



Temporal sequence discrimination within and across senses: do we really hear what we see?

Daniel Bratzke¹ · Rolf Ulrich¹

Received: 25 March 2019 / Accepted: 14 September 2019 / Published online: 20 September 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Previous evidence suggests that people “hear” visual stimuli when encoding temporal information. This suggestion is based on the observation that auditory distractor information can strongly affect discrimination performance for visual temporal sequences. The present study aimed to replicate and extend this finding by investigating sequence discrimination within and across the two modalities. In two experimental series, participants judged whether two subsequently presented temporal sequences, a standard sequence followed by a comparison sequence, were identical or not. In Experimental Series A, irrelevant distractor information was presented simultaneously with the standard sequence. In Series B, the distraction appeared in the retention interval between the standard sequence and the comparison sequence. The results showed that auditory distraction impaired performance irrespective of whether the target sequences were auditory or visual, whereas visual distraction only impaired the discrimination of visual target sequences. Furthermore, auditory distraction was always at least as effective as visual distraction, irrespective of standard modality. Generally, discrimination performance was much better for auditory than for visual sequences. Overall, the present results are consistent with the idea that people code visual temporal information in the auditory modality. Moreover, the present study also suggests that such cross-modal interference effects should be interpreted cautiously with respect to their underlying timing mechanism because of the basic differences in temporal sensitivity between the two modalities.

Keywords Cross-modal timing · Rhythm discrimination · Cross-modal encoding · Time perception

Introduction

In the processing of temporal information, a dominance of auditory over visual stimulation can usually be observed (e.g., Bausenhardt et al. 2014; Repp and Penel 2002; Burr et al. 2009; Fendrich and Corballis 2001; Ortega et al. 2014). This dominance supports the view that the auditory modality is most appropriate for the perception of temporal stimulation (modality-appropriateness hypothesis, e.g., Welch and Warren 1980). A study by Guttman et al. (2005) tested this hypothesis, asking whether people might even “hear” the temporal structure of visual information (cross-modal encoding hypothesis, see also Bratzke et al. 2012; Kanai

et al. 2011). In short, these authors observed that the encoding of visual temporal sequences was strongly disrupted by concurrent auditory distractor information. They concluded that visual temporal information is obligatorily and automatically encoded in the auditory domain.

Guttman et al.’s (2005) study comprised three experiments which all shared the basic procedure and the same participants. In each trial, two visual temporal sequences (a standard and a comparison sequence consisting of contrast reversals of gabor patches) of about 2-s duration were presented and participants had to discriminate whether the two sequences were the same or different. In Experiment 1, the two visual sequences were presented together with task-irrelevant auditory information (sequences of 2-ms auditory clicks) which could be either congruent or incongruent with the visual sequence. Here, discrimination performance was strongly impaired when the irrelevant distractor information was incongruent, and improved when it was congruent, compared to a no-sound control condition. In Experiment 2, the two visual sequences were presented only with

✉ Daniel Bratzke
daniel.bratzke@uni-tuebingen.de

Rolf Ulrich
rolf.ulrich@uni-tuebingen.de

¹ Department of Psychology, University of Tübingen,
Schleichstr. 4, 72076 Tübingen, Germany

Table 1 Overview of the experiments in the present study

Experiment	Experimental Series A			Experimental Series B		
	A1	A2	A3	B1	B2	B3
Target modality	Visual	Auditory	Auditory	Auditory/visual	Auditory/visual	Auditory/visual
Distractor modality	Auditory/visual	Auditory/visual	Auditory/visual	Auditory/visual	Auditory/visual	Auditory/visual
Target type	Shape	Frequency	Frequency	Waveform/shape	Waveform/shape	Waveform/shape
Distractor type	Frequency/color	Frequency/color	Waveform/shape	Frequency/color	Frequency/color	Frequency/color
Distraction	Encoding	Encoding	Encoding	Retention	Retention	Retention
Catch trials	–	25%	25%	33.3%	33.3%	33.3%
Number of frames	17	17	17	17	9	9
Frame duration (ms)	100	100	140	100	100	180
Number of changes	8	8	8	8	4	4

incongruent distractor information, both in the auditory and the visual modality. Specifically, the visual distraction consisted of changes in the nature of the visual change defining the sequence (contrast reversal or 90° orientation change of gabor patches) between the two sequences, whereas the auditory distractor information was the same as in Experiment 1. The results showed a much stronger impairment of discrimination performance due to the auditory than visual distraction. In Experiment 3, again only incongruent auditory distractors were used and these were presented together with the standard sequence, the comparison sequence, or both sequences. The results showed that comparison performance was most strongly affected by the auditory distractor information when it was presented together with the standard sequence, that is during the encoding of the standard sequence.

In our view, Guttman et al.'s (2005) results cannot be interpreted as clear evidence for the cross-modal encoding hypothesis. First, their between-sequences manipulation of the visual distractors in Experiment 2 is not comparable with the manipulation of auditory distractors within the standard sequence in their other experiments (for a similar argument, see also Grahn 2012). It is possible, and in our view also very likely, that concurrent distraction as in the latter case is much more effective in disrupting or interfering with perception and encoding of the target information than the between-sequences distraction in the former case. Second, in the result section of their Experiment 1, Guttman et al. refer to a follow-up experiment, in which participants had to discriminate auditory sequences in isolation, with congruent visual information, or with incongruent visual information. According to the authors, visual distraction had virtually no effect on the discrimination of the auditory sequences. The authors, however, do not report whether and, if so, how they controlled for the possibility that participants simply ignore the irrelevant information, which should be much easier for

visual than for auditory stimulation (e.g., by closing the eyes or looking away from the stimulation).

In the present study, we, therefore, set out to more rigorously test the cross-modal encoding hypothesis by employing the same basic paradigm as Guttman et al. (2005) but using (1) more comparable auditory and visual distractors, (2) varying target and distractor modalities (auditory vs. visual) orthogonally, and (3) controlling for the possibility that participants block out the irrelevant information. The present study comprised two series of experiments. In all experiments, participants had to discriminate two successive temporal sequences, which could be the same or different. In a first series of experiments (Experiments A1–3), the distractor information was presented together with the standard sequence (for example sequences, see Fig. 1). In a second series (Experiments B1–3), the distraction was presented in a retention interval between the standard and the comparison sequence. To prevent participants from blocking out the task-irrelevant information, catch trials were included in both series. Different types of visual and auditory stimuli were used to define temporal target and distractor sequences. Additionally, across the experiments, we varied the length of the target sequence (i.e., the number of frames), the duration of each frame, and the number of changes appearing within each sequence. In contrast to Guttman et al.'s study, we also employed conditions in which the irrelevant information was presented within the same modality as the target information. Table 1 gives an overview of the experiments in the present study. According to the cross-modal encoding hypothesis, especially the presence of auditory distractors should impair discrimination performance, and this impairment should be relatively independent of target modality.

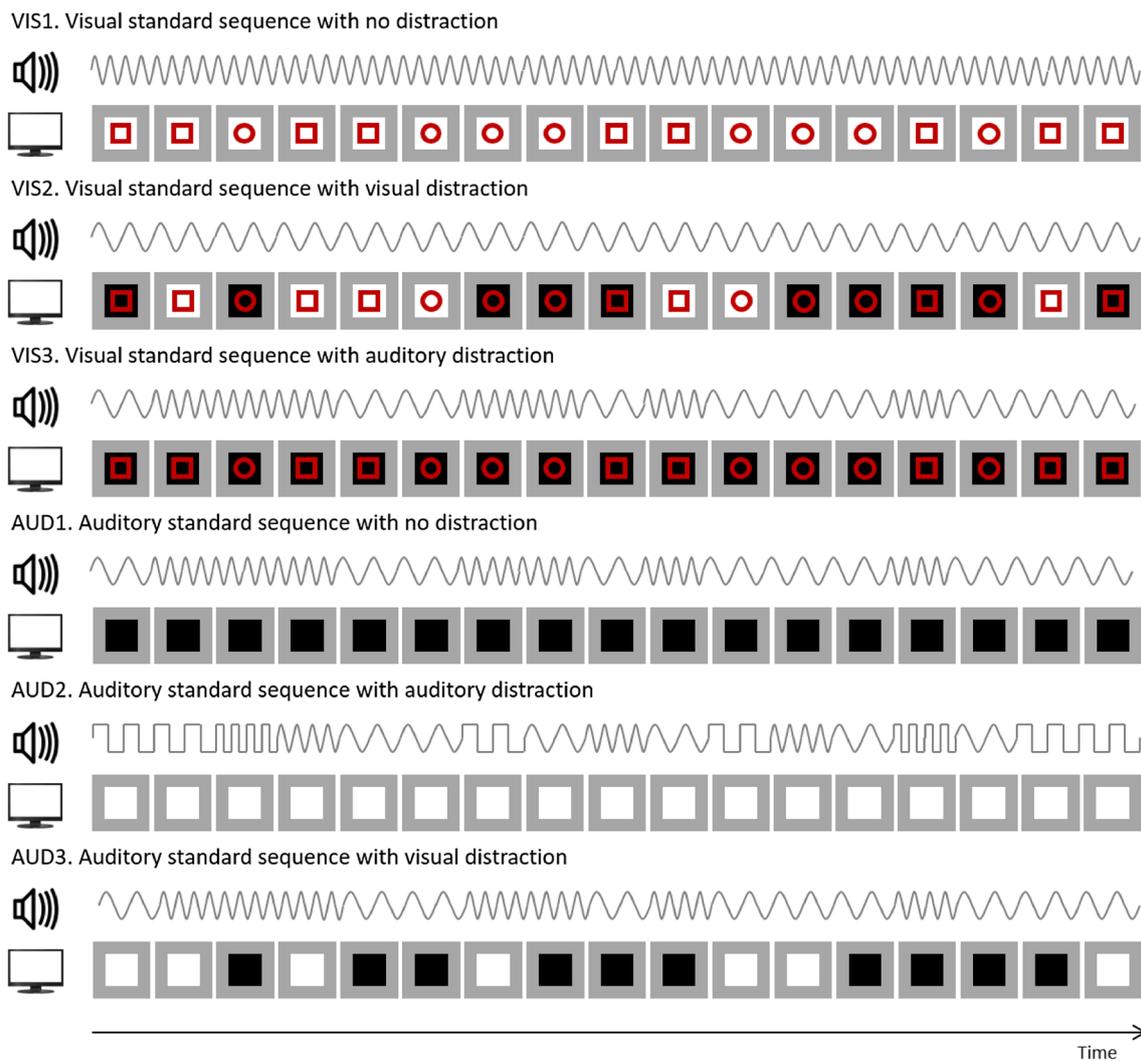


Fig. 1 Examples of standard sequences in Experimental Series A. Each sequence consisted of 17 subsequent frames with a duration of 100 ms/frame (Exp. A3: 140 ms/frame). In standard and distractor sequences, changes appeared on 8 positions according to a random point process, with the restriction of no more than 4 consecutive changes and no more than 4 frames without any change. The temporal sequence in standard and distractor sequences was always different. Visual target stimuli were red geometrical frames (square

vs. circle), and visual distraction was provided by the color (black vs. white) of the background square. Auditory target stimuli were tones of different frequency (400 vs. 800 Hz), and auditory distraction was provided by the waveform of the tones (sine vs. squarewave). In conditions without distraction (VIS1 and AUD1), and with distraction from the same modality (VIS2 and AUD2), the stimulus from the other modality did not convey any temporal sequence information

Experimental Series A: sequence discrimination with distraction during encoding

Experimental Series A comprised three experiments, which all employed task-irrelevant information presented together with the standard sequence. In all experiments of this series, the distractor information could be auditory or visual, whereas the target modality was always the same (visual in Experiment A1 and auditory in Experiments A2 and A3). In Experiment A3, the duration of each sequence frame was prolonged with the aim to especially increase the effect of

visual distraction.¹ As in Guttman et al.’s (2005) study, participants were asked to compare the standard sequence with a comparison sequence and to judge whether both were the same or different.

¹ A potentially stronger effect of visual distraction with prolonged frame duration was suggested by a pilot experiment ($N=10$), which explored visual and auditory sequence discrimination without distraction. This experiment showed an improvement of discrimination performance with a prolonged frame duration (100 vs. 140 ms) only for visual sequences.

Method

Participants

Ten volunteers participated in each experiment (Exp. A1: 7 females and 3 males, aged from 19 to 32 years, $M = 23.5$ years; Exp. A2: 6 females and 4 males, aged from 20 to 42 years, $M = 27.3$; Exp. A3: 5 females and 5 males, aged from 19 to 32 years, $M = 23.0$ years). They reported normal hearing and normal or corrected-to-normal vision, and received either course credit or payment. Before each experiment started, all participants gave written informed consent. Each experiment lasted about 1 h.

Apparatus and stimuli

The experiments were programmed in Matlab using the Psychophysics Toolbox (Version 3.0.8; Brainard 1997; Kleiner et al. 2007; Pelli 1997) on an Apple Mac Pro Computer, and were run in a sound-attenuated, dimly illuminated experimental booth. Participants sat in front of a CRT monitor (Samsung SyncMaster 1100MB, 150 Hz) with a viewing distance of about 50 cm.

Auditory stimuli were presented through headphones (Sennheiser HD 380 pro) at an intensity level of 70 dB(A) SPL. Visual stimuli were a red square frame ($4.4^\circ \times 4.4^\circ$ of visual angle, frame width: 0.9° of visual angle) and a red circle frame (with a diameter of 4.4° of visual angle, frame width: 0.9° of visual angle). Visual distractor information was provided by the color (black vs. white) of a larger square ($5.0^\circ \times 5.0^\circ$ of visual angle) that formed the background of the frame stimuli. In Experiments A2 and A3, there were additional catch stimuli. The visual catch stimuli were the capital letters A and E ($1.1^\circ \times 1.4^\circ$ of visual angle). Auditory catch stimuli were the spoken letters “A” and “E”. All visual stimuli were presented at the center of the screen against a gray background. Auditory stimuli were sine and squarewave tones of different frequencies (400 vs. 800 Hz).

The target and the distractor sequences were constructed following the procedure by Guttman et al. (2005). Within each sequence of 17 frames (Exp. A1 and A2: 100 ms/frame; Exp. A3: 140 ms/frame) changes of the respective stimulus (shape of the red frame, color of the background square, tone frequency, or tone waveform) appeared on 8 positions according to a random point process, with the restriction of no more than 4 consecutive changes and no more than 4 frames without any change.

In Experiment A1, the target modality was visual and the target sequence was defined by a change of the shape of the red frame (square vs. circle). Visual distractor information consisted of a color change of the background square (black vs. white). Auditory distractor information was defined by a change of the sine tone’s frequency (400 vs. 800 Hz). For

examples of these sequences, see Sequences VIS1–VIS3 in Fig. 1. In Experiments A2 and A3, the target modality was auditory and the target sequence was defined by a change of the tone’s frequency (400 vs. 800 Hz). The visual distractor was the same as in Experiment A1 (color change of background square), whereas the auditory distractor information consisted of a change of the tone’s waveform (sine vs. squarewave). In Experiments A2 and A3, the start (and final) tone of the standard sequence was either a sine or a squarewave tone. Accordingly, in trials without auditory distraction the whole standard sequence consisted of either sine or squarewave tones. For examples of these sequences, see Sequences AUD1–AUD3 in Fig. 1.

As in Guttman et al.’s Experiment 3, the distractor information was presented simultaneously with the standard sequence. In all trials, the background square (black or white) was presented during the encoding phase, irrespective of whether it conveyed distractor information or not. Similarly, in all trials, an auditory stimulus was presented during the encoding phase. That is, in trials without auditory target stimuli or distraction one of the two sine tones was played for the whole duration of the standard sequence. During the presentation of the comparison sequence, however, no other stimulus than the target sequence was presented. Accordingly, in Experiment A1, only the red frames were presented, and in Experiments A2 and A3 only the tone sequence was played. In Experiments A2 and A3, the comparison sequence was always presented with the waveform of the first (and last) tone of the standard sequence. In *same* trials, the target sequence and the comparison sequence were the same; whereas in *different* trials, the two sequences differed.

Procedure

Each trial started with the presentation of a fixation cross in the center of the screen. Then, participants started the presentation of the standard sequence with a press of the space bar. After an inter-stimulus interval of 1.6 s, the comparison sequence was presented. At the end of the trial, participants were asked to indicate whether the standard and the comparison sequence were the same (left control-key) or different (right control-key). After an inter-trial interval of 2 s, the next trial started. In case of an incorrect response, during the inter-trial interval, a feedback message was presented at the center of the screen. This message indicated that the response was incorrect and whether the sequences were the same or different.

In Experiments A2 and A3, there were also catch trials. In half of these trials, a visual catch stimulus was presented at the center of the screen (in black or white, always in contrast to the background color). In the other half, an auditory catch stimulus was presented. The catch stimuli were presented for the duration of one sequence frame at a

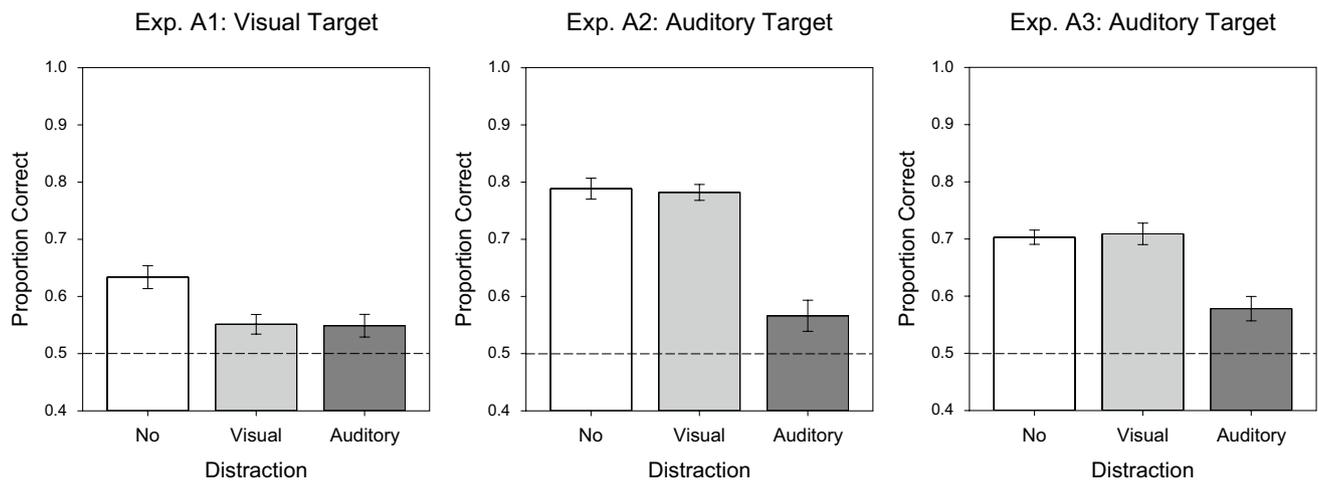


Fig. 2 Mean accuracy (proportion correct) as a function of distraction in Experimental Series A. Left panel: Experiment A1 with visual target sequences. Middle panel: Experiment A2 with auditory

target sequences. Right panel: Experiment A3 with auditory target sequences and prolonged frame duration (140 ms per frame). Error bars represent ± 1 within-subjects *SE*

random position of the sequence with the exception of the first and the last position. In catch trials, no comparison sequence was presented. Instead, after the presentation of the standard sequence, participants were asked to indicate with the corresponding key of the keyboard whether an “A” or an “E” had been presented during the encoding phase.

To ensure that participants did not take off the headphones during Experiment A1, at the end of the experiment they were presented with a spoken message which instructed them to report a randomly chosen number (“1” vs. “2”) to the experimenter.

Each experimental block consisted of 48 trials. In Experiment A1, there were 6 blocks; Experiments A2 and A3 comprised 8 blocks. In Experiment A1, within each block, each combination of distractor (no, visual, or auditory), sequence congruency (same vs. different), visual start distractor (black vs. white), and auditory start distractor (400 vs. 800 Hz) was presented twice and in random order. In Experiments A2 and A3, there were 25% catch trials (additional 96 trials). In these experiments, trial order was randomized across the whole experiment. In all experiments, the first block was considered practice.

Results and discussion

For all experiments, an ANOVA with the within-subjects factor distraction (no, visual, or auditory) was conducted on discrimination accuracy (proportion correct). Figure 2 depicts accuracy performance (proportion correct) as a function of distraction in Experiments A1–3. Within-subjects standard errors were calculated according to Morey (2008).

Exp. A1: visual sequence discrimination

Accuracy differed across the distraction conditions, $F(2, 18) = 6.45$, $p = 0.008$, $\eta_p^2 = 0.42$, with the highest accuracy when no distraction was present ($M = 0.63$, $SD = 0.10$), followed by very similar accuracies for visual ($M = 0.55$, $SD = 0.07$) and auditory ($M = 0.55$, $SD = 0.08$) distractions. Accordingly, Tukey contrasts indicated that the no-distraction condition differed significantly from the visual and the auditory distraction condition, both $p_s \leq 0.010$; whereas, the two conditions with distraction did not differ significantly from each other, $p = 0.991$. *T* tests indicated that accuracy differed from chance performance (0.5) in all conditions, $p_s \leq 0.039$.

Exp. A2: auditory sequence discrimination

Overall catch trial accuracy was very high ($M = 99.3\%$, $SD = 1.0\%$). Nevertheless, accuracy was slightly higher in auditory ($M = 100\%$, $SD = 0.0\%$) than in visual ($M = 98.5\%$, $SD = 2.0\%$) catch trials, $t(9) = 2.27$, $p = 0.049$. As one would expect, overall accuracy for auditory sequence discrimination was higher than in the visual Experiment A1. Accuracy again differed across the distraction conditions, $F(2, 18) = 38.01$, $p < 0.001$, $\eta_p^2 = 0.81$. However, the pattern differed from the one in Experiment A1. Accuracy was comparably high when no distraction was present ($M = 0.79$, $SD = 0.12$) and when the distraction was visual ($M = 0.78$, $SD = 0.11$), whereas it was clearly lower in the auditory distraction condition ($M = 0.57$, $SD = 0.05$). Tukey contrasts confirmed this pattern, indicating that the auditory distraction

condition differed from the two other distraction conditions, $p_s < 0.001$, while there was no significant difference between no distraction and visual distraction, $p = 0.895$. Again, accuracy differed from change performance in all conditions, $p_s < 0.001$.

Exp. A3: auditory sequence discrimination with prolonged duration/frame

Catch trial accuracy was somewhat lower ($M = 95.1\%$, $SD = 8.2\%$) than in Experiment A2, and again higher in auditory ($M = 98.6\%$, $SD = 4.4\%$) than in visual ($M = 91.6\%$, $SD = 12.6\%$) catch trials, $t(9) = 2.64$, $p = 0.027$. Accuracy again differed across the distraction conditions, $F(2, 18) = 17.02$, $p < 0.001$, $\eta_p^2 = 0.65$. Although overall accuracy was somewhat lower than in Experiment A2, the pattern was very similar. Accordingly, accuracy was comparably high when no distraction was present ($M = 0.70$, $SD = 0.11$) and when the distraction was visual ($M = 0.71$, $SD = 0.11$), whereas it was clearly lower in the auditory distraction condition ($M = 0.58$, $SD = 0.08$). Tukey contrasts again confirmed this pattern, indicating that the auditory distraction condition differed from the two other distraction conditions, $p_s < 0.001$, while there was no significant difference between no and visual distraction, $p = 0.971$. As in Experiments A1 and A2, accuracy differed from change performance in all conditions, $p_s \leq 0.006$.

Summary

The results pattern of Experimental Series A can be summarized as follows. First, overall discrimination performance was clearly better for auditory (Exps. A2 and A3) than for visual sequences (Exp. A1), even when no distraction was present. Second, auditory distraction impaired performance irrespective of whether the target sequences were auditory or visual, whereas visual distraction only impaired the discrimination of visual target sequences. This pattern confirms and extends the previous results by Guttman et al. (2005). Third, a prolongation of the frame duration within each sequence in Experiment A3, which we hypothesized could have strengthened the impact of the visual distraction, did not change the result pattern.

Experimental Series B: sequence discrimination with distraction during retention

Experimental Series B aimed to investigate the same research questions as Experimental Series A, but employed a slightly different methodological approach. In three

experiments, the distracting information was presented not during the encoding of the standard sequences, but in a retention interval between the standard and the comparison sequence (see also Bratzke et al. 2016; Rattat and Picard 2012). The intention behind this change was to examine whether the distraction only affects the encoding of temporal sequences or whether it might even disrupt the short-term retention of such sequences. Additionally, we used two different stimulus characteristics for each modality to define the target sequences. Visual sequences were defined by a change of the shape or the color of a visual object, whereas auditory sequences were defined by the frequency or the waveform of a tone. In contrast to Experimental Series A, all trials included the presentation of incongruent distraction, either from the same or the different modality. Finally, this time all participants performed visual and auditory sequence discrimination. Experiment B1 involved the same temporal sequence structure as in Series A (17 frames with 8 changes). In Experiments B2 and B3, the total duration of each sequence was shortened (9 frames with 4 changes), and in Experiment B3 the duration of each frame was increased to 180 ms. These changes were made with the aim to potentially increase the discriminability especially for visual sequences.

Method

Participants

Sixteen volunteers participated in each experiment (Exp. B1: 13 females and 3 males, aged from 19 to 29 years, $M = 23.6$ years; Exp. B2: 11 females and 5 males, aged from 20 to 54 years, $M = 28.4$ years; Exp. B3: 10 females and 6 males, aged from 19 to 29 years, $M = 21.8$ years). They reported normal hearing and normal or corrected-to-normal vision, and received either course credit or payment. In Experiment B1, one participant was replaced by another participant because of very poor catch trial performance ($< 67\%$). Before each experiment started, all participants gave written informed consent.

Apparatus and stimuli

The experimental setup was identical to Experimental Series A. Visual stimuli were squares and circles ($2.6^\circ \times 2.6^\circ$ of visual angle; black vs. white), which were presented at the center of the screen against a gray background. Auditory stimuli were sine and squarewave tones of different frequencies (400 vs. 800 Hz), played through headphones at an intensity level of 70 dB(A) SPL. Visual sequences were defined by a change of the visual object's shape or color. Similarly, auditory sequences were defined by a change of the tone's frequency or waveform. The construction of the

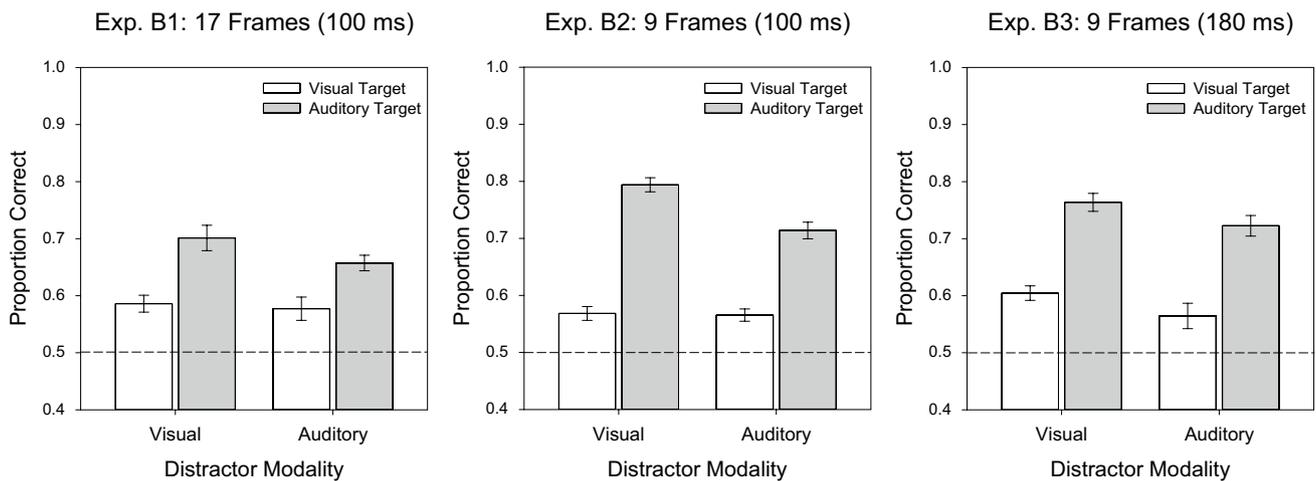


Fig. 3 Mean accuracy (proportion correct) as a function of target modality and distractor modality in Experimental Series B. In Experiments B2 and B3, the length of each sequence was reduced from 17

frames (8 changes) to 9 frames (4 changes). In Experiment B3, additionally, the duration of each sequence frame was prolonged (from 100 to 180 ms). Error bars represent ± 1 within-subjects SE

temporal structure of the target and the distractor sequences followed the procedure of Experimental Series A. In Experiment B2, the length of each sequence was reduced from 17 frames with 8 changes to 9 frames with 4 changes. In Experiment B3, additionally the duration of each sequence frame was prolonged from 100 to 180 ms. The catch stimuli were the same as in Experimental Series A.

Procedure

Each trial started with the presentation of a fixation cross in the center of the screen. Then, participants started the presentation of the standard sequence with a press of the space key. After that, the distraction sequence was presented, which was followed by the comparison sequence. The three sequences were separated by inter-stimulus intervals of 0.5 s. During these intervals, the fixation cross reappeared in the center of the screen. At the end of the trial, participants were asked to indicate whether the standard and the comparison sequence were the same (left control-key) or different (right control-key). After an inter-trial interval of 2 s, the next trial started. In case of an incorrect response, during the inter-trial interval, a feedback message was presented at the center of the screen.

As in Experimental Series A, there were also catch trials, which followed the same procedure as in Experimental Series A. In contrast to Experiments A2 and A3, the modality of the catch stimulus was always identical to the distractor modality. Each participant performed visual and auditory sequence discrimination. The different target modalities were tested in different halves of the experiment and modality order was balanced across participants. Each participant was assigned to one of the four possible combinations of

visual and auditory target and distractor stimuli (color/frequency–shape/waveform, color/waveform–shape/frequency, shape/frequency–color waveform, shape/waveform–color/frequency).

Within each half of experiment, each combination of distractor modality (visual vs. auditory), start target color/frequency (black vs. white/400 vs. 800 Hz), start target shape/waveform (square vs. circle/sine vs. square), start distractor frequency/color (400 vs. 800 Hz/black vs. white), and start distractor waveform/shape (sine vs. square/square vs. circle) was presented twice, plus additional 96 catch trials (a third of all trials). Experimental trials were subdivided into blocks of 48 trials. At the beginning of each half of experiment, there was a practice block of 24 trials (randomly drawn from the 96 possible experimental trials within each half).

Results and discussion

For all experiments, an ANOVA with the within-subjects factors target modality (visual vs. auditory) and distractor modality (visual vs. auditory) was conducted on discrimination accuracy (proportion correct). Figure 3 depicts accuracy performance (proportion correct) as a function of target modality and distractor modality in Experiments B1–3. Within-subjects standard errors were calculated according to Morey (2008).

Exp. B1: auditory and visual sequence discrimination

Catch trial accuracy was slightly higher for auditory ($M = 99.8\%$, $SD = 0.5\%$) than for visual ($M = 98.2\%$, $SD = 2.8\%$) catch stimuli, $t(15) = 2.39$, $p = 0.030$. Accuracy was higher for auditory ($M = 0.68$, $SD = 0.11$) than for visual sequences

($M = 0.58$, $SD = 0.09$), $F(1, 15) = 17.72$, $p < 0.001$, $\eta_p^2 = 0.54$. The main effect of distraction was not significant, $F(1, 15) = 2.31$, $p < 0.150$, $\eta_p^2 = 0.13$. The interaction between target modality and distractor modality was also not significant, $F(1, 15) = 2.21$, $p < 0.167$, $\eta_p^2 = 0.12$. In all conditions, discrimination performance was significantly better than chance performance, $p_s \leq 0.001$.

Exp. B2: auditory and visual sequence discrimination with reduced sequence length

Catch trial accuracy was similarly high for auditory ($M = 99.7\%$, $SD = 0.8\%$) and for visual ($M = 99.4\%$, $SD = 1.6\%$) catch stimuli, $t(15) = 0.61$, $p = 0.549$. Accuracy was again substantially higher for auditory ($M = 0.75$, $SD = 0.08$) than for visual sequences ($M = 0.57$, $SD = 0.07$), $F(1, 15) = 159.73$, $p < 0.001$, $\eta_p^2 = 0.91$. Thus, shortening the total sequence duration improved discrimination performance, if any, for auditory instead of visual sequences. In contrast to Experiment B1, the main effect of distractor modality and the interaction were also significant. Auditory distraction ($M = 0.64$, $SD = 0.11$) was more effective than visual distraction ($M = 0.68$, $SD = 0.13$), $F(1, 15) = 9.94$, $p = 0.007$, $\eta_p^2 = 0.40$, and this was the case for auditory but not for visual target sequences, $F(1, 15) = 19.06$, $p = 0.001$, $\eta_p^2 = 0.56$. Specifically, accuracy for auditory sequences was 0.79 ($SD = 0.06$) with visual distraction and 0.71 ($SD = 0.08$) with auditory distraction, whereas accuracy for visual sequences was virtually the same with auditory ($M = 0.57$, $SD = 0.07$) and visual ($M = 0.57$, $SD = 0.06$) distraction (see middle panel of Fig. 3). In all conditions, discrimination performance was significantly better than chance performance, $p_s \leq 0.001$.

Exp. B3: auditory and visual sequence discrimination with reduced sequence length and prolonged duration/frame

As in Experiment B1, catch trial accuracy was slightly higher for auditory ($M = 99.8\%$, $SD = 0.8\%$) than for visual ($M = 99.2\%$, $SD = 1.1\%$) catch stimuli, $t(15) = 2.42$, $p = 0.029$. Despite the prolongation of each frame's duration, discrimination performance for the two modalities was very similar to the pattern of Experiment B2. Accordingly, accuracy was again substantially higher for auditory ($M = 0.74$, $SD = 0.12$) than for visual sequences ($M = 0.58$, $SD = 0.09$), $F(1, 15) = 48.36$, $p < 0.001$, $\eta_p^2 = 0.76$. The main effect of distractor modality was also again significant, $F(1, 15) = 9.41$, $p = 0.008$, $\eta_p^2 = 0.39$, with on average higher accuracy for visual ($M = 0.68$, $SD = 0.14$) than for auditory ($M = 0.64$, $SD = 0.13$) distraction. In contrast to Experiment B2, however, this effect was very similar

for auditory and visual target sequences (see right panel of Fig. 3), $F(1, 15) < 0.01$, $p = 0.975$, $\eta_p^2 < 0.01$. In all conditions, discrimination performance was significantly better than chance performance, $p_s \leq 0.005$.

Summary

The results of Experimental Series B showed that auditory distraction was always at least as effective as visual distraction. As in Experimental Series A, discrimination performance was generally much better for auditory than for visual sequences. Neither a shortening of the sequences by reducing the number of frames and changes per sequence, nor a prolongation of each frame's duration from 100 to 180 ms was sufficient to change this basic pattern.

General discussion

The present study aimed to replicate and extend results by Guttman et al. (2005), which showed that visual sequence discrimination was strongly disrupted by auditory but not by visual distraction. As we have outlined in the Introduction, the visual distraction provided in their study, however, cannot be regarded as an adequate control condition (for a similar argument, see also Grahn 2012). We, therefore, attempted to create comparable auditory and visual conditions and investigated the effects of irrelevant distraction on temporal sequence discrimination within as well as across the two modalities.

Our results clearly replicate Guttman et al.'s (2005) finding that auditory distraction strongly impairs the discrimination of visual sequences. Furthermore, auditory distraction was generally at least as effective in impairing discrimination performance as visual distraction. Importantly, this was the case not only for auditory but also for visual target sequences. When participants had to discriminate auditory target sequences, visual distraction was almost completely ineffective (see Experiments A2 and A3). This result is especially remarkable since we carefully prevented participants from simply blocking out the distractor information by using catch trials. Moreover, the stronger impact of auditory than visual distraction was also observed when the distraction was presented in a retention interval between the standard and the comparison sequence (Experimental Series B). Together, these results demonstrate that auditory distraction is much more effective in impairing discrimination of temporal sequences than visual distraction, irrespective of whether the to-be-discriminated target sequences are provided in the auditory or visual modality.

An obvious feature of the present results is the much lower performance level for visual compared to auditory

sequences (for similar results see also Grahn 2012). In most cases, discrimination performance in the visual modality was not much better than chance performance. Even in Experiment A1, in which participants discriminated visual sequences without any specific distraction, discrimination accuracy was only 13 percentage points higher than the 50% chance level. In Guttman et al.'s (2005) study, the performance level for visual sequences was slightly higher (approx. 75–79%) than in our study but still much lower than the one for auditory sequences.² There are several possible reasons for these differences in the overall performance level between our and Guttman et al.'s study, as for example, the different stimulus material used to create visual (shape and color vs. contrast reversal of gabor patches) and auditory (frequency and waveform vs. brief auditory clicks) sequences, and the participant samples (Guttman et al.'s sample included the three authors).

In light of the rather poor discrimination performance for visual temporal sequences, it is perhaps not surprising that the same visual sequences were rather ineffective when they were used as distraction. In fact, we made several attempts to increase the impact of visual distraction. First, we used several visual stimulus characteristics to define the temporal structure of the sequences. Second, based on the results of a pilot experiment, we prolonged the duration of each sequence frame. Third, since sensory memory may be shorter for visual (e.g., Coltheart 1980) than for auditory (e.g., Cowan et al. 1984) stimuli (see Grahn 2012), we shortened the total duration of each sequence. All these efforts proved insufficient to increase the discriminability of the visual sequences. Thus, a fundamental problem in the study of cross-modal sequence discrimination, which is difficult, if not impossible to resolve, is the asymmetry in timing abilities for visual and auditory sequences with comparable temporal structures. In fact, this asymmetry is not specific to the discrimination of rapid temporal sequences but can be observed in a large variety of timing tasks and stimulus conditions (e.g., Glenberg et al. 1989; Grondin 1993; Repp and Penel 2002; Ulrich et al. 2006). There is some evidence that in bimodal duration perception, the auditory information is still dominant, even when discriminability for the visual and the auditory stimulation is approximately equalized by “smearing” the on- and offset of the auditory stimuli (Ortega et al. 2014). Whether this holds true also for temporal discrimination of rapid temporal sequences, however, remains an open question. Furthermore, we certainly cannot rule out that other visual stimuli might be more effective in

distracting from or interfering with auditory sequence discrimination than the ones we used in the present study.

Guttman et al. (2005) interpreted their results as evidence for the cross-modal encoding view. According to this view, visual temporal information is obligatorily and automatically encoded in the auditory system. While this view has received further support (Bratzke et al. 2012; Kanai et al. 2011; Grahn 2012, but see also McAuley and Henry 2010), we think that the present results and those of Guttman et al. do not provide strong evidence for this view, although they are certainly consistent with it. This is because the potential of a distractor sequence to cause interference with the processing of a target sequence may be related to the temporal sensitivities of the sensory modalities, in which the sequences are presented. Accordingly, the lower potential of a visual distractor sequence to interfere with the encoding or retention of an auditory target sequence than vice versa can be attributed to the poorer discriminability of such sequences in the visual than in the auditory modality. This problem hampers not only a proper interpretation of the present cross-modal interference effects but also the results of cross-modal transfer paradigms, as for example in the study of perceptual learning (Bratzke et al. 2012, 2014; Chen and Zhou 2014; Grondin and Ulrich 2011; Lapid et al. 2009).

In conclusion, the present study replicated the previous result by Guttman et al. (2005) that visual temporal sequence discrimination was strongly disrupted by incongruent auditory distraction. Furthermore, our results demonstrate that this disruption can also appear during short-term retention of the to-be-discriminated sequence. Moreover, the present study highlights why the interpretation of such results with regard to the underlying timing system is somewhat hampered.

Acknowledgements We thank Karolin Schuh for her assistance with data collection. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (UL 116/14-1).

References

- Bausenhardt KM, de la Rosa MD, Ulrich R (2014) Multimodal integration of time: visual and auditory contributions to perceived duration and sensitivity. *Exp Psychol* 61:310–322
- Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436
- Bratzke D, Seifried T, Ulrich R (2012) Perceptual learning in temporal discrimination: asymmetric cross-modal transfer from audition to vision. *Exp Brain Res* 221:205–210
- Bratzke D, Schröter H, Ulrich R (2014) The role of consolidation for perceptual learning in temporal discrimination within and across modalities. *Acta Psychol* 147:75–79
- Bratzke D, Quinn KR, Ulrich R, Bausenhardt KM (2016) Representations of temporal information in short-term memory: are they modality-specific? *Acta Psychol* 170:163–167

² In the results section of their Experiment 1, Guttman et al. (2005) refer to the results of a follow-up experiment, in which participants discriminated auditory sequences (in isolation, with congruent visual information, or with incongruent visual information) with an average accuracy of 94.6% correct.

- Burr D, Banks MS, Morrone MC (2009) Auditory dominance over vision in the perception of interval duration. *Exp Brain Res* 198:49–57
- Chen L, Zhou X (2014) Fast transfer of crossmodal time interval training. *Exp Brain Res* 232:1855–1864
- Coltheart M (1980) Iconic memory and visible persistence. *Percept Psychophys* 27:183–228
- Cowan N, Aaronson D, Glanzer M, Balota D, Clifton R, Pollatsek A, Crowder R (1984) On short and long auditory stores. *Psychol Bull* 96:341–370
- Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. *Percept Psychophys* 63:719–725
- Glenberg AM, Mann S, Altman L, Forman T, Prociase S (1989) Modality effects in the coding and reproduction of rhythms. *Mem Cogn* 17:373–383
- Grahn JA (2012) See what I hear? Beat perception in auditory and visual rhythms. *Exp Brain Res* 220:51–61
- Grondin S (1993) Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Percept Psychophys* 54:383–394
- Grondin S, Ulrich R (2011) Duration discrimination performance: no cross-modal transfer from audition to vision even after massive perceptual learning. In: Vatakis A, Esposito E, Giagkou M, Cummins F, Papadelis G (eds) *Multidisciplinary aspects of time and time perception*, vol 6789. LNAI. Springer, Berlin, pp 92–100
- Guttman SE, Gilroy LA, Blake R (2005) Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychol Sci* 16:228–235
- Kanai R, Lloyd H, Buetti D, Walsh V (2011) Modality-independent role of the primary auditory cortex in time estimation. *Exp Brain Res* 209:465–471
- Kleiner M, Brainard DH, Pelli D (2007) What's new in psychtoolbox-3? Perception, 36 ECVF Abstract Supplement
- Lapid E, Ulrich R, Rammsayer T (2009) Perceptual learning in auditory temporal discrimination: no evidence for a cross-modal transfer to the visual modality. *Psychon Bull Rev* 16:382–389
- McAuley JD, Henry MJ (2010) Modality effects in rhythm processing: auditory encoding of visual rhythms is neither obligatory nor automatic. *Atten Percept Perform* 72:1377–1389
- Morey R (2008) Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutor Quant Methods Psychol* 4:61–64
- Ortega L, Guzman-Martinez E, Grabowecky M, Suzuki S (2014) Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability. *Atten Percept Psychophys* 76:1485–1502
- Pelli D (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 4:437–442
- Rattat AC, Picard D (2012) Short-term memory for auditory and visual durations: evidence for selective interference effects. *Psychol Res* 76:32–40
- Repp BH, Penel A (2002) Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences. *J Exp Psychol Hum Percept Perform* 28:1085–1099
- Ulrich R, Nitschke J, Rammsayer T (2006) Crossmodal temporal discrimination: assessing the predictions of a general pacemaker-counter model. *Percept Psychophys* 68:1140–1152
- Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. *Psychol Bull* 88:638–667

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.