Tests of auditory–visual integration efficiency within the framework of the fuzzy logical model of perception

Dominic W. Massaro[a] and Michael M. Cohen
University of California, Santa Cruz, California 95060

(Received 12 October 1999; revised 22 February 2000; accepted 21 April 2000)

The fuzzy logical model of perception [FLMP, Massaro, *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT Press, Cambridge, MA, 1998)] has been extremely successful at describing performance across a wide range of ecological domains as well as for a broad spectrum of individuals. Because the model predicts optimal or maximally efficient integration, an important issue is whether this is the case for most individuals. Three databases are evaluated to determine to what extent a significant quantitative improvement in predictive ability can be obtained if integration is assumed to be somewhat inefficient. For the most part, there were no significant signs of inefficient integration. The previous differences found by Grant and Seitz [J. Acoust. Soc. Am. **104**, 2438–2450 (1998)] must be due to their measures of efficiency, which appear to be invalid and/or conflate information with integration efficiency. Finally, the descriptive ability of the FLMP is shown to be theoretically informative and not simply the model's ability to describe any possible outcome. © *2000 Acoustical Society of America.* [S0001-4966(00)01008-0]

PACS numbers: 43.71.An, 43.71.Ma [CWT]

I. TESTS OF AUDITORY–VISUAL INTEGRATION WITHIN THE FRAMEWORK OF THE FLMP

Grant and Seitz (1998) should be applauded for initiating an important study of the individual's ability to integrate auditory and visual (AV) information. As pointed out in their discussion, this assessment has important implications for treatment of hearing-impaired and even visual-impaired subjects (see Grant and Walden, 1995; Grant, Walden, and Seitz, 1998). However, before we can reach any conclusion about a person's ability to integrate, a more precise treatment within a validated model of integration must be carried out.

Grant and Seitz propose four measures of AV integration efficiency. The first is an integration efficiency (IE) measure from Braida's (1991) Pre-Labeling Model (PRE). From what we can tell, however, this measure uses only correct responses rather than all of the cells of the bimodal confusion matrix. We believe that all of the data should be used in measuring IE. The McGurk susceptibility measure is confounded with the quality of the auditory and visual information. A person will be more susceptible to McGurk effects to the extent she has poor auditory information and good visual information, independently of her IE (Massaro, 1998, pp. 12–14). The two measures of auditory delay are also debatable in that we might expect better integrators to be less susceptible to asynchrony of the two sources rather than more susceptible.

Their two measures of AV benefit certainly reflect the quality of the auditory and visual information. The measures $\mathrm{AV}-A$ and $(\mathrm{AV}-A)/(1-A)$ are measures of the benefit gained from the addition of visible speech. The measure $(\mathrm{AV}-A)/(1-A)$ adjusts for overall performance level and provides a coarse measure of the visual contribution. However, this measure is influenced by both the amount of visual information and the efficiency of integrating it with the auditory information. Thus, this measure is necessarily confounded with the amount of visual information the subject has available. Given these limitations, we propose that the Grant and Seitz (1998) analyses do not provide valid measures of integration efficiency.

II. A FORMAL MODEL OF INFORMATION INTEGRATION

We argue that any measure of integration efficiency requires a formal model of performance that specifies exactly how integration takes place. Performance of a given individual can then be assessed within the framework of the model to address the question of integration efficiency. We propose the fuzzy logical model of perception (FLMP) as an ideal model for this type of analysis. The assumptions central to the model are: (1) each source of information is evaluated to determine the 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. In a two-alternative task with /ba/ and /da/ alternatives, the degree of auditory support for /da/ can be represented by a_i, and the support for /ba/ by $(1-a_i)$. Similarly, the degree of visual support for /da/ can be represented by v_j, and the support for /ba/ by $(1-v_j)$. The probability of a response to the unimodal stimulus is simply equal to the feature value. For bimodal trials, the predicted probability of a response, $P(/\mathrm{da}/)$ is equal to

$$P(/\mathrm{da}/)=a_i v_j/[a_i v_j+(1-a_i)(1-v_j)]. \qquad (1)$$

[a]Electronic mail: massaro@fuzzy.ucsc.edu

TABLE I. The amount of support for each of the three alternatives as a function of the auditory and visual sources of information. These degrees of support illustrate that overall performance on the bimodal conditions can be significantly worse than overall performance on the auditory conditions.

Support for alternative	Auditory /ba/	Auditory /va/	Auditory /da/	Visual /ba/	Visual /va/	Visual /da/
/ba/	0.7	0.1	0.2	0.3	0.2	0.5
/va/	0.3	0.6	0.1	0.3	0.3	0.4
/da/	0.1	0.2	0.7	0.5	0.3	0.2

$$P(/\mathrm{ba}/|A_bV_b)=\frac{0.7\times 0.3}{(0.7\times 0.3)+(0.3\times 0.3)+(0.1\times 0.5)}$$

$$=0.21/0.35=0.60,$$

$$P(/\mathrm{va}/|A_vV_v)=\frac{0.6\times 0.3}{(0.6\times 0.3)+(0.1\times 0.2)+(0.2\times 0.3)}$$

$$=0.18/0.26=0.69,$$

$$P(/\mathrm{da}/|A_dV_d)=\frac{0.7\times 0.2}{(0.7\times 0.2)+(0.2\times 0.5)+(0.1\times 0.4)}$$

$$=0.14/0.28=0.50,$$

Average auditory unimodal $=(0.7+0.6+0.7)/3=2.0/3=0.67$,
Average bimodal $=0.6+0.69+0.50=0.60$.

In the course of our research, we have found that the FLMP accurately describes human pattern recognition. We have learned that people use many sources of information in perceiving and understanding speech, emotion, and other aspects of the environment. The experimental paradigm that we have developed also allows us to determine which of the many potentially functional cues are actually used by human observers (Massaro, 1998, Chap. 1). This paradigm has already proven to be effective in the study of audible, visible, and bimodal speech perception (Massaro, 1987, 1998; Massaro and Cohen, 1976; Oden and Massaro, 1978).

How can the FLMP be used to assess integration efficiency? As can be seen in Eq. (1), the auditory and visual sources of support are multiplied to give an overall degree of support for each response alternative. The value a_i representing the degree of auditory support is assumed to be the same on both unimodal auditory and bimodal trials. This same property holds for the visual support. This property and the multiplicative integration rule, followed by the relative goodness rule (RGR), entail the process to be optimal and thus maximally efficient (see Massaro, 1998, pp. 115–117; Massaro and Stork, 1998; Massaro and Friedman, 1990).

Grant and Seitz (1998) express an ''apparent belief that a second source of information can only improve overall performance.'' Although it seems unintuitive with an optimal integration rule, a second source can indeed lower performance, even though (in fact because) integration is operating efficiently. In order to demonstrate this possibility within the context of the FLMP, assume three alternatives, /ba/, /va/, and /da/, which are differentially supported by the auditory and visual sources by the parameter values given in Table I. Assume that a bimodal /da/ is presented and auditory /da/ supports /da/ 0.7, /va/ 0.1, and /ba/ 0.2. Visual /da/ supports /da/ 0.5, /va/ 0.4, and /ba/ 0.2. Correcting identifying /da/ on unimodal auditory and on unimodal visual trials would be 0.7 and 0.5, respectively. As shown in Table I, performance on bimodal trials would be equal to 0.50, which is well below performance of 0.7 on unimodal auditory trials. This somewhat surprising outcome can be true for overall performance in the task (see Table I). This theoretical demonstration, which is also found in empirical results, challenges Grant and Seitz's apparent belief that the FLMP predicts that a second source of information can only improve overall performance.

In our previous work, we have contrasted a fuzzy logical model of perception with a single-channel model (SCM) of perception. These represent integration and nonintegration models, respectively, and therefore a test between these models at the individual subject level indicates whether a person integrates the auditory and visual speech. The SCM is mathematically equivalent to a weighted averaging model (WTAV), which is an inefficient algorithm for combining the auditory and visual sources. This model predicts that two sources can never be more informative than one. Thus, previous contrasts of the two models have addressed the issue of integration efficiency. Given that the FLMP has consistently provided a significantly better description of a variety of results from several different types of experiments, people generally must be fairly efficient information integrators. In further simulations of the two-alternative task (Massaro, 1998, Chap. 10), however, we found that the fit of the FLMP fell slightly shy of a benchmark criterion indicating a perfectly accurate fit. The addition of decision noise (noise added at the response selection stage) with a standard deviation of 0.1 was necessary to bring the FLMP in line with what would be expected from the data being generated by this model. Thus, there is some hint that the perceivers might not be perfectly efficient integrators. In this letter, we pursue this issue more directly.

A direct way to measure integration efficiency in the FLMP is to determine whether there is any loss of information on bimodal trials relative to unimodal trials. A potentially inefficient integration model can be formalized within the FLMP framework. One simply assumes that reduced information from the auditory and visual sources can be present on bimodal trials relative to unimodal trials. In this case, the degree of auditory support on bimodal trials, a_{iB}, is compromised by the function

$$a_{iB}=w_a a_{iU}+(1-w_a)0.5. \tag{2}$$

An analogous function describes the visual information

$$v_{jB}=w_v v_{jU}+(1-w_v)0.5, \tag{3}$$

where w_a and w_v correspond to the weights given the auditory and visual feature values, respectively.

For tasks with two response alternatives or for models with features that lie between 0 and 1, the feature values represent more support for an alternative to the extent the value is greater than 0.5. The value 0.5 is completely ambiguous. Because the weights can lie between 0 and 1, the smaller the weight value the less the support is controlled by the unimodal feature value and the more it is controlled by 0.5 (complete ambiguity). A weight value of 1 makes the same prediction as the original FLMP.

Tests of efficiency, therefore, simply involve testing this new model and determining the weight values and also to what extent this model gives a better description of performance compared to the standard FLMP. We apply this model and two related models to two data sets given in Massaro (1998, Chaps. 2, 6, 10), and the data set described by Grant and Seitz (1998).

III. EXPANDED FACTORIAL DESIGN WITH TWO RESPONSE ALTERNATIVES

A typical manipulation is to systematically vary the ambiguity of each of the sources of information. We used synthetic speech to cross five levels of audible speech varying between /ba/ and /da/ with five levels of visible speech varying between the same alternatives in an expanded factorial design (Massaro, 1998). There were 24 observations at each of the 35 unique experimental conditions. Eighty-two subjects were instructed to listen and to watch the speaker, and to identify the syllable as /ba/ or /da/.

The mean observed proportion of /da/ identifications was computed for each subject for the 35 unimodal and bimodal conditions. Both the auditory and the visual sources of information had a strong impact on the identification judgments in both the unimodal and bimodal conditions. Most importantly, the auditory and visual effects were not additive in the bimodal condition, as demonstrated by a significant auditory–visual interaction. This result is consistently obtained in this type of experiment. It means that the influence of one source of information is greatest when the other source is neutral or ambiguous.

The FLMP gave a better description than the WTAV model for 94% of these 82 participants, with average root-mean-square deviations (RMSDs) of the individual model fits of 0.051 and 0.097, respectively. To further address the issue of efficiency, we tested the weighted bimodal FLMP given by Eqs. (2) and (3) against these same results. The average RMSD was 0.0449, smaller than the standard fit of 0.051. An analysis of variance on the RMSDs from the two models showed that this difference was significant, $p < 0.001$. This statistical result should not be surprising because the weighted bimodal FLMP is identical to the standard FLMP, but with two additional free parameters. Rather than conclude that some of the observers are inefficient, the better fit may simply reflect the addition of two free parameters rather than a real loss of information in the bimodal condition. We are confident that a goodness-of-fit measure which takes into account model flexibility (Myung and Pitt, 1997; Massaro *et al.*, in press) would not find a significant advantage of the inefficient integration model relative to the FLMP.

Given this possibility, we felt that a more appropriate model would be one that simply compromises the outcome of integration rather than the separate information values. In this case, the predicted compromised probability of a response, $P(/\mathrm{da}/)$ on bimodal trials is equal to

$$P(/\mathrm{da}/) = wP(/\mathrm{da}/) + (1-w)0.5, \tag{4}$$

where $P(/\mathrm{da}/)$ is equal to Eq. (1), and the a_i and v_j values are identical to those on the unimodal trials.

This more direct model of efficiency gave an average RMSD of 0.0491, which was very little improvement over the standard FLMP. Most of the estimated weight values were close to 1, indicating that there was very little loss of information in the bimodal condition relative to the unimodal conditions. It appears that only 6 or 8 of the 82 subjects appear to be inefficient integrators.

IV. EXPANDED FACTORIAL DESIGN WITH EIGHT RESPONSE ALTERNATIVES

For the second database, we replicated this same experiment with 36 subjects, given eight rather than just two response alternatives. Subjects were instructed to listen to and watch the talker, and to identify the syllable as /ba/, /da/, /bda/, /dba/, /tha/, /va/, /ga/, or ''other.'' The category other was to be used by the subject whenever none of the other seven responses seemed suitable. Although the test continua were between /ba/ and /da/, we obtained several other response alternatives (see Massaro, 1998, pp. 184–188). These judgments reflect the contribution of both auditory and visual speech, even when observers are permitted a larger permissible set of response alternatives.

The FLMP is tested against results with multiple response alternatives in the same manner as with just two response alternatives. With more than two alternatives, it is necessary to estimate a unique parameter to represent the degree to which each source of information supports each alternative. The fit of this model requires $5a_i$ and $5v_j$ parameters for each of the 8 response alternatives, for a total of 80 free parameters. This might seem like a large amount but the number of data points to be predicted has increased by the same factor. We are now predicting $35\times 8=280$ data points. The fit of this model to each of the 36 subjects produced an average RMSD of 0.0507. To assess whether the FLMP maintains its advantage with multiple response alternatives, we compared this fit with that of a single-channel model (or equivalently, a weighted averaging model). The fit of this competing model was about two times poorer, giving an RMSD of 0.1049. The FLMP gave a better description than the WTAV model for all but one of the 36 subjects.

To address the issue of efficiency, we tested the two weighted bimodal FLMPs against these same results. The predictions are made in the same way except now the neutral response probability is 1/8. For the two-weight model, the average RMSD was 0.0501, very close to the original average RMSD of 0.0507. The weight values were equal to 1 for most of the subjects and close to 1 for the others. For the one-weight model, the average RMSD was 0.0500. These weight values were equal to 1 for most of the subjects and close to 1 for the others.

V. IDENTIFICATION CONFUSION MATRIX WITH 18 ALTERNATIVES

Finally, we analyzed Grant and Seitz's (1998) 40 subjects in the vCv consonant identification task. We have carried out two different types of model descriptions of this type of experiment with A, V, and AV confusion matrices: modality analysis and feature analysis (Massaro and Cohen, 1999).

A. Modality-analysis implementation

As in the eight-alternative task, it is necessary to estimate a unique parameter to represent the degree to which each source of information supports each alternative. We use a_{Bi} to represent the degree to which the audible speech supports the alternative /ba/. The term v_{Dj} would represent the degree to which the visible speech supports the alternative /da/, and so on for the other response alternatives. Given both audible and visible speech, the total support for the alternative /ba/, $s(/\text{ba}/)$, would be

$$s(/\text{ba}/)=a_{Bi}v_{Bj}, \tag{5}$$

and so on for the other test conditions and the other alternatives.

As in the case of just two alternatives, the probability of a particular categorization is assumed to be equal to the relative goodness of match of that alternative relative to the sum of the goodness-of-match values of all possible response alternatives. With 18 stimulus–response alternatives, each test stimulus provides different degrees of support for each alternative. It is necessary to estimate 18 free parameters for each of the 18 test stimuli in each modality. Thus, 324 free parameters are required for the auditory modality and 324 for the visual modality. We are able to test the model by predicting 3×324 data points with 2×324 free parameters.

As in the previous tests, the standard FLMP and the two-weighted bimodal FLMPs were fit to the results. The predictions are made in the same way except that the neutral response probability is 1/18. The average RMSD for the two-weight model was equal to the average RMSD of the standard FLMP (0.0111). The weight values were equal to 1 for all of the subjects. For the one-weight model, the average RMSD was 0.0110. The weight values were equal to 1 for most of the subjects and close to 1 for the others.

B. Feature analysis implementation

The model test we have just presented makes no assumptions about the psychophysical relationship among the different test items. A unique parameter is estimated for each possible stimulus–response pairing. For example, a unique parameter is estimated to represent the amount of support a visual /b/ provides for the response alternative /d/. To test the psychological reality of various linguistic features and to reduce the number of free parameters, we have articulated the FLMP in terms of audible and visible support for these features (Massaro and Cohen, 1999). This formulation has the potential to save a large number of features, because it is assumed that a given feature in a given modality has the same impact regardless of what segment it is in. Following the tradition begun with Miller and Nicely (1955), we can define the 18 consonants by five features: voicing, nasality, place, duration, and frication (see Massaro and Cohen, 1999).

We assume that features for the 18 consonants are simply sensory primitives that distinguish speech categories. Although the features used in the following tests are chosen to be equivalent to the linguistic features, they should be thought of simply as handy labels for the underlying sensory features. Thus, for example, the auditory feature for place would not necessarily be equivalent to the parameter value for the visible feature for place. In fact, the feature values for one modality should be independent of the feature values for another modality. For example, we would expect that voicing and nasality would have informative feature values for auditory speech and relatively neutral feature values for visible speech. The place feature, on the other hand, would give relatively informative values for visible speech. Thus, the features at the evaluation stage are not linguistic, but perceptual.

Thus, each of the 18 syllables would be described by the conjunction of five features for unimodal speech and the conjunction of ten features for bimodal speech. Even though each feature is defined as a specific value or its complement (e.g., voiced or voiceless), its influence in the perception of visible speech is represented by a value between 0 and 1. The parameter value for the feature indicates the amount of influence that feature has. Therefore, if the /ma/ and /na/ prototypes are each expected to have a nasal feature and the calculated parameter value for this feature is 0.90, then the nasal feature is highly functional in the expected direction. Alternatively, if the calculated parameter value for the nasal feature is 0.50, then the nasal feature is not functional at all. Because of the definition of negation as 1 minus the feature value, a feature value of 0.5 would give the same degree of support for an alternative that has the feature, as it should for an alternative that doesn't have the feature. Finally, if the calculated parameter value is 0.20 then the nasal feature is functional but the opposite of the expected direction. Finally, it should be noted that the features are not marked in this formulation: absence of nasality is as informative as presence of nasality. Thus, if a nasal stimulus supports a nasal response alternatives to the degree 0.9, then a non-nasal stimulus also supports a non-nasal alternative to the degree 0.9.

The overall match of a test stimulus to each syllable prototype was calculated by combining the feature matches according to the assumptions of the FLMP. These constraints dictate that (1) the features are the sources of information that are evaluated independently of one another, and (2) the features are integrated multiplicatively to give the overall degree of support for a syllable alternative. Thus, the overall degree of support for /ba/, $s(/\text{ba}/)$, given the presentation of a /ba/ syllable, is

$$s(/\text{ba}/|/\text{ba}/)=a_v a_n a_p a_d a_f v_v v_n v_v v_d v_f, \tag{6}$$

where each feature value indexes a match between the feature in the stimulus and the corresponding feature in the /ba/ prototype. The feature a_v correspond to auditory voicing, v_n to visual nasality, and so on. A mismatch between the feature in the stimulus and the corresponding feature in the prototype would be indexed by $(1-f_i)$, where f_i corresponds to the modality's feature value. Thus, the support for the /ka/ prototype given presentation of a /ba/ syllable, is

$$\begin{aligned} s(/\text{ka}/)=&(1-a_v)a_n(1-a_p)a_d a_f(1-v_v)\\ &\times v_n(1-v_p)v_d v_f, \end{aligned} \tag{7}$$

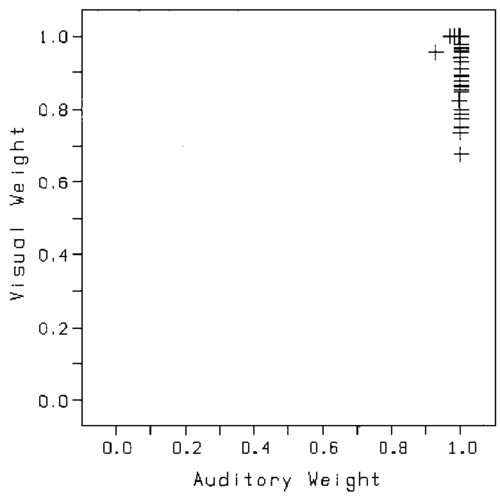

FIG. 1. The estimated auditory and visual weight values for the two-weight model for each of the 40 subjects in the Grant and Seitz (1998) 18-alternative task, for the feature-analysis implementation.

where $(1-f_i)$ indexes a mismatch between the feature in the /ba/ stimulus and the corresponding feature in the /ka/ prototype. This same formulization is used for the place feature which has six levels rather than just two.

After the overall degree of support for each syllable is calculated, the stimulus is categorized according to the RGR, which states that the relative probably of choosing an alternative is the goodness of match of that alternative divided by the sum of the goodness of match of all alternatives. Thus, this model implementation parallels the previous one in all aspects except in terms of the featural description of the stimulus and response alternatives. The FLMP can thus be tested against the confusion matrix by estimating the amount of information in each feature and the featural correspondence between the stimulus and response prototypes. Thus, five parameters are necessary to describe the auditory information and the same number to describe the visual.

The standard FLMP and the two weighted bimodal FLMPs were fitted to the results in the same manner, with the neutral feature value of 0.5. The average RMSD of the standard FLMP fit was 0.1001. For the two-weight and one-weight models, the average RMSDs were equal to 0.1000 and 0.1001, respectively, essentially equal to the fit of the standard FLMP. The auditory and visual weight values for each subject for the two-weight model are given in Fig. 1. These weights were equal to one for most of the subjects and close to one for the others. The distribution of the weight values for the one-weight model is shown in Fig. 2. These values were also very close to one.

In summary, our test of integration efficiency revealed very little support for the thesis that some individuals might be less-efficient integrators than others. This result clearly held for seven of eight model tests across three different data sets. The only hint of inefficient integration was for the data set of 82 subjects from our expanded factorial design with two response alternatives. Possible explanations for this outcome are given in Massaro (1998, pp. 313–318).

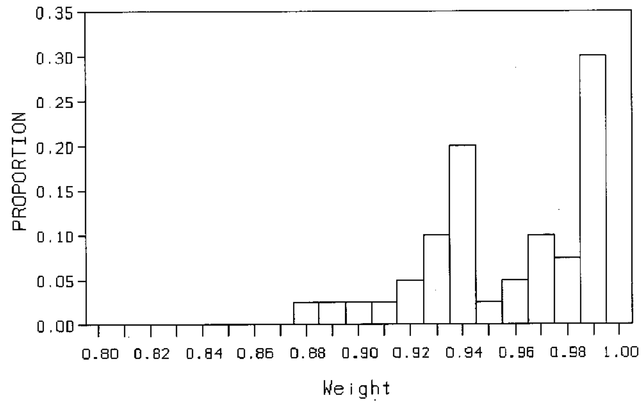

FIG. 2. The distribution of weight values for the one-weight model for the 40 subjects in the Grant and Seitz (1998) 18-alternative task, for the feature-analysis implementation.

VI. SENSITIVITY OF THE FLMP

A second important issue is raised by Grant and Seitz's (1998) claim that ''the consistently excellent fits achieved by the FLMP may also suggest that the model is less sensitive in recognizing subtle changes in integration efficiency'' (p. 2439). To support their claim, Grant and Seitz fit the FLMP to two data sets, one original set of data and a second that involved some modification of the first data set to give 16% better AV performance. The RMSDs of the fit to the two data sets were 0.013 and 0.020, respectively. Although the authors interpreted this small difference as a nonsignificant one, there was no justification for this conclusion and we believe that the observed difference is significant. To pursue this possibility, we made a similar modification of each of their 40 data sets and tested the FLMP against these two sets of data. To achieve the new hypothetical data sets, we generated a new set of bimodal results for each subject by differentially modifying the correct and incorrect proportions in the bimodal confusion matrix. Every correct cell along the negative diagonal was multiplied by 1.8. For the other cells, we multiplied their proportions by 0.2. Each entry in the final bimodal matrix was determined by normalizing each cell value by the total of all 18 cells in that row. The overall accuracy of this new set of bimodal results for each subject averaged about 14% more accurate than the original data set. Thus, the new data set has the original unimodal results and hypothetical bimodal results that are more accurate than would be expected from the FLMP.

If Grant and Seitz are correct, then the FLMP should fit these hypothetical results about as well as the original results. We have already reported the RMSD of the original data set, which was 0.0112. The average RMSD for the simulated data set was 0.0130, significantly larger that the fit to the original data, $F(1,28)=8.28$, $p=0.008$. As an additional test of sensitivity, we carried out the same contrast but with a parameter-free test in which the unimodal response probabilities were used to predict the bimodal judgments. The RMSD values for the original data set and the enhanced data set were 0.0522 and 0.0658, a significant difference, $p<0.001$. Thus, we can conclude that the FLMP is indeed sensitive to small differences in categorization behavior, and as demonstrated, provides a powerful framework for evaluating the efficiency of integration.

ACKNOWLEDGMENTS

The research was supported by grants from the National Institute of Deafness and Other Communicative Disorders (Grant No. PHS R01 DC00236), the National Science Foundation (Challenge Grant No. CDA-9726363), Intel Corporation, and the University of California Digital Media Innovation Program.

Braida, L. D. (**1991**). ''Crossmodal integration in the identification of consonant segments,'' Quarterly J. Exp. Psychol. **43**, 647–677.

Grant, K. W., and Seitz, P. F. (**1998**). ''Measures of auditory–visual integration in nonsense syllables and sentences,'' J. Acoust. Soc. Am. **104**, 2438–2450.

Grant, K. W., and Walden, B. E. (**1995**). ''Predicting auditory–visual speech recognition in hearing-impaired listeners,'' in Proceedings of the XIIIth International Congress of Phonetic Sciences, Vol. 3, pp. 122–129 (unpublished).

Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). ''Auditory–visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory–visual integration,'' J. Acoust. Soc. Am. **103**, 2677–2690.

Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Erlbaum, Hillsdale, NJ).

Massaro, D. W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT Press, Cambridge, MA).

Massaro, D. W., and Cohen, M. M. (**1976**). ''The contribution of fundamental frequency and voice onset times to the /zi/–/si/ distinction,'' J. Acoust. Soc. Am. **60**, 704–717.

Massaro, D. W., and Cohen, M. M. (**1999**). ''Speech perception in hearing-impaired perceivers: Synergy of multiple modalities,'' J. Speech Lang. Hear. Sci. **42**, 21–41.

Massaro, D. W., Cohen, M. M., Campbell, C. S., and Rodrigues, T. (**in press**). ''Bayes factor of model selection balidates FLMP,'' Psychonom. Bull. Rev.

Massaro, D.W., and Freidman, D. (**1990**). ''Models of integration given multiple sources of information,'' Psychol. Rev. **97**, 225–252.

Massaro, D. W., and Stork, D. G. (**1998**). ''Speech recognition and sensory integration,'' Am. Sci. **86**, 236–244.

Massaro, D. W. (submitted).

Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.

Myung, I. J., and Pitt, M. A. (**1997**). ''Applying Occam's razor in modeling cognition: A Bayesian approach,'' Psychonom. Bull. Rev. **7**, 79–95.

Oden, G. C., and Massaro, D. W. (**1978**). ''Integration of featural information in speech perception,'' Psychol. Rev. **85**, 172–191.