Journal of Experimental Psychology: Human Perception and Performance
1976, Vol. 2, No. 4, 538–543

How the Listener Integrates the Components of Speaking Rate

François Grosjean and Harlan Lane
Northeastern University

The rate of speaking in words per minute is a function of three independent variables, namely, articulation rate and the number and durations of pauses. The present study varies each of these components separately in a factorial design in order to determine how the listener combines them into a global impression of speech rate. A model was obtained by direct scaling and linear regression that accounts reasonably well for the estimates of apparent rate obtained in this and several other studies. It provides for a trading relation between articulation rate and pause rate in which the former is much more influential in determining the listener's perception of speech rate.

Previous studies of the listener's perception of speech rate have generally used the overall rate (words or syllables/min) as the independent variable. Hutton (1954) reported that category estimates of apparent rate were related to physical rate (in words/min) by a logarithmic function. In a magnitude estimation study, however, Lane and Grosjean (1973) found that apparent rate grows as a power of physical rate, with an exponent of 1.6. This exponent was shown subsequently to remain reasonably invariant across a range of rates, passages, and languages (Grosjean & Lass, in press).

Although investigators have long been aware that global rate is made up of several variables, specifically, articulation rate and the duration and frequency of pauses (Cotton, 1936; Kelly & Steer, 1949), few researchers have attempted to describe the relative contributions of these component variables to the listener's perception of rate. On the one hand, Goldman-Eisler (1968) proposed that "What is experienced as increase of speed in talking is therefore due largely to the closing of gaps and to the heightened continuity with which movements performed at a relatively constant rate succeed each other" (p. 26). On the other hand, Kelly and Steer (1949) asked judges to give category estimates of overall rate and of "sentence rate" (i.e., articulation rate) and concluded that the sentence rate method of describing extemporaneous speech is more highly related to audience judgment than is overall rate of speaking. The trouble with that conclusion, as Clevenger and Clark (1963) point out, is that Kelly and Steer did not study the effects of pausing. Articulation rate, pause time, and overall rate may interact in some way to produce the listener's perception of speech rate. Grosjean and Lane (1974) varied articulation rate and pause frequency in a magnitude estimation study and concluded that these two components of overall rate contribute about equally over the ranges employed. Pause duration, however, was held constant so its role remained unassessed.

Several studies have examined the effect of mechanical rather than natural alterations in rate on perception (Cartwright & Lass, 1976), preference (Cain & Lass, 1974; Foulke & Sticht, 1966; Lass & Goff, 1974), and comprehension (e.g., Orr, Friedman, & Williams, 1974; Foulke, Note 1). Once again, the global variable of words/min has been widely used, which makes it impossible to compare the effects of natural and mechanical alterations in rate, since each operates on the component variables differently, and the relative weights of

This research was supported in part by a Fulbright-Hays Award to the first author, who is currently on leave of absence from the Université de Paris VII.

Requests for reprints should be sent to François Grosjean, Department of Psychology, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115.

these variables in governing the perception of rate are unknown. Specifically, speech compressors delete or add brief segments of the speech stream, independent of its temporal structure; they maintain a constant phonation time ratio across all of the rates produced and, in achieving this, often extend the articulation rate beyond its natural limits without affecting the pause frequency or duration to the same extent. Speakers, in contrast, alter articulation rate only slightly, whereas they change pause time liberally, thus yielding substantial changes in the phonation time ratio (Lane & Grosjean, 1973).

We are interested, then, in determining how the listener combines articulation rate and the number and length of the pauses to produce a global impression of speech rate. Taking our cue from functional measurement (Anderson, 1973), we varied each component of speech rate separately, in a factorial design, while listeners gave estimates of the rates of the resulting passages. This technique allows us to determine subjective scales for each of the components and whether these components combine additively or multiplicatively, but it does not assign a numerical weight to the subjective scale value of each component. To obtain the weights, we regressed the subjective scale values of the passages on the listeners' estimates of the global rates of those passages. The result is a numerical model of information integration for the particular case of the perception of speech rate.

Method

Subjects

Subjects were 20 undergraduate students, native speakers of English with no reported speech or hearing defects, who served individually in sessions lasting 1 hour.

Materials

A female announcer, seated in a sound-treated room, read the "pop fan" passage (Lane & Grosjean, 1973) at a wide variety of rates. From the corpus thus recorded (Nagra IV tape recorder), 27 reading rates, ranging from 97 to 224 words/min, were obtained by selection and tape splicing. Each rate corresponded to a combination of one of three levels of articulation rate—4.0, 4.9, and 5.8 syllables/sec (±1.4%); of number of pauses—17, 6, and 3; and of mean length of pauses—.29, .53, and .76 sec (±1.4%). These values were chosen on the basis of three prior studies (Grosjean, 1972; Grosjean & Deschamps, 1975; Lane & Grosjean, 1973) that analyzed the temporal variables in readout and spontaneous speech. The mid value of each variable corresponded to the mean obtained across these studies, while the lower and upper values matched levels ±2 standard deviations from that mean. The experimental passage (also employed in Grosjean & Lane, 1974; Lane & Grosjean, 1973) contained the following 51 words of text (including contractions), which comprised 75 syllables (the number at each pause emplacement shows the lowest pause frequency that entailed pausing there; all of the higher pause frequencies also used that emplacement):

> as far as I know /17/ I'm a fairly normal /17/ fifteen year old /3/ neither a complete /17/ psychological case /6/ nor a cut /17/ above the others /3/ I listen /17/ to Radio Luxembourg /6/ my hair falls forward /17/ in the fashionable style /6/ and I wear /17/ polo neck /17/ sweaters /3/ but I don't /17/ consider myself /17/ a great /17/ pop fan.

The hierarchy of pause emplacement is that obtained in the autophonic experiment of Lane and Grosjean (1973). When obtaining the appropriate mean pause lengths by tape splicing, care was taken to respect the rank ordering of durations of each pause emplacement: The pause at the end of a sentence was made to last longer than that after a clause, which in turn was longer than that after a noun or verb phrase. Consequently, the mean pause durations cited above reflect varying ranges of pause length; for example, pause durations ranged from .68 to 1.04 sec with a mean of .76 sec for the passage read with a mid value of both articulation rate and number of pauses and a high value of the length of the pauses.

Procedure

The method of magnitude estimation (Stevens, 1956) was used to measure the perception of speech rate. Subjects were presented with the stimuli through binaural headsets (Telephonics TDH-39) supplied by a studio tape recorder (Crown SS 800). The experimenter first played a recording of the passage read with a mid value of the three variables (4.9 syllables/sec, 6 pauses, and .53 sec/pause). He assigned the numerical value 10 to this standard rate (167 words/min). Next he played 81 recordings of the passage (each of the 27 rates presented three times in random order), and the listeners assigned to each a number proportional to its apparent rate. The subjects reported their estimates on a teletypewriter terminal of a laboratory computer (Digital Equipment PDP-11).

Results and Discussion

An analysis of the variance in subjects' estimates showed significant main effects

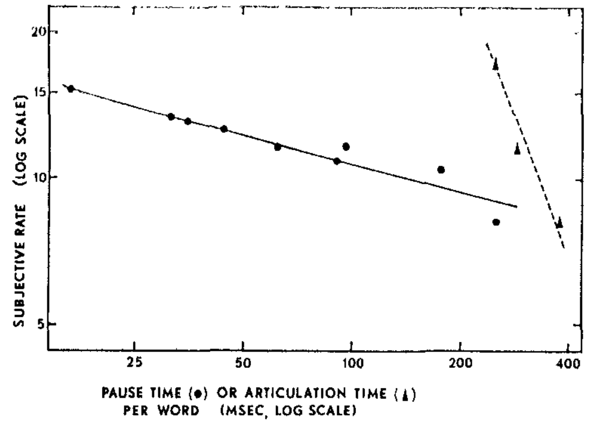

FIGURE 1. Growth of apparent reading rate as a function of the speaker's pause time and articulation time per word. (Twenty-seven reading rates—comprised of 3 levels each of articulation rate, pause frequency, and pause length—were estimated 3 times each by 20 subjects. Data points are arithmetic means; lines are fitted by the least squares method.)

for articulation rate, $F(2, 38) = 106$, $p < .01$; number of pauses, $F(2, 38) = 18$, $p < .01$; and length of pauses, $F(2, 38) = 43$, $p < .01$. There was a significant Number of Pauses × Length of Pauses interaction, $F(4, 76) = 7.2$, $p < .01$, but articulation rate did not interact significantly with either of the other two variables in determining estimates of reading rate. It seems likely, however, that articulation and pausing have independent effects only when extreme values of both are avoided. In the present study, the three levels of each variable were chosen at the mean and ±2 standard deviation values in a large corpus of spoken English. In our prior study (Grosjean & Lane, 1974), in which five levels of both articulation rate and pause frequency were used, we pushed the ranges to the speaker's limits (±3 and 5 standard deviations, respectively) and obtained a significant Articulation Rate × Pause Frequency interaction.

We want to choose a simple algebraic model that describes how our subjects integrated the three kinds of information in order to arrive at an overall impression of reading rate. Anderson (1973) has discussed several algebraic models in perception, including the linear, the multiplicative, and the multilinear. The latter type is consistent with the present results. Specifically,

$$E = W_1 A^m + W_2 (N \cdot \bar{L})^n, \qquad (1)$$

where A = the average phonation time per word, and $N \cdot \bar{L}$ = the average pause time per word (P). Equation 1 states that rate estimates are the sum of the weighted subjective scale values of articulation rate and pause rate (neglecting constants representing response variability and the intercept). The simple multiplicative interaction of pause number and length is an approximation to the present results, since only 31% of their interaction sum of squares was associated with the bilinear component. Once articulation rate and pause rate have been converted to their subjective scale values, the fit of the model can be computed by means of multiple regression.

Accordingly, Figure 1 presents subjective rate as a function of the physical articulation and pause rates. In the present study these ranged from .365 to .251 and from .253 to .017 sec/word, respectively. At the lowest reading rate, then, it took well over .5 sec, including phonation and pause, to produce a word; at the highest, it took about .25 sec [words/min = $60/(A + P)$]. The straight lines in these log–log coordinates, fitted by the method of least squares, have slopes of −2 and −.2, respectively. The former slope is somewhat steeper than the value of 1.5 obtained in our prior study with a larger range of articulation rates; the latter slope is the same as that reported previously. There is roughly an order of magnitude difference in the rate of growth of sensation on the two scales. With these exponents, the 27 values of A and P were converted to their subjective scale values and regressed on mean estimates, following Equation 1. The multiple correlation was .98.

The relation between the rate estimates predicted by the multilinear model and those obtained in the present study is shown by the first plot in Figure 2. The prediction equation was as follows:

$$E' = A^{-2} + 6P^{-.2} - 10, \qquad (2)$$

where A and P are given in sec/word. The linear relation between predicted and ob-

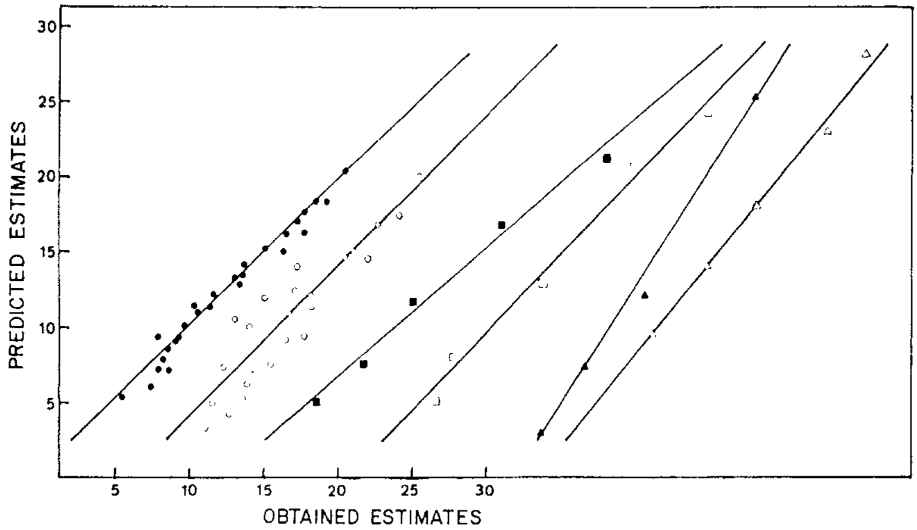

FIGURE 2. Relation between rate estimates predicted by a multilinear model and those obtained in six studies of apparent reading rate. (For clarity, successive curves have been translated by 0, 5, 10, 15, 25, and 30 units along the x axis.)

tained estimates (least squares slope .96) testifies that the model well fits the data on which it is based. Of greater interest is the generality of the model, shown here by using Equation 2 to fit the outcome of five published studies of apparent rate, which used a variety of stimulus arrays, languages, and procedures. The second study represented (Grosjean & Lane, 1974) also employed a factorial design, but length of the pauses was held constant, and there were five levels of the remaining two variables. Again, the estimates predicted by the model are a linear function of those actually obtained (each point is the mean of four estimates by each of 17 listeners), although there is somewhat more variability in prediction. The estimates obtained in the third study represented (Grosjean & Lass, in press), where articulation and pause rate covaried normally in five rates of reading, grew somewhat more rapidly than those predicted; the disparity was probably due to the rather small range of stimuli employed (96 to 230 words/min). The fourth set of data was obtained by the same investigators with a different passage and slightly broader range of rates (in both cases, 16 listeners gave four estimates of each rate). The multilinear model fits the outcome well. The result is less satisfactory for the fifth study (Grosjean & Lane, 1974), quite similar to the third but presenting a very wide range of rates (92–360 words/min; 8 listeners gave five estimates of each rate). Finally, a study (Grosjean & Lass, in press) conducted in French (five rates from 116 to 271 words/min, four replications, 16 listeners) yielded rate estimates that are in accord with the model. The median prediction slope in Figure 2 is 1.0, and the median correlation between predicted and obtained estimates is .99. Very small stimulus ranges tend to flatten the prediction slope; very broad ones tend to steepen it. We conclude that Equation 2, which made these predictions, provides a rather general account of how listeners process the three components of speech rate in order to arrive at an overall impression of reading rate.

Equation 2 describes a trading relation, in controlling the listener's impression of speech rate, between the time spent articulating and the time spent pausing. The present factorial design, in which these two components were decoupled from their natural covariation, allows us to determine iso-rate contours empirically; they are plotted in Figure 3. As was already apparent from Equation 2, a much larger change in pause rate is required to offset a change in articulation rate. As we pointed out in our earlier article (Grosjean & Lane, 1974), this undoubtedly reflects the considerable difference in the operating ranges

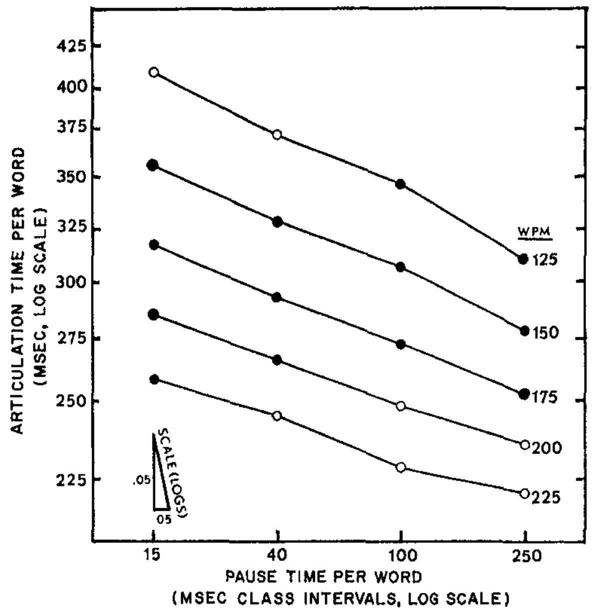

FIGURE 3. Iso-rate contours. (The unfilled points were obtained by extrapolation from families of curves obtained in this experiment; WPM = words per minute.)

of these two parameters of speech rate. When articulation rate spans 4 standard deviations (from the low to the high values in this experiment), the physical increase is only 1.5-fold, whereas the comparable standardized range for pause rate corresponds to a 15-fold increase in sec/word. The physical differences are reduced somewhat by the perceptual process: According to Equation 2, the corresponding ranges in perceived rate are 1.7:1 and 1.4:1, respectively; nevertheless, articulation rate remains the more influential variable.

When articulation and pause rate both span a dynamic range of 4 standard deviations, so that the global rate varies from 97 to 224 words/min, the increase in apparent rate is 3.6-fold, according to the model. In a word, the range of speaking rates we encounter is perceptually rather small, roughly 4:1. Quite the same prediction of apparent range (3.8:1) comes from the original subjective scale simply based on the global rate in words/min (raised to the 1.6 power). The multilinear model is preferable because it reflects how the listener integrates the independent components of speaking rate.

REFERENCE NOTE

1. Foulke, E. *Comprehension of rapid speech by the blind: Part III.* (Final Progress Report on Cooperative Research Project 2430). Louisville, Ky.: University of Louisville, 1963.

REFERENCES

Anderson, N. Algebraic models in perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception.* New York: Academic Press, 1973.

Cain, C. J., & Lass, N. J. Listening rate preferences of adults. In S. Duker (Ed.), *Time-compressed speech.* Metuchen, N. J.: Scarecrow Press, 1974.

Cartwright, L. R. J., & Lass, N. J. A psychophysical study of rate of continuous speech stimuli by means of direct magnitude estimation scaling. *Language and Speech*, 1976, *18*, 358–365.

Clevenger, T., & Clark, M. L. Coincidental variation as a source of confusion in the experimental study of rate. *Language and Speech*, 1963, *6*, 144–149.

Cotton, J. C. Syllabic rate: A new concept in the study of speech rate variation. *Speech Monographs*, 1936, *2*, 112–117.

Foulke, E., & Sticht, T. G. Listening rate preferences of college students for literary material of moderate difficulty. *Journal of Auditory Research*, 1966, *6*, 397–401.

Goldman-Eisler, F. *Psycholinguistics: Experiments in spontaneous speech.* London: Academic Press, 1968.

Grosjean, F. *Le rôle joué par trois variables temporelles dans la compréhension orale de l'anglais, étudié comme seconde langue.* Unpublished doctoral dissertation, Université de Paris VII, 1972.

Grosjean, F., & Deschamps, A. Analyse contrastive des variables temporelles de l'anglais et du français: Vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*, 1975, *31*, 144–184.

Grosjean, F., & Lane, H. Effects of two temporal variables on the listener's perception of reading rate. *Journal of Experimental Psychology*, 1974, *102*, 893–896.

Grosjean, F., & Lass, N. Some factors affecting the listener's perception of reading rate in English and French. *Language and Speech*, in press.

Hutton, C. L. *A psychophysical study of speech rate.* Unpublished doctoral dissertation, University of Illinois, 1954.

Kelly, J. C., & Steer, M. D. Revised concept of rate. *Journal of Speech and Hearing Disorders*, 1949, *14*(3), 222–227.

Lane, H., & Grosjean, F. Perception of reading rate by listeners and speakers. *Journal of Experimental Psychology*, 1973, *97*, 141–147.

Lass, N. J., & Goff, V. L. Consistency of speech rate evaluations of experienced and inexperienced listeners. In S. Duker (Ed.), *Time-compressed speech*. Metuchen, N. J.: Scarecrow Press, 1974.

Orr, D. B., Friedman, H. L., & Williams, J. C. Trainability of listening comprehension of speeded discourse. In S. Duker (Ed.), *Time-compressed speech*. Metuchen, N. J.: Scarecrow Press, 1974.

Stevens, S. S. The direct estimation of sensory magnitudes: Loudness. *American Journal of Psychology*, 1956, *69*, 1–25.

(Received January 6, 1976)