Sensory modalities are not separate modalities: plasticity and interactions
Shinsuke Shimojo* and Ladan Shams†

Historically, perception has been viewed as a modular function, with the different sensory modalities operating independently of each other. Recent behavioral and brain imaging studies challenge this view, by suggesting that cross-modal interactions are the rule and not the exception in perception, and that the cortical pathways previously thought to be sensory-specific are modulated by signals from other modalities.

Addresses
Division of Biology, Computation and Neural Systems, California Institute of Technology, B1 139-74, Pasadena, CA 91125, USA
*e-mail: sshimojo@caltech.edu
†e-mail: ladan@caltech.edu


Introduction
Cross-modal integration is performed on a vast level in the brain and contributes significantly to adaptive behavior in our daily life. Very little is known about how integration is achieved or its underlying neural mechanisms, however, because the overwhelming majority of studies on perception have focused on one sensory modality. Studying perception in an isolated single modality would be justifiable if different modalities processed sensory inputs independently of each other, as separate ‘modules’. But are sensory modalities really separate modules? A variety of evidence seems to counter this notion of modularity. In this review, we summarize the evidence for vigorous interaction among sensory modalities.

Plasticity across sensory modalities
Both animal and human studies suggest that sensory modalities in early stages of development are not as inherently distinct and independent as was previously thought. For example, in a study of cross-modal plasticity, Sur et al. [1] removed the superior colliculus of both the ferret and the hamster on the day of birth by direct ablation. They also deprived the medial geniculate nucleus or the ventrobasal nucleus from their normal sensory input by sectioning the major input pathways. The retina then invaded these thalamic nuclei, which under ordinary circumstances relay auditory and somatosensory signals to the cortices, respectively. They found that visual responses (i.e. responses triggered by light stimulation on the retina) were elicited from neurons in the auditory or the somatosensory cortex.

More recently, Sur and co-workers [2••] reported that these animals interpreted the activation of the rewired auditory cortex as visual sensation, indicating a functional relevance of the rewiring. There is also evidence for a cross-modal postnatal perceptual enhancement by prenatal sensory stimulation. When bobwhite quail chicks were prenatally exposed to an auditory, visual, tactile or vestibular stimuli, their postnatal auditory and visual responsiveness was enhanced, irrespective of the modality of prenatal stimulation [3–6].

Cross-modal plasticity has also been reported in humans that have had sensory deprivation in early life [7–13]. The typical finding in these studies has been that sensory deprivation in one modality starting from an early period of life causes the cortical area that is normally devoted to that modality to be used by some other modality. Early studies reported that visual event-related potentials tend to be enhanced in early onset deaf individuals [7,8].

Auditory-evoked potentials have also been recorded from posterior (occipital) regions in early and late blind subjects [9]. Although suggestive, however, these studies were neither able to identify exactly which part of the brain is responsible for the enhanced activity, nor able to examine functional relevance of the activity.

Studies that use a perceptual task have been more informative in this regard. For example, Uhl et al. [10] have provided evidence for posterior DC potentials in blind subjects engaging in a tactile reading task. Similarly, a positron emission tomography study (PET) by Sadato et al. [11] indicated the activation of primary and secondary visual cortical areas induced by Braille reading in early blind subjects, but the deactivation of the same areas relative to the rest condition induced by Braille reading in sighted subjects [11]. Simple tactile stimuli that did not require discrimination produced no activation of visual areas in either group.

Furthermore, the same researchers applied a train of pulses of transcranial magnetic stimulation to the occipital region of blind subjects engaged in the Braille reading task, and showed that indeed it degraded performance but selective to this task only [12].

Taken together, these animal and human studies indicate a surprising degree of neural plasticity in early stages of life, and clearly argue against the most stringent version of brain segregation and modularity of sensory modalities.

Interaction across modalities
In the phenomenological and psychophysical literature, a wealth of literature on mature perceptual systems indicates...
that there is vigorous interaction and integration among the sensory modalities — a far greater interaction than what is expected from a ‘naive modularity’ point of view. One of the most intriguing examples of cross-modal interaction is synesthesia, in which an involuntary conscious sensation (such as color) is induced by a stimulus in another modality (such as sound). Synesthesia can occur in normal, healthy populations, in brain-damaged or sensory-deafferented patients, or in people who are addicted to hallucinogenic drugs [14••].

Vision alters other modalities
In normal people, the effects of vigorous cross-modal interaction are made apparent under cleverly designed artificial conditions. The McGurk effect exemplifies such a condition [15]. The McGurk effect is a perceptual phenomenon in which vision alters speech perception (e.g., the sound ‘ba’ tends to be perceived as ‘da’ when it is coupled with a visual lip movement associated with ‘ga’ [15]).

The spatial location of a sound source can also be drastically influenced by visual stimulation. This effect is known as the ‘ventriloquist effect’ [16], and is experienced frequently in daily life when watching television or movies in which the voices are perceived to originate from the actors on the screen, despite a potentially large spatial discrepancy between the two.

It has been shown that tactile location, such as location of a finger pointing, can also be ‘captured’ by visual location [17]. All these effects emphasize the strong effect of visual signals on the other modalities, consistent with the commonsense notion that human is primarily a vision-dominated animal.

Sound alters the temporal aspects of vision
Although the best-known cross-modal effects are those of vision influencing other modalities, visual perception can also be altered by other modalities. All the earlier reports of alteration of visual perception by other modalities have been in the temporal domain. For example, the perceived duration [18] or rate [19–21] of a visual stimulus has been shown to be influenced by accompanying sound signals. More recently, Scheier et al. [22] have shown that visual temporal resolution can be either improved or degraded by sounds, depending on the temporal relationship. They showed that when two lights are turned on with a small temporal delay (in the ~60 to +60 ms range), the accuracy of temporal order judgment is better when a sound precedes and another follows the visual stimuli (the audio-visual-audio [A-V-V-A] time order). In contrast, the subjects’ performance becomes worse when two sounds are inserted between the two visual stimuli (V-A-V-A time order), as illustrated in Figure 1.

Sound alters other aspects of vision
The great body of behavioral findings on cross-modal interactions has been accounted for in a ‘modality appropriateness’ hypothesis [23]. This hypothesis postulates that the modality that is most appropriate or reliable with respect to a given task is the modality that dominates the perception in the context of that task. Vision has a higher spatial resolution, hence its dominance in spatial tasks (e.g., the ventriloquist effect and visual capture), whereas audition has a higher temporal resolution, hence its dominance in temporal tasks.

Alteration of vision by sound, however, turns out to be not limited to temporal aspects. The perceived intensity of a visual stimulus has been shown recently to be enhanced in presence of sound [24]. Moreover, the quality or structure of visual perception itself can be altered by auditory stimuli, according to the latest psychophysical studies.

Sekuler et al. [25] have shown that the presence of a sound can also alter the perceptual interpretation of an ambiguous visual motion event. Two identical visual targets moving across each other can be perceived either to bounce off or to stream through each other, as their trajectories are nearly identical (Figure 2a). Nonetheless, most observers report a perception of streaming, not bouncing motion.

If a brief sound is added at the moment that the targets coincide visually, however, visual perception is strongly biased in favor of bouncing motion [25]. The sound has to have a sharp onset to induce this effect. The ecological origins of
this phenomenon are intuitively obvious: it should be relat-
ed to the fact that the majority of collision events in the
natural environment yield synchronized cross-modal signals.

Other studies on this phenomenon have revealed some
unexpected properties. First, a transient sensory stimulus
biases visual perception toward bouncing irrespective of its
modality; for example, a brief visual flash or a brief tactile
vibration on an observer’s finger also induces bouncing
perception [26,27]. The stimulus, however, has to be syn-
chronized with the visual coincidence of two objects (the
effective time window was found to be in the range of –300
to +200 ms, –600 to +100 ms and –100 to +100 ms for audi-
tory, tactile and vision stimuli, respectively; see Figure 2b).
Thus, the same effect can be obtained as long as there is a
transient sensory stimulus that is
approximately synchronized with the critical visual event, regardless of the modality in
which this transient stimulus is given.

A recent study has shown that the alteration of vision by sound is not limited to the situations of ambiguity in the visual stim-
ulus. The ‘illusory flash effect’ reports that an illusion in which

Figure 2

Cross-modal modulation of visual motion perception. (a) The ambiguous motion display, in which two objects move across each other. Observers typically perceive an object as streaming with an X-shaped trajectory when there is no accompanying sound or an
accompanying sound that is not synchronous with the visual crossing. When there is a
sound (or a flash or a tactile vibration) synchronized with visual coincidence,
however, the perceptual dominance reverses and most observers now see the objects as
bouncing against each other. (b) Temporal tuning curves of the bounce-inducing effects
of a synchronous transient signal in auditory, visual or tactile modality. The percentage
increase in the perception of bounce is plotted against asynchrony between the
transient signal in each modality and the visual coincidence. Thus, qualitatively the same
effect can be obtained as long as there is a
transient sensory stimulus that is
approximately synchronized with the critical visual event, regardless of the modality in
which the transient stimulus is given.

A radical change in the phenomenological quality of perceiving
a non-ambiguous visual stimulus is induced by sound [31*].
When a single brief visual flash is accompanied by multiple
auditory beeps, the single flash is perceived as multiple flash-
es (see Figure 3). Control conditions, catch trials and many
other observations indicate that the illusory flashing phenom-
emon is indeed a perceptual illusion, and is not due to the
difficulty of the task or some cognitive bias (caused by sound).

The illusory double flash is perceptually very similar to the
physical double flash. Furthermore, the illusion is very
robust to the observer’s knowledge about the physical stim-
ulus, and to variations in stimuli parameters. The temporal
tuning of this effect was also measured by varying the relative
timing of visual and auditory stimuli. The illusory
flashing effect decreased at separations greater than 70 ms;
however, illusory flashing occurred as long as the beeps and
flash were within about 100 ms — consistent with the inte-
gration time of polysensory neurons in the brain [32,33]:

The alteration of vision by sound in this experiment was
found to be asymmetrical: alteration occurred strongly only
when a single flash was coupled with multiple beeps, and
not when multiple flashes were paired with a single beep.
In other words, strong alteration of vision by sound occurs
only when sound is more discontinuous and structured
than the visual stimulus. A similar phenomenon seems to
be at work in a study that investigated the effect of vision
on hearing [34]. Saldaña and Rosenblum’s [34] data sug-
uggest that only the discontinuous visual stimulus has a
strong effect on the perception of the sound.
Cross-modal interactions depend on stimuli structure

The direction of cross-modal interactions has been thought to be determined by the relative appropriateness of the modalities involved in the task, as mentioned above. The evidence discussed above, however, seems to indicate that the direction of cross-modal interactions depends, at least in part, on the structure of the stimuli; that is, the modality that carries a signal which is more discontinuous (and hence more salient) becomes the influential or modulating modality. Such a hypothesis would also account for the findings of streaming/bouncing motion studies discussed above, in which the transient, hence more discontinuous and structured, stimulus alters the perception of a continuous visual stimuli regardless of its modality (auditory, tactile or visual) [25–30].

Neural mechanisms

At what point along the perceptual processing pathway do these cross-modal interactions take place? Recent data from brain imaging studies suggest that they occur at brain sites that used to be considered as modality-specific. For example, Calvert et al. [35] carried out a functional magnetic resonance imaging (fMRI) study that showed that the primary auditory cortex is activated when a talking face is viewed in the absence of sound. The activation was observed specifically in speech or pseudo-speech type of lip movements, but not in other type of lip movements.

The same research group has also reported that activity in visual (V5) and auditory (BA 41/42) cortices after exposure to bimodal (audio-visual) speech is enhanced relative to activity after a unimodal stimulus [36*]. The enhancement was found to be contingent on congruency between the audio and visual signals. But as these studies deal only with speech perception, would their findings generalize to other types of stimuli?

Two very recent studies may have more general implications. An fMRI study has shown that tactile stimulation of a hand enhances activity in the visual cortex when the touched hand is on the same side as the visual stimulus [37**]. Likewise, an event-related potential (ERP) study also suggests that activity in the visual cortical areas is modulated by sound [38]. The modulation was found as early as 100 ms after the visual stimulus onset. This study used a brief flash and a brief beep as visual and auditory stimuli.

In contrast to speech stimuli, which inherently hinge on some higher level matching process (such as congruency detection), the stimuli used in the last two studies are very simple, and thus the results may have more general implications. In particular, the results of the ERP study, unlike any of the other findings discussed above, may have very general implications because they do not require congruency between the stimuli in the two modalities and do not seem to involve spatial attention mechanisms.

Together, the results of the studies summarized in this review challenge the common belief that ‘modality-specific’ cortices function in isolation from other modalities.

Conclusions

We have discussed a wide variety of evidence against the notion of strict modularity of sensory modalities. Both animal studies and human-deprivation cases provide evidence for a surprising degree of cross-modal plasticity in cortical processing. Psychophysical data indicate that interaction between modalities is the rule as opposed to the exception in brain function, and brain imaging and recording studies provide evidence against modularity and for interaction in areas traditionally thought to be unimodal.

Motivated by some of the recent psychophysical findings, we put forth a new hypothesis for multisensory interactions. We propose that the transient/discontinuous signals possess a special status in determining the direction of cross-modal interactions, with the transient stimulus strongly influencing the multimodal perception, regardless of the modality in which it occurs.

Acknowledgements

We apologize to the many researchers whose work was omitted owing to space constraints. This work was supported in part by NIH grant HD08506. We thank the section editors for their insightful comments.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- **of outstanding interest


The authors investigate whether activity in the auditory cortex, which has been rewired to receive input normally projected to the visual cortex, is interpreted by the animal as auditory or visual sensation. They find that the activation of rewired auditory cortex is interpreted as visual stimulation, which suggests that the modality of a neocortical area is intrinsic and can be significantly influenced by its extrinsic input.


This paper presents a visual illusion that is induced by sound. This constitutes the first report of a change in the phenomenological quality of the visual perception caused by another modality even when the visual stimulus is not ambiguous. One of the implications of this finding is that, like other modalities, vision is also malleable by other modalities.


Using fMRI, the authors find that the activity of auditory and visual cortices arising from exposure to audio-visual speech is enhanced relative to that arising from unimodal speech. Because the enhancement is specific to semantically congruent bimodal stimuli, the authors attribute the enhancement to backprojections from modality-specific areas that are involved in cross-modal binding.


The authors report an enhancement, measured by fMRI, in the activity of visual cortex owing to simultaneous tactile stimulation. On the basis of the connectivity of brain areas, they argue that this enhancement is mediated through backprojections from multimodal cortical areas.