Psychoacoustic Manipulation of the Sound-Induced Illusory Flash

Sonia Wilkie, Catherine Stevens, and Roger Dean

MARCS Auditory Laboratories, University of Western Sydney,
Locked Bag 1797, Penrith South DC, Sydney, Australia
s.wilkie@uws.edu.au, kj.stevens@uws.edu.au,
roger.dean@uws.edu.au
http://marcs.uws.edu.au

Abstract. Psychological research on cross-modal auditory-visual perception has focused on the manipulation of sensory information predominantly by visual information. There are relatively few studies of the way auditory stimuli may affect other sensory information.

The Sound-induced Illusory Flash is one illusory paradigm that involves the auditory system biasing other senses. However, little is known about the cross-modal illusion. More research is needed into the structure of the illusion that investigates the different conditions under which the Sound-induced Illusory Flash manifests and is enhanced or reduced.

The experiment reported here investigates the effect of new auditory stimulus variables on the Sound-induced Illusory Flash. The variables to be discussed concern forming a contrast in the auditory stimulus to emphasise the illusory percept. The auditory contrasts used were single pitched beeps versus those alternating in pitch by an octave, and the presentation of sound monaurally versus binaurally.

The ultimate aim is to develop the illusory effect as a basis for new intermedia techniques and creative applications for the temporal manipulation and spatialisation of visual objects.

Keywords: Sound-induced Illusory Flash, cross-modal illusion, multisensory interaction, auditory-visual perception, temporal bias, pitch interval, spatialisation.

1 Background

Research on cross-modal interactions in auditory and visual perception has focused predominantly on conditions where visual stimuli are used to manipulate auditory perception (e.g. Alais & Burr, 2004; McGurk & Macdonald, 1976). The results of these experiments suggest that vision is the dominant sense. However, the Sound-induced Illusory Flash (Shams, Kamitani & Shimojo, 2002) exploits the capacity of the auditory system to distort visual perception.

Conditions that give rise to this cross-modal illusion involve presentation of a visual stimulus consisting of a single white dot that is flashed once in the participant's peripheral visual field. This is accompanied by an auditory stimulus of multiple beeps

R. Kronland-Martinet, S. Ystad, and K. Jensen (Eds.): CMMR 2007, LNCS 4969, pp. 223–234, 2008.

of sound. The dual presentation of temporal stimuli appearing to emanate from a single source creates confusion regarding the number of physical flashes perceived and gives rise to the percept that the dot flashes almost as many times as there are beeps. This illusory percept appears to occur because of the superior temporal resolution of the auditory system over the visual system which, in this case, overrides visual information (see comments of Sinex, 1978).

The Sound-induced Illusory Flash is a relatively recent discovery (Shams, Kamitani & Shimojo, 2000). Whilst research (Shams, Iwaki, Chawla & Bhattacharya, 2005; Shams, 2005) has focused on neural mechanisms that underpin the illusion, only the initial studies (Shams, 2002; Shimojo & Shams, 2001; Shimojo, Scheier, Nijhawan, Shams, Kamitani & Watanabe, 2001) explore basic variables of auditory structure that give rise to the illusion. It is these structural variables that are of interest for further exploration of the illusion experience. Investigation of the illusion in this way provides new opportunities for creative applications and intermedia transmission techniques where auditory stimuli might influence visual perception in novel and interesting ways.

2 The Sound-Induced Illusory Flash and Auditory Stimulus Variables – A Review

The variables manipulated in previous research consist primarily of differing combinations of the number of auditory and visual stimuli presented (Shams, Kamitani, & Shimojo, 2002), with only minor adjustments to stimuli across experiments such as pitch frequency at 1kHz or 3.5kHz, the transmission of auditory stimuli through headphones or speakers, and background screen colour of grey or black. The flashes and the beeps were each presented with a constant inter-stimulus interval. The minute alterations of the stimuli were not considered by Shams and colleagues as significant variables, hence there has been little discussion of their effects or interactions.

Brief discussion of the visual stimulus concludes that the illusory percept is stronger when the dots are placed in the periphery rather than fovea (Shimojo, Scheier, Nijhawan, Shams, Kamitani & Watanabe, 2001), but there have been no published studies on the optimal spatial position in peripheral vision. The number of dots presented has been varied. However, multiple dots or correlations with auditory spatialisation have not been investigated.

2.1 Visual and Auditory Rhythms

Shams, Kamitani and Shimojo (2000) used constant intervals between beeps of constant length. This established a constant pulse but no rhythmic or metrical structure was superimposed. For this to be achieved, either some beeps would need greater intensity, or some other novel feature arranged in a recurrent pattern. A broad examination of research into auditory perception, and more specifically, auditory scene analysis (e.g., Bregman), reveals that rhythm of stimuli and its time scale may be important variables. In the context of the Sound-induced Illusory Flash it has only been shown that if the inter-beep time exceeds 70ms the flash illusion is diminished (Shams, 2002).

Whilst this research explores the elementary formations of rhythm, research has not investigated the inter-stimulus interval that would cause perceptual fusion between the auditory and visual stimuli, nor has it investigated the combination of various durations to form rhythmic motifs, or the potential for auditory rhythm motifs to create perception of visual rhythm. Shams, Kamitani & Shimojo (2002) simply noted that "Moderate manipulation of the relative and absolute timings of the auditory and visual stimuli .... do not disrupt the illusion" (p. 152).

To have confidence in this strong conclusion, there is a need for closer and systematic examination of the spatial disparity and inter-stimulus interval for both large and short durations and the effects on fusion.

2.2 Manipulation of Fine-Grained Time Scale

Time scale is an important variable that has been employed and manipulated in many illusory paradigms. The capture of sensory information may increase with time; therefore, misperception sometimes occurs from insufficient time to process perceptual information. Employment of micro-time scales as a variable includes microsound (Roads, 1996; Roads, 2001) with stimuli generated at 600ms or less; the Octave / Scale / Chromatic Illusion (Deutsch, 1981; Deutsch, 1975) involves auditory stimuli of 250ms; The Illusory Continuity of Tones (Bregman, 1999) consists of noise at 50ms or less; and the Auditory Driving of Visual Flicker (Shipley, 1964) comprises auditory stimuli at 150ms. For these examples, the micro time scale may limit perceptual processing resulting in an illusory percept. When the stimulus is presented on larger time scales, the provision of time allows for the acquisition of a greater amount of sensory information, that results in a more accurate representation of the physical stimulus and the illusions begin to fragment.

We have begun to explore the structural variable of auditory rhythm, a musical parameter, in the current research. We predicted that the formation of an auditory rhythmic pattern might influence the extent of illusion differently from the condition in which the auditory stimulus is merely pulsed. We next discuss other musical parameters that may influence the illusion.

2.3 Frequency and Pitch Interval

Frequency has been used only rarely as a stimulus variable for cross-modal manipulation even though Marks demonstrated clear correlations between the percepts of greyness, pitch height and loudness (Marks, 1974). Subjects, for example, stated that low pitches were greyer in colour than high. While pitch has hardly been manipulated in cross-modal illusions, it is at the core of several uni-modal auditory illusions. For example, the Octave Illusion (Deutsch, 1974; Deutsch, 1983), Scale Illusion (Deutsch, 1975) and the Chromatic illusion (Deutsch, 1988), pit the perceptual grouping principles of similarity of frequency and spatialisation against one another resulting in an illusory percept based on pitch proximity. It is possible that both of these variables – pitch relationship and spatial location – may translate to cross-modal illusions in which they influence the judgement of visual stimuli.

The variable of intervallic harmonic relationship can be used to distort perceived direction of motion of consecutive tones. This variable is most notably exploited in investigations of the Tritone Paradox (Deutsch, 1986) that employ multiple layered frequencies around the octave; and Shepard Tones (Shepard, 1964; Risset, 1972; Risset, 1986) that consist of multiple layered frequencies with partials related to the octave or augmented 5th. We argue that the main impact of pitch patterns upon the flash illusion will operate via the formation of a coherent auditory stream: conditions which accentuate such auditory streaming may facilitate the illusion. Since the octave relationship (frequencies in a 1 : 2 ratio) is the closest within tonal harmony, we chose this pitch pattern.

2.4 Stream Segregation and Spatialisation

The organisation of sound to encourage perception as coherent streams is achieved by using a contrast in the stimulus, so that similar, proximal, continuous frequencies, durations or timbres are grouped and associated as emanating from related sources (Bregman, 1999).

Stream segregation is often exemplified with the use of frequency separation, but may also be elicited by factors including a contrast of timbres, spatial proximity of sound, rhythms, amplitudes, and envelopes. These variables have been extensively researched by Bregman (1999) and van Noorden (1977). They have potential for rich creative application as a musical technique using monophonic instruments. For example, spatialisation can be manipulated by headphone presentation of a sound either monaurally (with the same signal in each channel) or binaurally (with the signal in only one or the other channel).

Our prediction was that a pitch or spatial pattern articulating the successive beeps into a metrical structure, superimposed on the basic pulse structure, would enhance the illusion.

2.5 Interim Summary

Thus, there is a need for investigation of new variables – pitch interval and spatialisation – and their effects on auditory-visual perception. In the experiment presented here, pitch interval and spatialisation were used for purposes of illusory emphasis. We anticipated that their application as variables to articulate auditory rhythm, should impact on the Sound-induced Illusory Flash and might enhance the illusion.

2.6 Aim

The aim of the experiment was to investigate the effect of pitch interval and spatialised presentation on the impact of a series of beeps of constant inter-stimulus interval on the Sound-induced Illusory Flash.

2.7 Design

The experimental design consisted of three independent variables. These were the number of auditory stimulus beeps, two, three, four and five; the pitch interval separation of the beeps with the Unison set at 261.5Hz and Octave separation of 261.5 and

523Hz; and the spatialisation of sound with monaural versus binaural presentation. Using headphones, the monaural condition transmitted the auditory stimulus through the left and right channels concurrently, whilst the binaural presentation alternated between the left and right channels. The number of beeps and pitch interval variable were presented within subjects and the spatial presentation variable between subjects.

The dependent variable concerned the visual stimulus and was the number of events perceived, that is, the perceived number of times the dot appeared.

The visual stimulus remained the same in the illusory trials, with the independent variables concerning only the aural stimuli.

2.8 Hypotheses

It was hypothesised that introducing a contrast in the auditory stimulus will cohere the series of individual beeps into a stream, drawing attention to and articulating individual beeps, thereby emphasising the illusory percept.

The contrast in the auditory stimulus is created by using a pitch separation of the octave, and spatial separation with individual beeps presented binaurally alternating between left and right channels.

It was hypothesised that the octave interval creates a greater illusory percept than the unison eliciting perception of two auditory fixation points corresponding to the high and low pitches (reflective of enhanced auditory stream coherence due to superimposition of rhythmic pattern upon pulse). The dot will be perceived to flicker accordingly and rhythmically, emphasising the illusory effect.

It was hypothesised that binaural spatial presentation creates a greater illusory percept than monaural presentation. Binaural presentation draws spatial attention to individual beeps and again articulates a rhythmic pattern that may further enhance the illusory effect.

3 Method

3.1 Participants

A sample of 40 participants naïve to the illusion were recruited. They were Psychology 1A students from the University of Western Sydney and received course credit for their participation. Participants were aged between 17 and 54 years (M = 21.18 years, SD = 6.68), with more female participants than male participants (36 female, 4 male). People reporting a hearing impairment, visual impairment (corrected to normal vision allowed), severe migraines or epilepsy were excluded from testing.

3.2 Stimuli

The visual stimulus consisted of a centrally located fixation point, and a single white dot positioned below the centre of the screen that was located in the participants' peripheral vision. The visual angle of the dot subtended 2° below the fixation point.

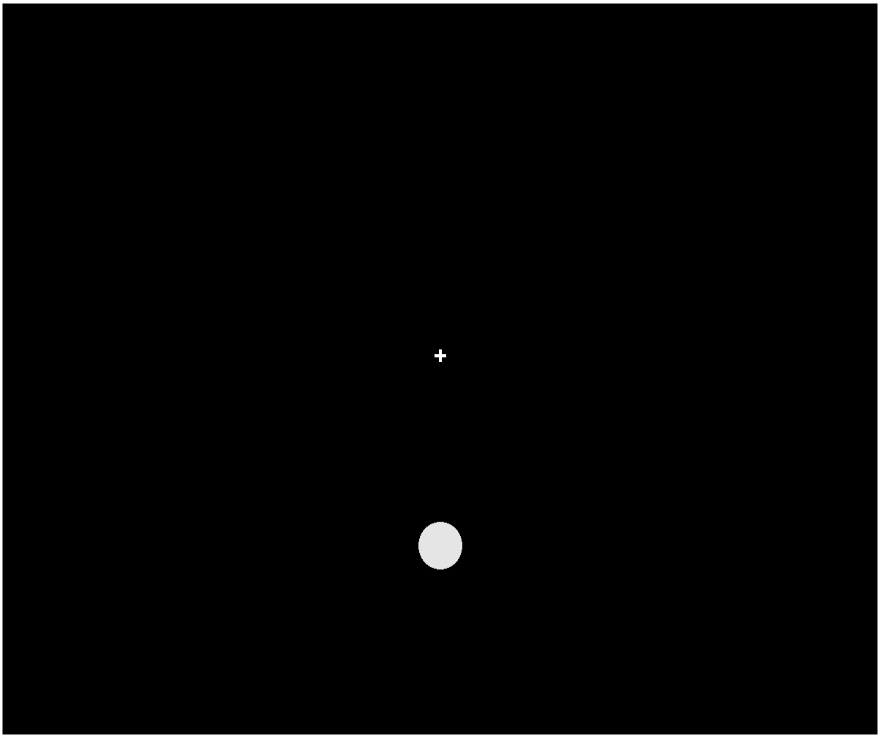

Fig. 1. Screen Capture of the visual stimulus. The fixation point is the centrally located cross. The dot appeared 2° below the fixation point for 17 ms.

The visual angle of the dot to fixation point, onset times, and durations of both the visual and auditory stimulus were derived from those used by Shams, Kamitani & Shimojo (2002). The auditory stimulus consisted of a sine tone generated every 50ms and lasting for a total of 7ms (attack 2ms, sustain 3ms, decay 2ms). The frequencies of the pitches used were the unison set at 261.5Hz and the Octave separation set at 261.5 and 523Hz. The visual stimulus dot was presented for 17ms, a duration that remained the same in all trials. However, the onset time of the dot presentation varied according to the number of beeps presented.

In 2-beep illusory trials the dot was presented 23ms after the auditory stimulus onset (Fig 2 illustrates this particular case); in 3-beep illusory trials the dot was presented 50ms after the auditory stimulus onset and at the same time as the second beep; in 4-beep illusory trials the dot was presented 73ms after the auditory stimulus; and in 5-beep illusory trials the dot was presented 100ms after the auditory stimulus onset and at the same time as the third beep.

Trials were also included where the dot was presented at the same time and as many times as the auditory beeps, i.e., two physical flashes with two auditory beeps; three physical flashes with three auditory beeps; etc. These trials were included as a

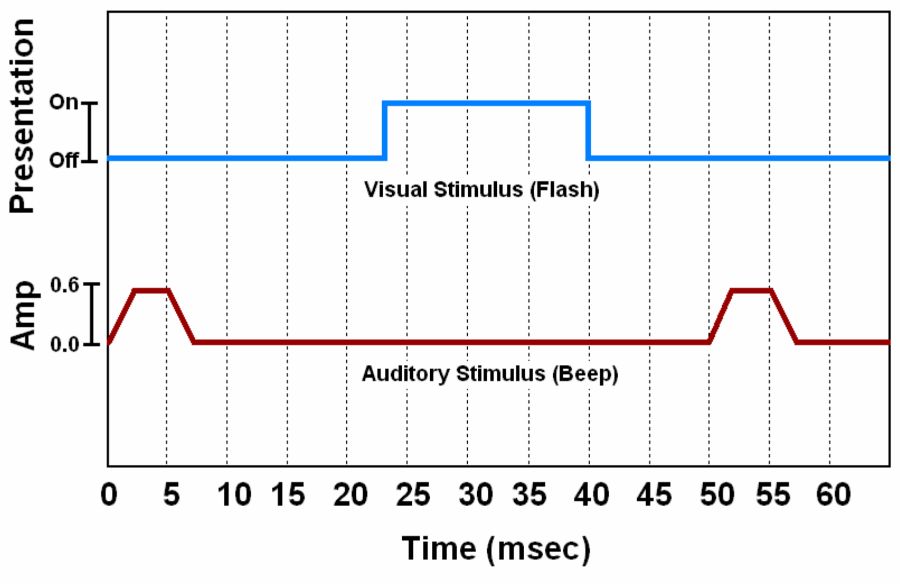

Fig. 2. Stimulus onset time for 2 beep illusory trial. The auditory stimulus is presented every 50ms and lasts for a duration of 7ms (attack 2ms, sustain 3ms, decay 2ms). The visual stimulus is presented 23ms after the auditory stimulus onset, and for a duration of 17ms.

rationale to disqualify participants with poor performance, however they may have biased the illusory percept by causing participants to expect multiple physical flashes. All trials were presented in a random order with equal chance of the stimulus being illusory or not. Our equal presentation ratio of illusory to physical flash trials is greater than that of the 15 illusory trials to 45 physical trials used by Shams, Kamitani & Shimojo (2002). Therefore, our expectation bias should be lesser than theirs.

Presented with 32 conditions in total (16 illusory flash conditions, 16 physical flash conditions), all participants were presented with trials using the variables of pitch interval (unison versus octave) and number of beeps (two, three, four and five presentations). The spatial presentation (monaural versus binaural) variable was divided into two blocks with participants receiving either monaural or binaural presentation.

3.3 Apparatus

Participants were located at computer workstations with their head positioned on a chin rest 40cm from the computer monitor and eyes level with the fixation point.

A Mac Pro G5 with a Diamond Digital cathode-ray-tube (CRT) monitor (85 Hz refresh rate) was used and sound was presented through AKG K271 headphones.

The software MAX / MSP was used to construct an application that generated the auditory and visual stimulus; presented the trials in a randomised and collected order (using the urn object); generated the questionnaire as an onscreen pop-up window; and collected the participants responses in a text file.

The presentation of the visual stimulus was recorded with a Vision Research Phantom V10 high-speed camera to determine its duration. The camera, recording with a frame rate of 1ms, showed that the presentation of the dot indeed occupied 17ms. However, the computer monitor's refresh rate presented horizontal contrasts in luminance, with complete fading of the dot at 8ms from onset, for a duration of 5ms.

This is perhaps the first high speed recording of the visual stimulus to determine the physical nature of the presentation and raises the question of the impact of the raster scan and refresh rate of the monitor on illusory percept of the illusion. However, as the stimulus is presented on such a micro time scale, the window of visibility should not impact on the temporal percept (Zele and Vingrys, 2004). It is also highly likely that our refresh rate is comparable to that used by Shams.

3.4 Procedure

Participants sat at a computer workstation and were informed of the experiment and procedure. They were given an information sheet, summarising both the experiment procedure and the ethics approval, and they signed a consent form. They were instructed to put on headphones, place their head on the chin rest, to focus on the fixation point and use their peripheral vision to count the number of times a dot was presented. Participants were assigned alternately to monaural or binaural auditory stimulus presentation.

The task required them to state on an onscreen multiple choice questionnaire the number times the dot was presented, ranging from one event to five events, within an 8 second time limit. Each trial lasted for 11 – 13 seconds. With a total of 96 trial presentations the experiment lasted for approximately 19 minutes.

4 Results

The data collected in the experiment refer to the number of flashes perceived. The experiment recovered the previously described illusion and, as hypothesised, pitch interval and binaural presentation enhanced the illusory effect. The mean perceived number of flashes, as a function of beep number, spatial presentation, and pitch separation, is shown in Figure 3.

We anticipated that the number of beeps would influence the perceived number of flashes. We performed a repeated measures analysis of variance that included two within subject variables - the number of beeps presented and the pitch interval, and one between subject variable of spatial presentation. As predicted, there was a main effect for number of beeps $F(3,36)=159.85$, $p<.001$. Our prediction that the use of pitches alternating at the octave would enhance the flash illusion was upheld : there was a main effect for pitch interval $F(1,38)=77.616$, $p<.001$, and an interaction between number of beeps x pitch interval $F(3,36)=15.09$, $p<.001$.

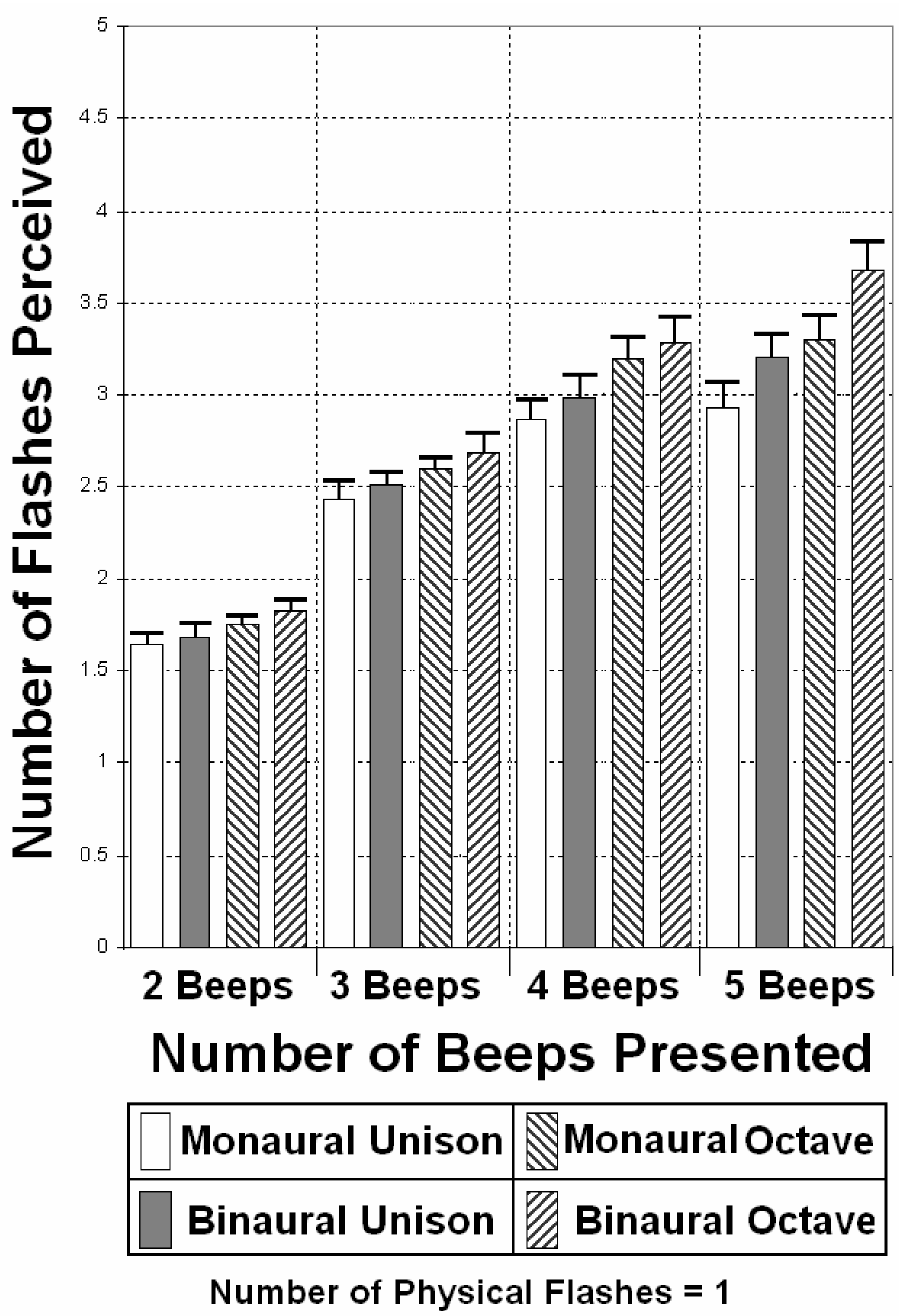

Fig. 3. The mean number of flashes perceived. Error bars refer to standard error of mean.

5 Discussion

The results indicate that modifying the auditory stimulus with the octave interval but not with binaural presentation enhanced the illusory effect of the Sound-induced Illusory Flash. That is, by creating a contrast between individual beeps with frequency separation using the octave interval segregated individual beeps which emphasised temporal information and influenced the visual event perceived. We interpret this as an influence of formation of rhythmic patterns.

However, the complementary rhythmic contrast that was created using spatial separation so that the binaural presentation alternated individual beeps between left and right channels did not produce a statistically significant effect.

One explanation for spatial separation not being as effective as pitch separation for the segregation of independent streams may be related to the visual stimulus. In this experiment the visual stimulus presented a single dot that was centrally located. The spatialisation of sound into two separate channels implies that there are multiple individual objects from which the sound source emanates. As the visual stimulus presented a single object that was centrally located, and not two objects (dots) spatially placed in association with the auditory stimulus (ie left and right), the visual information of the single object in this experiment may have overridden the auditory spatial information that would have created a rhythmic pattern. Therefore the lessened rhythmic articulation and emphasis on individual beeps produced very limited illusory effect.

The present results warrant further exploration of the auditory stimulus and systematic manipulation of variables of pitch separation and spatial separation, and their associations with the visual stimulus. It would be interesting to investigate the illusory percept if the visual stimulus involves multiple dots that are spatially separated and whether the auditory stimulus drives temporal spatial perception of the dots.

The outcomes of this research may also be useful for creative application. In relation to electronic arts, the principles of the illusion raise possibilities for use as a strobe effect. We (Dean, Whitelaw, Smith & Worrall, 2006, p317) have previously developed the concept of 'algorithmic synaesthesia' in which generated streams of sound and image share data, algorithmic processing, and segmentation, or all three features. The concept is primarily one of aesthetic and cultural cognition, and it was not claimed that it necessarily involves cross-modal perceptual illusion or true synaesthesia. However, the data here suggest that during creative applications of algorithmic synaesthetic processes, it may be possible to generate true cross-modal illusions, which might influence the cognitive and affective responses of viewers.

Acknowledgements. Thankyou to Dr. Freya Bailes for assistance with statistical analysis. The research reported here was supported by the University of Western Sydney Australian Postgraduate Award and Top-Up Award.

References

1. Alais, D., Burr, D.: The ventriloquist effect results from near-optimal bi-modal integration. Current Biology 14, 257–262 (2004)
2. Bhattacharya, J., Shams, L., Shimojo, S.: Sound-induced Illusory Flash Perception: Role of Gamma Band Responses. Cognitive Neuroscience and Neuropsychology 13, 14 (2002)
3. Bregman, A.: Auditory Scene Analysis. The MIT Press, Cambridge Massachusetts (1999)
4. Dean, R.T., Whitelaw, M., Smith, H., Worrall, D.: The mirage of real-time algorithmic synaesthesia: Some compositional mechanisms and research agendas in computer music and sonification. Contemporary Music Review 25, 311–327 (2006)
5. Deutsch, D.: An auditory illusion. Nature 251, 307–309 (1974)
6. Deutsch, D.: Two-channel listening to musical scales. Journal of the Acoustical Society of America 57, 1156–1160 (1975)
7. Deutsch, D.: The octave illusion and auditory perceptual integration. Hearing Research and Theory 1, 99–142 (1981)
8. Deutsch, D.: The octave illusion in relation to handedness and familial handedness background. Neuropsychologia 21, 289–293 (1983)
9. Deutsch, D.: A musical paradox. Music Perception 3, 27–280 (1986)
10. Deutsch, D.: The semitone paradox. Music Perception 6(2), 115–132 (1988)
11. Grove, R.: Thinking in Four Dimensions. Melbourne University Press, Melbourne (2005)
12. Marks, L.: On the associations of light and sound: the mediation of brightness, pitch, and loudness. American journal of psychology 87 (1974)
13. McGurk, H., Macdonald, J.: Hearing lips and seeing voices. Nature 264, 746–748 (1976)
14. Risset, J.-C.: Musical Acoustics. IRCAM, Paris (1972)
15. Risset, J.-C.: Pitch and rhythm paradoxes: Comments on 'Auditory paradox based on fractal waveform'. The journal of the Acoustical Society of America 80(3), 961–962 (1986)
16. Roads, C.: The Computer Music Tutorial. The MIT Press, Cambridge (1996)
17. Roads, C.: Microsound. The MIT Press, Cambridge (2001)
18. Shams, L.: Integration in the brain - the subconscious alteration of visual perception by cross-modal integration. Science and Consciousness Review 1, 1–4 (2002)
19. Shams, L., Iwaki, S., Chawla, A., Bhattacharya, J.: Early modulation of visual cortex by sound: an MEG study. Neuroscience Letters 378, 76–81 (2005)
20. Shams, L., Kamitani, Y., Shimojo, S.: What you see is what you hear. Nature 408, 788 (2000)
21. Shams, L., Kamitani, Y., Shimojo, S.: Visual illusion induced by sound. Cognitive Brain Research 14, 147–152 (2002)
22. Shams, L., Ma, W.J., Beierholm, U.: Sound-induced flash illusion as an optimal percept. Neuroreport 16, 17 (2005)
23. Shepard, R.: Circularity in judgements of relative pitch. Journal of the Acoustical Society of America 36, 2346–2353 (1964)
24. Shimojo, S., Scheier, C., Nijhawan, R., Shams, L., Kamitani, Y., Watanabe, K.: Beyond perceptual modality: auditory effects on visual perception. Acoustic Science and Technology 22(2), 61–67 (2001)
25. Shimojo, S., Shams, L.: Sensory modalities are not separate modalities: plasticity and interactions. Current opinion in neurobiology 11, 505–509 (2001)
26. Shipley, T.: Auditory flutter driving of visual flicker. Science 145, 1328–1330 (1964)

27. Sinex, D.G.: Cross-modality temporal resolution for auditory, vibrotactile and visual stimuli. Journal of the Acoustical Society of America 63(suppl. 1), 52 (1978)
28. van Noorden, L.P.A.S.: Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. Journal of the Acoustical Society of America 61(4), 1041–1056 (1977)
29. Zele, A.J., Vingrys, A.J.: Cathode-ray-tube monitor artefacts in neurophysiology. Journal of Neuroscience Methods 141, 1–7 (2004)