COGNITION AND EMOTION, 2000, *14* (3), 313–320

Fuzzy logical model of bimodal emotion perception: Comment on "The perception of emotions by ear and by eye" by de Gelder and Vroomen

Dominic W. Massaro and Michael M. Cohen
University of California, Santa Cruz, CA, USA

De Gelder and Vroomen (this issue) studied how emotions are perceived from information given by the face and the voice. Emotions in the face and the voice were presented under both unimodal and bimodal conditions. When participants identified the affect, they were influenced by information from both modalities. This result occurred even when they were instructed to base their judgement on just one of the modalities. These experiments and results provide an independent replication of similar studies published in Massaro (1998). Given this opportunity for a new set of tests, the fuzzy logical model of perception (FLMP) was fit to the new results provided by de Gelder and Vroomen. Of central interest is the nature of the bimodal performance as a function of the unimodal performance. The FLMP gave a good fit of performance. The description reveals that, although information differences exist across different instruction conditions, the information processing involved in pattern recognition appears to be the same and well-described by the FLMP.

The influence of facial information on speech perception has gained prominence in cognitive psychology beginning with the classic study of McGurk and MacDonald (1976). De Gelder and Vroomen (this issue) used these findings as a model for their studies of emotion perception. They asked participants to identify an emotion (e.g., happy or sad) given a photograph and/or an auditory spoken sentence. They found that their identification judgements were influenced by both sources of information,

Requests for reprints should be sent to Dr Dominic W. Massaro, Department of Psychology, University of California, Santa Cruz, Santa Cruz, CA 95064, USA; email: Massaro@fuzzy.ucsc.edu.

This research was supported, in part, by grants from the Public Health Service (PHS R01 DC00236), the National Science Foundation (23818, CDA-9726363), Intel Corporation, and the University of California, Santa Cruz. The authors thank Christopher Campbell for feedback on the article.

http://www.tandf.co.uk/journals/pp/02699931.html

even when they were instructed to base their judgement on just one of the sources. These results are particularly valuable to us because they represent an independent test of our theoretical framework by researchers other than ourselves. We use their results to test the fuzzy logical model of perception (FLMP), which to date has survived a number of empirical tests (Massaro, 1998; Massaro & Stork, 1998).

The FLMP provides an account of perception and pattern recognition in a wide variety of domains. Within the FLMP, perceptual recognition is viewed as having available multiple sources of information supporting the identification and interpretation of the environmental input. The assumptions central to the model are: (1) each source of information is evaluated to give the continuous degree to which that source specifies various alternatives; (2) the sources of information are evaluated independently of one another; (3) the sources are integrated to provide an overall degree of support for each alternative; and (4) perceptual identification and interpretation follows the relative degree of support among the alternatives. Figure 1 illustrates the stages of processing in the model.

The paradigm that we have developed permits us to determine how one source of information is processed and integrated with other sources of information. The results also inform us about which of the many potentially functional cues are actually used by human observers (Campbell & Massaro, 1997; Massaro, 1987, chapter 1; Massaro & Friedman, 1990). The systematic variation of properties of the signal combined with the quantitative test of models of speech perception enables the investigator to test the psychological validity of different cues. This paradigm has already

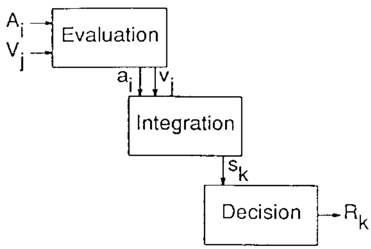

Figure 1. Schematic representation of the three processes involved in perceptual recognition. The three processes are shown to proceed left to right in time to illustrate their necessarily successive but overlapping processing. These processes make use of prototypes stored in long-term memory. The sources of information are represented by upper-case letters. Auditory information is represented by A_i and visual information by V_j. The evaluation process transforms these sources of information into psychological (or fuzzy truth, Zadeh, 1965) values (indicated by lower-case letters a_i and v_j). These sources are then integrated to give an overall degree of support, s_k, for each speech alternative, k. The decision operation maps the outputs of integration into some response alternative, R_k. The response can take the form of a discrete decision or a rating of the degree to which the alternative is likely.

proven to be effective in the study of audible, visible, and bimodal speech perception (Massaro, 1987, 1998). Thus, our research strategy not only addresses how different sources of information are evaluated and integrated, but can uncover what sources of information are actually used. We believe that the research paradigm confronts both the important psychophysical question of the nature of information and the process question of how the information is transformed and mapped into behaviour. Many independent tests point to the viability of the FLMP as a general description of pattern recognition. The FLMP is centred around a universal law of how people integrate multiple sources of information. This law and its relationship to other laws are presented in detail in Massaro (1998).

The assumptions of the FLMP are testable because they are expressed in quantitative form. One is the idea that sources of information are evaluated independently of one another. Independence of sources is motivated by the principle of category-conditional independence (Massaro & Stork, 1998): It is not possible to predict the evaluation of one source on the basis of the evaluation of another, so the independent evaluation of both sources is necessary to make an optimal category judgement. Sources are thus kept separate at evaluation, they are then integrated to achieve perception and interpretation.

Multiplicative integration yields a measure of total support for a given category identification. This operation, implemented in the model, allows the combination of two imperfect sources of information to yield better performance than would be possible using either source by itself. However, the output of integration is an absolute measure of support; it must be relativised, due to the observed factor of relative influence (the influence of one source increases as other sources become less influential, i.e., more ambiguous). Relativisation is effected through a decision stage, which divides the support for one category by the summed support for all other categories. An important empirical claim about this algorithm is that although information may vary from one perceptual situation to the next, the manner of combining this information—information processing—is invariant. With our algorithm, we thus propose an invariant law of pattern recognition describing how continuously perceived (fuzzy) information is processed to achieve perception of a category.

Given this framework, one emerging feature of the FLMP is the division of perception into the twin levels of information and information processing. The sources of information from the auditory and visual channels make contact with the perceiver at the evaluation stage of processing. The reduction in uncertainty effected by each source is defined as information. In the fit of the FLMP, for example, the parameter values indicating the degree of support from each modality correspond to information. These parameter values represent how informative each source of information is.

Information processing refers to how the sources of information are processed. In the FLMP, this processing is described by the evaluation, integration, and decision stages. Within this framework, we can ask what information differences exist among individuals and across different pattern-recognition situations. Similarly, we can ask whether differences in information processing occur. For example, we can look for differences in both information and information processing when participants are given different instructions in a pattern-recognition task.

In their first experiment, de Gelder and Vroomen asked participants to identify the person as happy or sad. The stimuli were manipulated in an expanded factorial design with an 11-step visual continuum between happy and sad and an auditory sentence that was read in either a happy or sad voice. Thus, there were 11 × 2 bimodal conditions, 11 visual-alone conditions, and 2 auditory-alone conditions, for a total of 35 unique stimulus conditions. The participants were instructed to watch the screen and to listen to the voice on each trial.

The FLMP was fit to the average results by estimating free parameters for the 11 levels of visual information and 2 levels of auditory information. Figure 2 gives the observed and predicted results. As can be seen in the

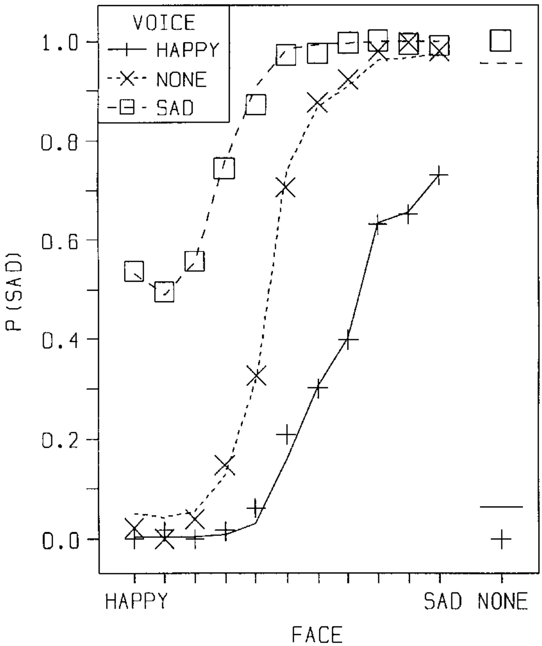

Figure 2. The points give the observed proportion of sad identifications in the auditory-alone, the factorial auditory-visual, and the visual-alone conditions as a function of the auditory and visual stimuli. The lines are the predictions of the FLMP.

TABLE 1
Parameter values for the 11 levels of the face and 2 levels of the voice for Experiment 1 (use both modalities) and Experiment 2 (ignore the voice)

Face	*Happy*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*Sad*
Exp. 1	.05	.04	.05	.13	.32	.74	.87	.91	.96	.96	.98
Exp. 2	.03	.04	.11	.17	.42	.61	.87	.94	.97	.97	.99

Voice	*Happy*	*Sad*
Exp. 1	.06	.96
Exp. 2	.30	.72

figure, the FLMP gives a good description of the average results with a root mean square deviation (RMSD) of .022. Table 1 gives the parameter values.

The same design was used in the second experiment except that the two auditory-alone trials were omitted. Observers were told to judge the face and to ignore the voice. The FLMP was fit to these new results by estimating a new set of free parameters for the 11 levels of visual information and 2 levels of auditory information. Figure 3 gives the observed and predicted

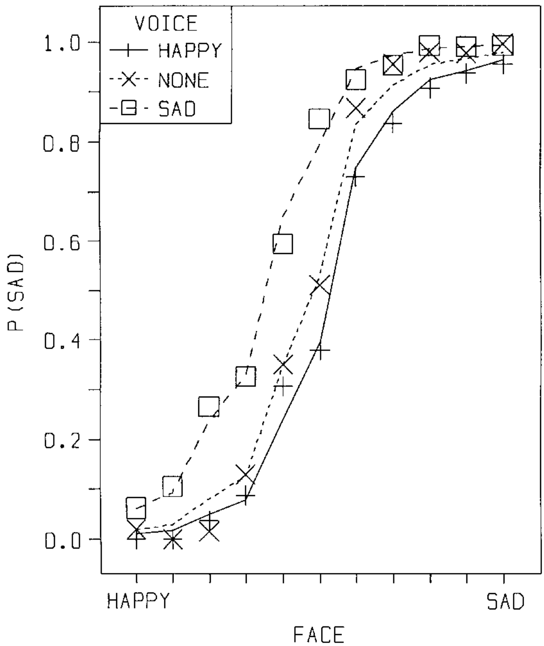

Figure 3. The points give the observed proportion of sad identifications in the factorial auditory-visual and the visual-alone conditions as a function of the auditory and visual stimuli. The lines are the predictions of the FLMP.

results. As can be seen in the figure, the FLMP gives a good description of the average results with a RMSD of .027. Table 1 gives the parameter values.

Comparison of the parameter values across the two experiments in Table 1 allows us to test the hypothesis that there are information differences in the two different instruction conditions. As can be seen in the table, the parameter values for the happy and sad voice are made much more neutral (closer to .5) in the situation in which participants were instructed to ignore the voice than in the situation in which they were told to use both modalities. The parameter values for the face were mostly similar across the two conditions. Thus, the FLMP is capable of describing the results by simply assuming that the information from the voice was attenuated when participants were instructed to ignore it. The good fit of the FLMP in both instruction conditions, however, indicates that the two sources are integrated in the same manner regardless of instructions.

In the third experiment aimed at having observers judge the voice and to ignore the face, a 7-step auditory sentence continuum was made between happy and afraid. The visual stimuli were happy and fearful photographs of the speaker of the sentences. For some reason the auditory-alone stimuli were not presented. Thus, there were only $7 \times 2 = 14$ experimental conditions. The FLMP was fit to the average results by estimating free parameters for the 7 levels of auditory information and 2 levels of visual information. Figure 4 gives the observed and predicted results. As can be seen in the figure, the FLMP gives a good description of the average results with an RMSD of .023. Table 2 gives the parameter values. Although a direct comparison between this experiment and Experiment 1 is not justified because of the different stimuli that were used, we can observe that the influence of the face was much smaller when participants were instructed to ignore it. Tables 1 and 2 show that the parameter values for the prototypical emotions were much attenuated in Experiment 3 relative to Experiment 1.

TABLE 2
Parameter values for the 7 levels of the voice and 2 levels of the face for Experiment 3

Voice	*Afraid*	*2*	*3*	*4*	*5*	*6*	*Happy*
Exp. 3	.08	.12	.16	.34	.70	.89	.94

Face	*Afraid*	*Happy*
Exp. 3	.36	.62

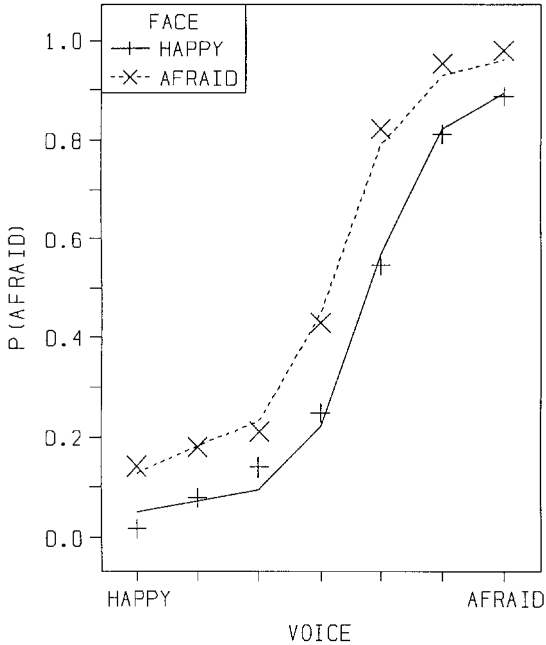

Figure 4. The points give the observed proportion of afraid identifications in factorial auditory-visual conditions as a function of the auditory and visual stimuli. The lines are the predictions of the FLMP.

In conclusion, we were successful in testing the FLMP against a new set of data from a new set of investigators. The framework and model provide a parsimonious account of several experimental manipulations. The distinction between information and information processing is a powerful concept and reveals how instructional differences can modulate performance in the task. This outcome replicates the findings in Massaro (1998, chapter 8) and adds to the body of results supporting a universal principle for pattern recognition.

Manuscript received 11 March 1999

REFERENCES

Campbell, C.S., & Massaro, D.W. (1997). Visible speech perception: Influence of spatial quantization. *Perception*, *26*, 627–644.

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

Massaro, D.W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle.* Cambridge, MA: MIT Press.

Massaro, D.W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review*, *97*, 225–252.

Massaro, D.W., & Stork, D.G. (1998). Speech recognition and sensory integration. *American Scientist*, *86*, 236–244.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, *8*, 338–353.