MULTISENSORY INTEGRATION IN VISUAL PATTERN RECOGNITION

Avi M. Aizenman, Jason M. Gold* & Robert Sekuler

Brandeis University & Indiana University*

INTRODUCTION

BACKGROUND Previously[1,] we examined sub-second short- term memory and learning of rapidly-presented sequences of random luminance contrasts. After viewing an 8-item sequence, subjects had much success in judging whether or not a sequence's final four items replicated its first four. Also, when a particular exemplar recurred intermittently during a session, performance improved steadily, showing long-term learning of that exemplar.

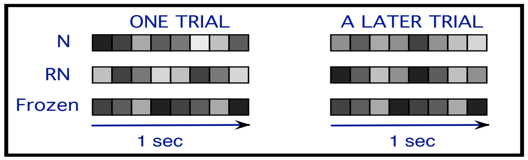
RATIONALE Synchronous co-modulation perceptually binds audio and visual signals, which improves perceptual precision.[2] We asked (i) whether co-modulated sequences enjoyed an advantage in short-term memory and longer-term learning, and (ii) whether subjects adopted the same strategy when judging visual or auditory sequences. As musical training influences performance with rapidly-presented stimuli,[3] we tested both musicians and non-musicians[4].

EXPERIMENT ONE

TASK. 24 subjects were given rapidly-presented 1-sec long sequence of eight items modulated at 8 Hz. Sequences were made up of (i) luminances, (ii) pure tones, or (iii) combinations in which audio and visual signals co-occurred synchronously.

UNIMODAL STIMULI

Visual. Sequences of varying luminances each presented at the same region of the display. Subjects judged whether the last four items in a sequence repeated the first four.

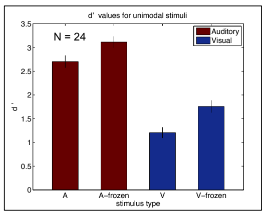


Audio. Sequences of pure tones.

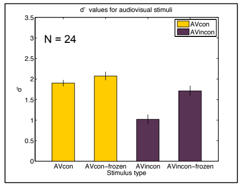
Frozen sequences. In every block of trials, one randomly-chosen Repeat sequence was preserved and made to recur intermittently during a block of trials. Changes in performance with this Frozen sequence allowed us to track learning.

UNIMODAL RESULTS. *d'* values were very much higher for Auditory stimuli than for Visual stimuli (p<.01)

In addition, *d'* values for recurring or Frozen sequences were higher than for stimuli presented just once (p<.05).



MULTIMODAL STIMULI.
Simultaneous sequences of visual and auditory stimuli. Subjects judged only whether halves of *visual* sequences repeated or not, while ignoring accompanying tones. Auditory Frequencies were each cross-modally matched[5] to luminance levels for Visual stimuli.



AV-congruent. Luminance changes were accompanied by cross-modally matched changes in auditory frequency.

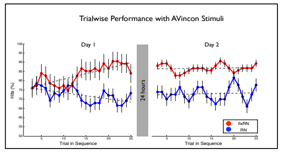
AV-incongruent. Luminance variation was *uncorrelated* with variation in the frequencies of accompanying tones.

MULTIMODAL RESULTS. *d'* values were significantly higher for AV-congruent stimuli than for AV-incongruent stimuli (p<.01).

MULTIMODAL RESULTS: FROZEN SEQUENCES.
Encountering the same Frozen sequence multiple times throughout a block of trials increased *d'* for both AV-congruent and AV-incongruent sequences.

Learning was much greater for Frozen AV-incongruent sequences, that is, sequences whose audiovisual components were independent of one another (p<.01).

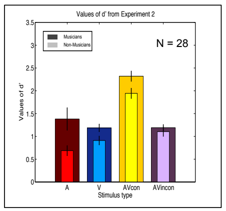
When subjects returned 24-hours later for a second testing session, learning they had acquired the previous day was fully intact.



EXPERIMENT TWO

Although individual Visual and Auditory items were matched cross-modally, Experiment One showed that rapidly-presented sequences of the two kinds of stimuli were not equivalent. To equate performance, Experiment Two altered the mapping of frequencies onto luminances, by shrinking the range of frequencies.

Also, as musical training improves pitch discrimination[6,] we tested 14 subjects who had musical training and 14 who had little or no training. "Trained" was defined as having played an instrument for six or more years.[4]



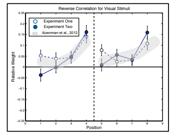
Narrowing the range of auditory frequencies equalized performance for the two types of unimodal sequences: mean *d'* was not different for Auditory and Visual stimuli (p >0.9).

Audio-visual congruence improved performance: mean *d'* was significantly higher for AV-congruent stimuli than for AV-incongruent stimuli (p < .00001).

Generally, musicians outperformed non-musicians (p<.02). Specifically musicians held an advantage with Auditory, Visual and AV-congruent sequences (p<.05).

REVERSE CORRELATION

Reverse correlation[7] characterized the relative weights subjects gave to each item in a sequence when judging whether the second four items did or not repeat the first four.

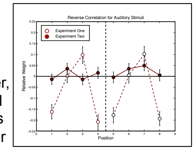


For unimodal Visual stimuli, subjects gave preferential weight to items occupying the fourth and eighth ordinal positions in an 8-item sequence.

This finding replicates our previous result with entirely different subjects.[1]

For unimodal Auditory stimuli, reverse correlations differed with the frequency range.

In Experiment One, the 1st, 3rd and 4th items received preferential weighting in each half sequence. However, Experiment Two's narrower frequency range produced no clear preferential weighting by ordinal position. This result tracked the reduced *d'* produced by the narrower frequency range.



CONCLUSIONS

1. Correlated (congruent) auditory and visual signals can boost performance on a nominally-visual task
2. Incongruence between auditory and visual signals can undermine performance when a sequence is unfamiliar. However, over repeated presentations, a consistent relationship between incongruent elements produces the strongest learning and retention.[8]
3. In general, subjects with musical training outperform subjects with little or no such training, including when rapidly-presented sequences are exclusively visual

SOME NEXT STEPS

1. With AV sequences, subjects were instructed to judge only whether visual items repeated, while ignoring auditory ones. Subjects could do this, although imperfectly. Would it be easier to filter out visual items in an AV sequence?
2. Subjects learned and retained certain kinds of sequences for 24 hours, with no loss. What are the limits to retention?
3. Does the impact of AV congruence depend upon a natural, pre-existing association, such as between auditory frequency and luminance?

REFERENCES

1. A Aizenman, SM Bond, JM Gold & R Sekuler (2012) Visual Sciences Society
2. GJ Thomas (1941) *J exp Psychology*; SH Lee & R Blake (1999) *Science*; GH Recanzone (2003) *J Neurophysiology*; D Alais & D Burr (2004) *Current Biology*
3. D Strait & N Kraus (2011) *Music Perception*
4. E Skoe & N Kraus, (2012) *J Neuroscience*
5. L Marks (1974) *Am J Psychology*
6. L Kishon-Rabin et al., (2001) *J Basic Clinical, Physiological and Pharmocology*
7. M Eckstein & A Ahumada (2002) *JoV*
8. L Shams & A Seitz (2008) *Trends in Cog Sci*

Supported by AFOSR , NIH and NSF. Abigail Noyce and Patrick J. Bennett contributed many helpful suggestions. Correspondence: VISION@BRANDEIS.EDU