THE CAPTURE OF VISION BY AUDITION:
EXAMINING TEMPORAL VENTRiloQUIsM
by
SHARON MOREIN-ZAMIR

BA, Ben-Gurion University of the Negev, Israel 2000
MA, Ben-Gurion University of the Negev, Israel 1997

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

(Department of Psychology, Cognitive Systems Programme)

We accept this thesis as conforming
 to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2002

© Sharon Morein-Zamir, 2002
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Psychology, Cognitive Systems Programme

The University of British Columbia
Vancouver, Canada

Date August 21st, 2002
Abstract

Four experiments were conducted to investigate whether sounds can draw the perception of lights further apart in time in a visual temporal order judgment task, thereby improving performance. In the first experiment, presenting a sound before the first light and after the second light improved performance relative to the baseline condition in which sounds appeared simultaneously with the lights. The second experiment ruled out a simple alerting account of this effect and indicated that the effect was due to the second sound trailing the second light. The third experiment extended the duration of the lags between the first sound and light, and the second light and sound, and found that performance returned to baseline level with lags of 450 milliseconds. In the final experiment, the second sound was found to improve performance only when the first sound was present, suggesting the importance of the pairing of the audiovisual stimuli. The results are interpreted as reflecting the temporal analogue of the classic spatial ventriloquist effect.
# Table of Contents

Abstract .............................................................................................................. ii
Table of Contents ............................................................................................... iii
List of Tables ........................................................................................................ iv
List of Figures ....................................................................................................... v
Dedication .............................................................................................................. vi

Introduction .......................................................................................................... 5

Experiment 1 ........................................................................................................ 6
Method .................................................................................................................. 6
Participants ......................................................................................................... 6
Apparatus ............................................................................................................. 6
Design and Procedure .......................................................................................... 6
Results .................................................................................................................. 7
Discussion ............................................................................................................ 8

Experiment 2 ........................................................................................................ 9
Method .................................................................................................................. 10
Participants ......................................................................................................... 10
Design and Procedure .......................................................................................... 10
Results .................................................................................................................. 10
Discussion ............................................................................................................ 11

Experiment 3 ....................................................................................................... 12
Method .................................................................................................................. 13
Participants ......................................................................................................... 13
Design and Procedure .......................................................................................... 13
Results .................................................................................................................. 14
Discussion ............................................................................................................ 14

Experiment 4 ....................................................................................................... 15
Method .................................................................................................................. 16
Participants ......................................................................................................... 16
Design and Procedure .......................................................................................... 16
Results .................................................................................................................. 17
Discussion ............................................................................................................ 17

General Discussion .............................................................................................. 18
The benefits of generalizing from the spatial to the temporal domain ............ 20
Concluding Remarks ............................................................................................. 24

Appendix ............................................................................................................... 26

References ........................................................................................................... 28

Footnotes .............................................................................................................. 33
List of Tables

Table 1: Mean Just Noticeable Differences (JNDs) in milliseconds for Experiment 1 ....34
Table 2: Mean Just Noticeable Differences (JNDs) in milliseconds for Experiment 2 ....35
Table 3: Mean Just Noticeable Differences (JNDs) in milliseconds for Experiment 3 ....36
Table 4: Mean Just Noticeable Differences (JNDs) in milliseconds for Experiment 4 ....37
List of Figures

Figure 1. A schematic illustration of the events in an example trial .................. 38

Figure 2. A schematic illustration of the conditions of Experiment 2 .................. 39

Figure 3a. Accuracy plotted as a function of SOA with fitted linear functions for Experiment 2 ................................................................. 40

Figure 3b. Accuracy plotted as a function of SOA with fitted linear functions for Experiment 3 ................................................................. 41

Figure 3c. Accuracy plotted as a function of SOA with fitted linear functions for Experiment 4 ................................................................. 42

Figure 4a. Accuracy plotted as a function of SOA with fitted linear functions for Experiment 2 ................................................................. 43

Figure 4b. Accuracy plotted as a function of SOA with fitted linear functions for Experiment 3 ................................................................. 44

Figure 4c. Accuracy plotted as a function of SOA with fitted linear functions for Experiment 4 ................................................................. 45
Introduction

Perception of the world is inherently a multisensory experience filled with various sights and sounds. The brain integrates this information from the different modalities to arrive at a better representation of the external environment. Normally, visual stimuli and sounds provide converging information, for example a person talking will both move their lips and produce sounds at the same time. Interestingly, multisensory interactions between vision and audition are frequently studied using intermodal conflict where the two modalities provide discrepant information. One of the most notable examples is the ventriloquist effect where the perceived location of a sound is biased by a visual stimulus presented at a different position (Howard & Templeton, 1966; Jack & Thurlow, 1973; Radeau & Bertelson, 1974). Much of the multisensory literature has focused on spatial interactions of this type. In addition, identity interactions such as the McGurk effect have also been studied (McGurk & MacDonald, 1976). In the McGurk effect the auditory perception of what is being heard (for example, /ba/) is altered by what is being seen (for example, /ga/). The final percept is a combination of the two (for example, /da/).

It is also the case that the senses can interact in time (i.e., not where or what are the events being perceived, but when they are perceived). The temporal relationships of stimuli from various modalities play an important role in the integration process. Indeed, synchrony between auditory and visual events is crucial to spatial ventriloquism, as the effect disappears when the audio-visual asynchrony exceeds approximately 300 milliseconds (ms) (Bertelson & Aschersleben, 1998; Bertelson, Vroomen, & de Gelder, 1997; Radeau & Bertelson, 1987). This is also the case in the McGurk effect, which fails to occur when the audio-visual asynchrony exceeds 200 to 300 ms (Massaro, Cohen, & Smeele, 1996; Munhall, Gribble, Sacco,
& Ward, 1996). As temporal relationships between the modalities are evidently important, multisensory temporal processing should be examined directly in its own right. The present study investigated temporal interactions between vision and audition, with the timing of lights and sounds varied so as to occasionally provide inconsistent information.

In the spatial domain, vision biases the location of a sound. The reverse, a sound influencing the localization of a visual stimulus, is either very weak or not present at all (Bertelson, 1998; Radeau & Bertelson, 1974). The exact reason for this is an ongoing research issue, although one key factor appears to be that vision provides better location information than sound (Welch, 1999; Welch & Warren, 1986). In particular, vision has a more precise registration of spatial location and is more appropriate in localization judgments (Welch & Warren, 1980). Therefore, when auditory and visual stimuli co-occur the perception of the sound location shifts in space towards the location of the visual stimulus.

In the identity domain, it is well established that vision biases audition (e.g., McGurk & MacDonald, 1976). In recent years, however, several studies have demonstrated that audition can also alter visual perception, making it better or even qualitatively different. Vroomen and de Gelder (2000) reported that the detection and localization of a visual target was improved by presenting an auditory oddball (a high tone embedded in a sequence of low tones) at the same time. Sekuler, Sekuler and Lau (1997) found that a click could promote the perception of two discs bouncing off each other, instead of continuing along their original trajectories and passing through each other. Additionally, Shams, Kamitani and Shimojo (2000) reported a visual illusion, where multiple beeps induced a brief unambiguous visual flash to be incorrectly seen as multiple flashes.
In the temporal dimension it is audition and not vision that is considered to be the more "appropriate" modality (Shipley, 1964; Welch, 1999; Welch & Warren, 1980). Shipley (1964) claimed that "the ear ... is an organ of eminently superior temporal resolution". Furthermore, the visual threshold for critical flicker fusion is substantially lower than the auditory threshold for critical flutter fusion (Welch, Duttonhurt & Warren, 1986). This has been interpreted as supporting the conclusion that audition has greater temporal acuity (Welch, 1999). In keeping with the notion of greater auditory temporal acuity, audition has been found to capture visual temporal perception in a phenomenon called auditory driving (Gebhard & Mowbray, 1959; Myers et al., 1981; Shipley, 1964). Here, changes in the rate of a repetitive clicking sound induce corresponding changes in the perceived rate of a repetitive flashing light (Gebhard & Mowbray, 1959; Welch et al., 1986). Thus, the auditory temporal perception 'drives' temporal visual perception. Auditory driving, however, has been criticized because it relies on subjective estimation and may reflect a response bias rather than a perceptual effect (Fendrich & Corballis, 2001).

In a recent investigation, Fendrich and Corballis (2001) attempted to overcome the inherent limitations of auditory driving, and to demonstrate a phase capture of a repeating flash by a repeating sound. They asked participants to report the location of a rotating visual marker on a clock face at the time of a flash (the marker rotated several times around the clock face on each trial). Clicks, heard with each rotation of the marker, influenced reports of the marker location at the time of the flash. When the clicks preceded the flash the position reported was earlier, and when the clicks followed the flash the position reported was later. The researchers determined that these capture effects indicated how temporal discrepancies between vision and audition could be reconciled, as the two events were drawn towards
temporal correspondence. Unfortunately, since participants were asked to report the position of a visual marker there was no direct measurement of temporal perception. Furthermore, as this finding could reflect a response bias, spatial cueing or even alerting rather than a temporal bias, it does not fully address the shortcomings inherent in the previous auditory driving research (McDonald, Teder-Salejarvi & Hillyard, 2000; Posner, 1978; Spence & Driver, 1997).

Consequently, although there is suggestive evidence to support auditory stimuli influencing the perceived timing of visual events, to date there has been no straightforward demonstration and the role of temporal components in crossmodal perception is still unclear. Furthermore, it is uncertain whether audition may exert any influence on visual perception when non-repetitive stimuli are used. The superiority of sound within the context of rhythm or repetitive presentations such as a ticking or fluttering does not necessitate that it dominate vision when simpler stimuli are used (Bertelson, 1998; Spence & Driver, 2000).

The present series of experiments set out to investigate whether auditory events can alter the perceived timing of visual events while avoiding the shortcomings described above. In particular, does the spatial ventriloquist effect have an analogue effect in the temporal domain? As a light can alter the localization of a target sound in space, the reviewed research would imply that in the temporal domain sounds should alter the perceived timing of target lights. A temporal order judgment task (TOJ) was used, in which participants were asked to determine which of two lights appeared first. Task-irrelevant sounds were presented either at the same time as the onsets of the lights, or at different times. Our aim was to introduce and explore the TOJ task as a means of studying temporal multisensory perception. Having ascertained that the
sounds do appear to shift the temporal perception of the lights, the nature of this phenomenon and its robustness was investigated.

Experiment 1

The working hypothesis for this experiment was that presenting task-irrelevant sounds before the first light and after the second light of the TOJ task might "pull" the two lights further apart in time, thereby improving performance. Presenting a sound simultaneous with the onset of each of the two lights provided a baseline measure of performance whereby the stimuli were equated but without a temporal discrepancy between the lights and sounds (see Figure 1). Several lags between the sounds and lights were chosen (75, 150 and 225 ms) to explore the robustness and time course of the effect. The lags between the sounds and lights were varied in an orthogonal manner to the Stimulus Onset Asynchronies (SOAs) between the two lights.

It is possible that just as the temporal relationship between the light and sound influences the spatial ventriloquist effect (Bertelson & Aschersleben, 1998; Bertelson, Vroomen, & de Gelder, 1997; Radeau & Bertelson, 1987), so the spatial relation between sounds and lights may influence the extent of temporal biasing. Consequently, any spatial confounds were removed by delivering the sounds from a speaker that was positioned immediately behind the visual display. The two lights were placed in front of the speaker, one at the level of the top of the speaker and the other at the level of the bottom of the speaker. Thus, the lights and sounds originated from approximately the same spatial location. In addition, the same sound was used throughout the entire series of experiments. This would ensure that the sound itself
and its location could not be used to perform the task of which light came first, and that the timing of the sounds would be the critical feature.

**Method**

**Participants.** Sixteen undergraduate students (13 female and 3 male) took part in Experiment 1 for course credit. The participants' mean age was 20 years (standard deviation of 1.3). All had normal or corrected-to-normal vision and hearing.

**Apparatus.** Two green light-emitting diodes (LEDs) 20 cm apart were placed below and above a central red LED. The red LED served as fixation and was positioned approximately at eye level. The three LEDs were positioned on a black screen mounted and centered on a loudspeaker (CoEntrant). A 486 PC running EXPE6 software (Pallier, Jeanin & Dupoux, 1997) controlled the onset and offset of the LEDs and noise bursts through a custom-made relay box. The sine-wave noise bursts were produced by a function generator (Dynascan 3010) by means of an Optimus MPA-40 amplifier. Manual responses were collected using two buttons positioned one above the other interfaced with the relay box. Instructions regarding breaks and the ending of the experiment were displayed on a 14 inch monitor also connected to the computer.

**Design and Procedure.** Four lags determined the temporal position of the sounds relative to the lights. The lags were 0, 75, 150 and 225 ms. The 0 ms lag was the baseline condition, where the sounds and lights were presented simultaneously. All other lags represent the time interval between the onset of the first sound and first light, and the interval between the onset of the second light and second sound. For example, in the 150 ms lag condition the first sound preceded the first light by 150 ms and the second sound trailed after the second light by 150 ms. The lags were manipulated independent of the SOAs between the two lights, which were 12, 24, 36,
48, 72, 96 and 144 ms. For each SOA the top or bottom light could appear first, and was presented 12 times under every lag condition. After 20 practice trials, participants completed the 672 randomly ordered trials, with five self-terminated breaks allowing for resting periods.

Participants were tested individually in a dark room and were seated approximately 57 cm from the display. Each trial began with the illumination of the red fixation LED for 500 ms. Following an interval varying randomly from 110 to 390 ms one of the green LEDs was turned on, and after a 12 to 144 ms SOA, the second LED was also turned on. Both LEDs remained lit until a response was made.

In addition, two sounds, each lasting for 5 ms (1800 Hz, 67 dB), were presented. Participants pressed the top button to indicate that the top light came first, and the lower button to indicate that the bottom light came first. The next trial began 1500 ms after a response. Responses to the lights were unspeeded and participants were informed that sounds did not predict in any way which light came first and should therefore be ignored.

Results

The just noticeable difference (JND) for each of the four conditions was computed as the dependent measure (see Table 1). The JND, calculated from the accuracy scores, measures the minimal SOA at which the temporal order of the two lights can be determined reliably. To calculate the JND for each participant, the percentage of "top light first" responses was computed across all the different light-light SOAs (ranging from -144 to 144 ms) for each sound-light lag (ranging from 0-225 ms), where the negative SOAs indicated that the lower light appeared first. A logistic function was then fitted to the data for each condition of each participant.
Finally, the JND was computed from the fitted function by calculating the difference between the SOAs at which 75% of the responses were "top light first" and 25% of the responses were "top light first", and dividing it by two (see Coren, Ward & Enns, 1999).

Discussion

The JNDs were submitted to a repeated-measures analysis of variance (ANOVA) with the lags between the sounds and lights as an independent factor (0, 75, 150 and 225 ms). The results indicated that the effect of lag was significant, $F(3,45)=6.69$, $p<.001$. Tukey post hoc analyses revealed that the mean JND for the lags of 75 and 150 ms were significantly lower than the baseline lag of 0 ms ($p<.01$ for both comparisons). The JND for the 225 ms lag was marginally lower than baseline ($p<.08$). There were no significant differences between the three lags of 75, 150 and 225 ms, showing that the magnitude of the effect appeared to be consistent ($p>.4$).

The crucial result in Experiment 1 was that the sounds improved visual TOJ performance relative to baseline. This is consistent with the hypothesis that the sounds pulled the two lights further apart in time. Thus, it seems as if the perceived time of the first light was "ventriloquised" by the preceding sound, and similarly, the perceived time of the second light was ventriloquised toward the time of the second sound. Performance improved with lags as short as 75 ms and approached significance with lags as long as 225 ms. This time range is in keeping with previous findings that visual and auditory speech information can be integrated with lags of
150 to 250 ms (Dixon & Spitz, 1980; Massaro & Cohen, 1993; Massaro et al., 1996), and that sounds affect the perception of moving visual stimuli with lags of 150 ms (Sekuler et al., 1997).

While these data are suggestive of a new multisensory phenomenon: “temporal ventriloquism”, it is important to note that an alternative explanation may account for the data. In particular, an alerting or warning effect triggered by the first sound could have influenced TOJ performance (Bernstein, & Edelstein, 1971; Posner, 1978; Posner, Nissen and Klein, 1976). Alerting involves a change in the internal state that follows the presentation of a warning signal, and results in a subsequent change in behavior (Fernandez-Duque & Posner, 1997). According to this account, presenting a sound before the first light simply alerted the participants that visual stimuli were about to be presented, thus enabling them to be better prepared to perform the task. Consequently, a sound before the first light leading to increased alerting may explain the improved TOJ performance.

Experiment 2

An alerting account predicts that performance on the TOJ task improves when the first sound precedes the two lights. The second experiment set out to test this prediction by varying independently the temporal relationships between the first sound and light, and the second sound and light (see Figure 2). Specifically, the first sound could appear at the same time as the first light or lead by 100 ms. Similarly, the second sound could appear at the same time as the second light, or trail by 100 ms. This manipulation served two purposes. First, it allowed the alerting hypothesis to be tested directly. Performance should improve only when the first sound precedes the
first light. Second, it is possible that even if alerting is affecting performance, temporal ventiloquiism is also playing a role. By controlling for any alerting effects of the first sound, one can examine whether there is any additional benefit to TOJ performance when the second sound trails the second light.

Method

The method was the same as in Experiment 1 with the following exceptions. 

Participants. Sixteen undergraduate Psychology students (14 female and 2 male) took part in the experiment for course credit. The mean age was 19.5 years old (standard deviation of 1.2 years of age). All had normal or corrected-to-normal vision and hearing. Two participants were replaced due to a failure to perform the task.

Design and Procedure. There were four conditions in a 2x2 design: the first sound simultaneous with the first light or leading by 100 ms, and the second sound simultaneous with the first light or trailing by 100 ms. Thus, the baseline condition and both sounds apart were similar to those used in the previous experiment. But in addition, a condition with the first sound leading and the second simultaneous, and a condition of first sound simultaneous and the second trailing, were also used (see Figure 2). The lag conditions and light-light SOAs were presented in random order as before.

Results

The JNDs for all cells were computed for each participant as in Experiment 1. Looking at the columns in Table 2, one sees that the JNDs are much lower than baseline (sounds simultaneous) when the second sound is trailing. In contrast, there does not seem to be an effect for the leading first sound. These observations were
confirmed when the JNDs were submitted to a two-way repeated measures ANOVA with the factors of First sound (simultaneous versus leading the first light) and Second sound (simultaneous versus trailing the second light). The results indicated that the effect of the First sound was not significant, $F<1$, and the Second sound was significant, $F(1,15)=11.1, p<.01$, with no interaction ($p>.15$).

Discussion

The present data indicate that alerting cannot account for the improved performance observed in Experiments 1 and 2. Contrary to an alerting account, presenting a sound before the first light does not improve performance relative to baseline. In fact, it seems as if the entire TOJ performance improvement is due to the presentation of the second sound after the second light. In terms of temporal ventriloquism, it would appear that the second sound biases the perception of the second light, but the first sound does not affect the first light. Why should this asymmetry exist?

A possible explanation is suggested by the fact that the temporal resolution of sound is superior to that of vision. When the sound precedes the first light, participants will have perceived the time of the first sound before a light has been presented or perceived. However, when the sound trails the second light, participants may well be encoding the second light and trying to mark its time of occurrence when the second sound appears. If this second sound overlaps with the ongoing temporal perception of the light, the perceived time of the second light may be biased by the temporal perception of this trailing sound. This explanation dovetails with the
observation of Dixon and Spitz (1980) who found that participants could integrate sounds and visual stimuli over a larger range of asynchronies when the sounds followed the visual stimuli compared to when the sounds preceded the visual stimuli (see also McGrath & Summerfield, 1985; Lewkowicz, 1996).

In summary, the current results indicate that alerting is not the underlying cause of the temporal ventriloquism effect. There is, however, an alternative explanation for the lack of an effect of the first sound and superior performance by the trailing second sound. If a sound that shortly precedes or is simultaneous with a light actually interferes with TOJ performance then one would expect precisely the present pattern of results (e.g., because the sound is distracting). Accordingly, the timing of the first sound would have no effect on performance because it was equally interfering with the first light when it was simultaneous and when it was leading by 100 ms. Furthermore, the timing of the second sound would lead to improved performance when it was trailing the second light because its ability to interfere with performance declined as the lag increased. The predictions of this alternative account are tested in the following experiment.

Experiment 3

This study was similar to Experiment 2 but tested a range of lags beyond 100 ms. The aim of this experiment was manifold. One aim was to test whether the lack of an effect of the first sound was due to interference. If the first sound interferes with the visual perception, then as the lag between the sound and light grows larger the negative influence of the sound should diminish and performance should improve. A second aim was to test whether the improved performance of the second sound was
also due to interference. If this were the case, then the second trailing sound would lead to improved performance across all lags. If improved performance were due to a multisensory mechanism however, performance should return to baseline levels as the second sound trails the second light at a range that does not afford audiovisual integration. The results of Experiment 1 are consistent with this latter prediction and suggest that the performance improvement produced by a trailing second sound began to weaken as the lag exceeded 150 ms. However in Experiment 1, the lags of both sounds were manipulated concurrently. Hence, a measure of the time course of the second sound alone is crucial to testing the predictions of the two accounts. Finally, it was examined whether the results of the Experiment 2 were robust and would be replicated within the context of a larger variety of conditions and additional practice.

Method

The method was the same as in Experiment 1 with the following exceptions.

Participants. Sixteen undergraduate students (11 female and 5 male) took part in the experiment for course credit. The mean age was 19.4 years (SD=1.3 years). All had normal or corrected-to-normal vision and hearing.

Design and Procedure. There were nine conditions varying the timing of the two sounds relative to the lights. One condition was baseline with sounds and lights occurring simultaneously. In four conditions the first sound preceded the first light by 100, 200, 450 and 600 ms while the second sound was simultaneous with the second light. In the remaining four conditions, the first sound was simultaneous with the first light and the second sound trailed 100, 200, 450 and 600 ms after the second light. The experiment comprised 2 sessions of 756 trials each, in which each combination of lag condition by light-light SOA was repeated 12 times, so that data from both
sessions provided the same number of measures per condition as the previous
experiments.

Results

The JNDs for all cells were computed for each participant as in the previous
experiments (see Table 3). A repeated-measures ANOVA revealed a significant effect
of sounds on performance, $F(8,20)=5.04$, $p<.01$. Tukey post-hoc analyses revealed
that the mean JND was significantly lower than baseline when the second sound
trailed the second light by 100 ms, replicating Experiment 2. No other conditions
were significantly different from baseline.

Discussion

The present experiment replicated our finding in Experiment 2 in that an
asynchrony between the first sound and light did not improve performance at any lag.
Most importantly, the data excluded an explanation that the sounds were having a
negative or interference effect. The first sound did not become ‘less interfering’ as the
temporal lag increased. This was not just a matter of a lack of statistical power, as
examination of Table 3 reveals that the JND when the first sound led by 200 ms is
actually larger, although not significantly so. In addition, examination of the time
course of the second sound excludes an interference hypothesis. When the second
sound trailed after the second light, the findings confirmed that improved performance
persisted to 100 ms but not at 200 ms and beyond. This finding negates the hypothesis
that the sound became less confusing as the lag increased. As noted previously, the
range in which improved performance was found is compatible with the range in
which many multisensory effects occur, presumably because the sounds and lights are
being integrated (e.g., Bertelson, 1998).

Finally, it is worth noting that the overall magnitude of the JNDs was much
lower than in the previous two experiments. It would appear that the second testing
session that was required because of the many conditions in Experiment 3 enabled
participants to improve performance, indicating that the visual TOJ task is subject to
practice effects. Consequently, the magnitude of the effect was somewhat smaller.

Experiment 4

The previous two experiments demonstrated that the second sound trailing the
second light is producing the performance improvement in the visual TOJ task. The
first sound seems to be having no effect. Experiment 4 questioned whether the second
sound by itself was sufficient to induce the phenomenon. Possibly, the context of
having two sounds is crucial to the temporal ventriloquism phenomenon. Without the
first sound, it is unclear whether the system would pair the 'second' sound with the
second light. This could be constituted as part of the crossmodal pairing between the
lights and the sounds (Bertelson, 1998). Is the pairing necessary for the temporal
ventriloquism to emerge?

To answer this question, the condition of first sound simultaneous with the
first light and second sound trailing the second light by 100 ms was examined in
detail. Specifically, a condition where a single sound presented simultaneous with the
first light and a condition where a single sound presented 100 ms after the second
light were added to the design. If crossmodal pairing is of no importance, then the
beneficial effect of the ‘second’ sound would emerge regardless of the presence of the first sound. Alternatively, if crossmodal pairing is a prerequisite for temporal ventriloquism, only when the first sound is present will a performance improvement be observed.

The presence of two sounds can be construed as a structural property of the stimuli. Structural properties refer to abstract attributes of the stimuli such as their position in space and their number (Radeau & Bertelson, 1987). Radeau and Bertelson (1987) replicated a study by Thomas (1941) in which they explored whether structural factors influence the ventriloquist effect. They examined the magnitude of the ventriloquist effect when the visual and auditory stimuli were continuous, a series of four slow pulses or a series of nine fast pulses, all of the same duration. They found that the ventriloquist effect was maximal when the biasing stimulus was of the same construct as the target stimulus. Given this importance of congruence between what is seen and heard in spatial ventriloquism, crossmodal pairing might also be important to temporal ventriloquism.

Method

The method was the same as in Experiment 1 with the following exceptions.

Participants. Sixteen undergraduate students from the same subject pool as Experiment 1 (12 female and 4 male) took part in the experiment for course credit. Mean age was 18.9 years with a standard deviation of 1. All had normal or corrected-to-normal vision and hearing.

Design and Procedure. Four conditions were tested: (1) baseline (sounds simultaneous with the two lights), (2) first sound simultaneous with the first light and second sound trailing the second light by 100 ms, (3) first sound simultaneous (no
second sound) and (4) no sound with the first light but a single sound trailing the second light by 100 ms.

Results

The JNDs for all cells were computed for each participant as in the previous experiments (see Table 4). The repeated-measures ANOVA resulted in a significant effect for sound conditions, $F(3,45)=9.9$, $p<.001$. Tukey post-hoc comparisons verified that first sound simultaneous-second sound trailing had a significantly smaller JND than the baseline condition and second sound trailing conditions ($p<.05$). No other comparisons reached significance.

Discussion

The results replicate the finding from Experiments 2 and 3 that performance improves when the second sound trails the second light by 100 ms. Crucially, if only a single sound trails the second light there is no improvement in performance. In fact, the JNDs for baseline and single sound trailing the second light are comparable. Therefore, the first sound appears to be necessary for the second sound to ventriloquise the onset of the second light. This notable finding would suggest a role for the overall ‘gestalt’ of the audio-visual stimuli in sound-light temporal capture. Thus, it appears that aspects such as the correspondence between the sounds and lights are crucial for the perceptual system’s pairing of the second sound and light
together (cf., Bertelson, 1998). In other words, in the context of two lights, two sounds are required.

The present results also provide additional evidence against the interference hypothesis that was tested in the previous experiment. If the first or second sound were indeed confusing, then the conditions with a single sound should improve performance beyond baseline. Neither of the two conditions (first simultaneous, or 'second' trailing) improved performance significantly relative to baseline.

General Discussion

The data from the present study demonstrate that sounds can alter the performance in a visual TOJ task. In Experiment 1 sounds presented before the first light and after the second light improved performance relative to baseline (two sounds presented simultaneously with the lights), as if the sounds pulled the perceived onsets of the lights apart in time. This effect was robust and was found to extend to lags of 150 ms, and at the lag of 225 ms the effect became marginal. Experiment 2 ruled out an alerting explanation, as only the second sound trailing the second light was found to improve performance. That is, in contrast to the alerting prediction, the first sound preceding the first light did not lead to the improved performance. In Experiment 3 the timing of the first light was again found to have no effect, even when presented up to 600 ms before the first light. The influence of the second sound disappeared as the lag grew beyond 100 ms. The results ruled out an interference account claiming that sounds preceding or simultaneous with the sounds result in a performance decrement. The findings of Experiment 3 are not in accord with this interference hypothesis as it would have predicted an effect for the first sound would emerge at longer lags, and the effect of the second sound would persist at all lags. The third experiment also
replicated Experiment 2, while providing the time course of the effect. Finally, in Experiment 4 a single sound trailing after the second light was found to be insufficient for the temporal capture to emerge. Although the timing of the first sound was not important, its presence was required, presumably so that a clear correspondence between the two lights and two sounds could emerge. Thus, the first sound would be associated with the first light being in close temporal proximity to it, and the second sound would be associated with the second light, being closer in time to it. In conclusion, our findings indicate that sounds can bias visual perception in the temporal domain (Welch, 1999). Most importantly, the current study ensured a direct measure of temporal perception and ruled out alerting and distractability of the sounds as the underlying mechanism for the effect. This study also allows us to generalize the conclusions regarding auditory temporal capture of vision to simple auditory stimuli with no rhythmic or linguistic factors.

Recently, an abstract by Scheier, Nijhawan and Shimojo (1999) detailing findings similar to some of the data in Experiment 1 came to our attention. In a single experiment, they found that sounds placed before the first light and after the second light improved TOJ performance. They concluded that the temporal placement of the sounds could change the temporal resolution between the lights, although a mechanism was not suggested. There are however, several noteworthy differences between the two studies. Most importantly, they used different baselines in which the auditory stimuli were unequated with the experimental condition. Because their baseline measures involved no sounds or a single sound, alternative accounts cannot be ruled out. In addition, the lags between the sounds and lights were quite different. Scheier et al. (1999) used a 40 ms lag, which would appear to be less than optimal given the time course of the data found in Experiments 1 and 2. In addition, the lights
in their presentation were placed on the horizontal plane making it a left/right choice
task and the spatial source of the sound was central, allowing for a spatial
ventriloquist effect to influence the results. It is also interesting to note that they found
no difference between their various baselines, in which a single sound could be
presented before or after the lights. This is in accordance with the findings of
Experiment 4 in which when only a single sound is present no temporal ventriloquism
will take place. Nevertheless, despite the above differences the similarity of the
general finding lends support to the notion that this effect is robust and replicable.

The benefits of generalizing from the spatial to the temporal domain

The current study emphasized that the temporal effect was analogous to spatial
ventriloquism, where lights influence the spatial localization of sounds. However,
there are several noteworthy differences between the two types of effects. Whereas
classic ventriloquism is typically measured by an absolute judgment of a single sound,
the temporal capture was presently measured by a comparative judgment between two
lights. Since the tasks are different, there may be additional processes in each that
affect behavior. Furthermore, the inherent characteristics of vision and audition
themselves are not symmetrical. Therefore, some issues that should be of concern for
one domain are not necessarily relevant for the other. For example, the addition of an
irrelevant light to a sound does not cause us to suspect increased alerting, unlike the
addition of an irrelevant sound to a light (Fernandez-Duque & Posner, 1997; Posner et
al., 1976).

Nonetheless, the analogy does provide several interesting avenues of research.
For example, the reverse spatial ventriloquist effect, a sound influencing the
localization of a visual stimulus is either very weak or not present at all (Bertelson, 1998; Radeau & Bertelson, 1974). It will be of interest to examine whether this is also the case for the temporal ventriloquist effect. If indeed audition has superior temporal resolution, one would not expect vision to influence auditory temporal perception (Welch & Warren, 1980). In particular, irrelevant visual events would not bias auditory temporal order judgments. At present there are conflicting studies regarding the outcome of such an experiment. Changes in the rate of a flashing light do not typically affect the perceived auditory flutter rate or are very weak when detected, indicating that auditory and visual driving are not symmetrical (Gebhard & Mowbray, 1959; Welch et al., 1986). In contrast, Fendrich and Corballis (2001) have found that a repetitive visual flash biased the perception of a click. However, this capture was smaller than the capture of the flash by the click, and was subject to confounds stated in the introduction section. It would be of interest to distinguish which of the outcomes would be the case for temporal ventriloquism so as to test the hypothesis that only the more appropriate, accurate modality influences the lesser one (Welch, 1999; Welch & Warren, 1980).

The ventriloquist effect has recently been determined to stem primarily from perceptual interactions between the modalities (Bertelson, 1999; Bertelson & Aschersleben, 1998). Can the temporal capture as seen in the present case, be classified as a sensory (or structural) or as a cognitive phenomenon (Welch, 1999; Welch & Warren, 1980)? Sensory factors are the inherent low-level properties of the stimuli, while cognitive factors include expectations and other top-down aspects (Bertelson, 1999). Several observations support the notion that the multisensory interactions found here are primarily structural. First, the sounds could not bias the actual response (which of the two lights appeared first) given that they were spatially
uninformative and that the same sound was used throughout the entire study. Second, the conditions and SOAs were always randomly ordered so that it would be implausible that participants developed implicit or explicit strategies even if they did become aware of the different contingencies between the sounds and lights. Finally, the experimental situation was simple and typically comprised of only two sounds and two lights with no complex or repetitive stimuli. Hence, context dependent expectations were kept to a minimum.

Crossmodal temporal capture was measured in the present study using JND, how well participants could temporally discriminate between two lights. Studies in other modalities and paradigms indicate that the capture is rarely if ever complete (e.g., Welch & Warren, 1980). For example, in ventriloquism, a light will influence the reported location of a sound even when the two are not perceptually fused (Bertelson & Radeau, 1981). It is unclear at present whether the lights were perceived to originate at the time the sounds appeared, thus being "perceptually fused" at all times (Bertelson, 1998). It is possible that there was a fusion between the second light and sound but that it was not always complete and the light could be perceived as merely appearing later than it actually did but not at the time of the sound. Unfortunately, the extent of the fusion could not be estimated in the present study. In addition, there is the related issue of whether participants actually phenomenologically perceived the second light as occurring later in time, or the time between the two lights as longer. In other words, to what extent did the sound truly capture or alter the temporal perception of the light? This is to some extent independent of fusion. Even if the fusion was not complete, the sound could have caused the perception of the second light to be shifted in time. The present design of the experiments did not allow us to directly test or measure this temporal 'shift' (see
the Appendix for a more detailed attempt to examine this issue). Both perceptual fusion and the temporal shift of perception should be examined in greater detail in the future.

There are many additional questions that remain to be answered. For example, if indeed the phenomenon described above is temporal ventriloquism, then introducing two sounds intervening between two lights should lead to a decrease in performance. Preliminary evidence suggests that this is precisely the case. Furthermore and in keeping with the finding of Experiment 4, a single sound midway between the two lights does not change performance relative to baseline. Again, it seems as if the entire gestalt of the stimulus set is important for the sounds to influence the perceived timing of the lights.

The present task could also clarify some of the general theoretical issues relating to multisensory interactions. For example, the importance of attentional and top-down mechanisms in determining perceptual capture or fusion are at present unclear. Evidence from the ventriloquist effect indicates that while the effect is in part perceptual (Bertelson & Aschersleben, 1998; Radeau & Bertelson, 1977), it is also subject to a strong top-down or cognitive biases (Bertelson & Radeau, 1981; Choe, Welch, Gilford & Juola, 1975). By examining whether performance in the temporal task is also subject to the same effects, the role of such cognitive factors can be better understood. Interestingly, visual driving of auditory stimuli was found to increase when participants were repeatedly asked to attend to the visual stimuli (Welch, 1999). Additionally, future research should question whether space plays an important role in determining the extent of temporal capture between modalities, in the same sense as time plays a role in determining the extent of spatial capture. However, it is important to remember that there are fundamental differences between how the brain deals with
spatial and temporal information. The extent to which the analogy of space and vision versus time and audition is true is controversial (Handel, 1988; Kubovy, 1988). All the same, the analogy can advance the research by allowing us to ask the same questions in the spatial and temporal domains and to examine whether the results from one generalize to the other.

Finally, the exact nature of the underlying processes responsible for the temporal capture is as of yet unclear. One possibility is a temporal cross capture mechanism of audition and vision that resolves crossmodal discrepancies (Fendrich & Corballis, 2001). The current effect could be mediated by such a mechanism. Thus, the onset of the irrelevant second sound may have altered the perceived onset of the second light, moving its perception to a later time. However, much remains to be determined regarding the exact nature of such a mechanism as well as the factors that instigate it. For example, a sound did not bias the perceived timing of the first light suggesting that such a mechanism cannot predate an already perceived event, at least not with the lags and stimuli used in the present study.

Conclusions

In conclusion, a new multisensory phenomenon of temporal capture has been demonstrated. In the temporal domain audition influences visual perception when the two sources of information are in conflict. The study of multisensory integration in time is as important as the study of spatial integration in order to understand how the brain achieves a coherent representation of the environment. Much research has focused on understanding the ventriloquist effect and spatial multisensory interactions
at the phenomenological, behavioural and neurophysiological levels. Possibly, the study of both spatial and temporal crossmodal interactions will lead to a greater understanding of the mechanisms underlying each.
Appendix

If the perception of the second light were shifted later in time because of the second sound, one would expect the point of simultaneity (PSS) to be shifted as well. The PSS is the lead-time by which one of the targets has to be presented in order for both of the targets to appear simultaneous to the observer (see Coren et al., 1999). For example, one can observe a PSS shift when judging the temporal order of a dim light and a bright light. The dim light has to be presented before the bright light in order for both to be perceived as simultaneous. Unfortunately, because of the particular design of the present experiment, such a shift in the PSS could not be computed in the typical manner. As only the second sound affected performance, it is as if one manipulated brightness but could only make the second light dimmer. Therefore, only half of the function plotting accuracy as a function of SOA can be obtained. Nonetheless, the accuracy data in Experiments 2 to 4 were plotted as a function of SOA collapsed across top and bottom light for both baseline and first sound simultaneous second trailing conditions. Then both a linear function and a logarithmic function were fitted to each of the conditions and the point where accuracy was .5 (analogous to the PSS) was extrapolated. Figures 3a, 3b and 3c display the linear functions for Experiments 2, 3 and 4, respectively. Figures 4a, 4b and 4c display the logarithmic functions for Experiments 2, 3 and 4, respectively. In all graphs only the baseline condition and the first simultaneous second trailing conditions are shown. The mean $R^2$ for the linear fits was .89 indicating a reasonable fit to the data. The mean $R^2$ for the logarithmic fits was .94, which was significantly higher than the linear fit, at the .05 level. For both linear and logarithmic functions, when the second sound is trailing the point where the graph crosses the x-axis is to the left of baseline. This is as if the second light (or 'dimmer' light) has to appear before the first light by a larger SOA compared to
baseline in order to be perceived as simultaneous. However, the effect is trivial in the case of the logarithmic functions, while on the order of 15 to 25 ms in the case of the linear functions. Therefore, only limited support for an effect of temporal capture on the PSS measure could be found.
References


Footnotes

1The JNDs were initially computed by fitting both logistic and linear functions. The results using the logistic and linear JNDs, as well as the logistic and linear slopes were the same, and thus only the first is reported. The logistic function provided a significantly better fit to the data than did the linear function (p<.05) and is used as our principle function for the remaining experiments. The JND and not slope data are presented due to their straightforward theoretical interpretation.

2For some participants in Experiments 2-4, a logistic function could not be fitted for all the conditions. In order not to lose the remaining data from those participants, the missing cell was calculated using a regression equation computed using the participants who had JND values for all their conditions. In Experiments 2 and 4, it was used to predict the value of one participant in the first sound simultaneous second sound apart condition (1.56% of the data in each experiment). In Experiment 3 it was used to predict the value of one participant in the sound preceding by 200 ms condition (0.7% of the data). In all cases the addition of the data did not alter the results of the analyses.
Table 1. Mean Just Noticeable Differences (JNDs) in milliseconds for Experiment 1. The onsets of the sounds were 0 to 225 ms before the first light and after the second light. Note that better performance is indicated by a smaller JND.

<table>
<thead>
<tr>
<th>Onsets of Sounds Relative to Lights</th>
<th>0 ms</th>
<th>75 ms</th>
<th>150 ms</th>
<th>225 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Baseline)</td>
<td>62</td>
<td>44</td>
<td>44</td>
<td>51</td>
</tr>
</tbody>
</table>
Table 2. Mean Just Noticeable Differences (JNDs) in milliseconds for Experiment 2.

The onset of the first sound could either precede the first light by 100 ms or occur simultaneously with the light. The onset of the second sound could occur simultaneously with the onset of the second light or lag 100 ms after it.

<table>
<thead>
<tr>
<th>Second Sound</th>
<th>Simultaneous</th>
<th>Trailing</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Sound</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simultaneous</td>
<td>73</td>
<td>61</td>
</tr>
<tr>
<td>Preceding</td>
<td>71</td>
<td>64</td>
</tr>
</tbody>
</table>
Table 3. Mean Just Noticeable Differences (JNDs) in milliseconds for Experiment 3.

The first sound could precede the first light by 600, 450, 200 and 100 ms while the second sound was kept simultaneous with the second light. The second sound could trail after the second light by 100, 200, 450 and 600 ms, while the first sound appeared simultaneously with the first light.

<table>
<thead>
<tr>
<th>Onset of Sounds Relative to Lights</th>
</tr>
</thead>
<tbody>
<tr>
<td>600 ms  450 ms  200 ms  100 ms  Baseline  100 ms  200 ms  450 ms  600 ms</td>
</tr>
<tr>
<td>Precedin Precedin Precedin Precedin Trailing Trailing Trailing Trailing</td>
</tr>
<tr>
<td>g       g       g       g</td>
</tr>
<tr>
<td>40      42      51      45      45      34      36      44      42</td>
</tr>
</tbody>
</table>
Table 4. Mean Just Noticeable Differences (JNDs) in milliseconds for Experiment 4. Sounds were presented simultaneous with the first light and trailing 100 ms after the second light.

<table>
<thead>
<tr>
<th>Onsets of Sounds Relative to Lights</th>
<th>Baseline</th>
<th>First Simul.</th>
<th>First Simul.</th>
<th>Second Trailing</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Both Simul.)</td>
<td>Second Trailing</td>
<td>80</td>
<td>60</td>
<td>70</td>
</tr>
</tbody>
</table>
Figure 1. A schematic illustration of the events in a trial: the lights are separated by an SOA ranging from 12 to 144 ms (see methods). In the baseline condition the sound onsets are simultaneous with the lights (i.e., 0 ms lag). In the sounds apart the first sound precedes the first light by a lag ranging from 75 to 225 ms, and the second sound trails after the second light by the same lag.
Figure 2. A schematic illustration of the conditions of Experiment 2, where the timing of the sounds were manipulated independently one from the other. In ‘first preceding’ condition, the first sound preceded the first light by 100 ms and the second was simultaneous with the second light. In the ‘second trailing’ condition, the first sound was simultaneous with the first light, and the second sound trailed by 100 ms after the second light. These two conditions depicted in Figure 1 were also included in the experiment, with the lags of both sounds apart also being 100 ms.
Figure 3a. Accuracy is plotted as a function of SOA between the two lights for baseline and first simultaneous second trailing in Experiment 2. The fitted linear functions indicate that the second sound trailing causes the second light to appear later in time.
Figure 3b. Accuracy is plotted as a function of SOA for baseline and first simultaneous second trailing by 100 and 200 ms in Experiment 3. The fitted linear functions indicate that the second sound trailing causes the second light to appear later in time.
Figure 3c. Accuracy is plotted as a function of SOA for baseline and first simultaneous second trailing in Experiment 2. The fitted linear functions indicate that the second sound trailing causes the second light to appear later in time.
Figure 4a. Accuracy is plotted as a function of SOA between the two lights for baseline and first simultaneous second trailing in Experiment 2. The fitted logarithmic functions indicate that the second sound trailing causes the second light to appear only slightly later in time, providing only limited support to the 'temporal shift' hypothesis.
Figure 4b. Accuracy is plotted as a function of SOA between the two lights for baseline and first simultaneous second trailing by 100 ms and 200 ms in Experiment 3. The fitted logarithmic functions indicate that the second sound trailing causes the second light to appear only slightly later in time, providing only limited support to the 'temporal shift' hypothesis.
Figure 4b. Accuracy is plotted as a function of SOA between the two lights for baseline and first simultaneous second trailing in Experiment 4. The fitted logarithmic functions indicate that the second sound trailing causes the second light to appear only slightly later in time, providing only limited support to the ‘temporal shift’ hypothesis.