

# Auditory grouping occurs prior to intersensory pairing: evidence from temporal ventriloquism

Mirjam Keetels · Jeroen Stekelenburg · Jean Vroomen

Received: 13 September 2006 / Accepted: 9 January 2007 / Published online: 6 February 2007  
© Springer-Verlag 2007

**Abstract** The authors examined how principles of auditory grouping relate to intersensory pairing. Two sounds that normally enhance sensitivity on a visual temporal order judgement task (i.e. temporal ventriloquism) were embedded in a sequence of flanker sounds which either had the same or different frequency (Exp. 1), rhythm (Exp. 2), or location (Exp. 3). In all experiments, we found that temporal ventriloquism only occurred when the two capture sounds differed from the flankers, demonstrating that grouping of the sounds in the auditory stream took priority over intersensory pairing. By combining principles of auditory grouping with intersensory pairing, we demonstrate that capture sounds were, counter-intuitively, more effective when their locations differed from that of the lights rather than when they came from the same position as the lights.

**Keywords** Multisensory perception · Auditory grouping · Intersensory pairing · Temporal order judgment · Temporal ventriloquism

## Introduction

Sense organs like the ears and eyes are continuously bombarded with information. Yet, observers perceive distinct objects or events. The way information is assigned to objects has, for vision, been described with

Gestalt principles like ‘similarity’, ‘good continuation’, or ‘common fate’, and similar principles have also been discovered for audition (Bregman 1990). It occurs, for instance, when a sequence of alternating high- and low-frequency tones is played at a certain rate. When the frequency difference between the tones is small, listeners group the tones into a single stream, but at bigger frequency differences, the sequence splits into two streams, one high and one low in pitch. Typically, these grouping principles apply within a single modality like vision or audition. However, sense organs not only work in isolation, they also have to cooperate to form a coherent multisensory representation of the environment. The notion on how information from different sense organs is assigned to a multisensory event is usually referred to as the ‘assumption of unity’. It states that as events from different modalities share more amodal properties, in particular space and time, it is more likely that they originate from a common object or source (e.g. Welch and Warren 1980; Bedford 1989; Stein and Meredith 1993; Radeau 1994; Bertelson 1999; Welch 1999). Following this notion, the assignment of information from different modalities to a single multisensory event will be reduced or absent when stimuli are too far apart in space or time, because in that case two objects or events will be perceived rather than a single multimodal one.

Here, we explored how principles of auditory grouping relate to intersensory pairing. Previous work on this topic suggests that auditory grouping may take priority over intersensory pairing. For example, Vroomen and de Gelder (2000) used a task in which participants had to detect a visual target in a rapidly changing sequence of visual distracters. They observed that a high tone embedded in a sequence of low tones

---

M. Keetels · J. Stekelenburg · J. Vroomen (✉)  
Department of Psychology,  
Tilburg University, Warandelaan 2,  
Tilburg, The Netherlands  
e-mail: j.vroomen@uvt.nl

enhanced detectability of a synchronously presented visual target. There was no such intersensory enhancement when the tone was embedded in a sequence of tones with the same frequency or when the tone was part of a melody. The cross-modal enhancement thus only occurred when the sound segregated from the sound sequence in which it was embedded. Similar results were obtained by Watanabe and Shimojo (2001). They explored how the ‘bounce illusion’ is affected by contextual auditory information. The bounce illusion is a cross-modal phenomenon in which a ‘collision’ sound presented near the crossover of two moving balls enhances the perception of the balls ‘bouncing’, whereas the absence of the sound results in a ‘streaming’ percept. The authors showed a reduction of the bounce illusion when the sounds were embedded in a sequence of similar sounds, as opposed to when the sounds were flanked by sounds of a different frequency.

Here we tested the generality of these findings by examining how auditory grouping affects auditory–visual (AV) pairing in the temporal domain using the so-called temporal ventriloquist effect (Scheier et al. 1999; Fendrich and Corballis 2001; Aschersleben and Bertelson 2003; Bertelson and Aschersleben 2003; Morein-Zamir et al. 2003; Vroomen and de Gelder 2004; Stekelenburg and Vroomen 2005; Vroomen and Keetels 2006). Temporal ventriloquism refers to the phenomenon that when a sound and light are presented at slightly different onset times (usually in the order of  $\sim 100$  ms), the sound will attract the temporal occurrence of the light. This phenomenon can be demonstrated in a visual temporal order judgment (TOJ) task in which participants are presented two lights at various stimulus onset asynchronies (SOAs) and judge, which came first. By presenting a sound before the first and after the second light, the just noticeable difference (JND) improves (i.e., participants become more sensitive), presumably because the two sounds attract the temporal occurrence of the two lights, and thus effectively pull the lights further apart in time (Scheier et al. 1999; Morein-Zamir et al. 2003; Vroomen and Keetels 2006). Judgments about which light came first are therefore more accurate if there is a  $\sim 100$  ms interval between the sounds and lights rather than when the sounds are presented simultaneously with the lights.

Here we asked what happens if the sounds that capture the onset of the lights are assigned to a stream of other sounds with which they form a well-formed sequence. If auditory grouping takes priority over intersensory pairing, one expects an improvement on the visual TOJ task only if the capture sounds segregate from the auditory stream. Alternatively, though,

audio–visual pairing might take priority over auditory grouping in which case observers should improve on the visual TOJ task no matter whether the sounds segregate or not.

In Experiment 1, these predictions were tested with two capture sounds that were embedded in a sequence of flanker sounds, which either had the same or a different frequency. When the frequency of the flanker and the capture sounds were the same, the sequence was heard as a single stream which, following previous findings (Vroomen and de Gelder 2000; Watanabe and Shimojo 2001), should prevent temporal ventriloquism to occur. When the flankers differed from the capture sounds, stream segregation was more likely to occur in which case the two sounds could possibly interact with the lights and thus improve performance on the visual TOJ task. Other stream segregation cues besides the frequency of the flanker and capture sounds were further explored in Experiment 2 (rhythm) and Experiment 3 (sound location).

### Experiment 1: Capture and flanker sounds with the same or different frequency

Participants performed a visual TOJ task in which they decided which of two lights appeared first. Two task-irrelevant high tones were presented either simultaneously with the lights (i.e., the  $\sim 0$  ms AV interval), or the first tone was presented  $\sim 100$  ms before the first light and the second tone  $\sim 100$  ms after the second light (i.e., the  $\sim 100$  ms AV interval). The two high tones were embedded in a sequence of other tones, which either had the same (high) or a different (low) frequency (see Fig. 1a for a schematic overview of the conditions).

#### Method

**Participants** Thirteen students from Tilburg University were given course credits for their participation. All reported normal hearing and normal or corrected-to-normal seeing. They were tested individually and were unaware of the purpose of the experiment. The experimental procedures were approved by the Institute and were in accordance with Declaration of Helsinki.

**Stimuli** Two auditory stimuli, a low (1,500 Hz) and a high (3,430 Hz) pure tone of 3 ms at 72 dB(A) were used that clearly differed in pitch. The sounds were presented via a hidden loudspeaker located at eye-level, at central location and at 90 cm distance. Visual

stimuli were presented by two red LEDs (diameter of 1 cm, luminance of 40 cd/m<sup>2</sup>), positioned 5° below and above the central loudspeaker. A small green LED was placed at the center of the loudspeaker and served as a fixation point (see Fig. 1b for a schematic set-up). Trials in the ~0 ms AV interval consisted of a sound sequence of 40 sounds in which the interval between the successive tones was equal to the SOA between the two lights. The lights were presented simultaneously with the 25th and 26th sound. Trials in the ~100 ms AV-interval consisted of a tone sequence of 15 sounds with the interval between the tones equal to the SOA of the lights plus 200 ms. The two lights were presented in the middle of the temporal gap, ~100 ms after the 10th and ~100 ms before the 11th sound.

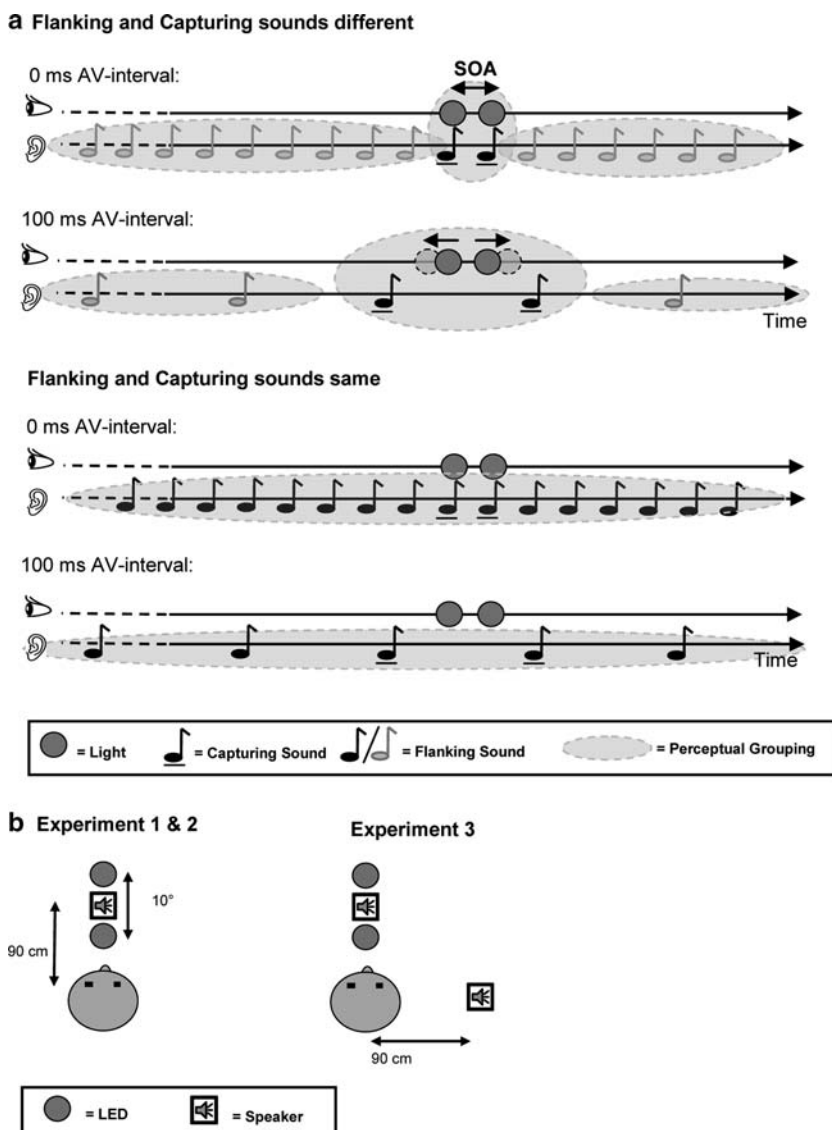
**Design** The experiment had three within-subjects factors: Frequency of the flanker sounds (same or

different frequency as the capture sounds), the AV-interval between the capture sounds and the lights (~0 or ~100 ms), and the SOA between the two lights (-75, -60, -45, -30, -15, +15, +30, +45, +60, and +75 ms; with negative values indicating that the lower light was presented first). These factors yielded 40 equi-probable conditions, each presented 20 times for a total of 800 trials (10 blocks of 80 trials each).

**Procedure** Participants sat at a table in a dimly lit and soundproof booth. The fixation light was illuminated at the beginning of the experiment, and participants were instructed to maintain fixation on this central green LED during testing. The participant’s task was to judge whether the lower or the upper light was presented first. Responses (unsped) were made by pressing one of two designated keys with the right thumb (lower light first) or right index (upper light first). Whenever a

**Fig. 1 a** A schematic illustration of a trial. Lights were presented with a particular SOA ranging between -75 and 75 ms, with negative values indicating that the lower light was presented first. Two capture sounds were presented either simultaneously with the lights (~0 ms AV-interval), or ~100 ms before the first and ~100 ms after the second light (~100 ms AV-interval). The capture sounds were embedded in a sequence of flanker sounds which either had the same or a different frequency (Exp1), rhythm (Exp 2), or location (Exp 3). Perceptual grouping of the sounds is illustrated by the grey ovals. Temporal ventriloquism only occurred when the capture sounds were different from the flanker sounds at the ~100 ms interval, as illustrated on the second line.

**b** Schematic set-up of Experiment 1, 2 and 3



response was detected, both LEDs were turned off and the next trial started after 2,000 ms. A practice block was included consisting of 16 trials in which the four longest SOAs were presented once in each condition. During practice, participants received verbal feedback (“Correct” or “Wrong”).

## Results and discussion

Trials of the practice session were excluded from analyses. The proportion of ‘up-first’ responses was calculated for each condition and converted into equivalent Z-scores assuming a cumulative normal distribution (cf. Finney 1964). For each of the four conditions, the best-fitting straight line was calculated over the ten SOAs. The lines’ slopes and intercepts were used to determine the point of subjective simultaneity (PSS) and the just noticeable difference (JND = 0.675/slope). The PSS represents the average interval by which the upper stimulus had to lead the lower one in order to be perceived as simultaneous. The JND represents the smallest interval between the onsets of the two lights needed for participants to correctly judge which stimulus had been presented first on 75% of the trials. Temporal ventriloquism was measured by subtracting the JND in the ~100 ms AV interval from the ~0 ms AV interval (see Table 1 for the average JNDs).

A 2 × 2 ANOVA with as within-subjects factors the frequency of the flanker sounds (same or different frequency as the capture sounds) and the AV-interval (~0 and ~100 ms) was conducted on the JNDs and PSSs. In the ANOVA on the PSSs, no effect was significant (all *P*’s > 0.30), which is in line with our expectations since no shift towards more ‘up’ or ‘down’ responses was expected. In the ANOVA on the JNDs, the important interaction between the AV-interval and the frequency of the flanker sounds was significant,  $F(1, 12) = 4.90$ ,  $P < 0.05$ . Separate *t* tests comparing

the ~0 ms AV interval with the ~100 ms AV interval showed that JNDs improved by 3.8 ms when the frequency of the capture and flanker sounds differed,  $t(12) = 3.06$ ,  $P < 0.01$ , but no significant difference was obtained (0.5 ms) when the capture and flanker sounds were the same,  $t(12) = 0.71$ ,  $P = 0.49$ . As predicted, temporal ventriloquism thus only occurred when the capture and flanker sounds differed. It seems therefore likely that segregation of the capture sounds from the flankers was necessary before the capture sounds could interact with the lights. When the capture and flanker sounds were the same, auditory grouping thus took priority over AV pairing.

## Experiment 2: Capture and flanker sounds with the same or different rhythm

To further explore the relation between auditory grouping and intersensory pairing, we presented the capture sounds in or out of rhythm with the flankers. Rhythm is, besides frequency, another important auditory segregation cue (Bregman 1990). It was expected that sounds presented out of rhythm would segregate from the sound sequence, thus enhancing performance on the visual TOJ task. Capture sounds presented in rhythm should not segregate from the auditory stream, and they should thus have no effect on the visual TOJ task.

## Method

**Participants** Twenty new students participated.

**Stimuli and design** Stimuli and design were as in Experiment 1, except that the time interval between the capture and flanker sounds was varied rather than their frequency. The auditory stimuli consisted of 5 m s

**Table 1** Mean just noticeable differences (JND) in ms, and standard errors of the mean (in parentheses) of Experiment 1 and 2

Experiment	AV-interval (ms)	Flanker sounds			
		Same as capture sounds		Different from capture sounds	
		JND	TVE	JND	TVE
Exp 1 (Frequency)	0	21.0 (0.8)	0.5	22.1 (1.1)	3.8*
	100	20.5 (1.1)		18.3 (0.7)	
Exp 2 (Rhythm)	0	29.1 (2.5)	-3.7	23.9 (1.9)	3.6*
	100	32.8 (2.5)		20.3 (0.8)	

Capture Sounds Presented at ~0 or ~100 ms audio–visual intervals; flanker Sounds with the same or different frequency (Exp 1) or rhythm (Exp 2) as capture sounds. The temporal ventriloquist effect (TVE) is the improvement in JND between the ~0 and ~100 ms audio–visual intervals

\* $P < 0.05$

sound bursts presented at 72 dB(A). When the capture sounds were presented in rhythm with the flankers, the same interval between consecutive sounds in the sequence was used as in Experiment 1 (i.e., SOA between the two lights +  $2 \times$  AV-interval). When the capture sounds were presented out of rhythm, the interval between the capture and flanker sounds was increased, such that there was a short pause before the first and after the second capture sound. In the  $\sim 0$  ms AV-interval condition, the two longer intervals were  $8 \times$  SOA of the lights, in the  $\sim 100$  ms AV-interval condition they were  $17 \times$  SOA of the lights +200 ms.

## Results and discussion

In the 2 (Capture sounds in or out of rhythm)  $\times$  2 (AV-interval  $\sim 0$  or  $\sim 100$  ms) ANOVA on the PSSs, no effect was significant (all  $P$ 's  $> 0.30$ ). The same ANOVA on the JNDs showed that the important interaction between AV-interval and rhythm was significant,  $F(1, 19) = 15.35$ ,  $P < 0.001$ . Separate  $t$  tests showed that the 3.6 ms temporal ventriloquist effect (lower JNDs in the  $\sim 100$  ms AV interval rather than  $\sim 0$  ms) of sounds presented out of rhythm was significant,  $t(19) = 2.37$ ,  $P < 0.05$ . Performance got actually worse ( $-3.7$  ms) when the capture sounds were presented in the same rhythm as the flanker sounds,  $t(19) = -2.15$ ,  $P < 0.05$ . Capture sounds thus again only improved performance on the visual TOJ task when they segregated from the flanker sounds.

### Experiment 3: Capture and flanker sounds from the same or different location

It is known that auditory stream segregation may also occur when there is a difference in the location of consecutive sounds (Bregman 1990). In Experiment 3, we therefore varied the location of the capture and flanker sounds. The sounds could emanate either from a central loudspeaker near the two lights, or from a lateral loudspeaker on the far left or far right. If intersensory pairing only occurs when the capture sounds segregate from the flankers, then temporal ventriloquism should be obtained when the locations of the capture and flanker sound differ, but not when they are the same.

This set-up also allowed us to explore whether spatial disparity between the capture sounds and lights affects intersensory pairing. The common notion on intersensory pairing states that commonality in space between the auditory and visual signal matters. However, in contrast with this notion, it has recently been shown that temporal ventriloquism may not be affected

by spatial discordance between the sounds and lights. In a study by Vroomen and Keetels (2006), it was shown that there were equal amounts of temporal ventriloquism when the two capture sounds came from the same or a different position as the lights, when the sounds were static or moved, or when the sounds and lights came from the same or opposite sides of fixation. Assuming that these results would be replicated in the present set-up as well, we expected sound location to be unimportant for intersensory pairing. Equal amounts of temporal ventriloquism were therefore expected from segregated sounds presented from the central location (near the lights) and the lateral location (far from the lights).

The notion that sound location matters for auditory grouping, but not for intersensory pairing also lead to a very counter-intuitive prediction. In case flanker sounds were presented near the central lights, there should more temporal ventriloquism by capture sounds presented from a lateral position than from central position, because only the lateral sounds segregate. With central flankers, there should thus be more temporal ventriloquism when the location of the sounds and lights differ, rather than when they are the same.

## Method

Participants, stimuli and procedures were the same as in Experiment 1, except for the following changes. Eighteen new students from the same subject pool participated. The auditory stimuli consisted of 5 ms sound bursts presented at 72 dB(A), presented from one of two loudspeakers (see Fig. 1b). One speaker was located at central location at eye-level and at 90 cm distance (as in Experiments 1 and 2), the other was located on either the far left or the far right (at  $90^\circ$  azimuth). Four within-subjects factors were used: Location of the two capture sounds (central or lateral), Location of the flanker sounds (same or different position as the capture sounds), the AV-interval between the capture sounds and lights ( $\sim 0$  or  $\sim 100$  ms), and the SOA between the two lights ( $-75$  to  $+75$  ms). The 80 conditions were each presented 20 times in 10 blocks of 160 trials each. In half of the blocks, the lateral speaker was on the left, in the other half it was on the right.

## Results

A  $2 \times 2 \times 2$  ANOVA with as within-subjects factors Location of the two capture sounds (central or lateral), location of the flanker sounds (same or different position as the capture sounds) and the AV interval

( $\sim 0$  and  $\sim 100$  ms) was conducted on the JNDs and PSSs. In the ANOVA on the PSS, there was an interaction between the location of the capture and flanker sounds,  $F(1, 17) = 6.31$ ,  $P < 0.025$ , indicating that there were slightly more ‘up’ responses in trials in which the capture and flanker sounds were presented centrally (mean PSS =  $-1.84$  ms) rather than in the other conditions (mean PSS =  $2.61$  ms), a finding for which there is no clear explanation.

In the ANOVA on the JNDs (see Table 2) there was no main effect of the location of the capture sounds,  $F(1, 17) = 1.67$ ,  $P = 0.21$ , indicating that JNDs were unaffected by whether the capture sounds were presented centrally (near the two lights) or laterally. Most importantly, there was an interaction between the AV-interval and the location of the flanker sounds,  $F(1, 17) = 5.11$ ,  $P < 0.05$ . Separate  $t$  tests confirmed that the temporal ventriloquist effect (better performance at  $\sim 100$  ms rather than  $\sim 0$  ms AV interval) was only significant when the flanker sounds came from a different position than the capture sounds. There was thus no temporal ventriloquism when the capture and flanker sounds came both from central or lateral positions (both  $P$ 's  $> 0.6$ ), while there was a 3.9 ms improvement for central capture sounds with lateral flankers,  $t(17) = 2.175$ ,  $P < 0.05$ , and a 3.1 ms improvement for lateral capture sounds with central flankers,  $t(17) = 2.55$ ,  $P < 0.025$ . These two improvements were not significantly different from each other,  $t(17) = 0.52$ ,  $P = 0.60$ . Moreover, as predicted, with central flankers, temporal ventriloquism by lateral capture sounds was bigger than that from central ones,  $t(18) = 1.798$ ,  $P < 0.05$ .

## General discussion

This study examined how principles of auditory grouping relate to intersensory pairing. Two capture sounds

that normally enhance performance on a visual TOJ task (i.e. temporal ventriloquism) were embedded in a sequence of flanker sounds which could differ in frequency (Exp. 1), rhythm (Exp. 2), or location (Exp. 3). In all experiments, we found that temporal ventriloquism only occurred when the capture sounds differed from the flankers, and there was thus no effect when flanker and capture sounds were the same. Presumably, when the capture sounds differ, they segregate from the auditory stream, and only then they can be paired cross-modally with the lights. When the two capture sounds do not differ from the flankers, they are perceptually grouped in an auditory stream, in which case they lose their saliency and cannot interact cross-modally anymore.

These results are similar to previous findings, which have shown that intersensory interactions do not occur when sounds that normally enhance performance belong to another auditory group (Vroomen and de Gelder 2000; Watanabe and Shimojo 2001; See also Sanabria et al. 2004a, b). The results also imply that a sound can only be assigned to a single event: it either belongs to the auditory stream, or it is paired with the lights, but it cannot be assigned to both simultaneously. In this respect, it is analogous to many of the well-known ambiguous figure-ground displays (e.g., the Face-Vase illusion or the Necker cube), where it is known that only one interpretation of the scene can be maintained.

Another important finding was that commonality in space between the capture sounds and lights did not affect temporal ventriloquism. The temporal ventriloquist effect was thus equally big for segregated capture sounds that were presented near the lights or far away from the lights. A few other studies have demonstrated before that spatial disparity between sound and vision may not affect intersensory interactions (Welch et al. 1986; Bertelson et al. 1994; Stein

**Table 2** Mean just noticeable differences (JND) in ms, and standard errors of the mean (in parentheses) of Experiment 3

Location of capture sounds	AV-interval (ms)	Location of flanker sounds			
		Same as capture sounds		Different from capture sounds	
		JND	TVE	JND	TVE
Central	0	28.2 (2.1)	-0.9	27.7 (2.0)	3.9*
	100	29.1 (2.3)		23.8 (1.4)	
Lateral	0	30.0 (2.2)	1.7	29.0 (1.8)	3.1*
	100	28.3 (2.0)		25.9 (1.8)	

Capture Sounds Presented at  $\sim 0$  or  $\sim 100$  ms audio–visual intervals from central or lateral location; flanker sounds presented from the same or different location as the capture sounds. The temporal ventriloquist effect (TVE) is the improvement in JND between the  $\sim 0$  and  $\sim 100$  ms audio–visual intervals

\* $P < 0.05$

et al. 1996; Colin et al. 2001; Murray et al. 2004; Vroomen and Keetels 2006). However, these studies always relied on null-effects, which entails the danger that they simply lacked the power to detect any effect of spatial disparity. Participants in previous studies might, for example, not have been able to perceive spatial disparity, or they might have learned to ignore it in the experimental task. Our findings, though, counter these arguments. By combining principles of auditory grouping with intersensory pairing, we were able to create a situation where the capture sounds were actually more effective when their locations differed from that of the lights rather than when they came from the same position as the lights. Within the same experimental situation, we thus demonstrated that sound location mattered for auditory grouping, but not for intersensory pairing. Such a finding makes it highly unlikely that sound location was not perceived or simply ignored. Rather, it becomes more likely that, at least in the temporal ventriloquist situation, commonality in space between sound and vision is not relevant for AV pairing.

This may, at first sight, seem unlikely, because after all, most natural multisensory events are spatially and temporally aligned, except for some minor variations in time or space that people are readily able to adjust (e.g. Vroomen et al. 2004). However, a critical assumption that underlies the idea of spatial correspondence for intersensory pairing is that space has the same function in vision and audition. This notion, though, is arguable as it has been proposed that the role of space in hearing is only to steer vision (Heffner and Heffner 1992), while in vision it is an indispensable attribute (Kubovy and Van Valkenburg 2001). If one accepts that auditory spatial perception evolved for steering vision, but not for deciding whether sound and light belong together, there is no reason why intersensory interactions would require spatial co-localization. Our results therefore have also important implications for designing multimodal devices or creating virtual reality environments, as they show that the brain can, at least in some cases, ignore intersensory discordance in space.

## References

- Aschersleben G, Bertelson P (2003) Temporal ventriloquism: crossmodal interaction on the time dimension. 2. Evidence from sensorimotor synchronization. *Int J Psychophysiol* 50:157–163
- Bedford FL (1989) Constraints on learning new mappings between perceptual dimensions. *J Exp Psychol-Hum Percept Perform* 15:232–248
- Bertelson P (1999) Ventriloquism: a case of crossmodal perceptual grouping. In: Aschersleben G, Bachmann T, Musseler J (eds) *Cognitive contributions to the perception of spatial and temporal events*. Elsevier, North-Holland, pp 347–363
- Bertelson P, Aschersleben G (2003) Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *Int J Psychophysiol* 50:147–155
- Bertelson P, Vroomen J, Wiegendaal G, de Gelder B (1994) Exploring the relation between McGurk interference and ventriloquism. In: *International congress on spoken language processing*, vol 2, Yokohama, pp 559–562
- Bregman AS (1990) *Auditory scene analysis*. MIT Press, Cambridge
- Colin C, Radeau M, Deltenre P, Morais J (2001) Rules of intersensory integration in spatial scene analysis and speechreading. *Psychol Belgica* 41:131–144
- Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. *Percept Psychophys* 63:719–725
- Finney DJ (1964) *Probit analysis*. Cambridge University Press, Cambridge
- Heffner RS, Heffner HE (1992) Visual factors in sound localization in mammals. *J Comp Neurol* 317:219–232
- Kubovy M, Van Valkenburg D (2001) Auditory and visual objects. *Cognition* 80:97–126
- Morein-Zamir S, Soto-Faraco S, Kingstone A (2003) Auditory capture of vision: examining temporal ventriloquism. *Cogn Brain Res* 17:154–163
- Murray MM, Michel CM, Grave de Peralta R, Ortigue S, Brunet D, Gonzalez Andino S, Schnider A (2004) Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage* 21:125–135
- Radeau M (1994) Auditory-visual spatial interaction and modularity. *Cah Psychol Cogn-Curr Psychol Cogn* 13:3–51
- Sanabria D, Correa A, Lupianez J, Spence C (2004a) Bouncing or streaming? Exploring the influence of auditory cues on the interpretation of ambiguous visual motion. *Exp Brain Res* 157:537–541
- Sanabria D, Soto-Faraco S, Spence C (2004b) Exploring the role of visual perceptual grouping on the audiovisual integration of motion. *Neuroreport* 15:2745–2749
- Scheier CR, Nijhawan R, Shimojo S (1999) Sound alters visual temporal resolution. *Invest Ophthalmol Vis Sci* 40:4169
- Stein BE, Meredith MA (1993) *The merging of the senses*. The MIT Press, Cambridge
- Stein BE, London N, Wilkinson LK, Price DD (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *J Cogn Neurosci* 8:497–506
- Stekelenburg JJ, Vroomen J (2005) An event-related potential investigation of the time-course of temporal ventriloquism. *Neuroreport* 16:641–644
- Vroomen J, de Gelder B (2000) Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J Exp Psychol Hum Percept Perform* 26:1583–1590
- Vroomen J, de Gelder B (2004) Temporal ventriloquism: sound modulates the flash-lag effect. *J Exp Psychol-Hum Percept Perform* 30:513–518
- Vroomen J, Keetels M (2006) The spatial constraint in intersensory pairing: no role in temporal ventriloquism. *J Exp Psychol Hum Percept Perform* 32(4):1063–1071
- Vroomen J, Keetels M, de Gelder B, Bertelson P (2004) Recalibration of temporal order perception by exposure to audio-visual asynchrony. *Cogn Brain Res* 22:32–35
- Watanabe K, Shimojo S (2001) When sound affects vision: effects of auditory grouping on visual motion perception. *Psychol Sci* 12:109–116

- Welch RB (1999) Meaning, attention, and the “unity assumption” in the intersensory bias of spatial and temporal perceptions. In: Aschersleben G, Bachmann T, Müsseleler J (eds) Cognitive contributions to the perception of spatial and temporal events. Elsevier, Amsterdam, pp 371–387
- Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. *Psychol Bull* 88:638–667
- Welch RB, DuttonHurt LD, Warren DH (1986) Contributions of audition and vision to temporal rate perception. *Percept Psychophys* 39:294–300