From CPC (Current Psychology of Cognition)
Volume 13
Number 1
February 1994

CPC, 1994, 13 (1), 97-102

COMMENTARY ON "AUDITORY-VISUAL SPATIAL INTERACTION" (M. RADEAU)

Modularity of information, not processing

DOMINIC W. MASSARO

University of California

Two people faced with a doughnut: one sees a hole; the other the doughnut. Two cognitive scientists faced with a body of evidence: one sees modularity; the other the antithesis. Radeau's paper and my reaction to it were reminiscent of the précis of "Speech perception by ear and eye" and the concomitant peer commentary (Massaro, 1989). Finding analogous processes in a wide variety domains, including speech, I argued against modularity of perceptual and cognitive functioning. Some commentators, on the other hand, saw the evidence supportive of modularity and others against it. We now have several instances in which we might agree on the evidence but not what it means with respect to modularity. Following my belief in a fine-grained analysis, perhaps we should stick with testing specific models and not debate whether these models are modular or not.

The framework of a Fuzzy Logical Model of Perception (FLMP) has guided our research and also helps clarify the modularity issue. It is assumed that patterns are processed through a sequence of processing stages: evaluation, integration, and decision. Continuously-valued features are evaluated, integrated, and matched against prototype descriptions in memory, and an identification decision is made on the basis of the relative goodness of match of the stimulus information with the relevant prototype descriptions.

Correspondence should be sent to Dominic W. Massaro, Program in Experimental Psychology, University of California, Santa Cruz, CA 95064, U.S.A. (bitnet: massaro@cats; internet: massaro@fuzzy.ucsc.edu).

The model allows an important distinction to be made between information and information processing. Information refers the attributes and characteristics of the stimulus world, serving as functional inputs to the evaluation operation in the FLMP. Information processing refers to how this information is processed. Information processing corresponds to the algorithmic nature of the evaluation, integration, and decision operations rather than the actual information that is being operated on. We argue that the information is clearly different across different domains like spatial location and speech, but that the information processing is identical across domains. If we must use the term modularity, there is no question of modularity of information across different domains. On the other hand, information processing appears to be highly similar, if not identical, across the different domains.

Radeau's findings

Radeau (1994) has expertly reviewed some impressive evidence for the integration of auditory and visual sources of information in locating objects in our environment. Using an aftereffect paradigm, Radeau and Bertelson (1974) found that both visual and auditory recalibration occurred after exposure to auditory-visual conflict. An important conclusion of Radeau's is that the visual information does not completely dominate or override auditory information in spatial localization. As the author points out, this result falsifies the common belief that vision dominates other modalities. This is a refreshing conclusion given the history of field. Scientists have usually believed that the visual information dominated completely. In this age of mutual concessions and peace, it is now time to recognize not only other countries but also the parallel contribution of the multiple sources of information. Perhaps the term "visual dominance" should be banned - even though this prohibition might be more difficult to enforce than gun control.

Recalibration in the FLMP

Although the domains described by the FLMP have always involved perceptual judgments of events with multiple sources of information, the model could be extended to describe the recalibration paradigm. Prototypes in the model have featural information about the ideal values for a given object, such as a speech syllable. Pairing two featural values that

disagree with one another usually leads to some unique judgment. The outcome of this judgment could, in turn, modify the ideal values for the sources of information influencing the judgment (see Friedman, Cohen, and Massaro, submitted, for how the prototype descriptions could be updated). For example, an auditory /ba/ paired with a visual /da/ often produces the percept /tha/. Thus, the feedback from the perceptual experience could modify the auditory and visual feature values for /tha/ to be more /ba/-like and /da/-like respectively. Thus, repeated pairing of these two sources of information should make the auditory source presented alone more /tha/-like and less /ba/-like. Similarly, the visual /da/ should be more /tha/-like after being paired repeatedly with auditory /ba/. Both of these sources presented together should also be perceived as more /tha/-like. Thus, the FLMP can account for recalibration (in a manner consistent with Radeau's interpretation) in a natural manner given its current assumptions.

Unfortunately, there seems to be no evidence for cross-modal recalibration in speech perception. Roberts and Summerfield (1981) found that adaptation of an auditory continuum was only dependent on the previous auditory inputs, not on the visual information the auditory inputs were paired with, or the perceptual experience resulting from the auditory and visual information. This negative result seems to contrast with the research on spatial localization. If adaptation was dependent the outcome of the perceptual judgment, then there should have been some modification of the ideal feature values, which would have influenced judgments along the auditory continuum presented alone. Given Radeau's positive findings in spatial location, it would be productive to develop additional tests of recalibration in the bimodal speech domain.

Commonalities across domains

It is easy to be impressed with the commonalities between processing spatial location and speech, as well as with other domains of processing. Integration of auditory and visual information in speech and spatial location is fairly robust with slight discrepancies in the temporal onsets of the two modalities. Bimodal syllables composed of the auditory and visible syllables /ba/ and /da/ were presented at five different onset asynchronies (Massaro & Cohen, in press). The second experiment replicated the same procedure but with the vowels /i/ and /u/. The results indicated that perceivers integrated the two sources of information at asynchronies of 200 msec or less. Massaro, Smeele, and Cohen

REFERENCES

Fisher, B. (1991). Integration of visual and auditory information in perception of speech events. Unpublished Doctoral Dissertation, University of California, Santa Cruz, CA.

Green, K. P., Kuhl, P. K., Meltzoff, A. M., & Stevens, E. B (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Perception and Psychophysics, 50,* 524-536.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1989). Multiple book review of Speech perception by ear and eye: A paradigm for psychological inquiry. *Behavioral and Brain Sciences, 12,* 741-794.

Massaro, D. W., Smeele, P.M.T., & Cohen, M. M. (in progress). Integration of bimodal speech as a function of onset differences between the auditory and visible speech.

Massaro, D. W., & Cohen, M. M. (in press). Perceiving asynchronous bimodal speech in consonant-vowel and vowel syllables. *Speech Communication.*

Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Cahiers de Psychologie Cognitive, 13,* 3-51.

Radeau, M., & Bertelson, P. (1974). The aftereffects of ventriloquism. *Quarterly Journal of Experimental Psychology, 26,* 63-71.

Radeau, M., & Bertelson, P. (1976). The effect of a textured visual field on modality dominance in a ventriloquism situation. *Perception and Psychophysics, 20,* 227-235.

Roberts, M. and Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception and Psychophysics, 30,* 309-314.

CPC, 1994, 13 (1), 103-109

COMMENTARY ON "AUDITORY-VISUAL SPATIAL INTERACTION" (M. RADEAU)

Modules, and the art of perceptual compromise

JOSE MORAIS

Université Libre de Bruxelles

Radeau's proposal to consider auditory-visual pairing in source localization and audio-visual speech as two examples of cross-modal modular systems is supported by a large amount of converging data. It thus appears very persuasive. Moreover, it is based on an approach to the architecture of the cognitive system which is also my own, so that when I was invited to write a comment on Radeau's paper I feared I would have nothing substantial to add to it or criticize in it. Given the long standing participation of both of us in the laboratory created by Paul Bertelson, the fact that I fundamentally agree with Radeau's theses is not surprising, and therefore needs not be developed here. However, for the fun of taking the opposite course to that of my friend Monique, and above all, because I have been noticing a disturbing mismatch between some proclaimed properties of modules and the assumed functions of those modules, I came to appreciate this opportunity to set down my doubts in writing.

Correspondence should be sent to José Morais, Laboratoire de Psychologie Expérimentale, Université Libre de Bruxelles, Avenue Adolphe Buyl 117, 1050 Bruxelles, Belgium.

(in progress) also varied asynchrony between the audible and visible speech using both natural and synthetic syllables in an expanded factorial design. The results showed no effect of asynchrony up to 250 msec. The FLMP provided the best description of this integration at all asynchronies. Other unpublished findings show that integration does break down if the temporal onsets of the two modalities differ by a half second or more. Similar to the findings in spatial location judgments, we may conclude that the integration of audible and visible speech is very robust across small changes in temporal occurrence.

A recent study in our laboratory provides the most convincing demonstration of analogous processes in spatial location and speech. Fisher (1991) studied bimodal speech perception and auditory location judgments in a situation in which the visual face and the auditory speech could come from different locations. Subjects were instructed report both where the sound came from, and what syllable was heard. Subjects were required to look at the display because there could also be visual-alone trials in which they had to lipread and to locate the image. There was an influence of the location of the visual image in the auditory location judgment. Similarly, there was an influence of the visual speech on the identification of the auditory syllable. These results replicated previous studies in both the spatial location and speech domains. The important new finding was no crosstalk between the speech and location dimensions. The speech information did not influence the location judgment and the location information did not influence the speech judgment.

The FLMP gave a good description of Fisher's (1991) results both for the integration of auditory and visual location information in location judgments and for the integration of auditory and visual speech information in syllable identification. These results reveal that people naturally integrate auditory and visual speech even when the two sources come from different spatial locations. Furthermore, spatial location judgments can be described by the same integration algorithm of the FLMP used to describe categorization. Research has also shown that the integration of audible and visible speech does not seem to be attenuated when there is a mismatch between the sex of the face and the sex of the voice (Green, Kuhl, Meltzoff, & Stevens, 1991). In all cases, the integration of auditory and visual information in speech is robust across discrepancies in spatial location, temporal asynchrony, and the sex of the face and voice. Radeau's review highlights a similar robustness in the integration of auditory and visual information in spatial location.

There is also a correspondence between the spatial location and speech domains in terms of cognitive and attentive influences. Radeau and Bertelson (1976) demonstrated that attention could not account for the relative influence of the auditory and visual sources of information in spatial localization. We have analogous results in speech perception. The contribution of a source of information is primarily a function of its information value (and the values of other sources) and this contribution is modulated very little by attention instructions (Massaro, 1987, Chapter 3).

Summary

Radeau has added considerable fuel to the flames of analogous processes across domains. She draws comparisons between intermodal and intramodal processing and between speech and spatial localization. For example, there is minimal impact of certain conceptual factors in both speech and spatial localization. The influence of a given source of information has less to do with attention than simply the informativeness of each source. Perceptual systems naturally interpret or view relatively synchronous data as relating to a single object or event. The distinction that Radeau draws between spatial location and speech is really in terms of information and not information processing. Spatial location depends primarily on synchronization whereas speech is also dependent on mouth configurations associated with speech sounds. In summary, Radeau's work nicely illustrates the value of looking for similarities and differences across different domains of inquiry. In the present case, the similarities far outweigh the differences.

ACKNOWLEDGEMENTS

The research reported in this paper and the writing of the paper were supported, in part, by grants from the Public Health Service (PHS R01 NS 20314), the National Science Foundation (BNS 8812728), and the graduate division of the University of California, Santa Cruz. Great appreciation to KTM for the stimulating discussion of recalibration during a beautiful fall lunch.