detection and discrimination performance were both enhanced by the presentation of a simultaneous sound, though only close to the noise level at which participants' target contrast thresholds started to increase with the increasing noise contrast. A further analysis of the psychometric function at this noise level revealed that the increase in sensitivity could not be explained by the reduction of participants' uncertainty regarding the onset time of the visual target. We suggest that this crossmodal facilitatory effect may be accounted for by perceptual enhancement elicited by a simultaneously-presented sound, and that the crossmodal facilitation was easier to observe when the visual system encountered a level of noise that happened to be close to the level of internal noise embedded within the system.

Keywords

Signal enhancement, temporal uncertainty, transient, audiovisual, multisensory facilitation

1. Introduction

Human perceptual systems, especially early vision, have traditionally been considered in terms of individual processing modules (e.g., Fodor, 1983; Pylyshyn,

---

* This article is part of the Multisensorial Perception collection, guest edited by S. Wuerger, D. Alais and M. Gondan.

** To whom correspondence should be addressed. E-mail: yi-chuan.chen@queens.oxon.org

 DOI:10.1163/187847611X603765

1999). This view was based, at least in part, on the fact that the cerebral cortex in humans is anatomically and functionally divided into discrete sensory processing areas, such as the well-known primary visual cortex, primary auditory cortex, and so on (see Felleman and van Essen, 1991). Recently, however, functional magnetic resonance imaging (fMRI) studies have demonstrated that activity in primary visual cortex can be modulated by the presentation of a synchronous sound (Noesselt *et al.*, 2010; Watkins *et al.*, 2007; Werner and Noppeney, 2010). Neurophysiological evidence such as this raises the possibility that auditory signals can enhance the early stages of visual information processing, either *via* direct neural connections between modality-specific brain areas and/or *via* feedback from multisensory brain areas (see Driver and Noesselt, 2008; Ghazanfar and Schroeder, 2006, for reviews). The present study was therefore designed to investigate whether the presentation of a synchronous sound can genuinely enhance human perceptual sensitivity regarding a visual target using a psychophysical technique; if true, three possible underlying mechanisms proposed by previous researchers (signal enhancement, transient boosting, and uncertainty reduction; see below) would be tested in order to provide the best explanation to the results reported here.

In human behavioral studies, crossmodal facilitatory effects have been demonstrated using various experimental paradigms. For example, people's detection sensitivity to the onset of a visual target can be enhanced by the presentation of a synchronous sound (e.g., Bolognini *et al.*, 2005; Frassinetti *et al.*, 2002; Lippert *et al.*, 2007). Furthermore, when the onset (or color change) of a visual target is embedded in those of a series of distractors (such as in a rapid serial visual presentation stream or a dynamic visual search display), participants' visual discrimination/identification performance can be enhanced by the presentation of a synchronous sound as well (e.g., Chen and Yeh, 2008, 2009; Ngo and Spence, 2010a, b; Olivers and van der Burg, 2008; Spence and Ngo, in press; van der Burg *et al.*, 2008; Vroomen and de Gelder, 2000). The mounting empirical evidence currently suggests that the crossmodal facilitation of visual perception elicited by the concurrently-presented sound is a robust phenomenon, and one that should extend beyond simplified experimental environments (cf. de Gelder and Bertelson, 2003).

Early studies suggested that the neural signals associated with the simultaneously-presented visual and auditory stimuli can be integrated, resulting in the mutual enhancement of both visual and auditory signals. Consequently, detection thresholds are lower (i.e., sensitivity is higher) for the bimodal stimulus as compared to the unimodal stimuli. This is known as the *signal enhancement hypothesis* (Bernstein *et al.*, 1970; Hershenson, 1962; see also Bolognini *et al.*, 2005; Frassinetti *et al.*, 2002). Evidence apparently supporting this hypothesis was reported by Stein *et al.* (1996). They reported that participants rated a light as appearing brighter when accompanied by a sound than when presented alone. Note that it is well known that the presentation of a sound will alert an observer and thus induce him/her to make a faster behavioral response (Nickerson, 1973). However, such an alerting effect is only ever reported in terms of a facilitation of response latency rather than in terms of a crossmodal enhancement of perceptual sensitivity (Spence, 2010).

On the other hand, recent studies have demonstrated that the transient signal provided by the onset (or abrupt change) of a synchronous sound can sometimes be critical to eliciting crossmodal facilitation, no matter whether the intensity (i.e., the amplitude) of the sound happens to consistently increase or decrease in time with the change of the visual target (see Anderson and Mamassian, 2008; Noesselt *et al.*, 2008). The onset, or abrupt change, of a stimulus is a transient feature that extends across modalities (see Downar *et al.*, 2000). Note that such transient features only provide information regarding the onset of the visual stimulus, but no information regarding the identity of the visual target. Accordingly, it should be predicted that only human visual detection performance, rather than human visual discrimination (or identification) performance, can be enhanced by the presentation of the sound. This is known as the *transient boosting hypothesis* (see Anderson and Mamassian, 2008).

Obviously, the presentation of a synchronous sound can also induce a more liberal response criterion in participants with respect to the visual target than when no sound is presented (e.g., Bolognini *et al.*, 2005; Frassinetti *et al.*, 2002; Lippert *et al.*, 2007). That said, Stein *et al.*'s (1996) results, though dramatic, have been criticized because of the confounding effect of response bias on participants' performance elicited by the presentation of the sound on a subset of the trials. When other researchers subsequently controlled for the response criterion shifting of participants, they were unable to replicate the finding that the presentation of a synchronous sound increased the perceived brightness of a visual target (see Arieh and Marks, 2008; Marks *et al.*, 2003; Odgaard *et al.*, 2003). Note that, *vice versa*, the presentation of a visual stimulus seems to genuinely increase the perceived loudness of a simultaneously-presented sound (Lovelace *et al.*, 2003; Odgaard *et al.*, 2004; though see Marks *et al.*, 2003). One of the methods that has been used to reduce such a response criterion shift has been to make the presentation of the synchronous sound uninformative with regard to the response that participants have to make (e.g., Vroomen and de Gelder, 2000). Nevertheless, there is another potentially important type of auditory modulation of human visual decisional processes which may also influence the visual sensitivity estimated from human behavior. That is, since the synchronous auditory stimulus signifies the onset time of the visual target, the temporal uncertainty of the participants' decision with regard to the time point at which the visual target occurred may plausibly be reduced. Observers therefore need to simply monitor the visual information on which their decision will be based in a specific time window cued by the onset of the sound. As a result, the estimated visual sensitivity will be higher when the participants' temporal uncertainty regarding the visual target is reduced (Earle and Rowe, 1971). This is known as the *uncertainty reduction hypothesis* (e.g., Lippert *et al.*, 2007). Notably, this possibility reflects a kind of decisional strategy by observers, rather than a genuine perceptual enhancement resulting from the presentation of the transient sound (i.e., the transient boosting hypothesis). In addition, when an observer's uncertainty regarding the visual target is reduced, his/her performance for detection and/or discrimination tasks in terms of a fitted psychometric function should reveal that the threshold is lower and the slope is shallower (see Pelli, 1985; Petrov *et al.*, 2006).

The experiment reported here was designed to investigate whether crossmodal facilitation resulting from the presentation of a synchronous sound on visual perception could be observed under conditions in which any possible shifting of participants' response criterion was controlled for. If crossmodal facilitation were to be obtained, further analysis would be conducted in order to see which of the three possible hypotheses (*signal enhancement*, *transient boosting* or *uncertainty reduction*) provides the best account for the results. Considering that people's performance in visual detection and discrimination tasks can provide a critical clue with which to differentiate the signal enhancement hypothesis from the transient boosting hypothesis, both tasks were included in the present study.

In this experiment, the visual target consisted of a tilted Gabor patch embedded within dynamic 2D white noise at one of seven different contrast levels (see Dosher and Lu, 1998; Lu and Dosher, 1998). The signal contrast thresholds required to detect the onset of the Gabor patch and to discriminate its orientation at each noise level were estimated. A two-alternative forced-choice (2-AFC) procedure was utilized in order to eliminate the influence of any response bias that might have been induced by the presentation of the sound (see Design and Procedure section). According to the signal enhancement hypothesis, the threshold contrast for the target Gabor should be lower in the sound present condition than in the sound absent condition in both the detection and discrimination tasks. On the other hand, according to the transient boosting hypothesis, crossmodal facilitation should only be observed in the detection task. If crossmodal facilitation is observed, any change in slope between the psychometric functions in the sound present and sound absent conditions would be tested in order to exclude the possible alternative of uncertainty reduction (see Hairol and Waugh, 2010; Yu *et al.*, 2002; see Table 1 for a summary).

2. Methods

2.1. Participants

The observers consisted of the two experimenters (YCC and PCH) and two naïve participants (NP1 and NP2). The naïve participants were paid and informed of their rights according to the ethical standards laid down in the 1990 Declaration of Helsinki. All of the participants had normal or corrected-to-normal vision and hearing by self-report.

2.2. Apparatus and Stimuli

The visual stimuli were presented on a 21 inch ViewSonic CRT monitor with a resolution of 1024 × 768 pixels and a frame rate of 120 Hz. The monitor was

Table 1.
Predictions concerning the crossmodal facilitatory effects revealed in psychometric functions elicited by the presentation of a simultaneous sound on human visual detection and discrimination performance. 'V' indicates the results predicted to be observed according to each hypothesis.

Task	Predictions	Hypothesis		
		Signal enhancement	Transient boosting	Uncertainty reduction
Detection	Lower threshold	V	V	V
	Shallower slope			V
Discrimination	Lower threshold	V		V
	Shallower slope			V

controlled by a personal computer with a VSG 2/5 graphics card (Cambridge Research Systems Ltd, UK) with pseudo 15 bit contrast resolution. The monitor was gamma-corrected using luminance measurements with an optical photometer interfaced with the VSG display calibration software (Cambridge Research Systems Ltd, UK). The auditory stimuli were presented from a pair of loudspeaker cones, one located on either side of the monitor and aligned with the center of the display in terms of their elevation. The participants were tested in a dimly-lit experimental chamber.

The visual target consisted of an achromatic Gabor patch. The Gabor consisted of a 2D grating multiplied by a Gaussian envelope, tilted either 10-degree clockwise or 10-degree counterclockwise of vertical. The Gabor was defined by the equation:

$$L(x, y, \theta) = L_0 + L_0 \times c \times \cos(2\pi f(x \sin\theta + y \cos\theta) + \rho) \times \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad (1)$$

where L_0 indicates the mean luminance, c the contrast of the Gabor, defined as the Michelson contrast of the grating, f the spatial frequency of the grating, θ the orientation of the grating relative to the vertical, σ the standard deviation of the Gaussian envelope and ρ the phase of the stimuli with respect to the center of the Gaussian window. The viewing distance was set at 57 cm, and the spatial frequency was 1 cpd with 0° phase angle. The space constant (σ) was 0.8° and consequently the bandwidth of the Gabor was about 0.65 octaves.

The visual noise stimuli consisted of random pixel gray elements sampled from an identical Gaussian distribution with a mean of 128 and variances of 38 ranging from 0 to 255. Seven noise energy levels were chosen by setting different levels of the Michelson contrast (0, 0.02, 0.04, 0.08, 0.12, 0.16, 0.25, equivelent to $-\infty$, -34, -28, -22, -18, -16, -12 dB, $\text{dB} = 20 \times \log(c)$, where c is the Michelson contrast of each noise level) of the sampled noise image. The size of noise element was $0.2 \times 0.2°$ of visual angle and the size of the noise frame was $11.4 \times 11.4°$. The mean luminance of the display background was 50 cd/m^2.

The auditory stimulus consisted of a 17 ms 1000 Hz pure tone (including 3 ms fade-in and fade-out). The sound was presented at a sound pressure level (SPL) of 60 dB from free-field loudspeakers. The onset of the Gabor and the sound were measured by an oscilloscope in order to ensure that their onsets were synchronized.

2.3. Design and Procedure

Three factors, sound (present or absent), noise level (7 levels) and task (detection *vs.* discrimination) were manipulated. There were two display sequences in each trial, one of which contained the target Gabor (see Fig. 1). The participants performed a dual 2-AFC task: first, they had to judge which sequence contained the Gabor (first or second, by pressing the Z or X key on the keyboard, respectively). Second, they

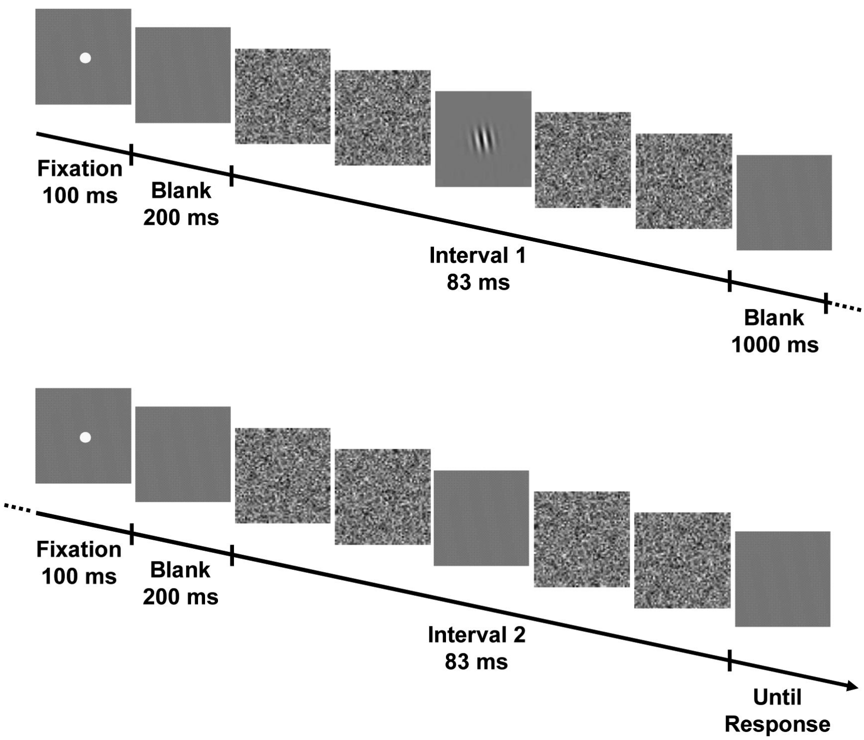

Figure 1. The sequence of stimulus displays presented in a trial in this study. Two intervals contained four noise frames, and either a visual target (a Gabor patch) or a blank frame. There were five blank frames interleaved between each of the above frames (not shown in the figure; see *Methods* for details). Note that size and time are not represented to scale in the figure. The participants performed a dual 2-AFC task, detecting which interval contained the target Gabor, and discriminating the orientation of the tilted Gabor. In this example, the correct response would have been to report ‘first interval’ and ‘counterclockwise’.

had to judge in which direction the Gabor was tilted (clockwise or counterclockwise, by pressing the left-arrow or right-arrow key, respectively). The participants were instructed to respond accurately but not speedily. Two signal contrast thresholds, one for the detection of the Gabor and the other for the discrimination of its orientation, were estimated for each noise level in the sound-present and sound-absent conditions. The method of constant stimuli with 7 different signal contrasts was used to generate the psychometric functions. The pure tone was presented in both display sequences in the sound-present condition, but in neither sequence in the sound absent condition. The presentation of the sound (i.e., a pure tone) therefore provided no information whatsoever with regards to either the detection or discrimination responses.

Each trial consisted of three parts: interval 1, interval 2 and the response. At the beginning of each interval, a fixation point (green dot, $0.25° \times 0.25°$, 100 ms) and a blank field with mean luminance that was the same as that of the background (200 ms) were presented. In order to manipulate the contrast of the Gabor and the noise independently, an interlaced presentation technique was used in each interval: In a 10 frame sequence (8 ms for each frame at a refresh rate of 120 Hz), a noise was presented in the first, third, seventh, and ninth frames. Either the target Gabor or a blank screen was presented in the sixth frame, leaving the remaining frames blank. These 10 frames constitute either one of the intervals with an 83 ms duration in a trial. Given that the 120 Hz refresh rate of the monitor lies beyond the temporal resolution of the human visual system (Kelly, 1974; Robson, 1966), participants may only perceive the noise and the target Gabor (if they were above threshold) but not the blank frames. In addition, the perceived stimulus contrast of the noise and target Gabor should be halved to reflect the temporal integration of the target (or the noise frame) and adjacent blank frame.

A blank frame with a duration of 1000 ms segregated the two display sequences. After the participants responded using the keyboard placed in front of them, the next trial started 1000 ms later. All of the conditions (7 signal contrasts $\times$ 7 noise contrasts $\times$ 2 sound conditions) were tested once in each block of trials. The order in which the trials were presented was randomized. The block was run 40 times. Testing was divided into six approximately one hour sessions conducted on different days.

3. Results

For each participant, the threshold *vs.* noise (TvN) functions in the sound present and sound absent conditions for both detection and discrimination performance were computed independently. At each noise level, the percentage correct in the detection and orientation discrimination tasks at each signal contrast level were fitted with a logistic function using a maximum likelihood procedure by using *Palamedes*

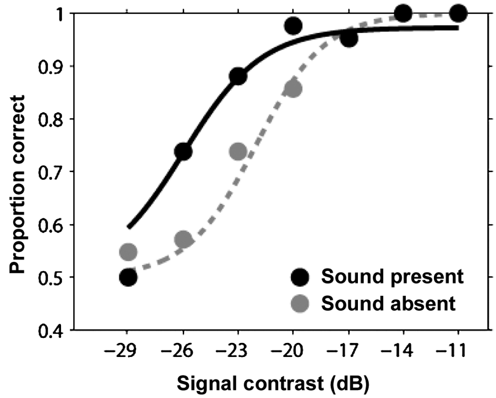

Figure 2. The psychometric functions in the sound present (black dots and solid line) and the sound absent (gray dots and dashed line) conditions of participant NP2's discrimination performance at the −22 dB noise level.

(Prins and Kingdom, 2009; see also Kingdom and Prins, 2010). The logistic function is given as:

$$L(x, \alpha, \beta) = \frac{1}{1 + \exp(-\beta(x - \alpha))}. \quad (2)$$

Thus, the psychometric function can be written as:

$$\Psi(x) = \gamma + (1 - \gamma - \lambda) \times L(x, \alpha, \beta), \quad (3)$$

where γ is the guessing rate (0.5), and λ is the lapsing rate which is set between 0 and 0.06; α is negatively related to the threshold and β is negatively related to the slope of the psychometric function. The signal contrast threshold was set where $L(x = \alpha; \alpha, \beta) = 0.5$. Given that the lapsing rate was set freely between 0 and 0.06, the threshold signal contrast would be the value corresponding to the proportion correct which is calculated by the following equation: $0.5 + (1 - 0.5 - \lambda) \times 0.5$. Accordingly, the threshold signal contrast would be the value corresponding to the proportion correct between 75% (when the lapsing rate was 0) and 72% (when the lapsing rate was 0.06). See Fig. 2 for an example of psychometric functions of NP2's discrimination performance (proportion correct against the signal contrast levels) in both sound-present and sound-absent conditions at the −22 dB noise level. The signal contrast threshold was then plotted as a function of the noise contrast in terms of dB values (i.e., the TvN function, see Fig. 3).

3.1. Results of the Grouped Data

The threshold of signal contrast (in dB) in each condition was submitted to a three-way analysis of variance (ANOVA) with the factors of task (detection *vs*. discrimination), sound (present *vs*. absent) and noise level (7 levels). The three main effects were all significant: The threshold was significantly lower in the detection than in the discrimination task ($F(1, 3) = 17.08$, $\text{MSE} = 5.34$, $p = 0.03$). The signal contrast threshold was lower in the sound present than in the sound absent condition

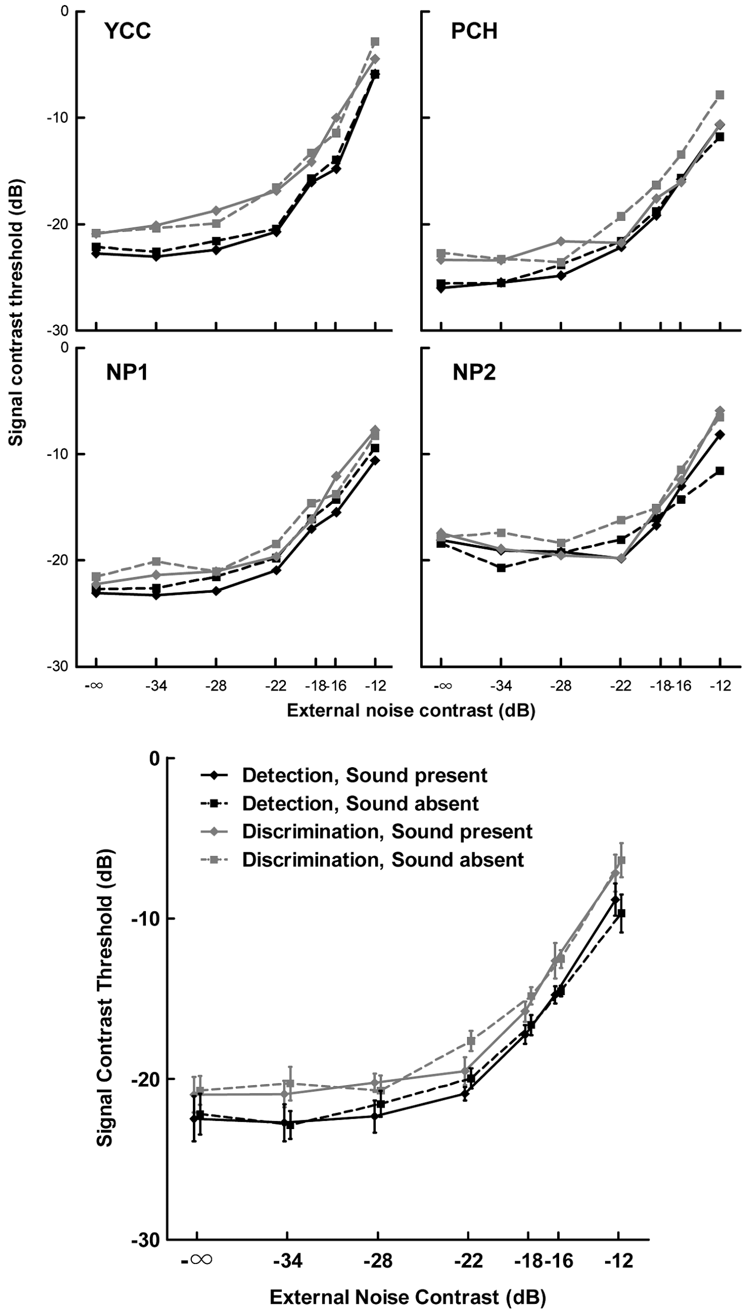

Figure 3. The signal contrast threshold was plotted as a function of the noise level (i.e., a TvN function) using a dB scale. The upper panels represent the results for the four individual participants, while the lower panel represents the average results for the four participants. The sound present and sound absent condition functions have been jittered slightly in the lower panel. The error bars indicate ±1 standard error of the means.

($F(1, 3) = 24.78$, MSE $= 0.45$, $p = 0.02$). The threshold increased with increasing noise levels ($F(6, 18) = 70.98$, MSE $= 5.81$, $p < 0.0001$). A *post-hoc* Tukey's test revealed that the signal contrast thresholds were similar in the first four noise levels (i.e., $-\infty$, -34, -28, -22 dB noise levels), and that they were all significantly different from those in the remaining three noise levels (i.e., -18, -16, -12 dB noise levels, all ps < 0.05). The signal contrast thresholds at the -18, -16, -12 dB noise levels all differed from each other (all ps < 0.05). None of the two- or three-way interaction was significant (all Fs < 1.48, ps > 0.24).

In order to verify the noise level at which the participants' signal contrast threshold for detection and discrimination tasks regarding the visual target was reduced by the presentation of a synchronous sound, planned simple main effects were conducted. The results revealed that the signal contrast threshold in the detection task was significantly lower in the sound present than in the sound absent condition at the -22 dB noise level ($F(1, 42) = 5.83$, MSE $= 0.71$, $p = 0.02$). Similarly, the signal contrast threshold in the discrimination task was also significantly lower in the sound present than in the sound absent condition at the -22 dB noise level ($F(1, 42) = 8.00$, $p = 0.007$). None of other simple main effect was significant (all Fs < 3.03, ps > 0.08). In summary, the results demonstrated that the participants' detection and discrimination sensitivity regarding the visual target were both enhanced by the presentation of a synchronous sound at the intermediate noise level of -22 dB.

3.2. The Psychometric Function at the Intermediate (−22 dB) Noise Level

In order to understand whether the crossmodal facilitatory effect in visual performance elicited by the presentation of a synchronous sound reflects reduced temporal uncertainty, the elevation of the threshold and slope between the psychometric functions from the sound absent to the sound present condition at the -22 dB noise level was compared (e.g., Hairol and Waugh, 2010; Petrov *et al.*, 2006; Yu *et al.*, 2002). A lower uncertainty should result in a lower threshold as well as a shallower slope (i.e., a lower slope value) of the psychometric function (see Pelli, 1985; Petrov *et al.*, 2006). That is, if the presentation of a synchronous sound reduced the participants' temporal uncertainty regarding the visual target (Lippert *et al.*, 2007), both the threshold and the slope of the psychometric function in the sound present condition should be systematically reduced as compared to those seen in the sound absent condition; namely, the correlation of threshold and slope elevation should be *positive* (Petrov *et al.*, 2006). On the other hand, according to the signal enhancement hypothesis, the presentation of a synchronous sound should only reduce the threshold without systematically changing the slope (Hairol and Waugh, 2010; Yu *et al.*, 2002).

The index of the threshold and slope elevation was calculated using the values (in percentage contrast) in the sound-present condition divided by those in the sound-absent condition following the methods in Hairol and Waugh's (2010) study. Values smaller than 1 indicate the threshold (or slope) was reduced, whereas values larger than 1 indicate that the threshold (or slope) was increased, by the presentation of the synchronous sound. In the detection task, the Pearson's correlation between the threshold and slope elevation due to the presentation of a synchronous sound was not significant (see Fig. 4(A), $r = -0.02$, $p = 0.98$, two-tailed). Similarly, in the discrimination task, the Pearson's correlation between the threshold and slope elevation due to the presentation of a synchronous sound was not significant either (see Fig. 4(B), $r = 0.43$, $p = 0.57$, two-tailed). Note that participant YCC's threshold in the discrimination task was increased by the presentation of the synchronous sound (i.e., the index of threshold elevation >1). When his data were excluded, the remaining three participants' results revealed a significant negative correlation between the threshold and slope elevation due to the presentation of a synchronous sound ($r = -0.998$, $p = 0.03$, two-tailed). In sum, given the fact that either null, or significant negative, correlations between threshold and slope elevations were observed in both the detection or the discrimination tasks, there is no obvious evidence revealing that the enhanced sensitivity resulting from the presentation of a synchronous sound could be attributed to temporal uncertainty reduction.

4. Discussion

In the present study, visual detection and discrimination performance were both enhanced crossmodally by the presentation of a synchronous sound at an intermediate visual noise level (the -22 dB noise level in this experiment). Furthermore, after examining the participants' psychometric functions for the visual target at the -22 dB noise levels, no evidence of the slope being systematically reduced with the threshold reduction (i.e., a positive correlation between the threshold and slope elevation) by the presentation of a simultaneous sound was observed.

In this study, the presentation of the sound was uninformative with regard to the response that a participant had to make in the dual 2-AFC task. Therefore, the crossmodal facilitation effects reported here cannot be accounted for by any form of response criterion shifting that might have been induced by the presentation of the sound (Arieh and Marks, 2008; Odgaard *et al.*, 2003). Given that the crossmodal facilitatory effect elicited by the presentation of a synchronous sound was observed in both visual detection and discrimination tasks, the transient boosting hypothesis does not provide a satisfactory explanation for our results since it predicts that crossmodal facilitation would only have been observed in participants' detection performance. Besides, the change of threshold and the slope of the psychometric functions were not positively correlated. That is, there was no evidence to suggest that the participants' uncertainty was systematically reduced along with their reduced threshold (i.e., improved sensitivity) regarding the visual target elicited by the presentation of the sound. Therefore, the uncertainty reduction hypothesis, though certainly intriguing, does not receive any univocal support from the results reported here.

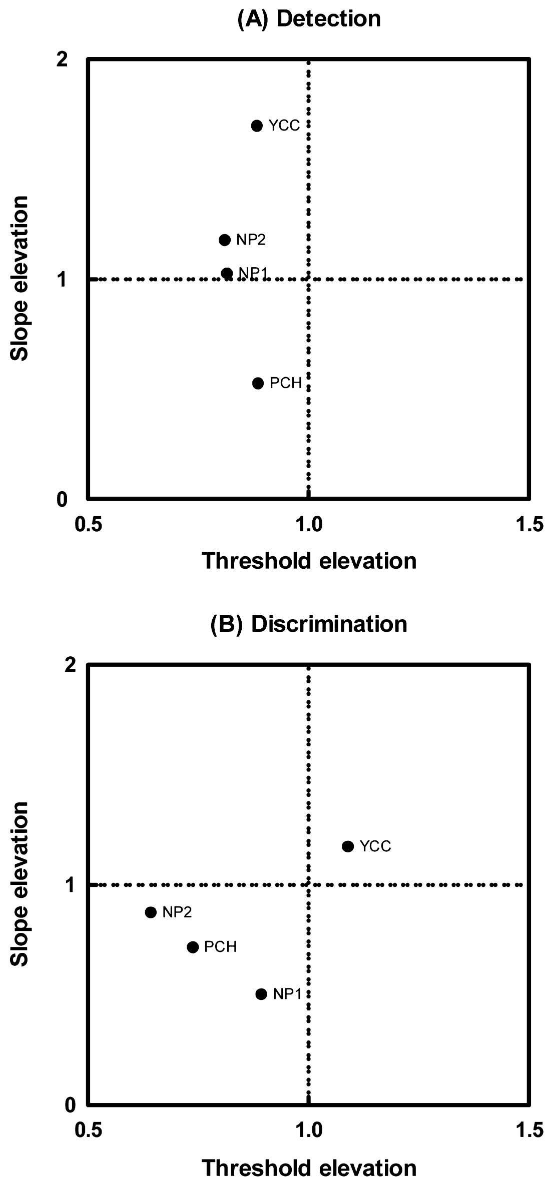

Figure 4. The threshold and slope elevation attributable to the presentation of a synchronous sound at the −22 dB noise level for (A) detection task; (B) discrimination task. Values smaller than 1 indicate the threshold (or slope) was reduced, whereas values larger than 1 the threshold (or slope) was increased, in the sound present condition as compared to the sound absent condition.

Before concluding that the crossmodal facilitation reported in the present study should be attributed to the enhancement of the visual signal by the presentation of a synchronous sound, two relevant issues need to be considered. First, a simple version of the signal enhancement hypothesis should predict that the visual signal can be enhanced by the presentation of a sound, so the signal-to-noise ratio (SNR) can be generally increased. However, the results of the present study are inconsistent with the simple signal enhancement hypothesis since crossmodal facilitation was observed only at the intermediate noise level. Second, if the presentation of the sound enhanced both the neural representation of the visual signal as well as the representation of the external noise, the crossmodal facilitatory effects should have been observed only when the noise level was low but perhaps not when it was high (see Lu and Dosher, 1998). However, the results revealed that the presentation of a synchronous sound did not induce any crossmodal facilitatory effect when the noise level was quite low (i.e., $-\infty$ and -34 dB noise levels in the present study) either. The two forms of the signal enhancement hypotheses, therefore, do not appear able to provide a comprehensive explanation for the present results.

The results reported here revealed that crossmodal facilitation was observed at intermediate noise levels. This is similar to the results reported by Ross *et al.* (2007; see also Ma *et al.*, 2009). In their study, the participants had to identify the spoken words, and so the seen lip movements were *informative* with regard to the identity of the target. The intensity of the target signal (the sound level of the spoken word) was fixed while that of background noise (auditory pink noise) was manipulated. Ross *et al.* observed the largest crossmodal facilitatory effects at a intermediate noise level. It may be noted that, in the present studty, the intermediate noise level (i.e., -22 dB) happened to be located at the turning point of TvN function (i.e., where the signal contrast threshold is about to increase). The turning point of a TvN function indicates the point of which the dominant noise for the participant's visual performance is about to transfer from the internal noise embedded in the sensory system itself to the noise originating from any external stimulation (see Lu and Dosher, 1998, 1999). In other words, this is the noise level at which the visual system is unstable in terms of either internal or external noise dominating visual information processing. One possible explanation, therefore, is that the presentation of a synchronous sound can genuinely enhance the perceived visual signal, and this effect is prone to modulate visual information processing only when the visual system itself is unstable. This notion, in turn, suggests that a certain range of noise levels needs to be tested when one plans to study crossmodal facilitation in terms of signal enhancement in human visual perception. A failure to observe any such crossmodal facilitation in terms of signal enhancement may therefore be attributable to a non-optimal noise level in the experimental setting.

In sum, the results of the present study reveal that the crossmodal facilitation of visual perception resulting from the presentation of a synchronous sound can still be observed in both visual detection and discrimination tasks when a participant's response criterion is well controlled. This crossmodal facilitation can be accounted for by the signal enhancement hypothesis only when the visual system encounters a noise level that happens to be close to its own internal noise. On the other hand, this crossmodal facilitation cannot be attributed to the presentation of the synchronous sound enhancing transient information or reducing the participants' uncertainty regarding the visual target. An important implication to emerge from these results is therefore that the visual system is susceptible to the input from another sensory modality (such as audition in the present study) when the factor determining the efficacy of visual information processing, such as the source of noise, is highly uncertain.

Acknowledgements

This research was supported by a joint project funded by the British Academy (CQ ROKK0) and the National Science Council in Taiwan (NSC 97-2911-I-002-038). Yi-Chuan Chen was supported by the Ministry of Education in Taiwan (SAS-96109-1-US-37). Su-Ling Yeh received supports from NSC 96-2413-H-002-009-MY3 and NSC 98-2410-H-002-023-MY3. YCC and PCH made an equal contribution to this study.

References

Anderson, T. S. and Mamassian, P. (2008). Audiovisual integration of stimulus transients, *Vision Res.* **48**, 2537–2544.

Arieh, Y. and Marks, L. E. (2008). Cross-modal interaction between vision and hearing: a speed-accuracy analysis, *Percept. Psychophys.* **70**, 412–421.

Bernstein, I. H., Rose, R. and Ashe, V. M. (1970). Energy integration in intersensory facilitation, *J. Exper. Psychol.* **86**, 196–203.

Bolognini, N., Frassinetti, F., Serino, A. and Ladavas, E. (2005). 'Acoustical vision' of below threshold stimuli: interaction among spatially converging audiovisual input, *Exper. Brain Res.* **160**, 273–282.

Chen, Y.-C. and Yeh, S.-L. (2008). Visual events modulated by sound in repetition blindness, *Psychon. B Rev.* **15**, 404–408.

Chen, Y.-C. and Yeh, S.-L. (2009). Catch the moment: multisensory enhancement of rapid visual events by sound, *Exper. Brain Res.* **198**, 209–219.

de Gelder, B. and Bertelson, P. (2003). Multisensory integration, perception and ecological validity, *Trends Cognit. Sci.* **7**, 460–467.

Dosher, B. A. and Lu, Z.-L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting, *Proc. Nat. Acad. Sci. USA* **95**, 13988–13993.

Downar, J., Crawley, A. P., Mikulis, D. J. and Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment, *Nat. Neurosci.* **3**, 277–283.

Driver, J. and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory specific' brain regions, neural responses, and judgments, *Neuron* **57**, 11–23.

Earle, D. C. and Lowe, G. (1971). Channel, temporal, and composite uncertainty in the detection and recognition of auditory and visual signals, *Percept. Psychophys.* **9**, 177–181.

Felleman, D. J. and van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex, *Cereb. Cortex* **1**, 1–47.

Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press, Cambridge, MA, USA.

Frassinetti, F., Pavani, F. and Lavadas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction, *Exper. Brain Res.* **147**, 332–343.

Ghazanfar, A. A. and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cognit. Sci.* **10**, 278–285.

Hairol, M. I. and Waugh, S. J. (2010). Lateral interactions across space reveal links between processing streams for luminance-modulated and contrast-modulated stimuli, *Vision Res.* **50**, 889–903.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation, *J. Exper. Psychol.* **63**, 289–293.

Kelly, D. H. (1974). Spatio-temporal frequency characteristics of color-vision mechanisms, *J. Opt. Soc. Am.* **64**, 983–990.

Kingdom, F. A. A. and Prins, N. (2010). *Psychophysics: A Practical Introduction*. Academic Press, London, UK.

Lippert, M., Logothetis, N. K. and Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound, *Brain Res.* **1173**, 102–109.

Lovelace, C. T., Stein, B. E. and Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection, *Cognit. Brain Res.* **17**, 447–453.

Lu, Z.-L. and Dosher, B. A. (1998). External noise distinguishes attention mechanisms, *Vision Res.* **38**, 1183–1198.

Lu, Z.-L. and Dosher, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise, *J. Optic. Soc. Amer. A* **16**, 764–778.

Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J. and Parra, L. C. (2009). Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space, *PLoS ONE* **4**, e4638. DOI: 10.1371/journal.pone.0004638.

Marks, L. E., Ben-Artzi, E. and Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination, *Intl J. Psychophysiol.* **50**, 125–145.

Ngo, M. K. and Spence, C. (2010a). Crossmodal facilitation of masked visual target identification, *Atten. Percept. Psychol.* **72**, 1938–1947.

Ngo, M. K. and Spence, C. (2010b). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli, *Atten. Percept. Psychol.* **72**, 1654–1665.

Nickerson, R. S. (1973). Intersensory facilitation of reaction time: energy summation or preparation enhancement? *Psychol. Rev.* **80**, 489–509.

Noesselt, T., Bergmann, D., Hake, M., Heinze, H.-J. and Fendrich, R. (2008). Sound increases the saliency of visual events, *Brain Res.* **1220**, 157–163.

Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H.-J. and Driver, J. (2010). Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity, *J. Neurosci.* **30**, 13609–13623.

Odgaard, E. C., Arieh, Y. and Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: sensory interaction versus response bias, *Percept. Psychophys.* **65**, 123–132.

Odgaard, E. C., Arieh, Y. and Marks, L. E. (2004). Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation, *Cognit. Affect. Behav. Neurosci.* **4**, 127–132.

Olivers, C. N. L. and van der Burg, E. (2008). Bleeping you out of the blink: sound saves vision from oblivion, *Brain Res.* **1242**, 191–199.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination, *J. Optic. Soc. Amer. A* **2**, 1508–1532.

Petrov, Y., Verghese, P. and McKee, S. P. (2006). Collinear facilitation is largely uncertainty reduction, *J. Vision* **6**, 170–178.

Prins, N. and Kingdom, F. A. A. (2009). *Palamedes: Matlab Routines for Analyzing Psychophysical Data*. www.palamedestoolbox.org.

Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception, *Behav. Brain Sci.* **22**, 341–423.

Robson, J. G. (1966). Spatial and temporal contrast-sensitivity functions of the visual system, *J. Optic. Soc. Amer.* **56**, 1141–1142.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C. and Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments, *Cereb. Cortex* **17**, 1147–1153.

Spence, C. (2010). Crossmodal attention, *Scholarpedia* **5**, 6309. DOI: 10.4249/scholarpedia.6309.

Spence, C. and Ngo, M.-C. (in press). Does attention or multisensory integration explain the cross-modal facilitation of masked visual target identification? in: *The New Handbook of Multisensory Processing*, B. E. Stein (Ed.). MIT Press, Cambridge, MA, USA.

Stein, B. E., London, N., Wilkinson, L. K. and Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis, *J. Cognit. Neurosci.* **8**, 497–506.

van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W. and Theeuwes, J. (2008). Pip and pop: non-spatial auditory signals improve spatial visual search, *J. Exper. Psychol. Human* **34**, 1053–1065.

Vroomen, J. and de Gelder, B. (2000). Sound enhances visual perception: crossmodal effects of auditory organization on vision, *J. Exper. Psychol. Human* **26**, 1583–1590.

Watkins, S., Shams, L., Josephs, O. and Rees, G. (2007). Activity in human V1 follows multisensory perception, *Neuroimage* **37**, 572–578.

Werner, S. and Noppeney, U. (2010). Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex, *J. Neurosci.* **30**, 7434–7446.

Yu, C., Klein, S. A. and Levi, D. M. (2002). Facilitation of contrast detection by cross-oriented surround stimuli and its psychophysical mechanisms, *J. Vision* **2**, 243–255.