Visual and Auditory Velocity Perception and Multimodal Illusions

Katherine S. Gasaway

Advisor: Paul M. Corballis, PhD

Reviewers: Paul M. Corballis, PhD, Eric Schumacher, PhD

Date: March 12th, 2008

Pages: 20

Keywords: Vision, audition, velocity, perception, illusion, ventriloquist, capture

Abstract

Although ambiguous and conflicting sensory information from different sensory modalities is common, people seldom experience perceptual ambiguities or conflict between senses. Just as the retinal nerve blind spot is “filled in” and seldom seen, conflicting or otherwise confusing sensory information is resolved in favor of the most appropriate modality, eliminating the confusion from conscious experience. The ventriloquism effect and auditory driving are two examples of perceptual phenomena arising from this sensory override. This research explores the hypothesis that velocity perception is subject to the same effects. Subjects were presented with two bimodal (auditory-visual) stimulus pairs and asked to determine which of the visual stimuli was moving faster. In a V2A2/V2A1 condition, participants responded significantly more frequently that the first visual stimulus was faster than in any non-target condition. This effect was not found for V2A2/V2A3 trials.

Visual and Auditory Velocity Perception and Multimodal Illusions

While realists and pragmatists have argued about the absolute nature of our perceptions, there is seldom any question about the role our senses play as the lens through which we experience the world (Baum, 2005). As a result of this unanimity and the numerous industrial and organizational applications of sensory research, the body of literature concerning both the physiology of the senses and the subjective experience of perception has grown rapidly. Motion detection is among the most well-studied aspects of perception. This research project is concerned with the union of visual and auditory motion information, and particularly with how velocity information from the two sensory modalities is combined.

Information from different sensory modalities is integrated for better interaction with the environment. As a result, perception can either be enhanced, impeded, or remain unaffected by this integration. At the single-neuron level in brain structures responding to more than one modality (so-called polymodal brain regions), when visual and auditory stimuli are presented in close temporal and spatial proximity, the neuronal response is typically enhanced (Frassinetti, 2002; Stein & Meredith, 1993). Conversely, separate auditory and visual stimuli produced a reduced or unchanged response (Stein & Meredith, 1993; Donovan, Lindsay, & Kingstone, 2004; Howard, Craske, & Templeton, 1965). In a recent experiment reported by Sanabria, Soto-Faraco and Spence’s 2005 paper, the effect of polymodal distractors was shown on a behavioral level. When participants were presented with a visual-tactile distractor and were asked to indicate in which direction a simultaneous auditory stimulus was moving, performance was significantly better when the distractor and target stimuli moved in the same direction rather than in opposite directions (Sanabria, Soto-Faraco & Spence, 2005). Similarly, in Shore, Barnes and Spence's examination of visual distractors on tactile information, it was found that performance on a tactile location discrimination task declined when a visual distractor provided incorrect information (i.e. the visual distractor was on the left, and the tactile target was on the right) and the onset for the distractor and target were temporally close (Shore, Barnes & Spence 2006). In Stanley and Matthews' 2004 experiment it was found that sensitivity to the direction of simulated auditory motion was significantly impaired when preceded by an invalid directional cue (Stanley & Matthews, 2004).

Although ambiguous and incomplete sensory information is common -- especially in the polymodal case, people seldom experience these occurrences as ambiguous. Rather, the conflicting information is usually perceived as a unified whole. Typically, biases in these reinterpretations of sensory information are thought to reflect the idea that when a conflict occurs, the perceptual system resolves the ambiguity in favor of the system with the information with the best resolution for the domain under consideration (Welch, DuttonHurt & Warren, 1986). This process is called *modality appropriateness (*Welch & Warren, 1986).

The "Deutsch illusion" illustrates a case where a unimodal ambiguity is resolved into more meaningful patterns. When a participant is presented with two different sets of auditory stimuli (one in each ear), the localization of those stimuli depends on the pitch relationships between the competing tones over the physical location of the sound (Deutsch, 1974). In Deutsch's 1974 experiment, participants were presented with continuous alternating 400Hz and 800Hz tones in both ears. When the 400Hz tone was presented in one ear, the 800Hz tone was presented in the other and vice versa. While it is easy to imagine the correct percept, none of the participants related it and instead reported hearing a single tone that oscillated pitch as it changed localization. For all but one participant the higher tone localized in the right ear and Deutsch suggested that this represented a localization of the higher pitch to the dominant hemisphere of the brain (Deutsch, 1974).

For multimodal ambiguities, the change in perception depends upon the kinds of information and on the specific sensory modalities involved. Most dramatically, vision provides more accurate spatial information than sound (Welch & Warren, 1986) and audition provides more accurate temporal information than vision (Shipley, 1964; Welch & Warren, 1980). For example, In the McGurk effect, what is heard is influenced by what is seen (McGurk & MacDonald, 1976). If a participant is presented with the syllable /ba/ as an auditory stimulus and sees the speaker say /ga/ the final auditory percept may be /da/ (McGurk & MacDonald, 1976).

Another salient example of vision affecting sound perception is the ventriloquist effect. Howard and Templeton (1966) describe it as the localization of sounds biased toward and by simultaneous visual stimuli (Howard & Templeton, 1966). When a ventriloquist speaks without moving his lips while syncing the movements of his dummy's mouth with his words, the sound will appear to come from the dummy's mouth although the visual and audio information are in conflict. This effect has been robustly and consistently reproduced and best presents itself when the audio and visual stimuli are spatially and temporally close (Alais, 2004; Bischoff, 2007; Lewald, 2003; Soto-Faraco, 2004; Vroomen, 2001; Vroomen & Bertelson, 2001; Wallace, Robertson, Hairston, Stein, Vaughn, & Schrillo, 2004).

Auditory driving, or auditory capture, is another polymodal illusion (Brooks, 2007; Bertelson, 2002; Fendrich & Corballis, 2001; Morein-Zamir, 2003; Spence, 2003; Vroomen, 2001; Vroomen, 2004). It can be defined as a phenomenon in which audition has been found to capture visual temporal perception. When temporal information is involved, audition provides the most accurate information. Part of the reason for this is that the physical mechanisms behind hearing take less time than the chemical and physical processes behind vision therefore, the physiology of the auditory system is faster than the physiology of the visual system (Spence & Squire, 2003).

Although audition results are more robust at around 16hz (Fendrich & Corballis, 2001), auditory driving can also be experienced at lower frequencies. For example, when a person has the turn signal on in their car and the person in the car in front of them does too, the two signals appear to be in sync or directly out of phase initially. After a few seconds, the disparity between the signals becomes apparent, but before that realization occurs, the click of the signal in the car drives the perception of the flash of the signal ahead of the driver. This is called “flutter driving flicker” (Fendrich & Corballis, 2001). While Bertelson contends that the opposite effect, “flicker driving flutter”, has not been found (Bertelson, 2002), the effect had been found although it is smaller than the reverse (Fendrich & Corballis, 2001). In their 2001 experiments, Fendrich and Corballis asked participants to make spatial and temporal judgments. In one condition, participants related when a flash occurred by indicating its position on a rotating marker resembling a clock face. In a trial where a click preceded or followed the flash, participant responses were significantly different from trials in which the click and flash were simultaneous and were biased toward the temporal location of the click. This is auditory driving. In another condition participants were asked to relate where the marker on the dial was when a click was presented. When a flash preceded or followed the click, responses were biased in the direction of the flash although the effect was smaller than that found in the first condition (Fendrich & Corballis, 2001).

Velocity information, by definition, is both spatial and temporal. The question at hand is whether either auditory driving or the ventriloquism effect plays a role in velocity perception. This study examines the possibility of auditory driving and whether earlier conclusions about the effect of auditory stimuli on visual perception with respect can be generalized to velocity perception.

Like all other human capacities, there are limits to motion perception. For visual motion to be perceived, the stimulus must traverse a distance large enough to be noticed. Also, the stimulus must travel within certain velocity constraints: if the motion is too slow, it will be imperceptible (Saberi & Perrott, 1989; Chandler & Grantham, 1991). If it is too fast, then instead of getting a motion percept, it will appear as if two stimuli were being presented (Spence, 2003). Similar limits on auditory limits on motion perception also stand. Auditory stimuli must traverse a minimum angle and not exceed a maximum velocity in order to be perceived as motion (Grantham, 1986; Perrot, 1997; Chandler, 1991; Saberi, 1990).

Certain difficulties are involved in digitally presenting sensory stimuli. A visual stimulus presented digitally will not be “real” motion but rather will consist of the rapid presentation of stationary images offset by small distances to create the illusion of movement. For this experiment the 3 by 3 pixel squares presented as visual stimuli were offset by 1, 2, or 3 pixels per screen refresh. Audio motion presentation can be physically represented with moving speakers (Chandler & Grantham, 1991), but there are disadvantages to this method including the difficulty of making a silent mechanism to convey the speakers from point to point and the special acoustical considerations that must be taken. Other studies have used multiple speakers generating short independent tones and independent LEDs to present apparent audio and visual motion (Sanabria, Spence, & Soto-Faraco, 2007, Perrott & Musicant, 1977), however, when studying velocity perception this configuration does not allow for enough variability in either visual or audio stimulus presentation (without prohibitively many speakers). To simplify these problems, headphones were used and motion in sound was created by modulating the volume of the sound in each headphone according to the "stereoscopic law of sines" (Grantham, 1986; Bauer, 1961).

Although auditory motion can be simulated either by interaural intensity differences (IIDs) or by interaural time differences (ITDs). Simulating auditory motion with IIDs involves the modulation of the amplitude of an auditory stimulus from one source to another to recreate the differences in stimulus intensity between the ears as the source of a sound moves. To use ITDs to simulate auditory motion auditory stimuli have to be presented at precisely different times from different sources to imitate the differences in arrival times of auditory information to each ear. We chose the former method because several reports have suggested that human observers are more sensitive to IIDs (Grantham, 1984; Rosenblum, Carello, & Pastore, 1987; Altman, Variahuina, Nitkin, Nikolay & Radinova, 1999; Stanley & Matthews, 2004).

The Colavita dominance effect refers to the phenomenon in which participants when presented with unimodal visual, unimodal audio, or bimodal audio-visual stimuli, fail to respond to the auditory component of stimulus sets significantly more frequently than the visual components (Koppen & Spence, 2007). Velocity, a frequently multi-modal percept, depends upon both spatial and temporal information, but the Colavita dominance effect suggests that participants may ignore the audio stimuli presented in this experiment. Participants were asked to respond which visual stimulus in two visual-audio bimodal stimulus pairs was moving faster. By examining the circumstances under which a participant judges a visual stimulus to be faster than the other when the visual stimuli are held constant and a paired audio stimulus is varied, evidence for auditory driving may be found. Under consideration are the speeds of the audio stimulus simultaneously presented and the difference in the speeds of the visual stimuli in the trial. As little research has been conducted on this issue, velocity is both spatial and temporal, and audition and vision have been shown to affect each other under varying conditions, I am hesitant to state a concrete hypothesis about whether auditory driving will be found. Rather, this experiment will explore whether audition can affect vision under these conditions.

To the best of my knowledge, this is the first foray into the subject of perceptual illusions in velocity perception and therefore it is necessary to treat this experiment as a pilot study – a test of the methods and standards that should be applied in future research. Should the results be significant, future research could expand on the limited scope of this experiment.

Method

Participants

N = 10 undergraduate students at Georgia Tech received course credit for participating in this study (M age = 19.2 , SD = 0.95; 6 male). All participants had normal or corrected to normal hearing and vision and gave informed consent before participating in this experiment. The researchers received Institutional Review Board permission for all experiments performed.

Stimuli and Apparatus

Participants were seated approximately 57cm from a 45.72 cm 1024x768 pixel monitor with an 85Hz refresh rate. An adjustable chin-rest was used to minimize head movement. The experiment was programmed in Presentation 11.0 and run on a Dell computer and a flat-screen Dell monitor.

The visual stimuli were 3x3 pixel white squares presented in sets of 125 per trial. The pixels occupied and traversed a field subtending 42.7º visual angle by 33.5□. All visual stimuli moved from left to right. Velocity was operationalized as pixel displacement per screen refresh. There were 85 screen refreshes per visual stimulus presentation for a running time of 1000 ms. There were three levels of visual stimuli. In the three levels, V1, V2 and V3, dots were shifted 1,2, or 3 pixels per screen refresh for a potential 3.8, 7.6 and 11.4 degrees visual angle displaced per dot, respectively. Dots were programmed to change from white to black (the background the visual stimuli were presented against was black) somewhat randomly to prevent participants from tracking individual dots.

The auditory stimuli were wav files of pure 500Hz tones (duration 1000ms) generated using Audacity. Velocity was operationalized as interaural intensity differences (IIDs). For all audio stimuli the amplitude in left ear is initially large and then gradually decreased over the duration of the stimulus. In the right ear, the audio stimulus is initially small and is gradually increased. The dB is 5dB, 7.5dB, and 10dB for A1, A2 and A3, respectively. All audio stimuli were presented through headphones.

Procedure

Participants were seated, and the chair height was adjusted so that their chin rested comfortably in the chin-rest. The participant was then given written and verbal instructions. The investigator watched while the participant put on the headphones and made sure that they were on correctly (right ear to right headphone). The participant began by pressing enter to start the practice section and fixating on the red cross in the center of the screen. During each trial participants saw two bimodal pairs of stimuli: one visual and audio pair and then, after a 1000ms break, another pair (see figures 2 and 3). Immediately after the second pair, a prompt was displayed to ask participants to respond with which of the visual stimuli was moving faster. Participants had 5000ms to respond. The next trial would begin immediately after the participant’s response or after that time was allowed to expire. Each participant ran one session of 3 blocks with 81 trials per block for a total of 243 trials.

Analysis

The within-subject variable examined in this experiment was the pattern of responses to target cases with two levels. Each target case consisted of first a V2-A2 pair and then either a V2-A1 pair or a V2-A3 pair (see figures 1 and 2). The V2-A2 pair was used as the first stimulus pair in all target conditions to eliminate the possibility that the first set of stimuli could automatically determine the correct answer. Ideally more pairings would have been used, but that is a consideration for future research. Although all possible combinations of the three levels of stimuli were run in the experiment, most of these combinations served as distracters. The V2-A2/V2-A2 condition represented a true neutral in which the only variant would be the participant's preference. The target data were analyzed with respect to V2-A2/V2-A2 conditions only. Non-target caseswere not examined. An Analysis of the Variance (ANOVA) was used to examine the evidence for differences among the levels. A sign test was used to examine the direction of individual participant responses.

Results

The primary dependent variable investigated in this experiment was participant response in trials consisting of a V2-A2 pair followed by a V2-A1 or V2-A3. An ANOVA showed a significant difference between participant performance in target trials and neutral trials ($F(df\ 2) = 5.535$, $p < .01$). A post hoc comparison, Tukey's HSD, showed significant differences in participant responses between neutral trials and V2-A2/V2-A1 trials ($p = 0.003$). A significant difference was not found between V2-A2/V2-A3 trials and neutral trials ($p > .05$). Table 1 relates the distribution of participant responses across experimental conditions.

Discussion

The present experiment was conducted to provide new information about the factors that underlie velocity perception. The results of the experiment suggest that although the Colavita effect suggests that auditory stimuli may be ignored by participants, audition plays an influential role in velocity judgments. An examination of the variance showed evidence for a difference between responses in the target conditions and neutral conditions. A Tukey's HSD post hoc analysis showed a significant difference in participant performance in conditions where a slower auditory stimulus was presented second in target trials. Participants were likely to respond the first visual stimulus was faster to a degree outside of chance. A similar difference was not found for the other target condition in which a faster auditory stimulus was presented second.

This asymmetrical result is interesting, however, the limited scope of this experiment necessitates further investigation before the result can be considered generalizable. An unfortunate effect of the necessitated expediency of the timeline for this experiment was that there was not time to run all the desired parameters. Ideally a larger range of target cases would have been examined and participants would have been run twice to get a large enough individual dataset to properly examine individual differences. Also, when participant responses to neutral conditions are examined, some individuals showed significant biases toward one answer or another (Table 1). While a participant responding correctly would show 50% of their responses allocated to each option, nearly all participants favored one response or another. Also across both target conditions and the neutral condition, only one participant showed the expected influence.

It is possible that the asymmetrical effect found is an artifact of the auditory or visual velocities explored, the size of the differences among the stimuli, or individual differences among participants. While the individual participant results do not support the presence of an effect of auditory driving, there may still be an overall effect. As with any initial foray into a subject, replication with different participants and expansion of the stimulus set will be necessary.

Although it was originally the researcher's intent to test for the presence of the ventriloquism effect, that question remains uninvestigated and significant. As Fendrich and Corballis's study suggests, the influence of audition on vision seen here may only be a smaller effect to be seen in the light of a larger effect (Fendrich & Corballis, 2001). To test for ventriloquism effect an experiment could be created using similar parameters to the current experiment. Instead of holding the visual stimulus constant and varying the auditory stimulus, the auditory stimulus would serve as the standard and the visual stimulus would be varied. If participant responses were significantly biased toward the value of the visual stimulus, evidence for the ventriloquism effect would be found. If the pattern shown in Fendrich and Corballis' experiments generalizes to velocity perception, then the ventriloquist effect would be larger than the auditory driving effect. If this were found to be the case, it may say something about the overall nature of the sensory perception systems. It would be necessary to attempt to locate when and where the integration of these kinds of information takes place to establish further causality.

Bibliography

Alais, D., Burr, D. (2006) The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*. 257-262.

Altman, J.A., Variaguina, O.V., Nitkitin, N.I., Radionova, E.A. (1999) Lateralization of a moving auditory image: Interrelation of interaural time and intensity differences. *Journal of the Acoustical Society of America, 105*(1) 366-376.

Bauer, B.B. (1961). Phasor analysis of some stereophonic phenomena. *Journal of the Acoustical Society of America*, 33, 1536-1539.

Baum, H.M. (2005). Understanding Behaviorism. MA: Blackwell Publishing.

Bertelson, P., Aschersleben G., Temporal ventriloquism: crossmodal interaction on the time dimension: 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology. 50*, 147-155.

Bischoff, M., Walter, B., Blecker, C.R., Morgen, K., Vaitl, D., Sammer, G. (2007). Utilizing the ventriloquism-effect to investigate audio-visual binding. *Neuropsychologica, 45*. 578-586.

Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S., Blanke, O. (2007). Auditory motion affects visual biological motion processing. *Neuropsychologica. 45*, 523-530.

Chandler, D.W., Grantham, D.W. (1991). Minimum audible movement angle in the horizontal plane as a function of stimuus frequency and bandwidth, sourth azimuth, and velocity. *Journal of Acoustic Society of America, 91*(3), 1624-1635.

Deutsch, D. (1974). An auditory illusion. *Nature, 251*, 307-309.

Donovan, C.L., Lindsay, S.D., Kingstone, A. (2004) Flexible and abstract resolutions to crossmodal conflicts. *Brain and Cognition. 56*, 1-4.

Fendrich, R., Corballis, P.M.(2001). The temporal cross-capture of audition and vision. *Perception and Psychophysics, 63*(4). 719-725.

Frassinetti, F., Bolognini, N., Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research. 174,* 332-343.

Grantham, D. W. (1984). Discrimination of dynamic interaural internsity differences. *Journal of the Acoustical Society of America, 76*(1). 71-76.

Granthan, D.W. (1986). Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *Journal of the Acoustical Society of America, 74*(6). 1939 –1949.

Howard, I.P., Craske, B., Templeton, W.B. (1966). Human Spatial Orientation. New York: Wiley.

Koppen, C., Spence, C. (2007). Seeing the light: exploring the colavita visual dominance effect. *Experimental Brain Research*, 180. 737-754.

Lewald, J., Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research, 16*. 468-478.

McGurk, H., MacDonald, J. W. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Morein-Zamir, S., Soto-Faraco, S., Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Reseach. 17,* 154-163.

Perrot, D.R., Musicant, A.D. (1977). Minimum auditory movement angle: Binaural localization of moving sound sources. *Journal of the Acoustical Society of America, 62*(6), 14631466.

Rosenblum, L.D., Carello, C., Pastore, R.E. (1987). Relative effectiveness of three stimulus variables for locating a moving sound source. *Perception, 16*(2). 175-186.

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science, 145*, 1328-1330.

Saberi, K., Perrot, D.R. (1990). Minimum audible movement angles as a functin of sound source trajectory. *Journal of Acoustics Society of America*, *88*(6), 2639-2644.

Sanabria, D., Spence, C., Soto-Faraco, S. (2007). Perceptual and decisional contributions to audiovisual interaction in the perception of apparent motion: A signal detection study. *Cognition*. 102, 299-310.

Shore, D.I., Barnes, M.E., Spence, C. (2006). Temporal aspects of the visuotactile congruency effect. *Neuroscience Letters. 392*, 96-100.

Soto-Faraco, S., Spence, C., Kingstone, A. (2004). Cross-Modal Dynamic Capture: Congruency Effects in the Percption of Motion Across Sensory Modalities. *Journal of Experimental Psychology*. *30*(2), 330-345.

Spence, C. Squire, S. (2003). Multisensory Integration: Maintaining the Perception of Synchrony. *Current Biology*. 13, R519-R521

Stanley, R.M., Matthews, N. (2004). A comparison of cues for auditory motion judgments. *Cognitive Neuroscience Society Abstracts*, p114, C131.

Stein, B.E., Meredith, M.A. (1993). The merging of the Senses. MA: The MIT Press.

Vroomen, J., Bertelson, P., de Gelder, B. (2001) The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception and Psychophysics*, *63*(4). 651-659.

Vroomen, J. de Gelder, B. (2004). Temporal Ventriloquism: Sound modulated the flash-lag effect. *Journal of Experimental Psychology: Human Perception and Performance, 30*(3). 513-518.

Wallace, M.T., Robertson, G.E., Hairston, W.D., Stein, B.E., Vaughan, J.W., Schrillo, J.A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*. 158, 252-258.

Welch, R.B., DuttonHurt, L.D., Warren, D.H. (1986. Contributions of audition and vision to temporal rate perception. *Perception and Psychophysics, 39,* 294-300.

Welch, R.B., Warren, D.H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*(3), 638-667.

Participant Responses Across Target Conditions

Participant	V2A2/V2A1	V2A2/V2A2	V2A2/V2A3
1	0.45	0.31	0.64
2	0.22	0.5	0.73
3	0.55	0.72	0.53
4	0.3	0.45	0.36
5	0.45	0.7	0.29
6	0.43	0.53	0.38
7	0.46	0.67	0.38
8	0.31	0.25	0.71
9	0.64	0.32	0.46
10	0.58	0.67	0.46

Table 1 – Proportion of participant “second condition faster” responses across both target conditions and the neutral condition.

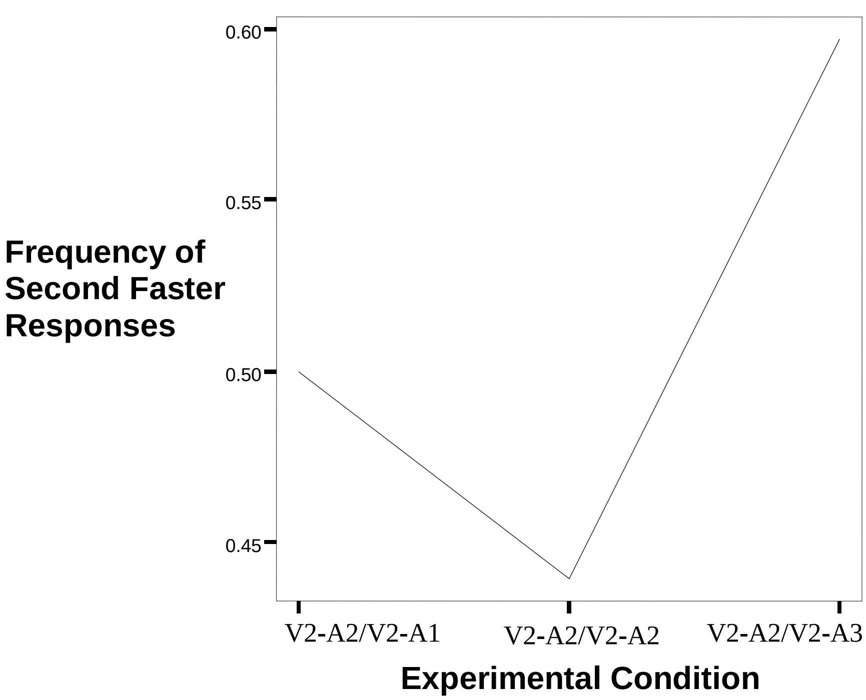

Figure 1 – Frequency of “second visual stimulus faster” responses across V2-A2/V2-A1, V2-A2/V2-A2, and V2-A2/V2-A3 conditions.

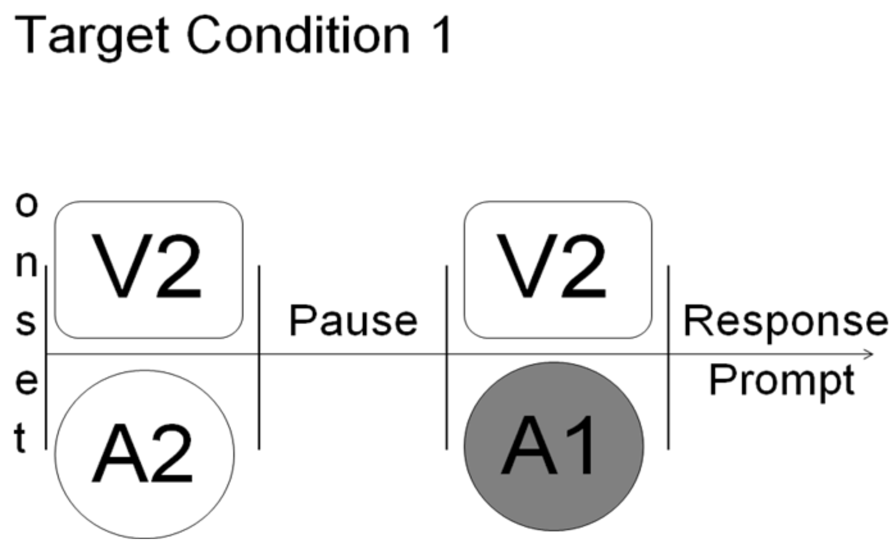

Figure 2 – Target condition 1(V2-A2/V2-A1)

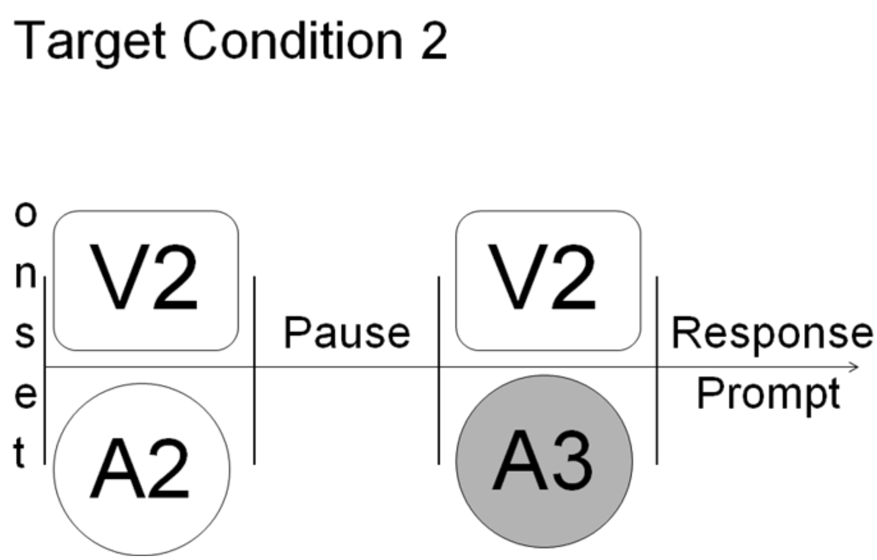

Figure 3- Target condition 2 (V2-A2/V2-A3)