Crossmodal integration: a good fit is no criterion

Massaro’s Fuzzy Logical Model of Perception (FLMP) is one of the dominant approaches to interoceptive integration. It was originally developed for the situation where listeners hear tokens from a /ba/-/da/ continuum while viewing a face articulating /ba/ or /da/. More recently, FLMP has been extended to other situations, such as ventriloquism and the bimodal perception of emotion. Massaro has argued that FLMP provides an adequate and universal model of perception, in particular about how information across modalities is combined. In this letter, however, we present several examples showing that FLMP does not enlighten the underlying perceptual processes. Massaro makes a strict distinction between information and information processing. FLMP is concerned only with the latter: “The model makes specific assumptions about how information is combined and Massaro’s extensive work suggests that, in almost all cases, FLMP fits the data more accurately than do alternative models that rely on different assumptions (additive or categorical).” For Massaro, this is the signature of a universal law by which information is integrated: it is the crux of what he refers to as information processing. However, with this exclusive emphasis on information processing, a disregarded issue is whether there are content-based constraints on what sources of information do or do not integrate.

Reading versus lip-reading
The other case of reading versus lip-reading. From a behavioral, developmental and neurological perspective, there are many reasons why reading is unlike lip-reading. And of course, Massaro is well aware of them, and he would agree that the information in the different modalities and the perceptual processes, for example, are different when an auditory token is combined with a different lip-read token, while other researchers have failed to obtain perceptual effects (finding only biases) when written text is combined with speech.

Massaro also claims that FLMP can be applied to the ventriloquist scenario in which subjects are asked to judge the apparent origin of a sound when presented with a visual signal that originates from a different location. Subjects had to understate the distance between the auditory and visual signals and sometimes even fuse them. It has been argued that this kind of crossstalk involves a decision about what is variously called, “pairing”, “unity”, or “object-identity”. The basic idea is that the perceptual system is required to decide whether auditory and visual information originate from a single source. In order to make this decision, the spatial proximity of the information sources and the similarity of the temporal pattern are thought to be used. As a consequence, as the distance between an auditory and a visual stimulus increases there will be two opposing effects on the overall bias of the perception of the auditory stimulus being displaced towards the visual attractor. Firstly, the proportion of trials on which an interaction occurs will decrease with increasing separation because the pairing decision is supported in fewer trials. On the other hand, the size of the attraction on those few trials in which there is pairing will increase with increasing separation. The overall effect of this might be that the crossmodal bias decreases when the distance between sound and light is increased. For example, when Bermant and Welch measured the visual bias on audition using separations between stimuli 60, 20 and 10 degrees, they obtained a decrease in the bias from 57 to 17 to 12%, respectively. Bertelson and Radeau also found that the visual bias decreased as the distance increased. Models that allow the pairing decision predict that the size of the ventriloquist effect should be increased when the sound and light sources move so far apart that the crossmodal interactions no longer be supported. Sound and vision are then treated as separate events without any interaction. In FLMP, the ventriloquist effect should increase when distance increases, because the farther apart the auditory and visual stimuli, the more the visual signal supports a distant location. The (weighted) average of the auditory and visual information should thus move away from the auditory location. How then is it possible to conclude that the FLMP provides a good description of the data? The only requirement to be that the FLMP does not make a prediction. While FLMP is a very flexible tool, it fits data retrospectively by adjusting true values until there is a satisfying fit.

Ventriloquism

Thus, when the visual signal moves away from the sound source and the bias decreases, one may obtain a good fit by decreasing the visual support for the more distant location. The crucial point, however, is that the truth values are meaningless because there is no guarantee that there is a correspondence with the perceptual mechanisms that lead to these truth values. Thus, the fact that fuzzy sets of mathematics can describe aspects of results does not enlighten us about the underlying mechanisms; a good fit is therefore no criterion to accept FLMP as an adequate theory of perception.

Acknowledgements
We would like to thank Paul Bertelson for his insightful comments on a previous version of this paper.

Jean Vroomen and Beatrice de Gelder
Tilburg University, Department of Psychology, Warandelaan 2,
PO Box 90153, 5000 Le Tilburg, The Netherlands.
tel: +31 13 466 2370 fax: +31 13 466 2394 e-mail: j.vroomen@tue.nl

References

Reply to Vroomen and de Gelder
Given that I am sympathetic to Vroomen and de Gelder’s commentary, I can only hope that they have failed to read my lips or mistake my written word rather than misunderstand what they have read. Admittedly, my short review article could be read out of context and the reader could easily believe that I have gone beyond the evidence given (in the same way that Vroomen and de Gelder’s commentary1 could be read out of context and the reader could easily believe that I have gone beyond the evidence given). We use and promote our information-processing framework primarily because it encourages the investigator to determine the stage (level in Vroomen and de Gelder’s terms) of processing responsible for various findings.

I will show that their two main points can be easily pursued within our framework, after a short qualification of the origins of the FLMP. Vroomen and de Gelder state that, ‘Originally, it (the FLMP) was developed for the situation where listeners hear tokens from a /ba-/ /da/ continuum while viewing a face articulating /b/ or /d/’. They are correct that the FLMP was not derived simply to describe pattern recognition. FLMP was not derived simply to describe pattern recognition; it is important to note that the FLMP was not derived simply to describe pattern recognition. It was developed for the situation where listeners hear tokens from a /ba-/ /da/ continuum while viewing a face articulating /b/ or /d/.

The model was originally developed to account for the interaction of several auditory cues in speech perception and for various sources of information in sentence comprehension. By assessing the model, it is important to note that the FLMP was not devised simply to describe speech perception by ear and eye, but rather to describe pattern recognition more generally.

Lip-reading versus reading
First, Vroomen and de Gelder question whether written text is operating at the same stage as visible speech when these sources are separated combined with auditory speech. It is intuitively, somehow, to believe that the influence of visible speech is more real than the influence of written text. However, it is worth noting a couple of caveats. First, Vroomen and de Gelder should not dismiss the positive finding of Frost et al.3 as simply a bias because we now know that biases can be truly perceptual. This possibility was pointed out long ago by Paul Bertelson4 when investigators tried to dismiss his ventriloquism effect as a response bias when analyzed within the context of signal detection theory (for further discussion of the important distinction between perceptual bias and decision bias, see Ref. 9). How would one test whether the two types of visual input operate differently? Our experiment is simply a first step along that road. Contrasting different models of performance should then follow. It is straightforward to formulate a model based on the interpretation proposed by Vroomen and de Gelder in which the visual input has its influence on decision rather than perception (Ref. 15, Chapter 2). The outcome of these tests would speak to the issue of analogous processes in reading and lip-reading.

Vroomen and de Gelder1 state that ‘the issue is not whether reading and lip-reading interact at all with speech, but whether they interact at the same processing level, and whether FLMP allows one to distinguish between the various forms that this interaction may take’. This question has always been of central interest to us and is why I argue for the formalization and testing of alternative models. The post-perceptual guessing model, the auditory dominance model, and the ‘faces’ model have all been tested as alternatives to the FLMP, primarily because they assume different forms of interaction of the two sources of information.

Additional experiments can be generated to test these models in more rigorous theoretical explanations. One important source of evidence comes from the nature of the judgments that are given. Specifications of the model are of interest in combination responses, such as /bd/; this is contrary to the story of /bd/, it should be noted that when and how often these combinations occur is highly variable and unpredictable. In an early study with open-ended alternatives, Radeau et al.10 found no combination responses (Ref. 11, and see Ref. 12, pp. 53–54). In our studies with /bd/ as one of the specified response alternatives, we have found up to 80% (Ref. 13) and as low as 10% combination responses (Ref. 10, p. 146) when a visual /b/ is paired with an auditory /d/. In order to understand whether these combination responses should be equivalent in the reading and lip-reading conditions, however, it is first necessary to understand the FLMP. Our interpretation has been that a visual /b/ is paired with an auditory /d/ to provide two sources of information that are consistent with /bd/. A visual /b/ looks a lot like a visual /b/, and an auditory /d/ is somewhat similar to auditory /d/. Thus, /bd/ can be a reasonable percept given these two sources. This explanation also predicts very few /b/ judgments when a visual /b/ is paired with an auditory /d/. This is consistent with the story of /bd/. These types of constraints probably do not occur in the reading situation, however, because the written letter activates some speech-like representation without actually providing a speech stimulus. This situation is more analogous to the