Lecture 8: Multi-Sensory Interactions

Part 02

Audio-Visual Interactions

Relevance

Ahrens, A., Lund, K.D., Marschall, M. and Dau, T., 2019. Sound source localization with varying amount of visual information in virtual reality. *PloS one*, *14*(3), p.e0214603.

Liu, J. and Ando, H., 2008, May. Hearing how you touch: Real-time synthesis of contact sounds for multisensory interaction. In *2008 Conference on Human System Interactions* (pp. 275-280). IEEE.

Metatla, O., Correia, N.N., Martin, F., Bryan-Kinns, N. and Stockman, T., 2016, May. Tap the ShapeTones: Exploring the effects of crossmodal congruence in an audio-visual interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1055-1066).

Tidoni, E., Gergondet, P., Kheddar, A. and Aglioti, S.M., 2014. Audio-visual feedback improves the BCI performance in the navigational control of a humanoid robot. *Frontiers in neurorobotics*, *8*, p.20.

Rébillat, M., Boutillon, X., Corteel, É. and Katz, B.F., 2012. Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments. *ACM Transactions on Applied Perception (TAP)*, *9*(4), pp.1-17.

Chan, V.Y.S., Jin, C.T. and van Schaik, A., 2012. Neuromorphic audio-visual sensor fusion on a sound-localising robot. *Frontiers in neuroscience*, *6*, p.21.

Learning Objectives

To provide a description of:

Unisensory and multisensory perception

The Ventriloquism effect- spatial and temporal

Visual and auditory dominance

The McGurk effect

Colavita Effect- auditory and tactile

Crossmodal dynamic capture

Learning Outcomes

To be able to define Unisensory and multisensory perception

To be able to describe and provide examples of key experiments used to understand the Ventriloquism effect- spatial and temporal

To be able to explain what meant by visual and auditory dominance and to be able to provide examples of experiments

To be able to describe the McGurk effect

To be able to describe the Colavita Effect- auditory and tactile and to able to describe some of the key experiments used to study this effect

To be able to describe the crossmodal dynamic capture to be able to describe some of the key experiments used to study this effect

How a Visual Stimulus May Affect Auditory Perception

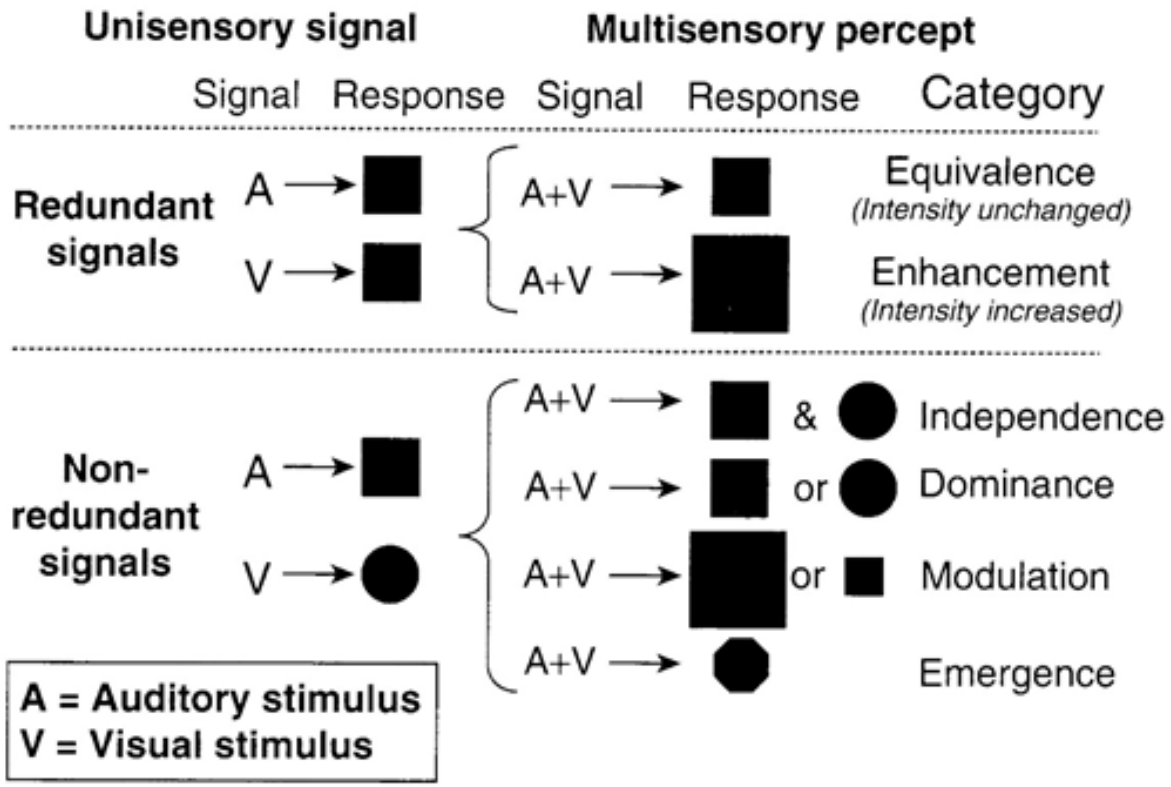


Fig. 12.1 Schematic figure illustrating some of the key ways in which the presentation of a visual stimulus can influence a person's auditory (multisensory) perception. Redundant signals are presented in the upper part of the figure, non-redundant signals in the lower part. The left part of the figure shows perceptual responses to the two signals (auditory and visual) when presented separately. The right part of the figure shows the multimodal perceptual response. In this figure, shape is used to signify the perceived identity of a stimulus, while size gives a schematic indication of the perceived intensity of the stimulus. Results showing equivalence or independence would suggest that there is little interaction between the senses. However, the majority of the studies reported in this chapter show one of the other forms of multisensory interaction. Enhancement has been shown in studies of audiovisual speech perception; dominance has been shown in studies of the Colavita effect; modulation has been shown in studies of spatial ventriloquism; while emergence has been demonstrated by studies of the McGurk effect. (Adapted and redrawn from Partan and Marler, 1999.)

Ventriloquism Effect

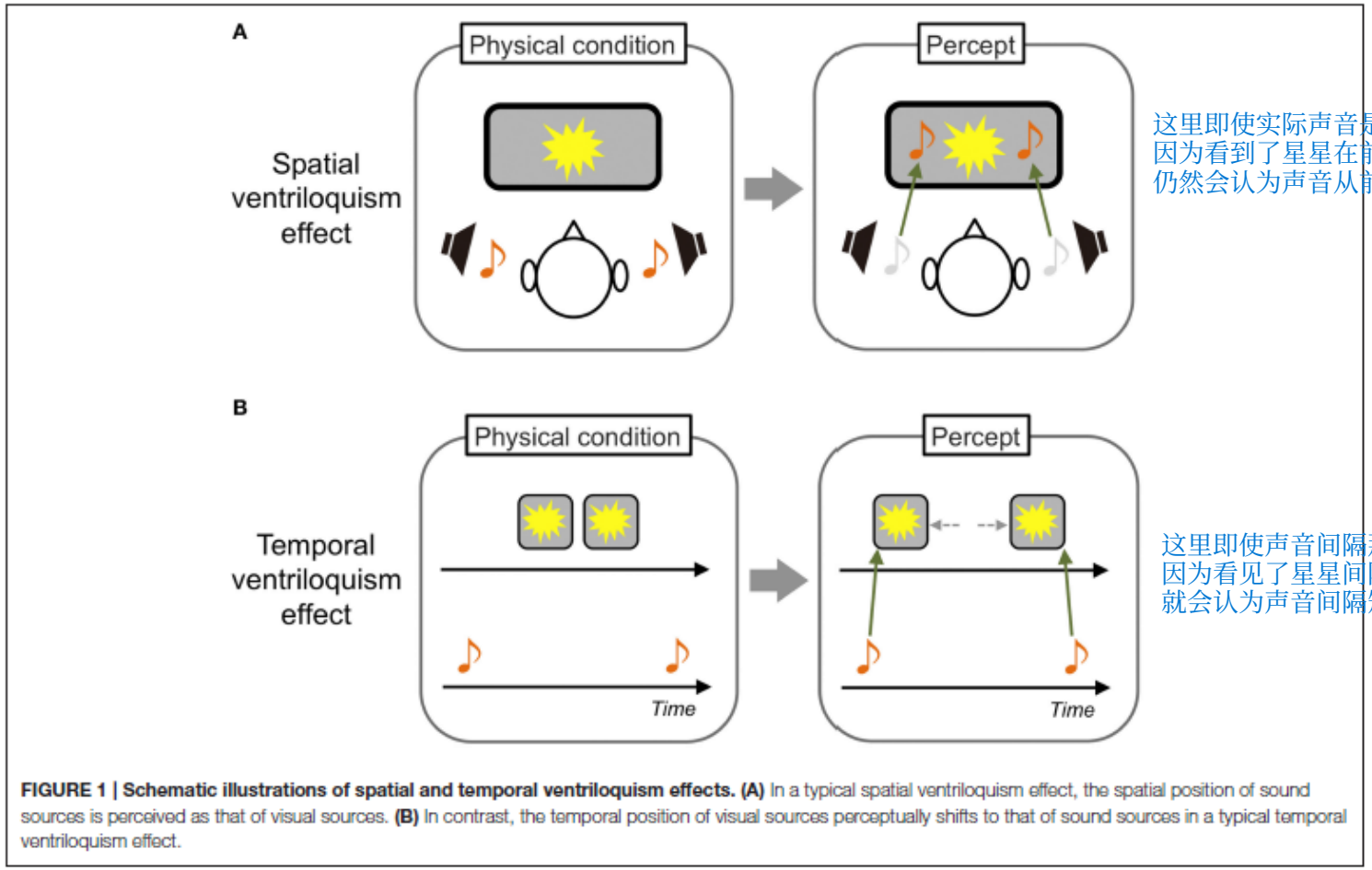


FIGURE 1 | Schematic illustrations of spatial and temporal ventriloquism effects. (A) In a typical spatial ventriloquism effect, the spatial position of sound sources is perceived as that of visual sources. **(B)** In contrast, the temporal position of visual sources perceptually shifts to that of sound sources in a typical temporal ventriloquism effect.

这里即使实际声音是从耳朵两边来的
因为看到了星星在前面，
仍然会认为声音从前面传过来的

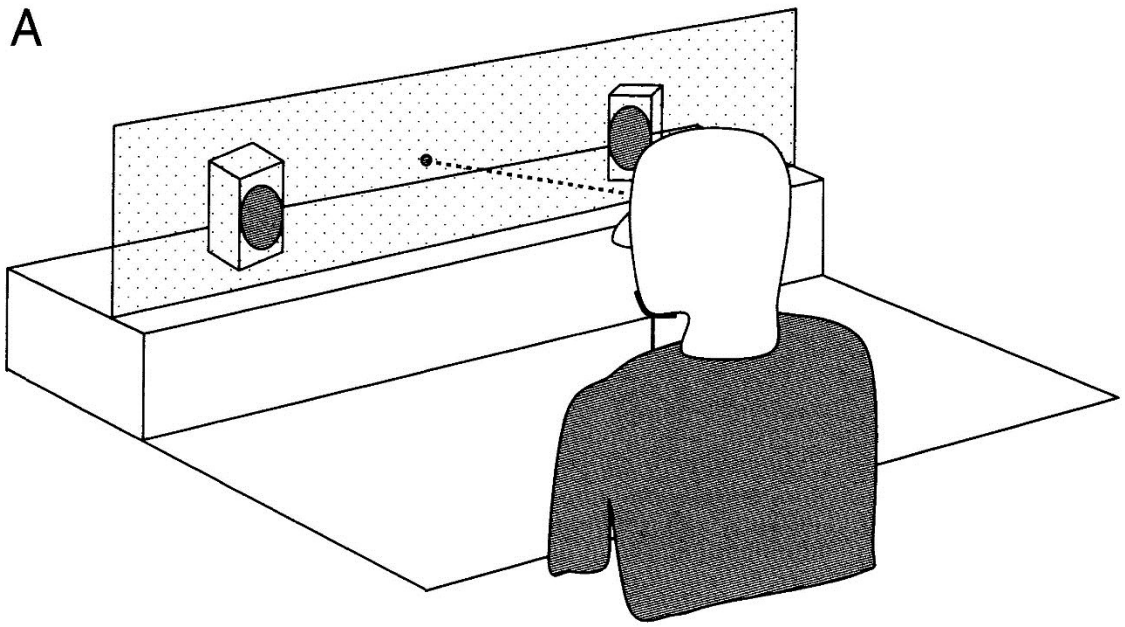
这里即使声音间隔那么长
因为看见了星星间隔短
就会认为声音间隔短

Example of dominance of vision over audition occurs when a conflict is introduced between the spatial/temporal origin of auditory and visual stimuli.

From: Hidaka, S., Teramoto, W. and Sugita, Y., 2015. Spatiotemporal processing in crossmodal interactions for perception of the external world: a review. *Frontiers in integrative neuroscience*, *9*, p.62.

Ventriloquism Effect: Experimental Study

Spatial ventriloquism effect



Study by Bertelson and Aschersleben (1998):

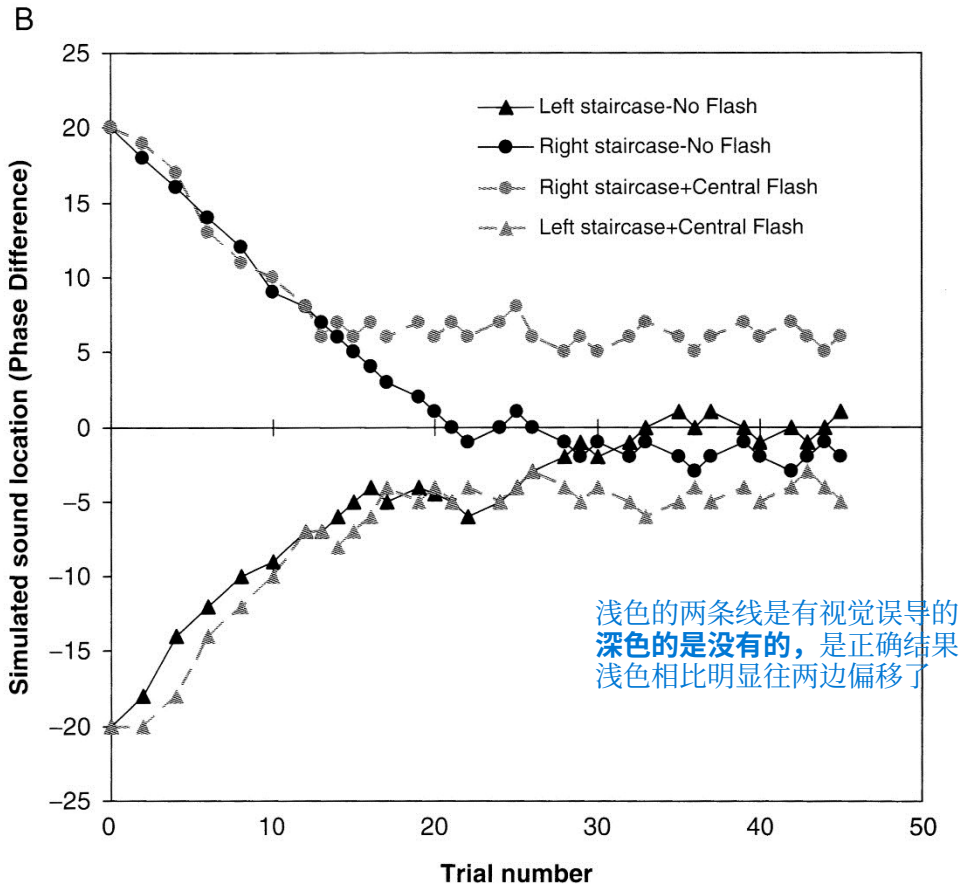
Sound presented right/left of fixation.

Participant had to decide whether the sound had been presented from the left or right of central fixation.

Oxford Handbook of Auditory Science: Hearing. C.J. Plack (2010). Chapter 12.

Ventriloquism Effect: Experimental Study

Spatial ventriloquism effect



Study by Bertelson and Aschersleben (1998):

Multiple interleaved psychophysical staircases- starting from left/right

Visual stimulus presented at fixation in synchrony with the sound on several of the different staircases

Sound presented right/left of fixation.

Participant had to decide whether the sound had been presented from the left or right of central fixation.

Participants' responses would end up alternating between left and right as the sound approached the centre (point of uncertainty).

Oxford Handbook of Auditory Science: Hearing. C.J. Plack (2010). Chapter 12.

Does Vison Always Dominate Audition?

Alais, D. and Burr, D., 2004:

> "When visual localization is good, vision does indeed dominate and capture sound.
>
> However, for severely blurred visual stimuli (that are poorly localized), the reverse holds: sound captures vision.
>
> For less blurred stimuli, neither sense dominates and perception follows the mean position."

From: Alais, D. and Burr, D., 2004. The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, *14*(3), pp.257-262.

Three of the visual stimuli used in the study by Alais and Burr (2004).

Does Vison Always Dominate Audition?

Alais, D. and Burr, D., 2004:

"Observers were required to localize in space brief light "blobs" or sound "clicks," presented either separately (unimodally) or together (bimodally).

In a given trial, two sets of stimuli were presented successively (separated by a 500 ms pause) and observers were asked to indicate which appeared to be more to the left, guessing if unsure.

Visual stimuli were low-contrast (10%) Gaussian blobs of various widths, back-projected for 15 ms onto a translucent perspex screen (80 × 110 cm).

Auditory stimuli were brief (1.5 ms) clicks presented though two visible high-quality speakers at the edge of the screen, with the apparent position of the sound controlled by interaural time differences".

From: Alais, D. and Burr, D., 2004. The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, *14*(3), pp.257-262.

要求受试者判断第二个刺激（视觉或听觉）**是否比第一个更向左**

Does Vison Always Dominate Audition?

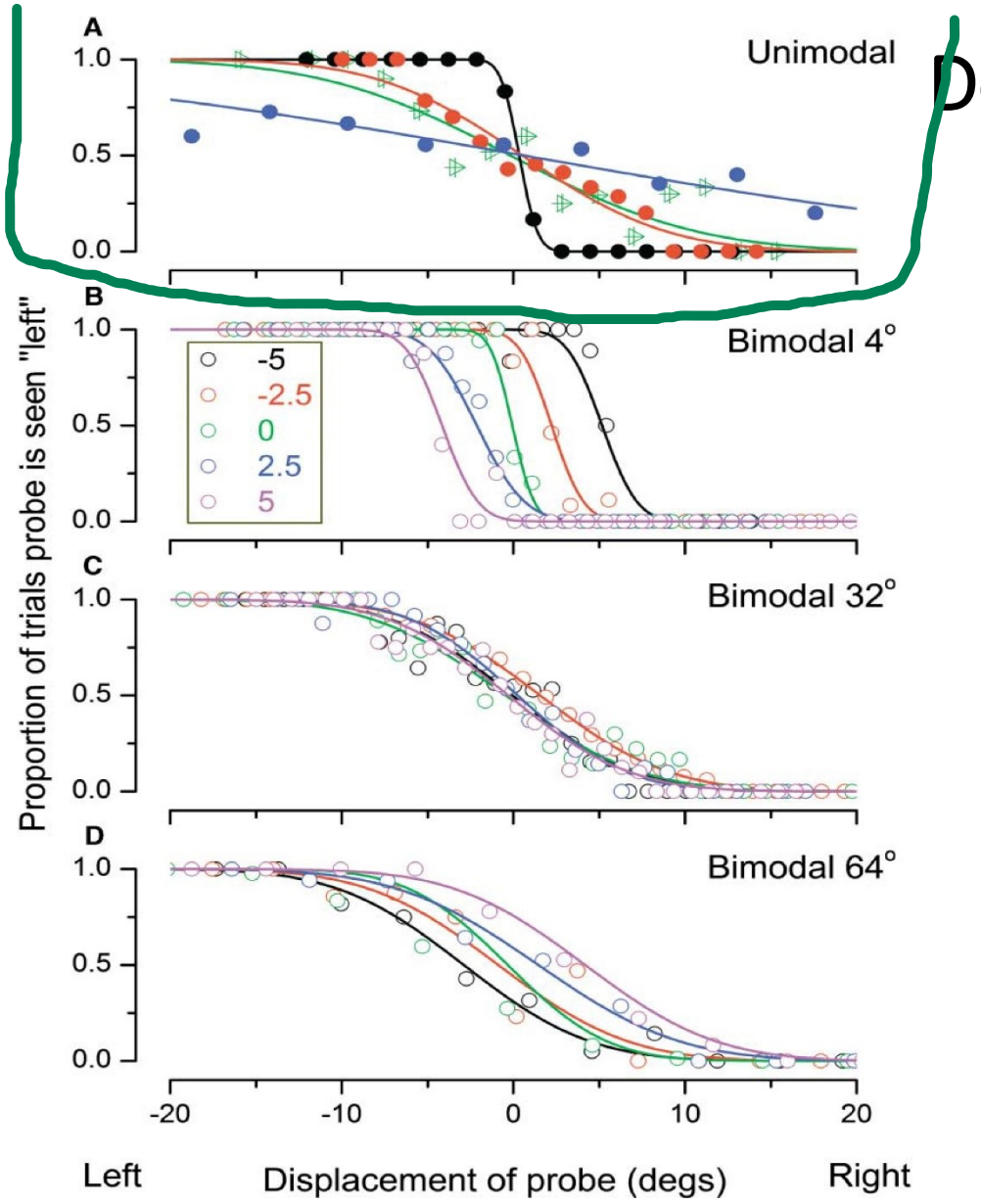


Figure 1. Unimodal and Bimodal Localization of Visual and Auditory Stimuli

Alais, D. and Burr, D., 2004:

Psychometric functions: Proportion of trials in which the second stimulus was seen to the left of the first, as a function of actual physical displacement.

A: Localising an auditory click (green speaker-shaped symbols) or visual blobs of various Gaussian space constants (2σ = 4°, black; 2σ = 32°, red; or 2σ = 64°, blue).

B–D: Psychometric functions for localizing bimodal presentations of the click and blob together (click centered within the blob), for blob widths 4° (B), 32° (C), or 64° (D).

不同颜色的曲线代表视觉刺激大小不同的情况

横坐标声源刺激从中心位置向左或向右偏移的角度。

bimodel后面跟着的xx°代表blob光圈多大

From: Alais, D. and Burr, D., 2004. The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, *14*(3), pp.257-262.

Does Vison Always Dominate Audition?

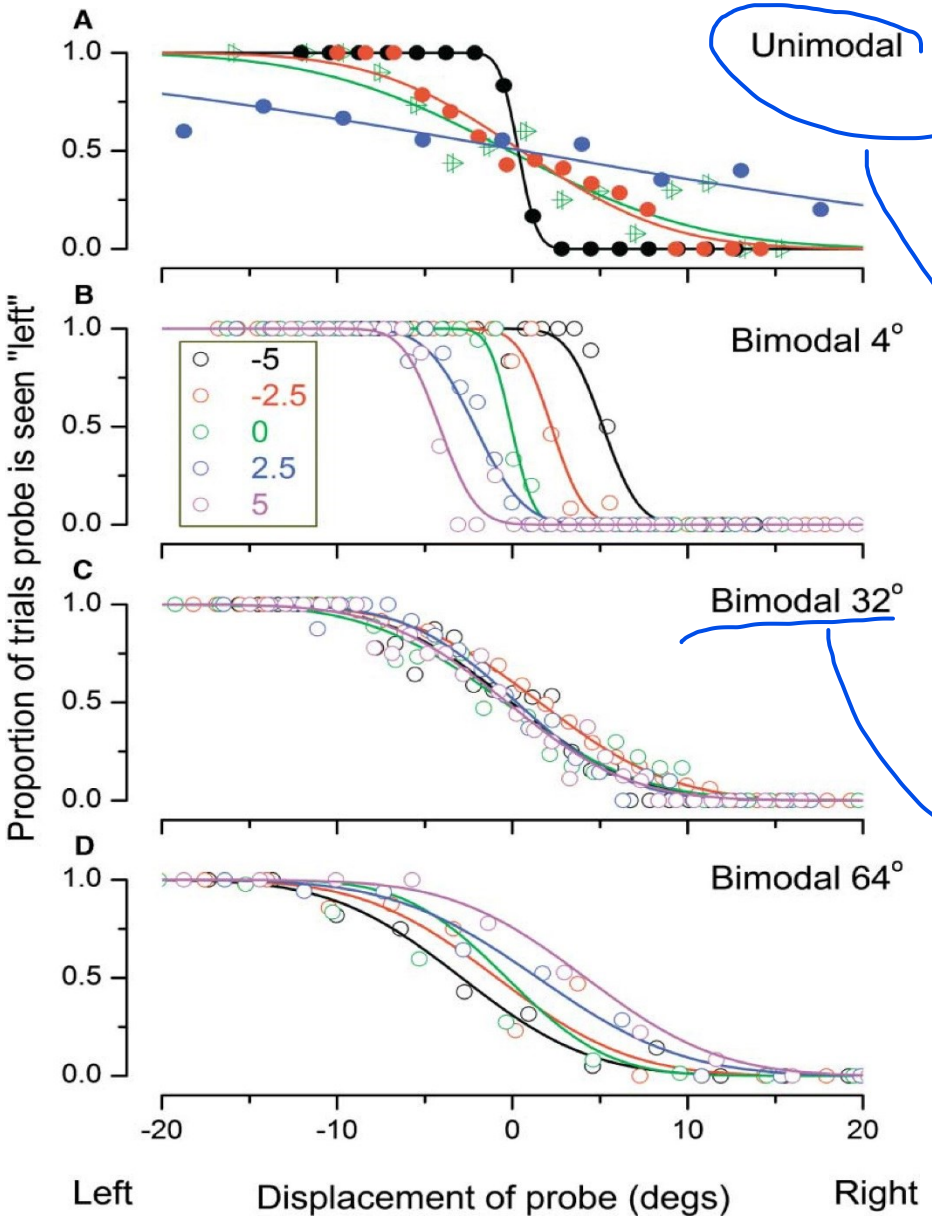


Figure 1. Unimodal and Bimodal Localization of Visual and Auditory Stimuli

Alais, D. and Burr, D., 2004:

For small blob widths (4°; Fig. 1B), curves displaced systematically in the direction of the side where the visual stimulus was displaced) - vision dominated perceived position of the incongruent stimuli (the ventriloquist effect).

For large blobs (64°; Fig. 1D), the reverse result; Curves displaced in the opposite direction, indicating that click sound dominated perceived position, “inverse ventriloquist effect.”

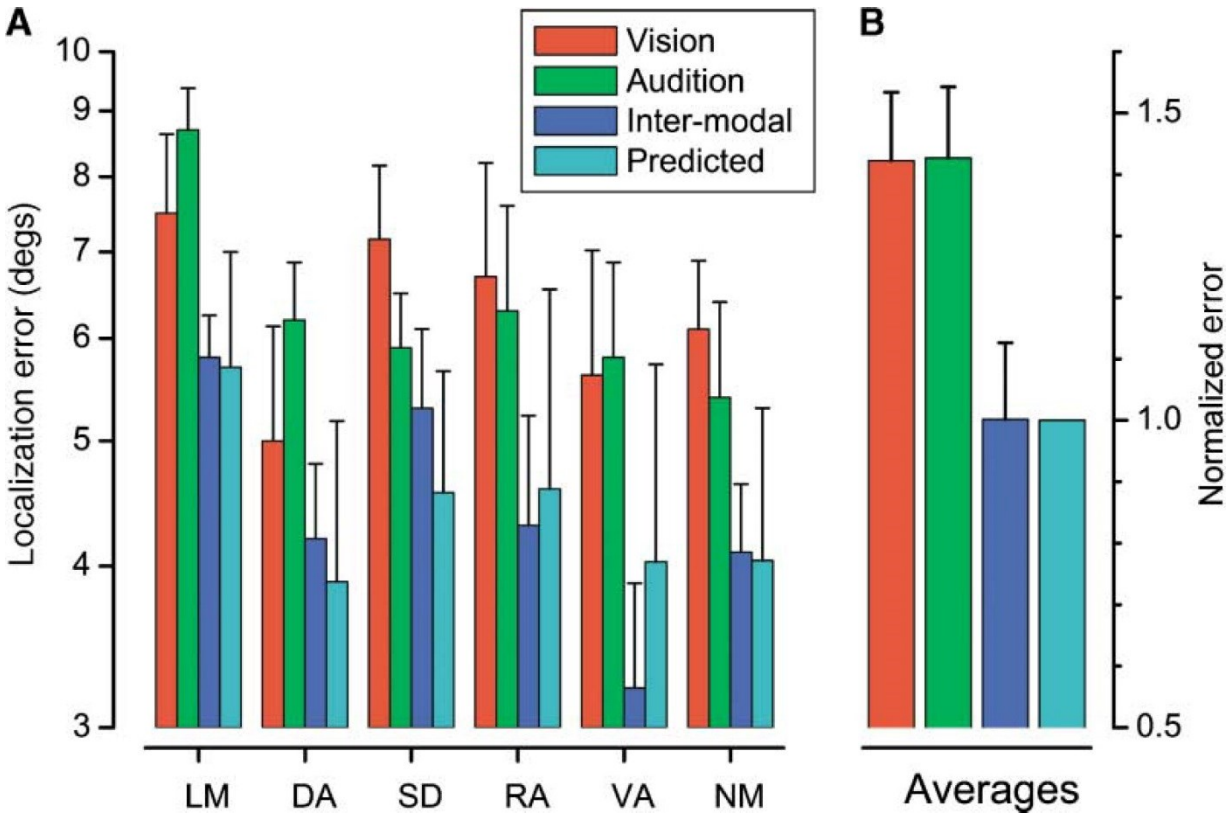
For mid-sized blobs (32°; Fig. 1C), curves tended to cluster together, suggesting that the perceived position depended on an average of the two modalities.

单模态（单独视觉或听觉）和双模态（视觉和听觉结合）

随着视觉刺激大小的增加，受试者将视觉信息的影响考虑得更多，听觉判断受视觉刺激的“牵引”效果影响更显著。

From: Alais, D. and Burr, D., 2004. The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, *14*(3), pp.257-262.

Does Vison Always Dominate Audition?



Alais, D. and Burr, D. (2004).

Fig. shows auditory, visual, bimodal, and predicted bimodal thresholds for one blob size (the one yielding the most similar auditory and visual thresholds and hence the largest predicted improvement) for each subject, and for the group mean.

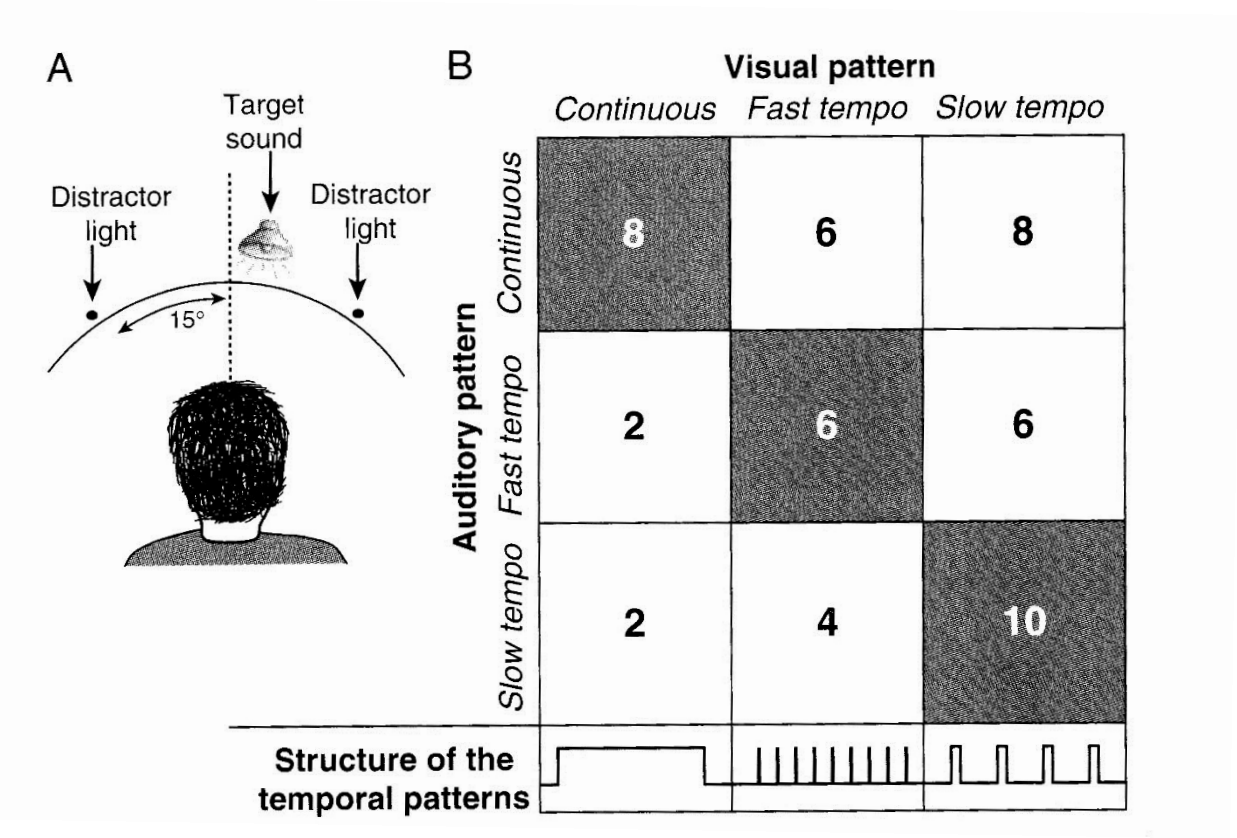
All participants: bimodal localization was better than either unimodal localization.

Averaged data: Bimodal thresholds are lower than the average of the visual and auditory thresholds by a factor of 1.425.

Overall, results support a model in which visual and auditory information is combined by minimizing variance, leading to an improvement in discriminating bimodal spatial location.

From: Alais, D. and Burr, D., 2004. The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, *14*(3), pp.257-262.

Ventriloquism Effect: Experimental Study



what about temporal correlation of auditory & visual stimuli on ventriloquism effect?

Radeau and Bertelson (1987) showed that correlation between the temporal profile of auditory and visual stimuli also influences the visual capture of audition seen in the spatial ventriloquism effect.

Oxford Handbook of Auditory Science: Hearing. C.J. Plack (2010). Chapter 12.

Ventriloquism Effect: Experimental Study

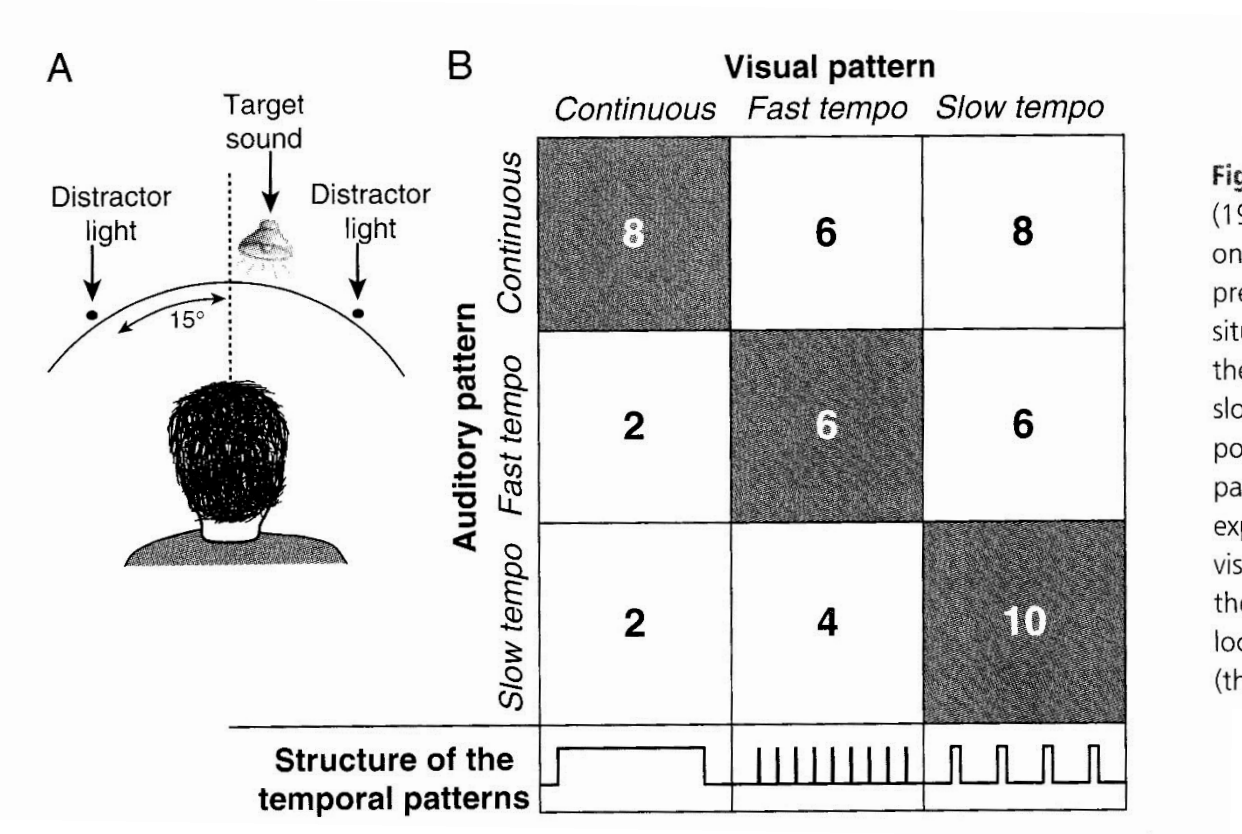
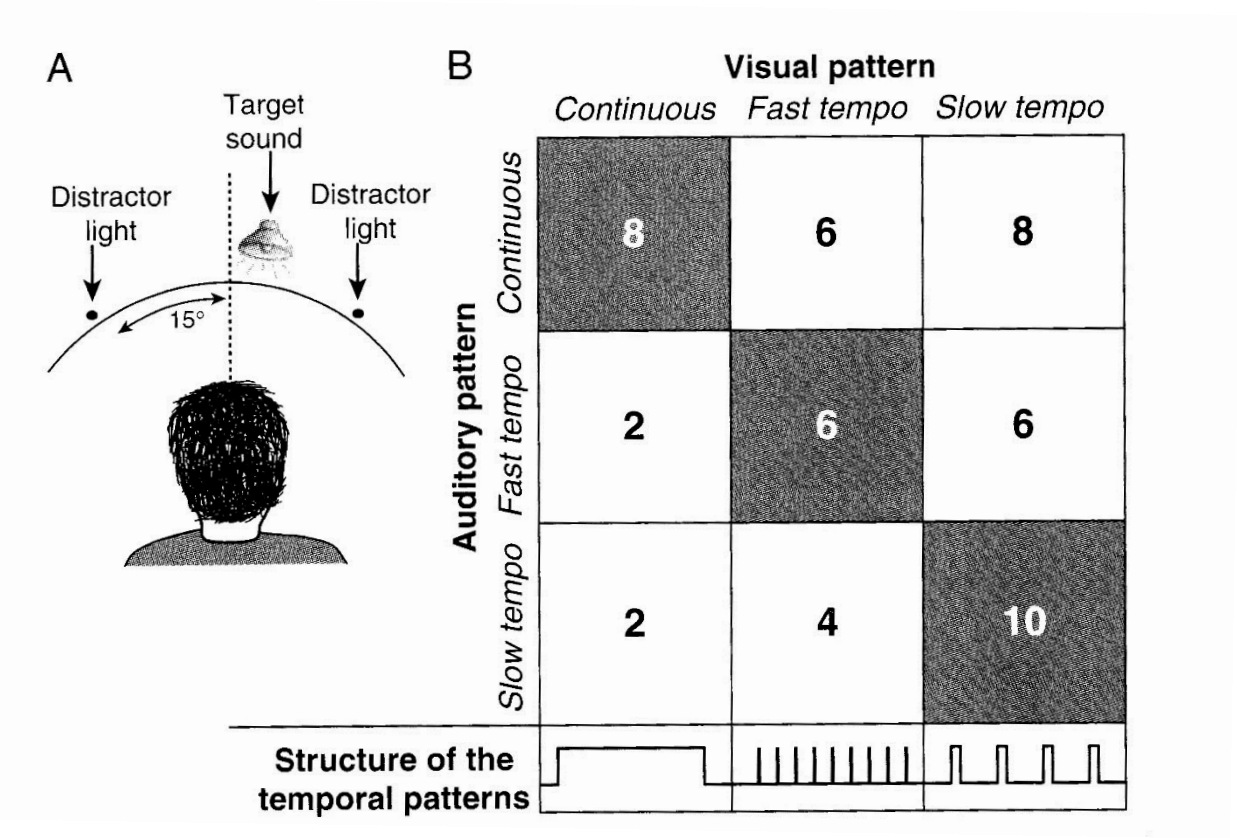


Fig. 12.4 (A) Schematic diagram illustrating the experimental set-up used in Radeau and Bertelson's (1987) study of the influence of temporal correlation (or patterning) of auditory and visual stimuli on the spatial ventriloquism effect. In the conditions of interest to us here, the auditory target was presented 1°, 3°, or 5° to one or the other side of the midline, together with a distractor light situated 15° to either the left or right of the midline. On each trial, the participants judged whether the sound had been presented to the left or right. The temporal pattern (continuous, fast tempo, or slow tempo) of the sounds and the lights was varied systematically. (B) The table highlights the nine possible combinations of three auditory and three visual temporal patterns presented to participants. The value in each cell provides a measure of the visual biasing of auditory localization, expressed as the difference between the number of 'right' judgements on those trials where the visual distractor was presented on the right versus on the left (thus, the bigger the value, the larger the visual biasing of auditory localization). Note that vision typically influences perceived sound location maximally when the temporal pattern presented in both sensory modalities coincided (the shaded cells in the table). (Adapted and redrawn from Spence, 2007, Figure 3.)

Oxford Handbook of Auditory Science: Hearing. C.J. Plack (2010). Chapter 12.

Ventriloquism Effect: Experimental Study

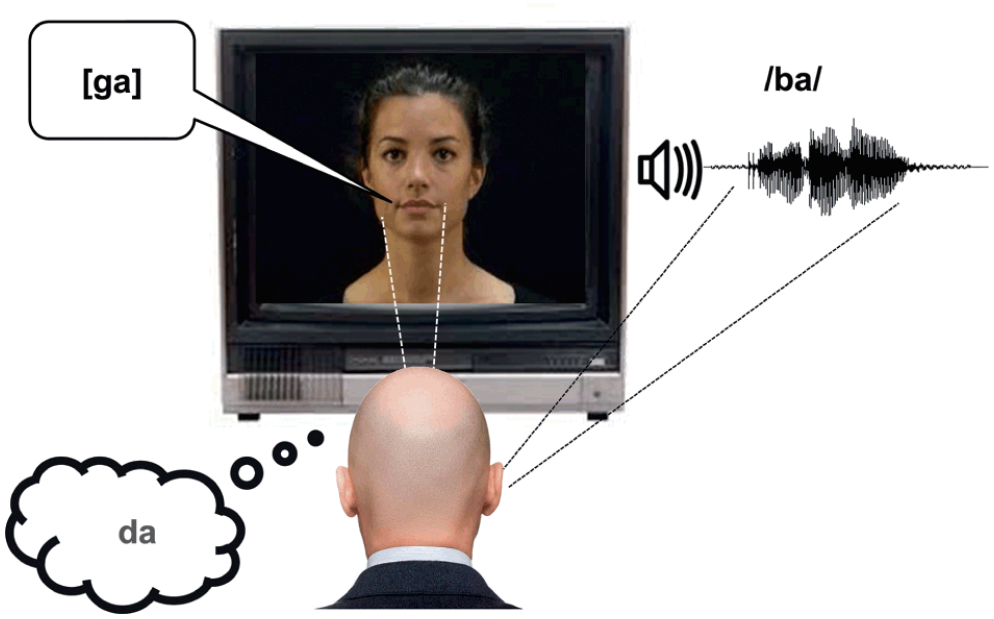


Results:

Visual biasing of the perceived location of the auditory stimulus was larger when the temporal configuration of the stimuli presented in the 2 sensory modalities matched than when it did not.

Oxford Handbook of Auditory Science: Hearing. C.J. Plack (2010). Chapter 12.

Audio-visual McGurk Effect



Hear a “ba” whilst simultaneously see synchronized lip movements associated with another syllable (e.g., “ga”), most participants report hearing another syllable (e.g., “da”) that is different from what had been presented in either modality.

i.e., phoneme people hear can be altered by the lip movements they see.

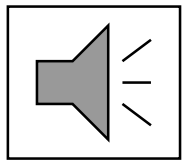
Method: Intersensory conflict

Where? Auditory cortex?

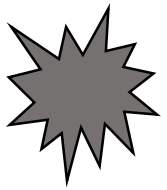
From Multisensory Interactions in the Real World. Salvador Soto-Faraco,Daria Kvasova, Emmanuel Biau, Nara Ikumi, Manuela Ruzzoli, Luis Morís-Fernández, Mireia Torralba. Published online by Cambridge University Press: 08 August 2019

Colavita Effect (audio-visual)

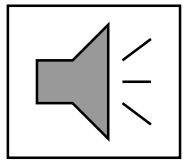
STIMULI



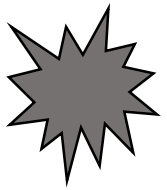
OR



Unimodal
(majority)



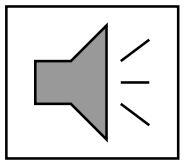
AND



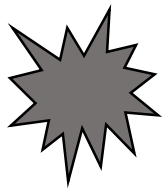
Bimodal
(minority)

RESPONSE

Button if:



Button if:



Colavita 1974:

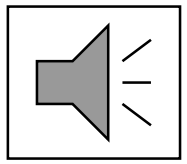
Participants: Presented with an unpredictable sequence of auditory (tone) and visual (flash) events.
Instructions: Press “tone button” when hear tone, press another 2nd button “flash button” when see flash.

Trials presented:
Majority unimodal - only tone or only flash.
Minority (~5/35) bimodal – simultaneous presentation of tone and flash.

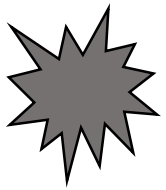
Unimodal and bimodal presentations randomised.

Colavita Effect (audio-visual)

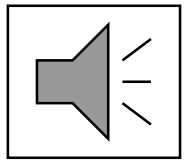
STIMULI



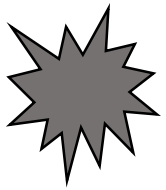
OR



Unimodal
(majority)



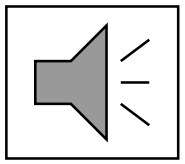
AND



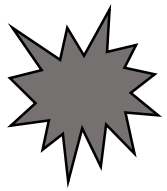
Bimodal
(minority)

RESPONSE

Button if:



Button if:

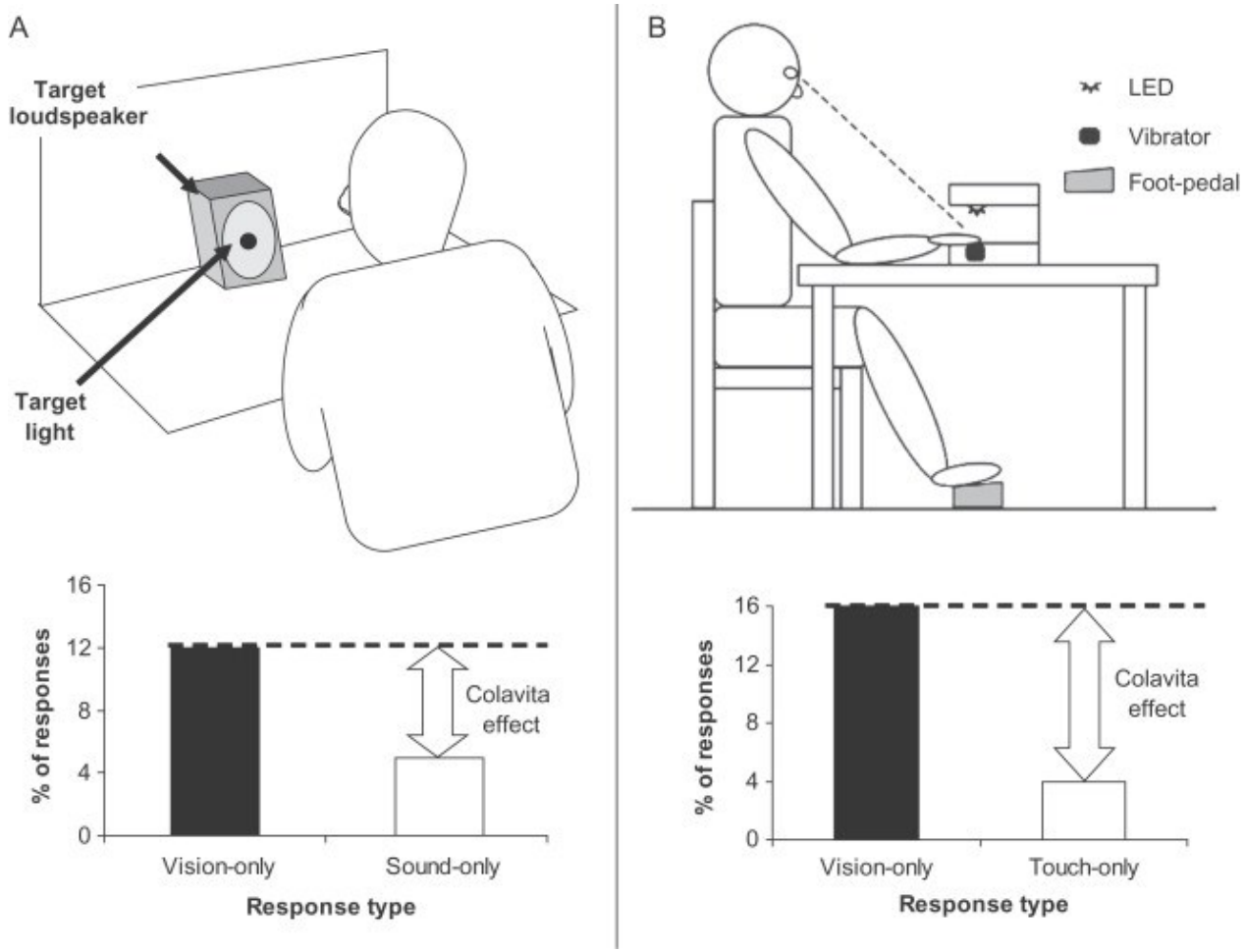


Result:

Participants easily respond to unimodal tones very quickly (fast response latency)
Participants failed to respond to sound on majority of bimodal presentations, instead only pressed button for flash .

Why? Possibly presentation of a visual stimulus slows responses to a sound

Colavita Effect- Auditory and Touch



Koppen and Spence (2007)

Hartcher-O'Brien et al. (2008))

Fig shows the experimental set-up used in an audiovisual Colavita effect (study A), and a visuotactile Colavita effect (study B).
Visual stimulus consisted of - illumination of the loudspeaker cone used to present the auditory stimuli in study A and illumination of the finger where the vibration was presented in study B.

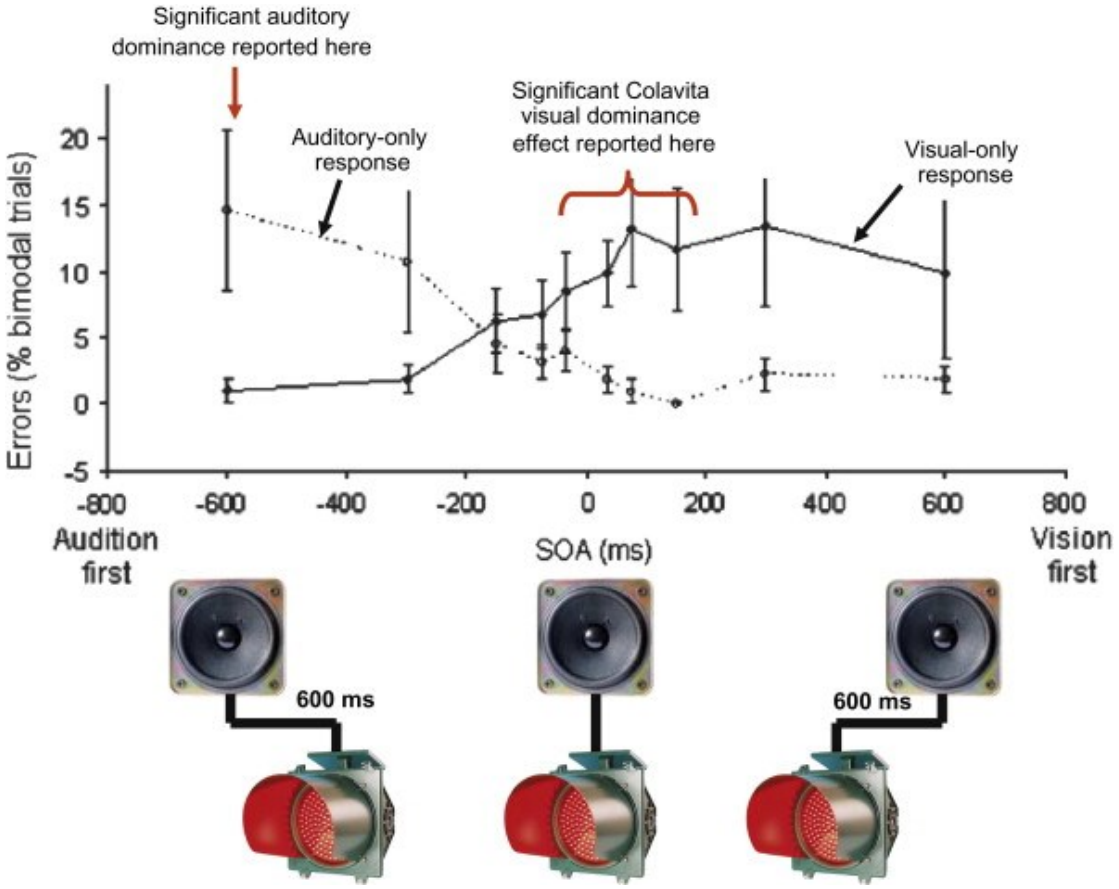
[Target stimuli presented from exactly the same spatial location in both studies.

Study A & B: Show visual dominance effect.

Plots - % bimodal trials in which the participants made either a visual-only or an auditory- (tactile-) only response.
However magnitude of Colavita effect is smaller than that reported in Colavita's early studies).

From: Spence, C., 2009. Explaining the Colavita visual dominance effect. *Progress in brain research*, *176*, pp.245-258.

Colavita Effect

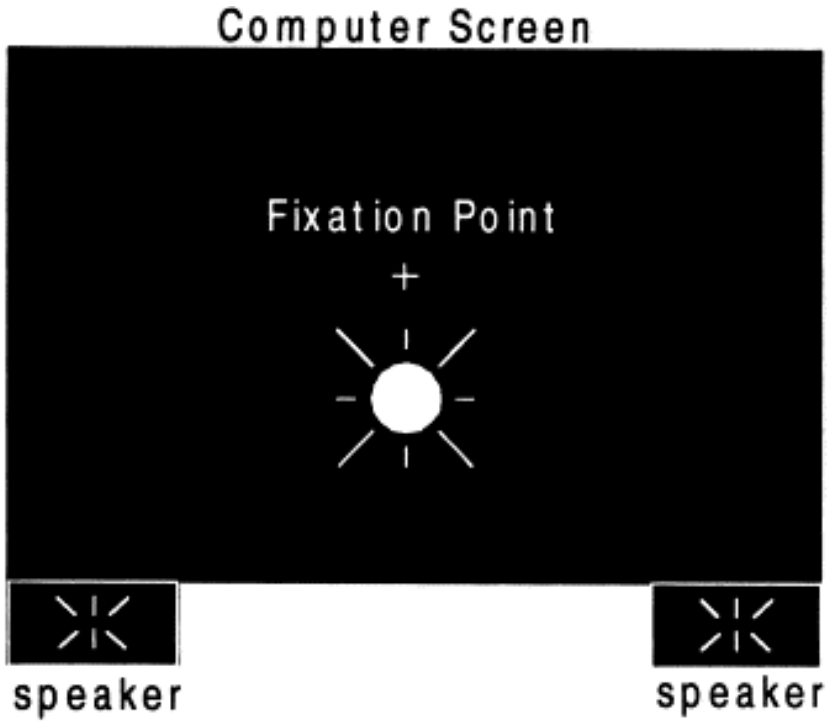


From: Spence, C., 2009. Explaining the Colavita visual dominance effect. *Progress in brain research*, *176*, pp.245-258.

Koppen and Spence 2007:
Colavita study in which the stimulus onset interval (SOA) between the auditory and visual targets on the bimodal target trials was varied randomly between 10 values (±600, ±300, ±150, ±75, and ±35 ms; negative values indicate auditory stimulus was presented before visual stimulus).

Colavita visual dominance effect observed at majority of SOAs, auditory dominance was observed when the auditory component of the bimodal target was presented 600 ms before the visual component.
(a similar smaller trend was observed at both 300 and 150 ms)

Flicker-Flutter Illusion



Flicker-Flutter illusion- visual perception can be altered by signals of other modalities (audio)-> cross-modal interaction.

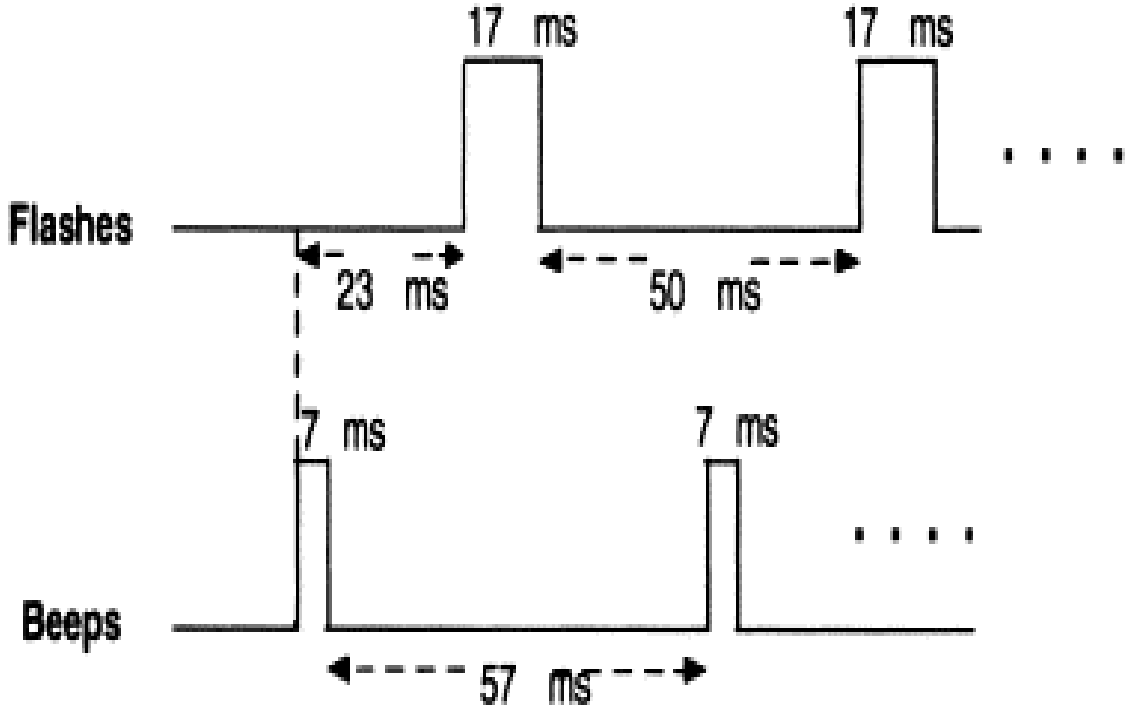
i.e., when a single flash of light is accompanied with multiple beeps, it is perceived as multiple flashes.

Shams, L., Kamitani, Y. and Shimojo, S. (2002). Experiment 1: Fig. 1. Stimulus configuration for Experiments 1 & 2.

White uniform disk is displayed against a black background at some eccentricity below the fixation point which is at the centre of the screen. Approximately at the same time some beeps are played from two speakers directly beneath and to the sides of the screen.

Shams, L., Kamitani, Y. and Shimojo, S., 2002. Visual illusion induced by sound. Cognitive brain research, 14(1), pp.147-152.

Flicker-Flutter Illusion



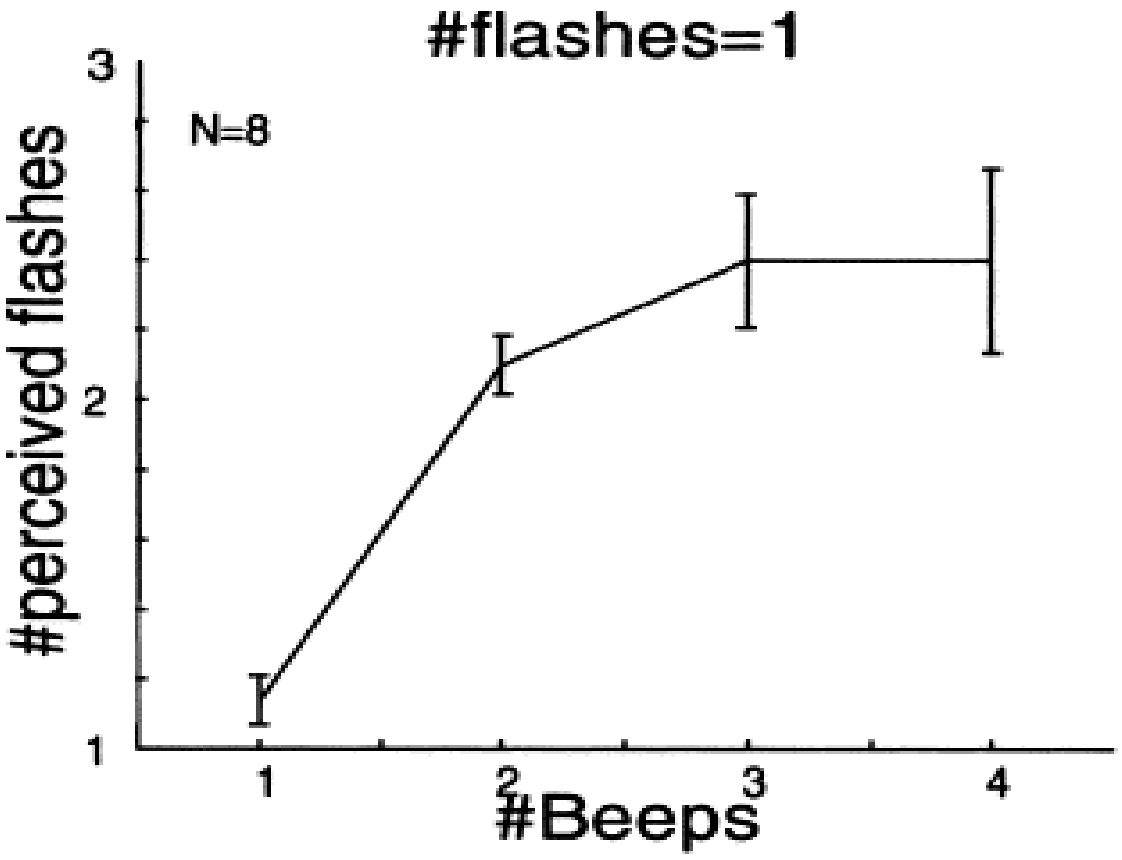
Shams, L., Kamitani, Y. and Shimojo, S. (2002).

Experiment 1: Fig. 2. Temporal profile of the stimuli in Experiment 1.

This diagram shows the relationship between the timing of the beep(s) and flash(es) as well as the time duration and spacing of the signals. In each trial there were one or more (up to four) flashes accompanied with zero or more (up to four) beeps.

Shams, L., Kamitani, Y. and Shimojo, S., 2002. Visual illusion induced by sound. Cognitive brain research, 14(1), pp.147-152.

Flicker-Flutter Illusion



Results:
Figure shows the data for trials in which **a single flash** was presented.

Number of perceived flashes plotted against number of beeps in each trial (averaged across observers. Participants report seeing one flash (veridical value) when number of accompanying beeps is one.

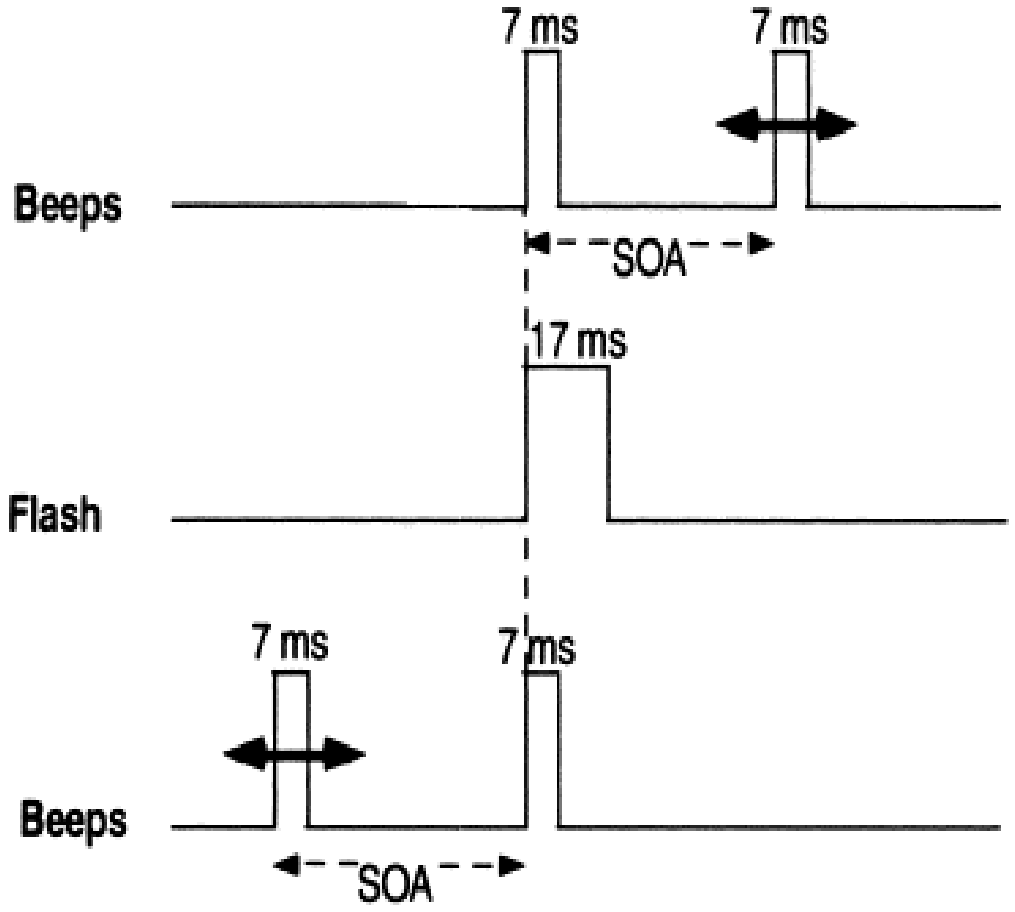
Participants report seeing two or more flashes when the flash is accompanied with two or more beeps.

Perceived number of flashes in trials with a single flash and two, three, or four beeps is significantly greater than that of trials with single flash and one (or no) beep ($P<0.001$).

Result of illusion trials -> Multiple beeps change the percept of a single flash into multiple flashes.

Shams, L., Kamitani, Y. and Shimojo, S., 2002. Visual illusion induced by sound. Cognitive brain research, 14(1), pp.147-152.

Flicker-Flutter Illusion



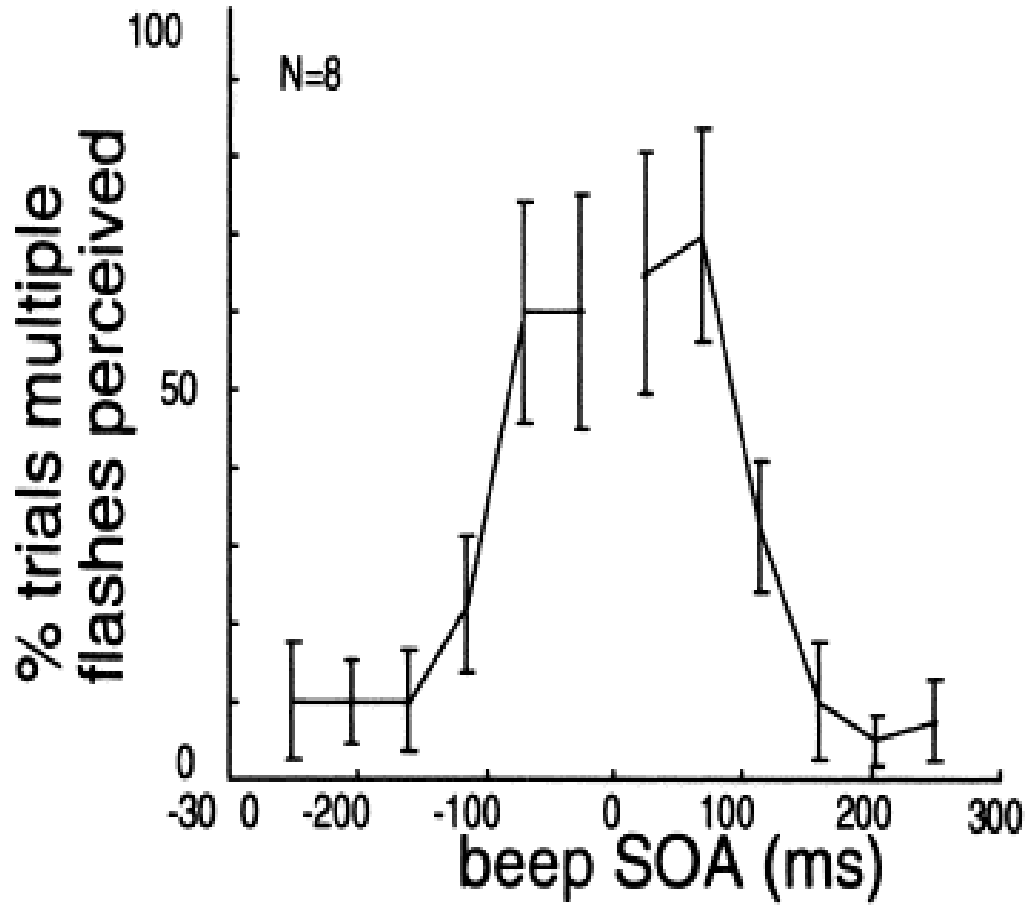
Shams, L., Kamitani, Y. and Shimojo, S. (2002).

Experiment 2: Fig. 2. Temporal profile of the stimuli in Experiment 2.
Same stimulus configuration as in Experiment 1, but:

In each trial one flash was ‘accompanied’ by two beeps.
One beep was always physically simultaneous with the flash, while timing other beep varied from trial to trial with stimulus onset asynchronies (SOAs): 25, 70, 115, 160, 205, 250 ms either before or after the flash.
Participant's task was to judge number of flashes seen on the screen in a 2-AFC paradigm (one or more flashes).

Shams, L., Kamitani, Y. and Shimojo, S., 2002. Visual illusion induced by sound. Cognitive brain research, 14(1), pp.147-152.

Flicker-Flutter Illusion



Shams, L., Kamitani, Y. and Shimojo, S. (2002). Experiment 2.

Results:
The horizontal axis represents the timing of the variable-time beep from the flash. Zero denotes the time of the flash and positive and negative numbers denote the time of the variable beep when it occurs after or before the flash, respectively. The vertical axis is a measure of the strength of the illusion.

The illusion remains strong within 115 ms of the flash.

Shams, L., Kamitani, Y. and Shimojo, S., 2002. Visual illusion induced by sound. Cognitive brain research, 14(1), pp.147-152.

Crossmodal Dynamic Capture

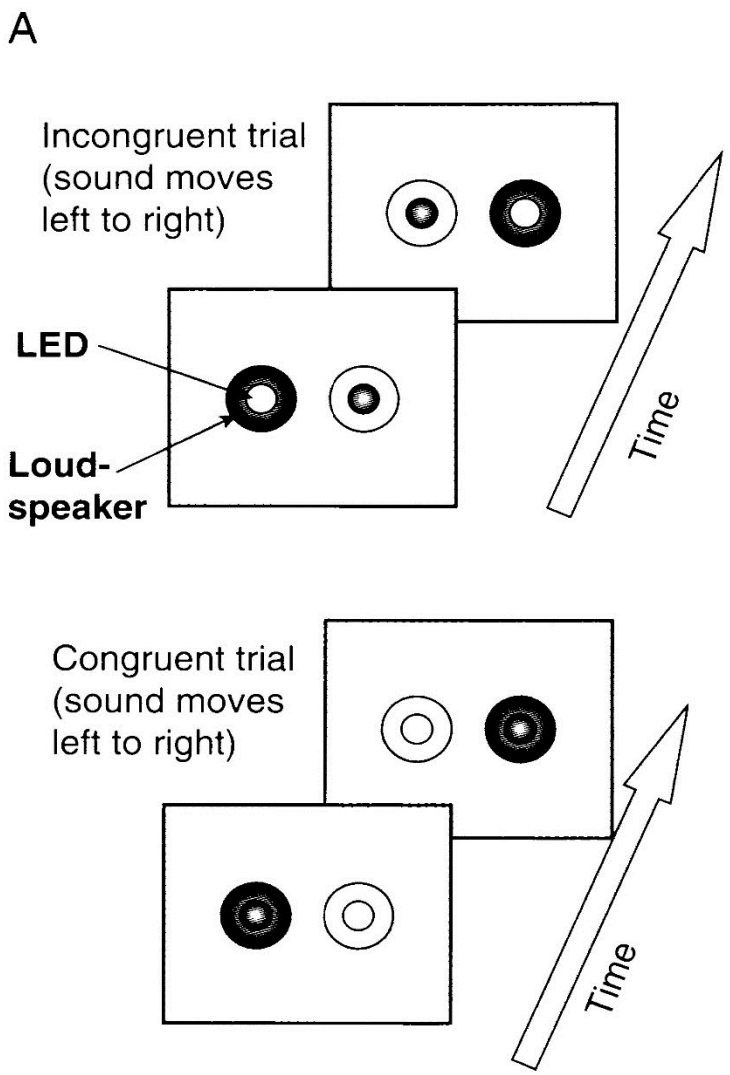
Soto-Faraco et al. (2002)

Visual capture of audition more pronounced for moving stimuli than static stimuli.

Participants judge direction of movement of the auditory apparent motion stream (leftwards or rightwards while simultaneously trying to ignore an irrelevant visual apparent motion stream moving in:
Same direction (congruent)
Opposite direction (incongruent).

Oxford Handbook of Auditory Science: Hearing. C.J. Plack (2010). Chapter 12.

Crossmodal Dynamic Capture

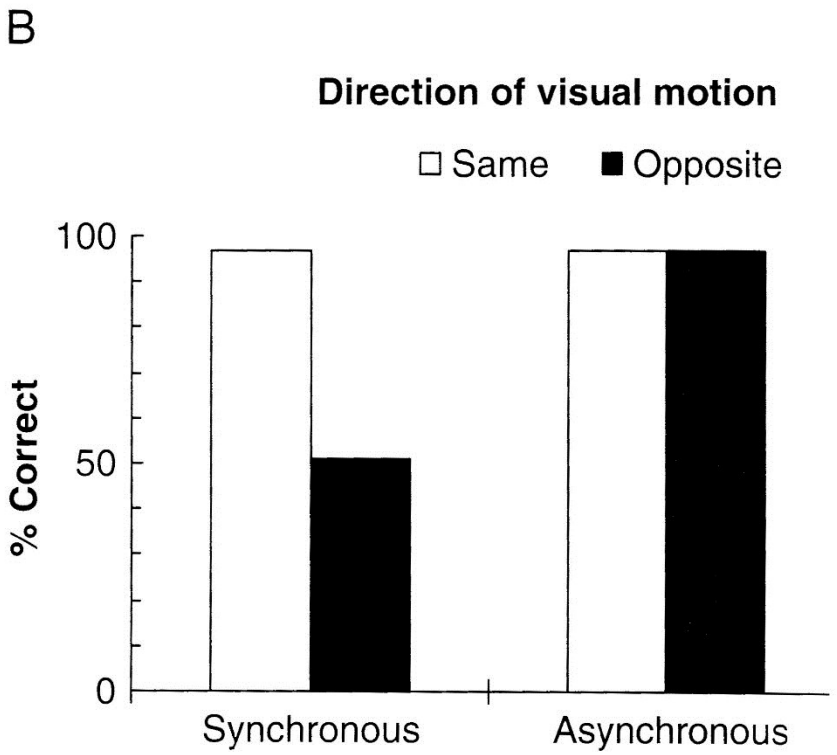


Soto-Faraco et al. (2002)

Fig. 12.7 (A) Schematic illustration of the stimuli used in Soto-Faraco *et al.*'s (2002) study of the audiovisual crossmodal dynamic capture effect. On each trial, a sound (100 ms duration) was presented sequentially from each of two loudspeakers (at an SOA of 150 ms), and each LED was also illuminated sequentially with the same timings. The order of presentation (i.e. left or right first of the stimuli in either sensory modality was entirely unpredictable. The participants' task involved trying to discriminate whether the sound moved from left-to-right (as in the example illustrated in (A)), while trying to ignore the apparent movement of the visual distractors which could move in conflicting (or incongruent, top) or congruent direction (bottom). Grey rectangles represent the stimuli presented at a given moment in the trial, with the inner circles representing the LEDs (white represents off, shaded represents lit) and outer circles representing the loudspeakers (same coding

Oxford Handbook of Auditory Science: Hearing. C.J. Plack (2010). Chapter 12.

Crossmodal Dynamic Capture



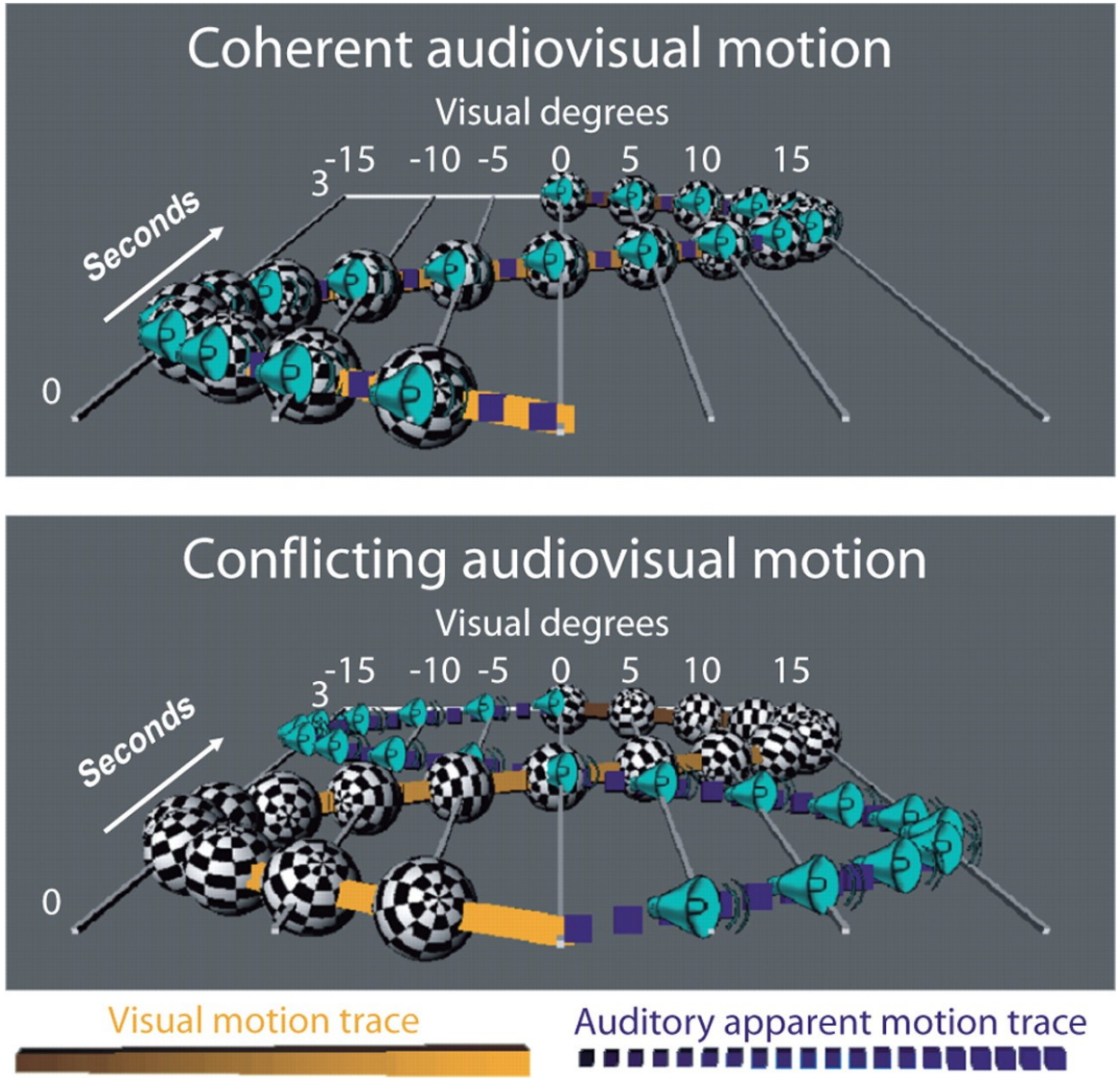
Soto-Faraco et al. (2002)

Results:
% of correct sound motion direction-discrimination responses for those trials in which visual apparent motion was presented in conflicting or congruent direction, either synchronously (example shown in experimental set up on previous slide) or asynchronously (not shown here) in time with the sounds.

Found that: Ability to discriminate direction of auditory apparent motion impaired when visual apparent motion stream moved in opposite, rather than same direction (at least when target and distractor streams were presented simultaneously).

Oxford Handbook of Auditory Science: Hearing. C.J. Plack (2010). Chapter 12.

Crossmodal Dynamic Capture



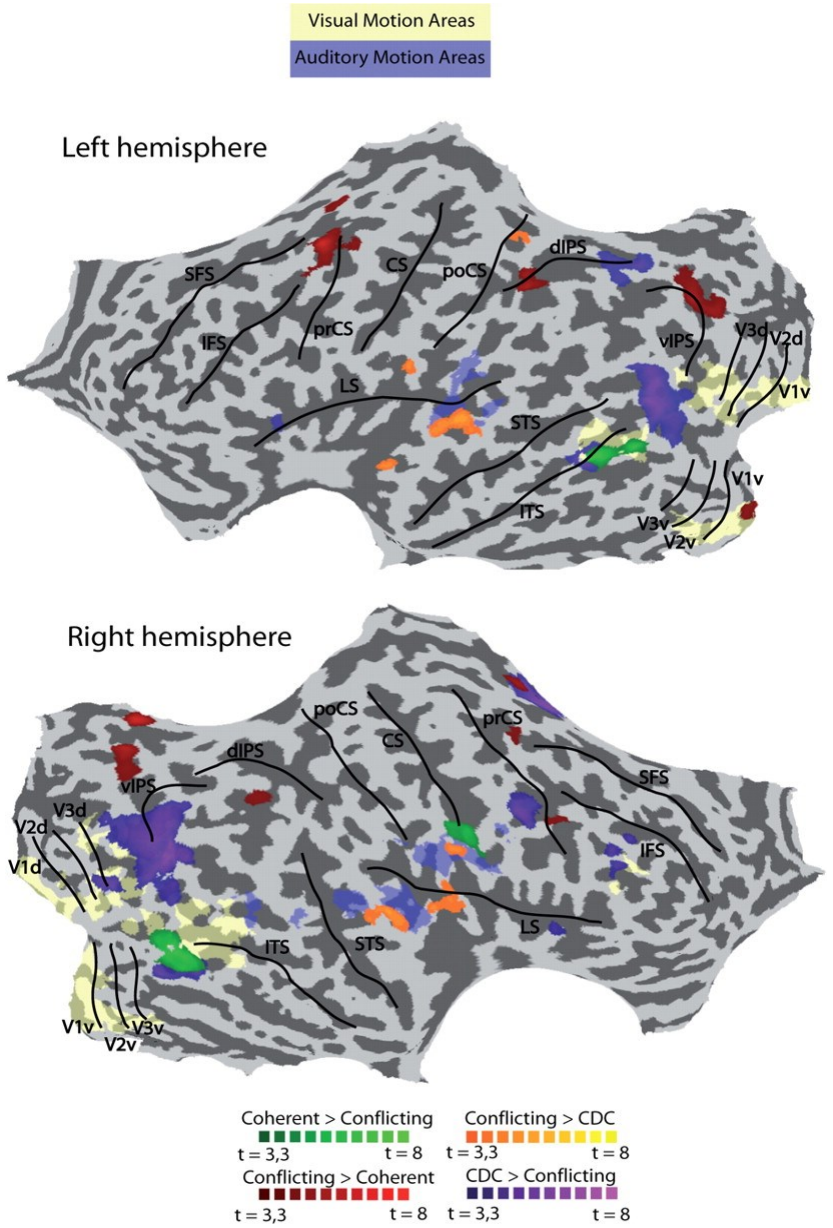
From: Alink, A., Singer, W. and Muckli, L., 2008. Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *Journal of Neuroscience*, *28*(11), pp.2690-2697.

Alink, A., Singer, W. and Muckli, L., (2008)

Stimulation of continuous visual motion and auditory apparent motion.
Spheres represent location of visual stimuli, and speakers represent perceived location of the auditory stimuli over time during a trial.
Auditory stimuli presented with a spatial resolution of 5°.
Stimuli locations shown for the 20 intervals during which an auditory stimulus was presented for 100 ms.
Interstimulus interval was 50 ms.

Top: Coherent trial - audiovisual stimuli move coherently
Bottom: Conflicting trial (bottom) in which motion direction was opposite across senses.

And measured fMRI



Crossmodal Dynamic Capture

Alink, A., Singer, W. and Muckli, L., (2008)

fMRI Results:

Neural correlates of crossmodal dynamic capture can be observed in brain areas dedicated to perceptual analysis of motion in each sense modality.
I.e., increase In neural activity in visual motion area and deactivation of auditory motion areas.

Found that: Early visual and auditory motion areas as well as frontal areas are affected by motion coherency and/or cross-modal dynamic capture - suggests that both the perceptual and the decisional stage of motion processing are involved in the integration of motion across modalities.

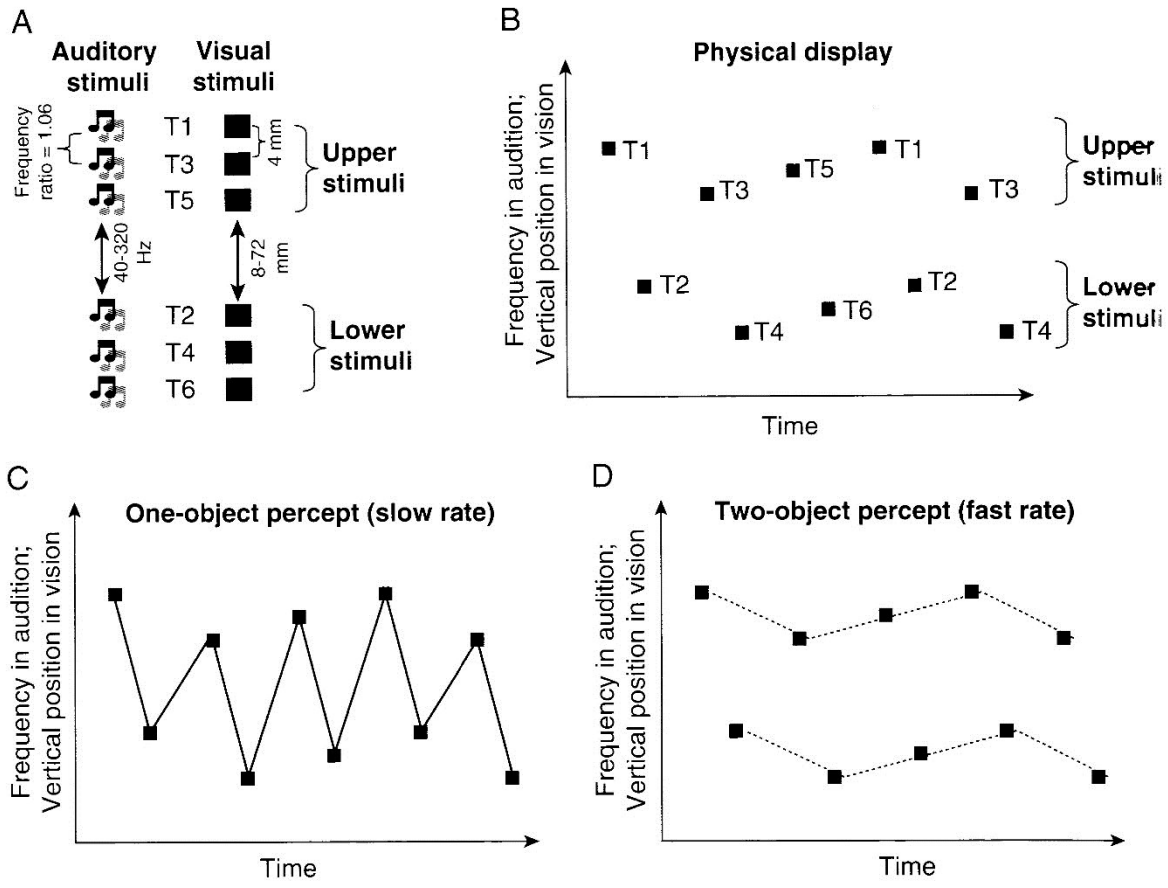
From: Alink, A., Singer, W. and Muckli, L., 2008. Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *Journal of Neuroscience*, *28*(11), pp.2690-2697.

Crossmodal Perceptual Grouping

O’Leary and Rhodes (1984):

Investigated whether the grouping of stimuli presented in 1 sense modality influences how the stimuli presented in another modality are grouped.

Crossmodal Perceptual Grouping

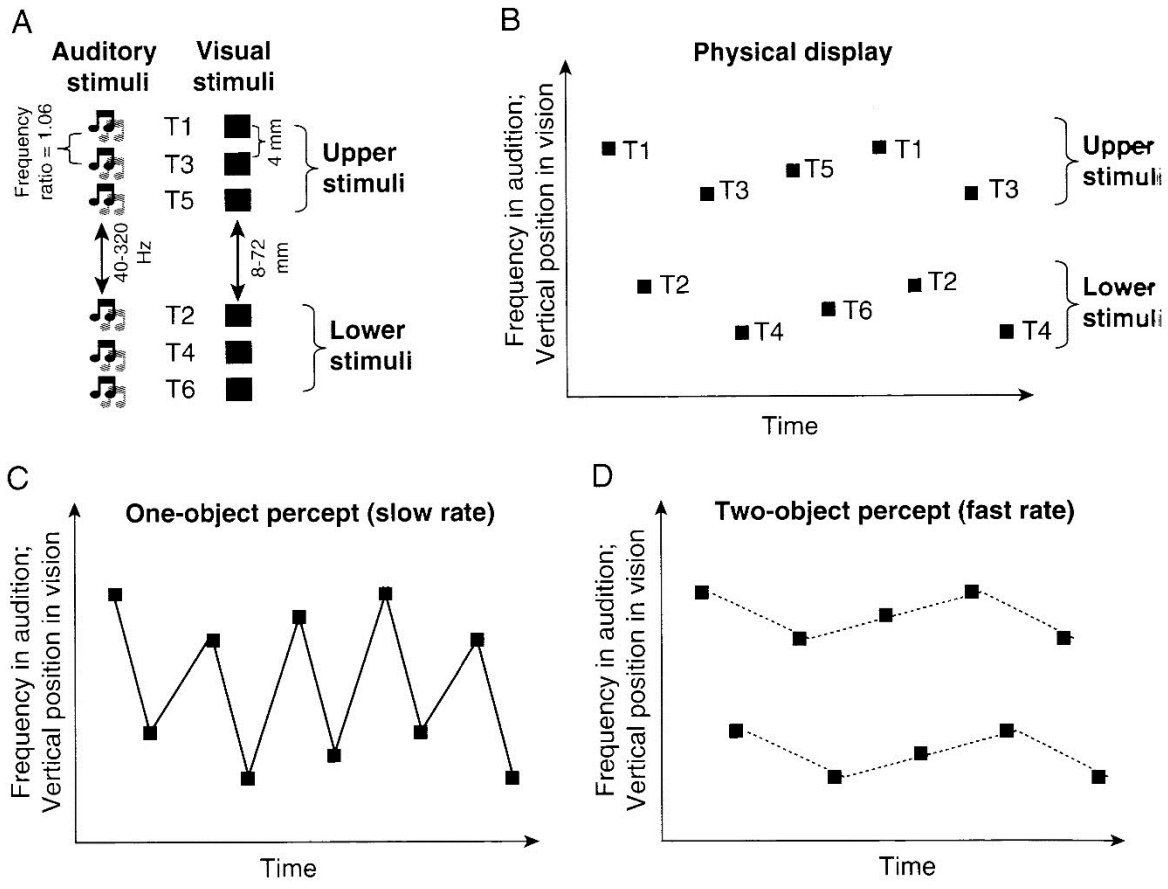


Method:
T1-T6:Temporal order in which stimuli presented.

Participants heard a repeating sequence of six tones (high and low frequency tones in alternation).
Also presented with visual stimuli.
One group of three visual stimuli presented in upper part of visual display, rest presented in lower part of visual display.

Oxford Handbook of Auditory Science: Hearing. C.J. Plack (2010). Chapter 12.

Crossmodal Perceptual Grouping



Result:

Perceptual organisation in visual modality can affect perceptual organisation of stimuli presented in auditory modality.

Slow rates: Perceived single stream (for audio or visual)

Fast rates: Perceived two concurrent streams; One stream in upper frequency for audio, or upper spatial position for visual. Other stream in lower frequency for audio, or lower spatial position for visual.

Oxford Handbook of Auditory Science: Hearing. C.J. Plack (2010). Chapter 12.

Overall Summary

Unisensory and multisensory perception
The Ventriloquism effect- spatial and temporal
Visual and auditory dominance
The McGurk effect
Colavita Effect- auditory and tactile
Crossmodal dynamic capture

Resources

Essential:

Sensation and Perception. E. Bruce Goldstein 8th edition: Pages 316-319 (McGurk Effect).

Oxford Handbook of Auditory Science: Hearing. C.J. Plack (2010). Chapter 12

Hidaka, S., Teramoto, W. and Sugita, Y., 2015. Spatiotemporal processing in crossmodal interactions for perception of the external world: a review. *Frontiers in integrative neuroscience*, *9*, p.62.

Alais, D. and Burr, D., 2004. The ventriloquist effect results from near-optimal bimodal integration. *Current biology*, *14*(3), pp.257-262.

Spence, C., 2009. Explaining the Colavita visual dominance effect. *Progress in brain research*, *176*, pp.245-258.

Shams, L., Kamitani, Y. and Shimojo, S., 2002. Visual illusion induced by sound. Cognitive brain research, 14(1), pp.147-152.

Alink, A., Singer, W. and Muckli, L., 2008. Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *Journal of Neuroscience*, *28*(11), pp.2690-2697.