Visual events modulated by sound in repetition blindness

YI-CHUAN CHEN AND SU-LING YEH
National Taiwan University, Taipei, Taiwan

Repetition blindness (RB; Kanwisher, 1987) is the term used to describe people's failure to detect or report an item that is repeated in a rapid serial visual presentation (RSVP) stream. Although RB is, by definition, a visual deficit, whether it is affected by an auditory signal remains unknown. In the present study, we added two sounds *before*, *simultaneous with*, or *after* the onset of the two critical visual items during RSVP to examine the effect of sound on RB. The results show that the addition of the sounds effectively reduced RB when they appeared at, or around, the critical items. These results indicate that it is easier to perceive an event containing multisensory information than unisensory ones. Possible mechanisms of how visual and auditory information interact are discussed.

If a wind shear is encountered by an airplane, the captain will turn on the "fasten seatbelt" sign, the onset of which is usually accompanied by a sound to avoid this sign's being ignored. Is an event containing multisensory information easier to perceive than a unisensory one? In this study, we examined whether a sound can ameliorate the unobserved visual event by adopting the repetition blindness (RB) paradigm.

RB is the failure to report the second occurrence of a repeated item in a rapid serial visual presentation (RSVP). For example, the sentence "I prefer green *tea* for *tea* time" may be misreported as "I prefer green *tea* for time," even though such an error leads to an ungrammatical sentence. RB is considered a general phenomenon in visual perception, since it has been demonstrated with many different visual stimuli, including English words (Harris & Morris, 2001), Chinese characters (Yeh & Li, 2004), colors and forms (Kanwisher, Driver, & Machado, 1995), and pictures (Bavelier, 1994).

By definition, RB is a visual deficit. Nevertheless, it is still unknown whether the factors affecting RB are confined to the visual modality or whether, instead, stimuli from other modalities can also influence RB. To date, there have been no studies in which this issue has been directly examined, although some previous studies have provided a few insights. For example, Kanwisher and Potter (1989) had their participants either see or hear sentences and found RB but not repetition deafness. They concluded that RB is a modality-specific phenomenon that is confined to vision (but see Miller & MacKay, 1994). Soto-Faraco and Spence (2002) found no evidence for a cross-modal repetition deficit when the two critical items, C1 and C2, were presented in different modalities (e.g., visual C1 and auditory C2). This implies that the source of modality could be taken as a part

of episodic representations, so that the factors affecting RB might be restricted to *vision* only.

Despite the foregoing, other studies have suggested that the perception of visual events can be affected by auditory input. For example, two visual streams crossing each other can be perceived as streaming or bouncing, depending on whether a sound is added (Sekuler, Sekuler, & Lau, 1997; Shimojo, Watanabe, & Scheier, 2001). Another example is the attentional blink (AB) for the visual target when it is preceded by an auditory target (Arnell & Larson, 2002). The proposed mechanism is that the auditory stimulus may occupy attentional resources or distract attention away from the visual stimuli (Arnell & Larson, 2002; Shimojo et al., 2001; but see Duncan, Martens, & Ward, 1997). By inference from these studies, it seems that adding sounds to an RSVP stream should *suppress* the visual process, due to lack of attention.

Recalling the seatbelt example mentioned earlier, however, makes us wonder whether there is a third possibility: Sound enhances the visibility of the originally unobserved visual event (i.e., enhances the visibility of C2 when it suffers from RB). To examine this, we added two different beep sounds (S1 and S2) to the visual stream.¹ If the added sounds can enhance the visibility of C1 and C2 in RSVP, they should be perceived more accurately, thus lessening RB. The onset time of S2 was manipulated in order to probe the effective time window of the sound facilitatory effect.

METHOD

Participants

Five groups of 20 undergraduates at National Taiwan University participated in each of the five sound conditions in exchange for course credit. They were all native speakers of Mandarin and were naive concerning the purpose of the experiment. All of the participants had normal or corrected-to-normal vision and hearing.

S.-L. Yeh, suling@ntu.edu.tw

Figure 1. (A) Examples of the rapid serial visual presentation sequence. In the repeated trial, C1 and C2 were the same character (銘). In the unrepeated trial, all the displays were the same as those in the repeated trial, except that another character (輔) was substituted for C1. The control trial contained only two characters. (B) The five sound conditions. The time scales of panels A and B are matched. C1, Critical Item 1; C2, Critical Item 2; F, filler character; S1, Sound 1, 500 Hz; S2, Sound 2, 800 Hz.

Stimuli

The stimuli were presented on a 15-in. calibrated EIZO color monitor and were controlled by a personal computer. The refresh rate of the monitor was set at 70 Hz. The participants sat at a viewing distance of 60 cm from the monitor in a dimly lit chamber.

Each trial consisted of seven items in an RSVP sequence, with three Chinese characters and four symbols all in white and presented one at a time for 43 msec in the center of a black background. All the Chinese characters that were selected were ones commonly used. The three characters included one filler character (F) and two critical characters (C1 and C2). To avoid any unwanted priming effects, the characters in each trial were not orthographically similar, except for repeated items, and did not have any obvious semantic relationship. Furthermore, the pronunciations did not sound as meaningful two- or three-character words, nor did they rhyme.

The F and C2 were in the Chia font and subtended a visual angle of $1.15^\circ \times 1.24^\circ$ (width $\times$ height). C1 was in the Fong font ($1.53^\circ \times 1.43^\circ$). C1 and C2 were chosen to be different in both font and size in order to avoid the possible confounding of visual masking caused by physically identical characters appearing at the same position. A fixation cross ($1.15^\circ \times 1.24^\circ$) was shown at the beginning and

end of each trial. Symbols (*) in each trial were randomly selected from a set of 30 signs, such as &, ∇ , $\equiv$, and \$, without repeated symbols occurring in the same trial. The sizes of the 30 symbols ranged from $1.21^\circ \times 0.80^\circ$ to $1.43^\circ \times 1.24^\circ$. In all the trials, the presented order in RSVP was *, C1, F, *, C2, * (see Figure 1A). The auditory stimuli (S1 and S2; 500 and 800 Hz, respectively) appeared simultaneously with the onsets of two visual items and lasted for 10 msec each.

Design

The two factors of sound condition (five conditions of beep sounds accompanying C1 and C2) and repetition (repeated or unrepeated) were manipulated. The five sound conditions were as follows. The *no-sound* condition was solely a visual presentation and served as the baseline for comparison. In the remaining four conditions, S1 was always presented at the onset of C1, and S2 was presented before C2 (-86 msec), at the *onset* of C2, at the *offset* of C2, or after C2 ($+86$ msec). Except for the onset time of S2, other items were the same in the four conditions, except for the no-sound condition (Figure 1B). Different participants were assigned to these five conditions.

The factor of repetition was in a yoked design; repeated trials in one version for half of the participants were switched to be unrepeated trials in the other version for the other half of the participants, and thus each participant saw a given C2 once only. In repeated trials, C1 and C2 were identical characters. In unrepeated trials, another character was substituted for C1. The character frequencies between the repeated and the unrepeated C1s did not differ, nor did their stroke counts differ ($ps > .05$; based on the corpus of Tsai, 1996). Each version consisted of 20 critical trials; half were repeated trials, and the other half were unrepeated.

There were also 10 control trials that contained two different characters in RSVP. The purpose of adding these control trials was to reduce the proportion of repeated trials, thus discouraging the participants from strategically filling in three characters when RB did occur. Thus, there were 30 trials in total, and their presentation order was completely randomized for each participant.

Procedure

The participants initiated a trial by pressing the space bar. At the start of each trial, a warning sound (1000 Hz) was presented for 150 msec, followed by the fixation cross in the center of the screen for 500 msec. The RSVP sequence was then presented from the same location as the fixation cross. After the sequence, the fixation cross reappeared and remained on the screen until the participant pressed the space bar to initiate the next trial. The participants were asked to write down all of the characters on each trial. They were informed that some trials contained three characters, whereas some contained only two, and if a character appeared twice, they were to write it down twice. No feedback was given to the participants concerning the accuracy of their responses during the experiment.

A practice session consisting of seven trials was conducted before the main experiment, four trials containing three characters and the others containing two characters. These trials were constructed with no repeated characters, and the characters were not used in the main experiment. All the participants fulfilled the requirement of reporting at least three trials correctly (one of which had to be a three-character trial) before proceeding to the formal experiment.

RESULTS

Accuracy was calculated on the basis of the percentage of trials on which both C1 and C2 were reported correctly,

regardless of the order of report (Park & Kanwisher, 1994). See Figure 2 for the results. A two-way ANOVA was conducted on the factors of sound condition and repetition.

The ANOVA showed that the main effects of both sound condition [$F(4,95) = 5.470$, $MS_e = 0.085$, $p < .001$] and repetition [$F(1,95) = 55.774$, $MS_e = 0.024$, $p < .0001$] were significant. The participants responded less accurately for repeated than for unrepeated trials (66.8% vs. 83.1%), indicating an overall RB effect.

The interaction between sound condition and repetition was significant [$F(4,95) = 2.979$, $MS_e = 0.024$, $p < .05$]. The analysis of the simple main effects showed that the accuracy for the five sound conditions was significantly different for repeated trials [$F(4,190) = 8.113$, $MS_e = 0.055$, $p < .0001$], but not for unrepeated trials [$F(4,190) = 1.741$, $p > .1$]. A post hoc Tukey test showed that, for repeated trials, the accuracy of the -86 -msec, onset, and offset conditions was significantly higher than that of the no-sound and $+86$ -msec conditions [$q(190,5).05 = 3.860$, $MS_e = 0.055$, $C(.05) = 20.2\%$]. RB effects for each sound condition were strong for the no-sound [$F(1,95) = 26.240$, $MS_e = 0.024$, $p < .0001$, $d = 0.88$]² and $+86$ -msec [$F(1,95) = 28.381$, $p < .0001$, $d = 0.91$] conditions, weak for the -86 -msec [$F(1,95) = 4.198$, $p < .05$, $d = 0.38$] and offset [$F(1,95) = 5.080$, $p < .05$, $d = 0.39$] conditions, and only marginally significant for the onset condition [$F(1,95) = 3.789$, $p = .054$, $d = 0.33$].

The error pattern for control trials was also analyzed to see whether C2 was reported twice in these trials—for example, whether 乖, 裕, and 裕 were reported when, in fact, only 乖 and 裕 were presented. The percentage of trials on which the participants reported C2 twice were 1%, 0%, 0.5%, 0%, and 1% for the five conditions, respectively [$F(4,95) = 0.725$, $p > .5$]. This indicates that the enhanced accuracy of repeated trials in the -86 -msec, onset, and offset conditions should not be attributed to this guessing bias.

Figure 2. Accuracy (percent correct when both C1 and C2 were reported correctly) and standard errors for repeated trials (gray bars) and unrepeated trials (white bars) under the five sound conditions.

DISCUSSION

In this study, we found that RB can be reduced when two critical items are accompanied by different sounds. We replicated the conventional RB effect in the no-sound condition, while observing a reduced RB effect in the -86 -msec, onset, and offset conditions. The $+86$ -msec condition revealed a magnitude of RB similar to that in the no-sound condition.

Regarding the three possibilities mentioned earlier of how sound would affect RB (*independent, suppressed, or enhanced*), our results favor the *enhanced* account, since the addition of the sounds reduced the magnitude of the RB effect by facilitating character identification. Even though processing the RSVP stream is demanding enough to prevent influences from peripheral sounds (Santangelo, Belardinelli, & Spence, 2007), in this study, the task-irrelevant sounds were not filtered out but, rather, still affected the demanding processing of the central stream. However, unlike the cross-modal AB effect (Arnell & Larson, 2002), we demonstrated that the presence of the sound enhanced, rather than suppressed, visual perception. Indeed, there is mounting evidence showing sound facilitatory effects in visual detection (e.g., Bolognini, Frassinetti, Serino, & Lavavas, 2005; McDonald, Teder-Saljarvi, & Hillyard, 2000) and visual filling in (Sheth & Shimojo, 2004). Nevertheless, these studies adopted a single LED light or a color patch as the visual target. Our results thus extend the sound facilitatory effect shown in the literature in that it can be obtained not only in the simple display, but also in a rapid and complex visual presentation (i.e., RSVP), which is more similar to our daily visual world.

Bolognini et al. (2005) reported that sound could enhance the sensitivity of visual perception, whereas Park and Kanwisher (1994) found that C2 sensitivity was reduced when it suffered from RB. Combining these two findings, we suggest that the effect of added sounds in the present study was to boost the lowered sensitivity to the C2, making it more likely to be immune from RB. These results imply that multisensory stimuli have a better opportunity to gain visual awareness (Kanwisher, 2001) than do unisensory ones. Our results may thus have important implications for the improvement of deficits in the perceiving of sequential events for right-parietal-damaged patients (Battelli, Pascual-Leone, & Cavanagh, 2007).

The processing of visual items in an RSVP stream can be understood within the type–token binding framework (e.g., Bowman & Wyble, 2007; Chun, 1997; Kanwisher, 1987), which suggests that each visual event activates two kinds of representations: type and token. A type is a long-term representation used for recognition, whereas a token is an episodic representation used to specify spatial and temporal information. To successfully report a visual event, the activated type needs to link to its token, a process called *token individuation*. Kanwisher (1987) suggested that RB results from the failure to individuate the second token of the same activated type, due to the difficulty of binding one type with two tokens within a short interval. This hypothesis has gained much support from

behavioral (e.g., Chun, 1997; Morris & Harris, 1999; Park & Kanwisher, 1994) and event-related potential (Schen- dan, Kanwisher, & Kutas, 1997) studies. Attention is required either to facilitate the activation of the type node (Bowman & Wyble, 2007) or to serve as a supplementary component for type–token binding (Chun, 1997).

The cross-modal facilitatory effect in RB we found here can be explained as follows. The presentation of the sounds may covertly orient participants' attention to focus on the upcoming visual items (McDonald et al., 2000), or it may boost the alertness of participants (Robertson, Mattingley, Rorden, & Driver, 1998); namely, sounds can modulate visual attention in the temporal domain. Another possibility is that sounds help temporal segregation of the RSVP stream, since they are dominant in the temporal domain; not only does audition have its advantage in the representations of temporal structure (Guttman, Gilroy, & Blake, 2005; Welch, DuttonHurt, & Warren, 1986), but also the discontinuous nature of sounds makes them more dominant than visual streams (Shimojo & Shams, 2001). C1 and C2 can be more distinguishable and, thus, suffer less RB if they are segregated into two groups, rather than appearing within the same group in the spatial domain (Chun & Cavanagh, 1997; Kanwisher & Potter, 1989), so that they are segregated in the temporal domain (Morein-Zamir, Soto-Faraco, & Kingstone, 2003). In this case, the temporal structure of the visual stream and the sounds should be coordinated, which may occur after the stimuli within each modality have been fully processed. A further possibility is visual–auditory integration's occurring at an early perceptual stage. The different sounds adopted in this study may have integrated with the repeated characters and modified their similarity, which is analogous to the demonstration of the reduction of RB by making C1 and C2 more visually distinct (e.g., Bavelier, 1994; Chun, 1997). Although the *temporal coincident rule* with a particular time window of cross-modal integration (Stein & Meredith, 1993) may account not only for the onset condition, but also for the -86 -msec and offset conditions, this integration account, however, presupposes a close connection between S2 and C2. Yet there is no a priori reason why this should be the case.

Previous studies have shown that sound can improve attentional deficits for spatial neglect patients through a nonspecific alertness effect and a spatially specific cross-modal integration effect (Frassinetti, Pavani, & Lavavas, 2002; Van Vleet & Robertson, 2006). Similarly, all of the above-mentioned mechanisms may be involved in the cross-modal facilitatory effect in RB, but each may favor a certain time window. For example, orienting requires time to develop (Nakayama & Mackeben, 1989), so the attentional boost account may best explain the -86 -msec condition. On the other hand, temporal coincidence is critical for the cross-modal integration account (Bolognini et al., 2005). To tease these possible mechanisms apart, however, manipulations of factors other than the temporal relationship between sounds and visual items are required, especially for the cross-modal integration account. For example, the effect we observed here is enhancement, rather than a new representation, as in the McGurk ef-

fect in speech perception (McGurk & MacDonald, 1976). Systematic investigations about whether pitch difference of sounds is critical may help to clarify these issues.

Nevertheless, regardless of mechanism, we have shown in this study that RB can be reduced by adding task-irrelevant sounds in RSVP. Multisensory stimuli, as compared with unisensory stimuli, have a better chance of breaking through visual awareness and making invisible stimuli visible. Hearing a sound accompanying the "fasten seatbelt" sign may, indeed, serve as a better warning signal than does seeing the visual notice alone.

AUTHOR NOTE

This research was supported by Grants NSC93-2413-H-002-017 and NSC94-2752-H-002-008-PAE from the National Science Council of Taiwan. We thank Catherine Caldwell-Harris, Anne Hillstrom, Steven Most, Shinsuke Shimojo, and Charles Spence for their helpful comments on earlier drafts of the manuscript. Correspondence concerning this article should be addressed to S.-L. Yeh, Department of Psychology, National Taiwan University, No. 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan, 106 (e-mail: suling@ntu.edu.tw).

REFERENCES

- ARNELL, K. M., & LARSON, J. M. (2002). Cross-modality attentional blinks without preparatory task-set switching. *Psychonomic Bulletin & Review*, **9**, 497-506.
- BATTELLI, L., PASCUAL-LEONE, A., & CAVANAGH, P. (2007). The "when" pathway of the right parietal lobe. *Trends in Cognitive Sciences*, **11**, 204-210.
- BAVELIER, D. (1994). Repetition blindness between visually different items: The case of pictures and words. *Cognition*, **51**, 199-236.
- BOLOGNINI, N., FRASSINETTI, F., SERINO, A., & LADAVAS, E. (2005). "Acoustical vision" of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, **160**, 273-282.
- BOWMAN, H., & WYBLE, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review*, **114**, 38-70.
- CHUN, M. M. (1997). Types and tokens in visual processing: A double dissociation between the attentional blink and repetition blindness. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 738-755.
- CHUN, M. M., & CAVANAGH, P. (1997). Seeing two as one: Linking apparent motion and repetition blindness. *Psychological Science*, **8**, 74-79.
- DUNCAN, J., MARTENS, S., & WARD, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, **387**, 808-810.
- FRASSINETTI, F., PAVANI, F., & LADAVAS, E. (2002). Acoustical vision of neglected stimuli: Interaction among spatially converging audiovisual inputs in neglect patients. *Journal of Cognitive Neuroscience*, **14**, 62-69.
- GUTTMAN, S. E., GILROY, L. A., & BLAKE, R. (2005). Hearing what the eyes see. *Psychological Science*, **16**, 229-235.
- HARRIS, C. L., & MORRIS, A. L. (2001). Identity and similarity in repetition blindness: No cross-over interaction. *Cognition*, **81**, 1-40.
- KANWISHER, N. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, **27**, 117-143.
- KANWISHER, N. (2001). Neural events and perceptual awareness. *Cognition*, **79**, 89-113.
- KANWISHER, N., DRIVER, J., & MACHADO, L. (1995). Spatial repetition blindness is modulated by selective attention to color or shape. *Cognitive Psychology*, **29**, 303-337.
- KANWISHER, N., & POTTER, M. C. (1989). Repetition blindness: The effects of stimulus modality and spatial displacement. *Memory & Cognition*, **17**, 117-124.
- MCDONALD, J. J., TEDER-SÄLEJÄRVI, W. A., & HILLYARD, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, **407**, 906-908.
- MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.
- MILLER, M. D., & MACKAY, D. G. (1994). Repetition deafness: Repeated words in computer-compressed speech are difficult to encode and recall. *Psychological Science*, **5**, 47-51.
- MOREIN-ZAMIR, S., SOTO-FARACO, S., & KINGSTONE, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, **17**, 154-163.
- MORRIS, A. L., & HARRIS, C. L. (1999). A sublexical locus for repetition blindness: Evidence from illusory words. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 1060-1075.
- NAKAYAMA, K., & MACKEBEN, M. (1989). Sustained and transient components of focal attention. *Vision Research*, **29**, 1631-1647.
- PARK, J., & KANWISHER, N. (1994). Determinants of repetition blindness. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 500-519.
- ROBERTSON, I. H., MATTINGLEY, J. B., RORDEN, C., & DRIVER, J. (1998). Phasic alerting of neglect patients overcomes their spatial deficit in visual awareness. *Nature*, **395**, 169-172.
- SANTANGELO, V., BELARDINELLI, M. O., & SPENCE, C. (2007). The suppression of reflexive visual and auditory orienting when attention is otherwise engaged. *Journal of Experimental Psychology: Human Perception & Performance*, **33**, 137-148.
- SCHENDAN, H. E., KANWISHER, N. G., & KUTAS, M. (1997). Early brain potentials link repetition blindness, priming and novelty detection. *NeuroReport*, **8**, 1943-1948.
- SEKULER, R., SEKULER, A. B., & LAU, R. (1997). Sound alters visual motion perception. *Nature*, **385**, 308.
- SHETH, B. R., & SHIMOJO, S. (2004). Sound-aided recovery from and persistence against visual filling-in. *Vision Research*, **44**, 1907-1917.
- SHIMOJO, S., & SHAMS, L. (2001). Sensory modalities are not separate modalities: Plasticity and interactions. *Current Opinion in Neurobiology*, **11**, 505-509.
- SHIMOJO, S., WATANABE, K., & SCHEIER, C. (2001). The resolution of ambiguous motion: Attentional modulation and development. In J. Braun, C. Koch, & J. L. Davis (Eds.), *Visual attention and control circuits* (pp. 243-264). Cambridge, MA: MIT Press.
- SOTO-FARACO, S. (2000). An auditory repetition deficit under low memory load. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 264-278.
- SOTO-FARACO, S., & SPENCE, C. (2002). Modality-specific auditory and visual temporal processing of deficits. *Quarterly Journal of Experimental Psychology*, **55A**, 23-40.
- STEIN, B. E., & MEREDITH, M. A. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.
- THALHEIMER, W., & COOK, S. (2002). *How to calculate effect sizes from published research articles: A simplified methodology*. Retrieved May 22, 2006 from work-learning.com/effect_sizes.htm.
- TSAI, C. H. (1996). *Frequency and stroke counts of Chinese characters*. Retrieved August 31, 1999, from technology.chtsai.org/charfreq/.
- VAN VLEET, T. M., & ROBERTSON, L. C. (2006). Cross-modal interactions in time and space: Auditory influence on visual attention in hemispatial neglect. *Journal of Cognitive Neuroscience*, **18**, 1368-1379.
- WELCH, R. B., DUTTONHURT, L. D., & WARREN, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics*, **39**, 294-300.
- YEH, S. L., & LI, J. L. (2004). Sublexical processing in visual recognition of Chinese characters: Evidence from repetition blindness for subcharacter components. *Brain & Language*, **88**, 47-53.

NOTES

1. When two repeated sounds were presented within a short interval, the performance of the second sound either was facilitated (Soto-Faraco & Spence, 2002) or deteriorated (Soto-Faraco, 2000), as compared with un-repeated sounds. Hence, we avoided using repeated sounds in our design.
2. The effect size was calculated by using Cohen's *d* (Thalheimer & Cook, 2002).