



Shared content across perceptual modalities: lessons from cross-modal illusions*

by

Casey O'Callaghan

Philosophy Dept., Bates College, Lewiston, ME, U. S.

Contacto / correspondence: *cocallag [-at-] bates.edu*
<http://www.bates.edu/~cocallag>

Electroneurobiología 2006; **14** (2), pp. 211-224; URL
<<http://electroneubio.secyt.gov.ar/index2.htm>>

Copyright © 2006 del autor / by the author. Este trabajo es un artículo de acceso público; su copia exacta y redistribución por cualquier medio están permitidas bajo la condición de conservar esta noticia y la referencia completa a su publicación incluyendo la URL (ver arriba). / This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's full citation and URL (above).

* Acknowledgements. This is a version of a paper presented on 5 April 2006 at *Toward a Science of Consciousness 2006* in Tucson, Arizona. I thank the members of that audience for helpful questions and comments. David Chalmers and David Rosenthal, in particular, raised important questions about whether cross-modal influence is merely causal or whether, as I suggest, its explanation calls for appeal to content shared across modalities. Adam Morton and Jeffrey Speaks also provided helpful comments and suggestions in correspondence. I hope to address these questions and to incorporate suggestions in future revisions to this material. I also wish to thank Mariela Szirko for comments and suggested revisions for the *Electroneurobiología* version of this paper. In particular, Professor Szirko encouraged me to consider Crocco's and Mondolfo's interesting work in this area.

ABSTRACT: Each perceptual modality cannot fully be understood in isolation from the others. The recently discovered sound-induced flash illusion is a visual illusion induced by sound (Shams *et al.* 2000, 2002). A single flash paired with multiple beeps is perceived as multiple flashes. The illusion is characterized by its discoverers as being induced by audition as a result of "cross-modal perceptual interactions" (2002:147).

Alva Noë has recently challenged on independent grounds what he calls the "snapshot conception" of visual experience according to which perception presents discrete snapshot-like contents that represent a scene "in sharp focus and uniform detail from the center out to the periphery" (Noë 2004, ch. 2). On the basis of a discussion of cross- and inter-modal perceptual effects, I argue in this paper that what I dub the "composite snapshot" conception of overall perceptual experience fails. Cross-modal and inter-modal illusions, including the sound-induced flash illusion and the more familiar ventriloquist illusion (in which vision influences sound localization) suggest that the influence of one modality upon the phenomenological and perceptual content of another modality requires for its explanation appeal to a dimension of shared content across perceptual modalities.

The cross-modal illusions thus demonstrate that a visuo-centric focus in theorizing about perception and perceptual content threatens to blind us to the nature and character of perceptual experience. Such effects indicate that individual modalities cannot fully be understood in isolation from the others – even vision and visual content are illuminated by considering the non-visual modalities. Abandoning both the visuo-centric focus in theorizing about perceptual experience and the composite snapshot conception of experience also contributes to resolving puzzles about the other modalities. For instance, auditory perception plays a role in situating subjects in a world of objects and events. Auditory perception, that is, reveals not only a world of sounds but also furnishes information about the things and happenings that generate those sounds. How could audition, whose proper objects are sounds, include object-involving content? Appeal to a shared dimension of content among perceptual modalities makes this question tractable. Common content among modalities, appeal to which is required to explain cross-modal effects, could ground an explanation for how audition might furnish genuinely perceptual awareness of objects and happenings and not mere inferential or otherwise non-perceptual awareness. In short, attention to cross- and inter-modal effects and illusions enhances our understanding of the phenomenological and per-

ceptual contents of experience by encouraging us to move beyond characterizing perceptual content as a composite of modality-specific contents.



1. Visuocentrism

Spanish neurobiologist Juan Cuatrecasas portrayed the human being as an "optical animal". Fittingly enough, philosophical thinking about perception has been driven primarily by attention to vision and to visual examples. Discussions of Mary the blind color scientist, spectrum inversion, the waterfall illusion, blindsight, and change and inattentive blindness are just a few examples in which vision has furnished not only the puzzle cases that any philosophical theory of perception must deal with, but also has guided the intuitions that shape any such theory. I do not want to attack this visuo-centric focus directly. I will argue that it is problematic by suggesting that thinking about other modalities apart from vision bears fruit not only by challenging or confirming what we learn through thinking about vision, but also by adding new puzzles that shape thinking about perception.

But this does not go far enough toward abandoning visuocentrism. I will also claim that simply shifting to thinking about the other modalities ultimately fails to reveal the most significant implications of considering multiple modalities in developing and evaluating theories of perception and perceptual content.

2. A puzzle from the case of sounds

Let me start off by presenting a puzzle that emerges from thinking about sounds and audition. It is clear that, in a relatively innocuous sense, sounds are the immediate objects of auditory experience – whatever else you hear, such as cars or crashes, you hear it in virtue of hearing a sound. But, auditory experience is what I have elsewhere (O'Callaghan, forthcoming *a* and forthcoming *b*) described as *object- and event-involving*. You learn on the basis of auditory experience that

the glass has broken, that there's a bell in the room, or that the train is passing. According to some views in evolutionary neurobiology, reference to objects was achieved before sensory modalities became diversified (Crocco, 2004). In any event, one plausible view about how you learn this is that you *hear* the train, the bell, the breaking of the glass.

The experience seems to you to be an experience of a train, a bell, or a glass breaking, a fact recognized by Plato, Aristotle and many other scholars in Antiquity and Medieval times. In fact, we speak about and classify sounds in just these terms. So, you hear a sound, and by or in hearing that sound, you hear the object or event that is its source.

Granted, this awareness feels less "direct" or more "secondary" than your awareness of the sound or the awareness of the apple you enjoy on the basis of your awareness of its color and shape (sometimes called "primary intention" in ancient portrayals), but there is still a sense in which it seems to one that one enjoys auditory awareness of a train, a bell, or a glass breaking in virtue of hearing their sounds. The puzzle is this: How could auditory experience, whose proper objects are sounds distinct from ordinary objects and events, furnish perceptual awareness of things like trains, bells, and breakings?

The puzzle raises two closely related questions about the content of auditory perception. The first is, "How *mediated* is one's awareness of ordinary objects and events in audition?" The second is, "How *rich* is the content of auditory experience?"

One might argue that the apparent perceptual awareness of ordinary objects and events is a mere illusion, and that the sound mediates consciousness of non-auditory objects and events only modulo some inferential or otherwise cognitive connection. Though the phenomenology of audition seems for all the world to furnish experiential awareness of things and happenings beyond sounds (why else do we reflexively act to orient toward or to avoid the source of a sound), perhaps the puzzle depends on missing the crucial cognitive step.

If all you are aware of is a sound and its qualities, and any consciousness of ordinary object and events is mediated by some further

non-perceptual cognitive states, then the puzzle dissolves apart from the question how it could strikingly *seem* that you are aware of objects and events in audition. But since the *seeming* requires explanation, a version of the puzzle persists.

If, however, the content of audition is very rich and audition can represent, e.g., something like *train oncoming*, or *glass breaking*, in a way that is mediated only by perceptual states, then the puzzle is at its most pressing.

If the truth is somewhere in between, and audition furnishes awareness of things like *source* or *object* or *event*, then the puzzle still arises. How could an extra-auditory object or event be among the objects of auditory perceptual experience when sounds are in the first case the things we hear? How is it possible for the non-auditory features of an object or event to be among the contents of auditory perception, whose immediate proper objects are sounds? How could auditory perception ever represent the presence of an ordinary object or event?

This question is closely tied to questions surrounding intermodal feature binding. How is it that one experiences the movement of a speaker's lips and the sound of her voice to share a common source? In ancient times this question had bearing on the unity of the human subject. The Western evolution of the issue was historically reviewed by Rodolfo Mondolfo in two important and interrelated works (1932, 1955). From a work co-authored by one of his disciples (Ávila and Crocco 1996, p. 744) I take the following summary of the puzzle's birth:

"*Gorgias' splitting, the Danaëan gift*. The starting point in the Western thought, for these researches on the unifying function of the experiencing, was the extreme form reached by the sensualist phenomenism in Gorgias (-Vth century). Along with reducing every possible sapience to sensation, he added that it is not communicable (the noematic *Unübertragbarkeit* pointed by Prof. Born: the one due to structurelessness, not that due to cadacualtez); not only from one experiencing to other (e. g., from yours to ours) but, also, even from each set of sense's sentiences (in any of their thetic modes) to any simultaneous other. So, the personal experiencing inside any single organism

was postulated as *multiple*, because of the separation of the different sensations into stanch compartments, mutually incommunicable.

Like the blindness for the *noetic* incommunicability of cadacualtic availabilities, this atomization is typical of every sensualist phenomenism, and a consequence of it, as too often evinced, for example, in the French sensualism at the XVIIIth century (with Diderot, and specially with Condillac) and parallel Eastern developments. It offered itself to Plato's especial reflection, as in *Theaetetus* 184 b sq., where he refined his critique of sensualist empiricism."

There, Plato denied and rejected that each sense modality could enjoy by itself a direct and exclusive apprehension or grasping of its own sensations. To clarify the need of a *unifying conspection* ("*binding*"), Plato advanced the comparison with the Danæan gift, which Prof. [Christfried] Jakob often recalled when recounting the history of the understanding of the sensations' conspectivity. Inside the wooden horse of Troy, each Danæan warrior remained distinct and separate. But the functional *purpose*, or systemic *finality*, of both Greek warriors and separated animal senses, requires a mutual unifying binding: one adjoining agencies previously apart. Bare sharing of a receptacle is not sufficient to explain why qualities available through different modalities are presented in experience as features of the same environmental particular. "

3. The composite snapshot conception of perceptual experience

I want to suggest that the puzzle just described ultimately has its source in the visuo-centrism I mentioned at the outset. In fact, the puzzle stems from a conception underwritten by the visuo-centric focus in thinking about perception. Some explanation is in order.

Alva Noë has recently challenged what he's called the "snapshot conception" of visual experience on empirical and phenomenological grounds. According to the snapshot conception, visual experience presents as a richly detailed snapshot-like scene before the eyes. It's colored and crisp and object-presenting from the center out to the periphery.

Whether or not Noë's criticisms are on the mark, it's fair to say that the traditional empiricist conception of *overall* perceptual experience is what we might call the "*composite* snapshot conception" of experience, with an emphasis on "composite". Whether or not the snapshot conception is correct, the composite snapshot conception is that perceptual experience is comprised of a set of discrete modality-specific experiences superimposed to create one's total perceptual experience at a time.

That is, vision has a certain content characterized by colors and shapes (and perhaps "visual objects"; compare Lewis's "color mosaic"), audition has a content characterized by sounds and their pitches (compare Strawson's purely auditory experience which he says could not ground perception of space, and so could not ground the self-other distinction required for object or event perception), smell has a content characterized by olfactory qualities, and so on for each of the perceptual modalities which, physiologists say, in humans number beyond a full score.

Whatever their number, each modality, according to this traditional empiricist picture, delivers from its unique perspective a discrete snapshot of the world that is qualitatively distinct from each of the others. Vision could not share elements of audition's snapshot and vice versa. The sum total of these snapshots, a sort of composite snapshot, constitutes and exhausts the content of one's total perceptual experience.

The traditional conception seems to stem from thinking of the senses as distinct systems or channels of awareness of the external world. They are understood to involve separate processes, and to work in isolation from each other perhaps until some relatively late stage. In addition, each modality is thought to deliver an experience with a distinctive qualitative character that could not be created by any other modality. Each of these modalities delivers an experiential ingredient for one's total perceptual experience.

The lesson of this paper is that this traditional story is false in important respects and incomplete in others. I want to suggest that an important class of perceptual effects that have gone relatively unrecognized or unappreciated by philosophers gives us good reason to

think that the composite snapshot conception of experience is incorrect.

But the illusions that I'll discuss don't have merely negative implications. I also want to suggest that they provide the ingredients for the beginning of a solution to the puzzle about audition I described above. Finally, they illuminate perception in a significant respect and teach us what we could not have otherwise learned with attention restricted to vision (or any other individual modality, for that matter). The modalities cannot even be understood individually in isolation from each other. Perception is very much the result of integrating, weighing, comparing, and extracting significant information from the senses considered collectively, and is not a mere assembling of discrete snapshots from each modal perspective.

4. Cross-modal illusions

The class of perceptual effects I have in mind are ones in which what is perceived in one modality affects what is experienced in another. One example, the ventriloquist illusion, has been well studied since the 19th century. Work in the second half of the 20th century has confirmed various ways in which the visual location of a stimulus affects perceived auditory location. The effect is neither cognitive nor inferential, but results from cross-modal perceptual interactions. Similar cross-modal connections are revealed in the fascinating McGurk effect in speech perception (McGurk and MacDonald, 1976; Wright and Wareham, 2005), an auditory illusion produced by a visual experience. In the McGurk effect, a subject is presented with simultaneous audio and video of a talker recorded saying, for example, the syllable "ma", and videotaped while saying the word "ka". The subject's visual experience of the talker producing an open-lip sound seems to override the auditory experience of a closed-lip "ma" syllable. Certain visual-tactile effects such as visual capture also demonstrate cross-modal perceptual interaction.

Each of these effects, however, could be explained in terms of vision's dominance over some other modality. Perhaps visuo-centrism

is vindicated by vision's dominance in perception over the other modalities?

Not so. Ladan Shams and her colleagues have recently discovered a class of illusions in which audition affects vision. In the "sound-induced flash illusion" subjects presented with a single visual flash and double auditory beep have the same visual experience as when presented with a double visual flash accompanied by a double beep. That is, the double auditory beep affects visual content.

A single flash accompanied by multiple beeps is perceived as multiple flashes. This phenomenon clearly demonstrates that sound can alter the visual percept qualitatively even when there is no ambiguity in the visual stimulus. (152)

Three features of this result are significant. First, it is not cognitive or inferential or based on some strategy adopted to respond to an ambiguous or conflicting experience. Shams *et al.* (2002) maintain that audition influences the phenomenology of vision as a result of cross-modal *perceptual* interactions.

Second, these and many other cross-modal effects are pre-attentional. "...Cross-modal interaction reorganizes the auditory-visual spatial scene on which selective attention later operates." (Bertelson and deGelder, p 165)

Finally, a semantic contribution from familiar bimodal contexts isn't necessary to generate the effect. It appears to be a perceptual effect that takes place at a relatively low level. The effect is not the result of something that's just learned for particular contexts, or for which specific bimodal experience is required. It is an audition-induced phenomenological change in the character of visual experience that persists through shifts in setting and stimulus characteristics.

"We present the first cross-modal modification of visual perception which involves a phenomenological change in the quality – as opposed to a small, gradual, or quantitative change – of the percept of a nonambiguous visual stimulus. We report a

visual illusion which is induced by sound: when a single flash of light is accompanied by multiple auditory beeps, the single flash is perceived as multiple flashes. We present two experiments as well as several observations which establish that this alteration of the visual percept is due to cross-modal perceptual interactions as opposed to cognitive, attentional, or other origins." (2002: 147)

5. Explaining cross-modal illusions

What are the consequences of cross-modal illusions for philosophical thinking about perception and perceptual content? Since these effects are systematic and persistent, to explain the influence of one modality upon what is experienced in another modality in a way that captures the *environmental* or *adaptive* significance of correlations across one or more modalities requires appeal to some common factor that makes principles for grouping and organizing stimuli across the modalities intelligible.

This fact is reflected in what have been called *unity assumptions* for cross-modal interactions. For example, when an incongruence (spatial or temporal) between stimuli from different modalities is relatively limited and when concordance surpasses some threshold, a common environmental source likely accounts for both stimuli. The perceptual system's response results in cross-modal biases, recalibrations, or illusions. The visual and auditory stimuli are treated as evidence of some single environmentally significant entity or event and a perceptual "unit" is formed according to principles analogous to those involved in Gestalt formation from vision and from audition (*cf.* Bregman). The difference is that the principles are not limited to a single modality, but deal with the integration of information from the different sensory systems. These principles appeal to assumptions about a common environmental object or event that gives rise to *both* environmental stimuli. The important point is that these assumptions are not specific only to a particular modality; rather, they amount to either modality-independent or multi-modal assumptions about environment particulars.

They are, in effect, modality-independent assumptions about the sources of sensory stimulation. It is precisely because these grouping principles capture genuine regularities in the world of objects and events that awareness across different modalities constitutes genuine perceptual awareness of objects and events in the world.

But there's still a gap between influences across the modalities at the subperceptual level and the failure of the composite snapshot conception at the level of conscious perceptual awareness. Subperceptual auditory processing might result in illusory visual experiences without this showing anything about the content (its nature or richness) of the overall perceptual experience or the appropriateness of the composite snapshot conception of experience. What's needed is a bridge between claims about the influence of one modality upon what's experienced in another and claims about the respective contents of each individual modality.

I believe such a connection exists. The grouping and binding principles I've mentioned appear systematically to affect or to determine modality-specific content. For example, a principle that slightly out-of-sync visual and auditory stimuli close enough in time probably originate from a common source, along with general deference to audition on the temporal dimension (it's better than vision on this dimension), might result in a visual experience that comports with the auditory stimuli even when that visual experience differs from what it would have been in absence of the auditory stimulus. In the bi-modal case, the visual and auditory experiences ultimately end up the way they do *because* in general such visual and auditory stimulation very likely share a common environmental cause – a common source object or event. Explaining the effect any other way fails to capture *why* it's useful for the perceptual system to try to reconcile divergent stimuli.

That is, the perceptual system deploys principles designed to track, in a causally or counterfactually dependent way, the kinds of ordinary objects and events that lead to auditory and visual stimuli. But notice that this assumes modality-independent or multi-modal characterizations of such objects and events.

Describing these operations, therefore, involves attributing to perception some traction on ordinary objects and events in a sense

that goes beyond the modality-specific notions of "visual object" or "auditory event" deployed within a given modality. The idea is that experience is shaped by multimodal organizing principles, and such principles track ordinary objects and events, so audition and vision involve a dimension of multi-modal content that cannot be characterized in purely auditory or purely visual terms.

It is therefore plausible to think that we have good reason to ascribe a dimension of modality-independent or multi-modally characterized content to vision and to audition, beyond a mere causal interaction. In fact, the very same amodal content might be shared by vision and audition. So, it seems fair to suppose that the object- or event-involving character of a given modality stems from underlying multi-modal principles and content with potential for sharing across modalities.

But, even in the case of vision, such content cannot be captured by *purely* visual principles, and requires appeal to relations to audition and other modalities. Likewise, the content of audition might involve a level of content shared with vision. If so, then we have a foothold on the solution to the puzzle about audition set out earlier.

Audition has an object- or event-involving character because modality-independent or multi-modal principles shape auditory experience and ground a level of content that cannot be characterized in purely auditory terms. We hear sources, objects, and events, and not just sounds, pitches, and timbres, because the senses do not act as isolated systems that deliver neat modality-specific contents from which we learn to infer the presence of ordinary objects and events.

What I'm suggesting is that a convincing explanation of the cross-modal effects requires appeal to a dimension of perceptual content shared across the modalities. If that's right, then any snapshot that arrives within a specific modality is itself already a multi-modal photo infused with information shaped by and gleaned from the other modalities. There is no separating off without remainder the purely auditory content or even the purely visual content. Even the content of vision itself cannot be thoroughly understood in complete isolation from the other modalities.

Not only does the traditional empiricist conception that likens perceptual experience to a composite of discrete modality-specific snapshots fail as a characterization of perceptual experience, but its failure reveals an important flaw in the focus from which it stems. The tendency to take vision as an independent and representative paradigm for theorizing about perception is not only incomplete, but the visuo-centric thinking it leads to threatens to blind us to the nature and character of perceptual experience.

References

- O'Callaghan, Casey (forthcoming, *a*). "Sounds," in T. Bayne, A. Cleermans, and P. Wilken, eds., *Oxford Companion to Consciousness*, Oxford University Press.
- O'Callaghan, Casey (forthcoming, *b*). *The World of Sounds: A Philosophical Theory*. Oxford University Press.
- Shams, Ladan, Kamitani, Y., and Shimojo, S. (2000). "What you see is what you hear." *Nature*, Vol. 408, pp.788.
- Shams, Ladan, Kamitani, Y., and Shimojo, S. (2002). "Visual illusion induced by sound." *Cognitive Brain Research*, Vol. 14, pp. 147-152.
- Noë, Alva (2004). *Action in Perception*. Cambridge, MA. The MIT Press.
- Crocco, Mario (2004). "¡Alma e' reptil! Los contenidos mentales de los reptiles y su procedencia filética." *Electroneurobiología* 12 (1),1-72.
- Mondolfo, Rodolfo (1934) *El infinito en el pensamiento de la antigüedad clásica (L'infinito nel pensiero del Greci)* (Le Monnier, Firenze; also Imán, Buenos Aires, 1952).
- Mondolfo, Rodolfo (1955). *La comprensión del sujeto humano en la cultura antigua*, ch. IV: "La actividad sintética del sujeto reconocida como condición del conocimiento" (Imán, Buenos Aires, 1955; and EUDEBA, Buenos Aires, 1968).
- Ávila, Alicia and Crocco, Mario (1996). *Sensing: A New Fundamental Action of Nature*. *Folia Neurobiológica Argentina*, vol. X: Institute for Advanced Study, Buenos Aires.
- McGurk, Harry and MacDonald, John "Hearing lips and seeing voices", *Nature* 264, 746-748 (1976). See also videos:

<http://ramil.sagum.net/item/mcgurk-effect>

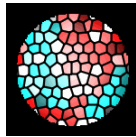
http://www.media.uio.no/personer/arntm/McGurk_english.html

Wright, Daniel and Wareham, Gary (2005); *Mixing sound and vision: The interaction of auditory and visual information for earwitnesses of a crime scene*, *Legal and Criminological Psychology*, Vol 10(1), pp. 103–108.

Bertelson, P. and de Gelder, B. "The psychology of multimodal perception." In C. Spence and J. Driver, eds., *Crossmodal Space and Crossmodal Attention*, pages 141–177. Oxford University Press, 2004.

Bregman, Albert S. *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press, Cambridge, MA, 1990.

Copyright © 2006 del autor / by the author. Este trabajo original constituye un artículo de acceso público; su copia exacta y redistribución por cualquier medio están permitidas bajo la condición de conservar esta noticia y la referencia completa a su publicación incluyendo la URL original (ver arriba). / This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's full citation and original URL (above).



revista

Electroneurobiología

ISSN: 0328-0446