Perception & Psychophysics
1983, 33 (5), 501-505

Consonant/vowel ratio: An improbable cue in speech

DOMINIC W. MASSARO
and MICHAEL M. COHEN
University of California, Santa Cruz, California

The idea of the consonant duration relative to the vowel duration as a cue to voicing of postvocalic consonants is almost three decades old. Port and Dalby (1982) offer what they believe to be convincing support for such a view. We argue that the acceptance of this idea is premature and probably incorrect. In every instance that the concept of consonant/vowel (C/V) ratio is contrasted with an alternative view, the latter provides a significantly better description of the results. This alternative view considers consonant and vowel duration as independent cues to voicing, following the evaluation and integration of cues within the fuzzy-logical model of speech perception (Massaro & Cohen, 1976, 1977; Massaro & Oden, 1980a, 1980b; Oden & Massaro, 1978).

The first relevant study was carried out by Denes (1955), who varied vowel duration and the final consonant duration in the perception of the voicing difference between the pronunciations of the word *use*. Without any formal analysis, Denes interpreted the results in terms of the ratio of consonant duration to vowel duration, serving as a cue to voicing of the final consonant. The perceived voicing should decrease systematically with increases in the C/V ratio. However, in an earlier paper (Massaro & Cohen, 1977), we provided a quantitative test of the C/V ratio hypothesis against the observed results of Denes (1955). We also tested the idea that vowel duration and consonant duration provide independent cues to voicing of the consonant. The fit of the latter model was twice as good as that of the former, even though the better model required only three-fifths as many free parameters.

In the present note, we evaluate the C/V ratio model against the data of Derr and Massaro (1980) and Port and Dalby (1982). The present tests offer as much flexibility as possible for the quantification of the C/V ratio idea. In addition, we contrast the C/V ratio model against the predictions of the fuzzy-logical model, assuming independent vowel and consonant duration cues. The fuzzy-logical model does a significantly better job of describing the results than does the C/V ratio model.

Port and Dalby (1982) asked subjects to identify synthetic speech stimuli as *dibber* or *dipper* in Experiment 1, and as *digger* and *dicker* in Experiment 2. The initial vowel duration and the medial consonant silent closure duration were orthogonally varied in a factorial design. The right panels of Figures 1 and 2 plot the proportion of voiced judgments as a function of closure duration; vowel duration is the curve parameter. As can be seen in the figure, both variables had a strong influence on the identification results. If the C/V ratio is the critical stimulus parameter for voicing of the medial stop, then the likelihood of a voiced response should vary *only* as a function of this ratio; the actual durations of the vowel and consonant closure should not matter. The plot of the results as a function of the C/V ratio is shown in the left panels of Figures 1 and 2. Port and Dalby (1982) concluded that the judgments were determined primarily by the C/V ratio. They based this conclusion on two types of analyses. First, the C/V boundary values (as 50% crossovers) did not change much with changes in vowel duration (cf. left panels of Figures 1 and 2). Second, a multiple regression analysis showed that C/V ratio alone accounted for almost as much variance (61% and 66%) as did the absolute values of vowel and consonant duration (63% and 70%). However, Port and Dalby did not provide a quantitative test of the role of C/V ratio, and did not contrast this idea against other extant models of the speech perception process. To provide such a comparison, we formalized the C/V ratio model and contrasted its predictions against the fuzzy-logical model.

According to the fuzzy-logical model, speech perception involves feature evaluation, prototype matching, and pattern classification operations. Vowel duration and consonant duration are assumed to be independent cues at feature evaluation. During prototype matching, the cue values are inserted in prototypes in long-term memory and integrated (conjoined) together to arrive at a goodness of match of the stimulus with each prototype. For the present speech contrast, prototypes for the voiced and voiceless alternatives are defined as the conjunction ($\wedge$) of vowel duration (V) and closure duration (C) information.

Voiced: long vowel and NOT(long closure)

$$= V\wedge(1-C).$$

Voiceless: NOT(long vowel) and long closure

$$= (1-V)\wedge C.$$

This research was supported by a grant from the National Institute of Mental Health (MH-35334) to Dominic W. Massaro. We would like to thank Dr. Port for providing us the data for analyses. Requests for reprints should be sent to Dominic W. Massaro, Program in Experimental Psychology, University of California, Santa Cruz, California 95064.

Copyright 1983 Psychonomic Society, Inc.

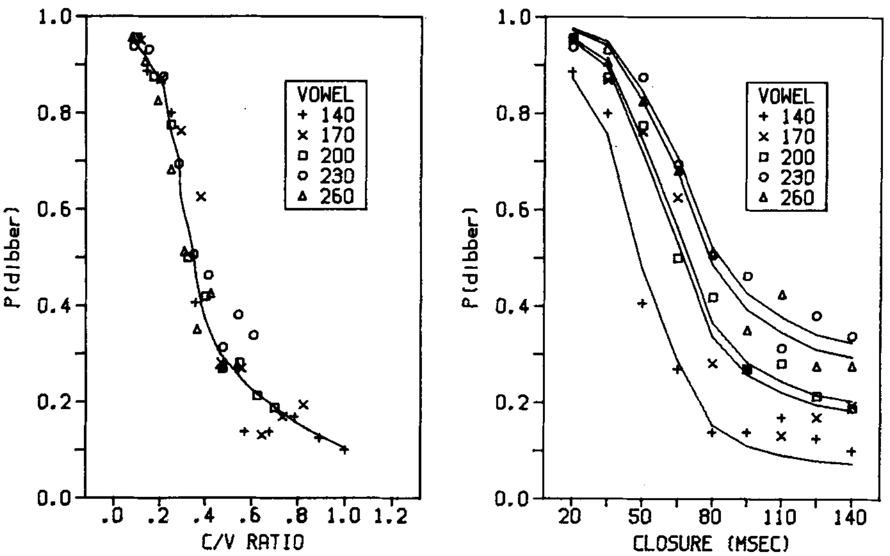

Figure 1. Left panel: Proportion of *dibber* judgments as a function of the closure/vowel ratio. Right panel: Proportion of *dibber* judgments as a function of closure duration and vowel duration. (Results from Port & Dalby, 1982.)

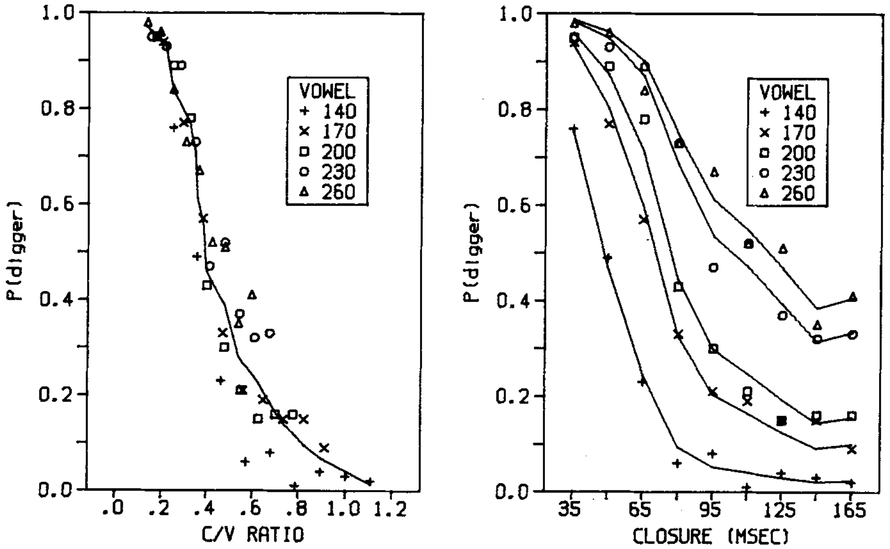

Figure 2. Left panel: Proportion of *digger* judgments as a function of the closure/vowel ratio. Right panel: Proportion of *digger* judgments as a function of closure duration and vowel duration. (Results from Port & Dalby, 1982.)

Subjects evaluate the degree to which the vowel and closure durations are long and enter these values into the prototype definitions. The cues are treated independently and given fuzzy-truth values between 0 and 1. The conjunction operation is defined as multiplication, and negation is defined as the additive complement. To arrive at a particular judgment, the relative goodness of match is taken following the logic of Luce's (1959) choice rule. In this case, the probability of a voiced judgment, P(Voiced), is equal to

$$\mathrm{P(Voiced)} = \frac{V \times (1 - C)}{V(1 - C) + [(1 - V) \times C]}.$$

The idea of independent consonant and vowel duration cues should not be interpreted as a statistical independence of the effects of these two variables. Multiplying the cues together at the prototype matching operation does not violate the principle of independent cues at the feature evaluation operation. The feature value of one cue remains independent of the feature value of another cue. Conjunction of two or more cues is, by definition, an interactive process. That is, the outcome of the conjunction depends on both values, which means that the values can interact statistically in their effects on performance.

To test the fuzzy-logical model, it is necessary to estimate a unique parameter for each unique value of vowel and closure duration to arrive at a particular set of results. Given five levels of vowel duration and nine levels of closure duration, 14 parameters are necessary to predict the 45 data points. The parameters were estimated by finding those values that minimized the squared deviations between the predicted and observed group results, using the minimization routine STEPIT (Chandler, 1969).

According to the C/V ratio model, equivalent C/V ratios should produce equivalent identification probabilities. However, Port and Dalby did not specify how the identification probabilities should change with changes in the C/V ratio. It is reasonable to assume that increases in the value of the C/V ratio will decrease the likelihood of a voiced response. Accordingly, the probability of a voiced response was assumed to be a monotonic function of the physical C/V values. Because 14 parameters were used in the description by the fuzzy-logical model, 14 parameters were used to estimate the monotonic changes in the identification probabilities with changes in the physical C/V values. With 14 parameters, it was possible to estimate the optimal starting point and ending point for 13 connected linear segments. The y ordinate corresponding to the proportion of identifications was divided into 13 equal intervals. Fourteen values along the x abscissa were then determined to give a monotonic function of 13 connected line segments. The 14 values were chosen to minimize the squared deviations between the predicted line and the observed points, using the minimization routine STEPIT (Chandler, 1969).

The measure of goodness of fit used to compare the two models was the root mean squared deviation (RMSD) between predicted and observed values. The C/V ratio model gave RMSD values of .063 and .078 for Port and Dalby's Experiments 1 and 2, respectively. The fuzzy-logical model gave RMSD values of .041 and .030. Accordingly, the results cannot be taken as evidence for C/V ratio, since the fuzzy-logical model does an even better job.

Although the fuzzy-logical model provides a better description, the C/V ratio model does a respectable job of describing the data. However, pooled group results may not be as revealing as individual subject data. In Port and Dalby's study, each subject was tested only 10 times on each unique stimulus condition. Given the small number of observations, the fit of the models to individuals is not reasonable. Luckily, an earlier study, carried out by Derr and Massaro (1980) for other purposes, provides reasonably reliable individual data.

Derr and Massaro (1980) synthesized sounds along a continuum between /jus/ and /juz/ (the noun and verb pronunciations of *use*). Four levels of steady-state vowel duration were factorially combined with four levels of frication duration. The fundamental frequency contour during the vowel was also varied systematically, but the current analysis of the results is pooled over this variable. Two experiments were carried out. In the identification task, the subjects identified the stimuli as one of the two alternatives. In the rating task, the subjects rated the degree to which the stimuli represented one alternative or the other. The 7 subjects in the identification task and the 10 subjects in the rating task each gave 60 observations at each of the 16 stimulus conditions.

The two models were fit to the individual results and to the group data, as in the analysis of the Port and Dalby (1982) study. The rating responses were translated into values lying between 0 and 1, corresponding to the linear distance along the rating scale. This value can be considered to be equal to the degree of voicing and treated identically to the percentage of voiced judgments in the identification task (see Derr & Massaro, 1980). In the fit of the fuzzy-logical model, eight parameters were estimated for the vowel duration and closure duration values. For the C/V ratio model, the eight parameters were used to derive a 7-segment piecewise linear transformation function. Table 1 presents the RMSD values for the individual subjects, the average RMSD, and the

Table 1
Root Mean Squared Deviations (RMSD) Values for the C/V Ratio Model and the Fuzzy-Logical Model Fit to the Identification Judgments of Derr and Massaro (1980)

	Model	
Subject	C/V Ratio	Fuzzy-Logical
1	.18	.03
2	.27	.05
3	.10	.03
4	.08	.05
5	.10	.04
6	.22	.02
7	.12	.01
Mean	.15	.03
Group Data	.14	.01

Table 2
Root Mean Squared Deviations (RMSD) Values for the C/V Ratio Model and the Fuzzy-Logical Model Fit to the Rating Judgments of Derr and Massaro (1980)

	Model	
Subject	C/V Ratio	Fuzzy-Logical
1	.23	.05
2	.19	.05
3	.12	.05
4	.09	.04
5	.11	.03
6	.07	.04
7	.16	.02
8	.28	.02
9	.11	.03
10	.16	.06
Mean	.15	.04
Group Data	.14	.04

RMSD for the group data for the two models for the identification results. Table 2 presents the same information for the rating results.

The results clearly support the fuzzy-logical model for both the identification and rating tasks in the Derr and Massaro study. The fuzzy-logical model is able to describe the group identifications within one percentage point, whereas the RMSD for the C/V ratio model is 14 times as large. The fit to the rating results is about four times better for the fuzzy-logical model than for the C/V ratio model. The differences between the models can be seen in Figures 3 and 4. The right panel of Figure 3 plots the average proportion of voiced judgments as a function of frication duration; vowel duration is the curve parameter. Figure 4 gives the corresponding results for the average rating judgments. The lines in the right panels of Figures 3 and 4 give the predictions of the fuzzy-logical model. As can be seen in the figures, the model gives a good description of the results. If the C/V ratio is the critical stimulus parameter for voicing, however, then the likelihood of a voiced response should vary *only* as a function of this ratio. The left panels of Figures 3 and 4 show that this is not the case, producing a poor fit by the C/V ratio model.

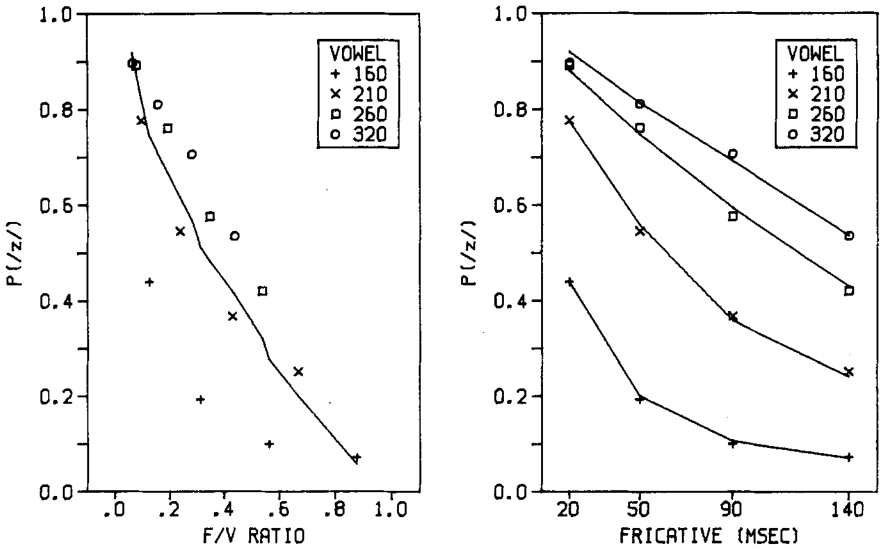

Figure 3. Left panel: Proportion of /z/ judgments as a function of the fricative/vowel duration ratio. Right panel: Proportion of /z/ judgments as a function of the fricative duration and vowel duration. (Results from Derr & Massaro, 1980.)

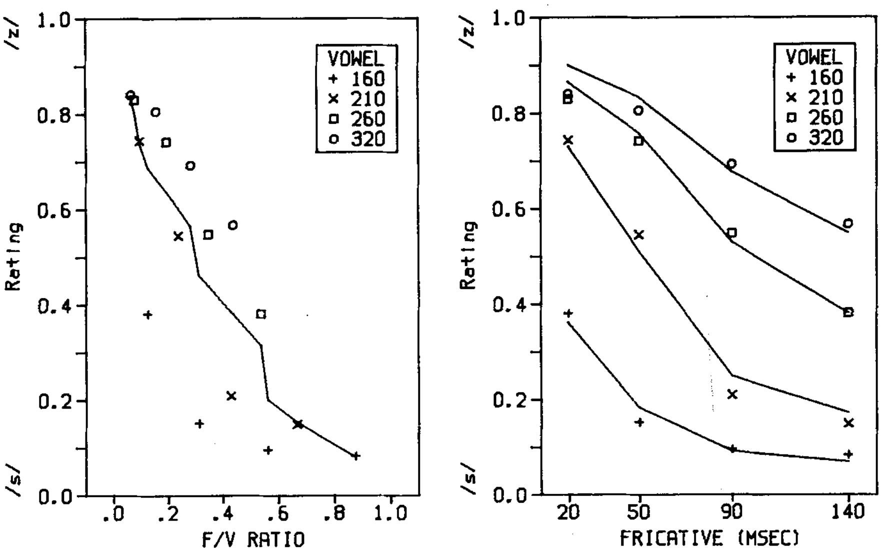

Figure 4. Left panel: Rating judgments as a function of the fricative/vowel duration ratio. Right panel: Rating judgments as a function of the fricative duration and vowel duration. (Results from Derr & Massaro, 1980.)

Why do Port and Dalby (1982) and we reach such different conclusions? They depended on the analyses of category boundary shifts and the use of multiple regression. First, they measured the C/V ratio boundary values (as 50% crossovers), using a probit analysis of nonasymptotic portions of the identification functions. They found that the C/V ratio values did not change very much with changes in vowel duration. However, the probit analysis can be misleading because the identification functions are not explained but are reduced to single boundary values. Since two identification functions can differ significantly even with similar boundary values, the identification functions provide a much more discriminating test between the C/V ratio and fuzzy-logical model.

The second analysis used by Port and Dalby was multiple regression, which showed that the C/V ratio alone accounted for almost as much variance as did an independence model assuming separate values of vowel and consonant duration. If the independent values of vowel duration and consonant duration are critical for the perception of voicing, why didn't the multiple regression analysis better discriminate between the C/V ratio and the independence model? The answer is that the independence model treating vowel duration and closure duration as independent variables does not accurately represent the description of the fuzzy-logical model. It should be noted that Port and Dalby (1982) did not equate the independence model with the fuzzy-logical model; however, we make the following points, because readers might interpret an independence model as equivalent to the fuzzy-logical model. In multiple regression, the physical values of the cues are the determining factor and the final contribution of one cue is independent of the contribution of the other cue. Both of these assumptions violate the interpretation given by the fuzzy-logical model. First, the influence of a cue is determined by its psychological feature value, and second, the multiplicative combination of cues followed by the relative goodness rule leads to the least ambiguous cue's having the most influence on the judgment. Accordingly, a multiple regression treating the physical durations as independent variables cannot be used as an evaluation of the fuzzy-logical model. We believe that contrasting models should be tested by directly comparing the relative accuracy of their quantitative predictions against the observed results.

Port and Dalby (1982) also evaluated for an influence of speaking rate, since a nice feature of C/V ratio is that it could be defined independently of speaking rate. Hence, C/V ratio could be a potential cue that does not require any normalization due to speaking rate. However, speaking tempo of a context sentence did influence the voicing judgments even when they are plotted against C/V ratio. If C/V ratio is a cue to voicing, it does not seem to function independently of sentence rate. The C/V ratio must vary with speaking rate, since decreasing the speech rate does not necessarily lengthen the vowel and consonant segments by the same proportion. A simple proportional normalization does not describe normalization results in production or perception (Miller, 1981; Nooteboom, 1981). Accordingly, if perception mirrors the information in the speech signal, then the C/V ratio needed for voicing must vary with speech rate. Any mechanism proposed to account for the normalization of C/V ratio could also be used with independent vowel and consonant cues to voicing. Massaro and Oden (1980b) discuss how normalization for speech rate can be described within the framework of the fuzzy-logical model.

REFERENCES

Chandler, J. P. Subroutine STEPIT finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, **14**, 81-82.

Denes, P. Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, 1955, **27**, 761-764.

Derr, M. A., & Massaro, D. M. The contribution of vowel duration, F_0 contour, and frication duration as cues to the /juz/-/jus/ distinction. *Perception & Psychophysics*, 1980, **27**, 51-59.

Luce, R. D. *Individual choice behavior*. New York: Wiley, 1959.

Massaro, D. W., & Cohen, M. M. The contribution of fundamental frequency and voice onset time to the /zi/-/si/ distinction. *Journal of the Acoustical Society of America*, 1976, **60**, 704-717.

Massaro, D. W., & Cohen, M. M. Voice onset time and fundamental frequency as cues to the /zi/-/si/ distinction. *Perception & Psychophysics*, 1977, **22**, 373-382.

Massaro, D. W., & Oden, G. C. Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America*, 1980, **63**, 81-97. (a)

Massaro, D. W., & Oden, G. C. Speech perception: A framework for research and theory. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 3). New York: Academic Press, 1980. (b)

Miller, J. L. Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives in the study of speech*. Hillsdale, N.J: Erlbaum, 1981.

Nooteboom, S. G. Speech rate and segmental perception or the role of words in phoneme identification. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland, 1981.

Oden, G. C., & Massaro, D. W. Integration of featural information in speech. *Psychological Review*, 1978, **85**, 172-191.

Port, R. F., & Dalby, J. Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics*, 1982, **32**, 141-152.

(Manuscript received October 25, 1982;
revision accepted for publication January 31, 1983.)