

# Mathematical treatment of context effects in phoneme and word recognition

Arthur Boothroyd

Graduate School, City University of New York, 33 West 42nd Street, New York, New York 10036

Susan Nittrouer

Boy's Town National Institute, 555 North 30th Street, Omaha, Nebraska 68131

(Received 20 May 1987; accepted for publication 8 March 1988)

Percent recognition of phonemes and whole syllables, measured in both consonant–vowel–consonant (CVC) words and CVC nonsense syllables, is reported for normal young adults listening at four signal-to-noise (S/N) ratios. Similar data are reported for the recognition of words and whole sentences in three types of sentence: high predictability (HP) sentences, with both semantic and syntactic constraints; low predictability (LP) sentences, with primarily syntactic constraints; and zero predictability (ZP) sentences, with neither semantic nor syntactic constraints. The probability of recognition of speech units in context ( $p_c$ ) is shown to be related to the probability of recognition without context ( $p_i$ ) by the equation  $p_c = 1 - (1 - p_i)^k$ , where  $k$  is a constant. The factor  $k$  is interpreted as the amount by which the channels of statistically independent information are effectively multiplied when contextual constraints are added. Empirical values of  $k$  are approximately 1.3 and 2.7 for word and sentence context, respectively. In a second analysis, the probability of recognition of wholes ( $p_w$ ) is shown to be related to the probability of recognition of the constituent parts ( $p_p$ ) by the equation  $p_w = p_p^j$ , where  $j$  represents the effective number of statistically independent parts within a whole. The empirically determined mean values of  $j$  for nonsense materials are not significantly different from the number of parts in a whole, as predicted by the underlying theory. In CVC words, the value of  $j$  is constant at approximately 2.5. In the four-word HP sentences, it falls from approximately 2.5 to approximately 1.6 as the inherent recognition probability for words falls from 100% to 0%, demonstrating an increasing tendency to perceive HP sentences either as wholes, or not at all, as S/N ratio deteriorates.

PACS numbers: 43.71.Gv, 43.71.Es

## INTRODUCTION

The accuracy of speech recognition depends not only on the sensory data generated from the stimulus itself, but also on the context within which the stimulus occurs. Under difficult listening conditions, for example, words are recognized more easily when they are presented in sentences rather than in isolation or in carrier phrases. This phenomenon was first investigated in depth by Miller *et al.* (1951), who found that, at certain signal-to-noise (S/N) ratios, the accuracy of recognition of words increased by as much as 30 percentage points when sentence context was added. More recently it has been shown that the effect of sentence context can be controlled by changing the extent to which the specific context limits the number of semantically plausible alternatives (Giolas *et al.*, 1970; Duffy and Giolas, 1974; Kalikow *et al.*, 1977). This phenomenon provides the basis of the speech in noise (SPIN) test, developed by Kalikow *et al.* for clinical evaluation of a subject's ability to take advantage of sentence context during word recognition tasks.

When word recognition is measured in a closed-set task, the set itself provides contextual information. In their 1951 study, Miller *et al.* showed that the S/N ratio for 50% recognition (after correction for random guessing) varied from –14 dB, for a set size of two, to –4 dB, for a set size of 256.

Extrapolation of their data suggests that open-set word recognition gives scores that are equivalent to closed-set identification for a set size of approximately 30 000 words.

Another example of a context effect is the improvement of word recognition in sentences that occurs when the perceiver is given prior knowledge of sentence topic (a situation that exists in most conversations). This effect has been demonstrated for lipreading by Hnath *et al.* (1985).

A more subtle kind of context is that provided by the properties of the lexicon from which words are drawn. It has been shown, for example, that real words, presented in isolation, are more easily recognized than are nonsense syllables and that words with a high frequency of occurrence are more easily recognized than are words with a low frequency of occurrence (Giolas and Epstein, 1963; Hirsch *et al.*, 1954; Howes, 1957; Owens, 1961; Pollack *et al.*, 1959; Rosenzweig and Postman, 1957; Savin, 1963; Schultz, 1964).

What all of these contexts have in common is that they influence the *a priori* probability for the stimulus pattern being presented. When the *a priori* probability is increased, the probability of correct recognition also increases. Exactly how the perceiver uses knowledge of *a priori* probability is not clear. Attempts to explain the phenomenon in terms of more effective guessing have been unable to account for the magnitude of the effect. A more successful approach is that

of Broadbent (1967) who, in investigating the word-frequency effect, invoked a signal detection model. He showed that the observed data could be explained in terms of a biasing of acceptance criteria in favor of words with a higher *a priori* probability. More recently, Massaro and Oden (1980) have developed what they call a “fuzzy logical model” to account for the combined effect of two or more clues to speech recognition.

One of the difficulties encountered in studying context effects is their quantification. Two methods have commonly been used. The most obvious is to calculate the difference in percent correct recognition under two conditions, one with the source of contextual information present and the other with it absent. Unfortunately, this method is sensitive to “ceiling” and “floor” effects. If the score with the context absent is close to 100%, then there is very little room for improvement when context is introduced. If the score with context present is close to 0%, then there is little opportunity for the score to fall when context is removed. To a certain extent, the ceiling and floor effects can be overcome with the use of masking, or other degradation, to ensure that scores remain in a range where context effects can be observed. This is the approach taken in the SPIN test of Kalikow *et al.* (1977) that was mentioned earlier. Even with this approach, however, the magnitude of the difference may not accurately reflect the magnitude of the context effect. It cannot be assumed, for example, that an increase from 20%–30% is equivalent to an increase from 70%–80%, in terms of the use being made of contextual information.

The other common approach to quantification is to determine the increase of S/N ratio required to keep recognition probability at 50% when contextual information is removed. When comparing the recognition of familiar and unfamiliar words, for example, Rosenzweig and Postman (1957) showed that an increase of 4 dB in S/N ratio was needed for each tenfold decrease in frequency of word occurrence. This procedure avoids ceiling and floor effects. It also permits the comparison of data obtained with different types of material. The results, however, cannot easily be incorporated into more extensive mathematical models. Moreover, the method does not lend itself easily to the investigation of context effects in nonauditory speech perception — by lip-reading, for example.

A third approach has been used recently by Bilger (1984). This involves measuring the probabilities of recognition with and without context, at a single S/N ratio, and calculating the difference between their *z* transforms. This approach helps to reduce the ceiling and floor effects associated with percent correct scores but the resulting metric is not easily incorporated into more extensive mathematical treatments. Furthermore, it does not necessarily reflect the true magnitude of the context effect. The difference between the *z* transforms for 20% and 30%, for example, is the same as that for 70% and 80%. There is no reason to assume, however, that the relative contributions of stimulus and context to these differences are identical.

An alternative approach to the quantification of context effects, based on the predictions of simple probability theory, has been proposed by Boothroyd (1978, 1985) (see, also,

Schiavetti *et al.*, 1984). In this approach, two equations are developed. The first relates the probabilities of recognition of speech units with and without context. The assumption is made that the effect of context is quantitatively equivalent to adding statistically independent channels of sensory data to those already available from the speech units themselves. Under this assumption, it can be shown that the logarithms of the error probabilities for contextual and intrinsic channels are additive. Thus

$$\log(1 - p_c) = \log(1 - p_i) + \log(1 - p_x), \quad (1)$$

where  $p_c$  is the probability of recognition of a speech unit in context,  $p_i$  is the probability of recognition without context (i.e., in isolation or in nonsense material), and  $p_x$  is the probability of recognition from context effects alone. Since the speech unit and its context must be perceived under the same constraints (e.g., distortion or masking), it is further assumed that  $\log(1 - p_x)$  is proportional to  $\log(1 - p_i)$ , in which case Eq. (1) reduces to

$$\log(1 - p_c) = k \log(1 - p_i), \quad (2)$$

where  $k$  is a constant. From this we derive

$$p_c = 1 - (1 - p_i)^k \quad (3)$$

and

$$k = \log(1 - p_c) / \log(1 - p_i). \quad (4)$$

In applying Eq. (4) to the data of Miller *et al.* (1951), Boothroyd (1978) found that the value of  $k$  was fairly constant over a wide range of values of  $p_i$ . Similar results have been reported in studies of contextual effects on the intelligibility of the speech of deaf subjects (Boothroyd, 1985; Schiavetti *et al.*, 1984). These findings suggest that, by expressing the effect of context as the ratio of the logarithms of the error probabilities for the context and no-context situations, we obtain a dimensionless quantity that represents the magnitude of the context effect and that is independent of the degree and type of degradation in the speech signal.

The second equation relates the recognition probabilities of wholes to the probabilities for the parts of which the wholes are constructed. If we first assume that the recognition of a whole requires the recognition of several parts, and that the probabilities of recognition of the parts are equal and statistically independent, then

$$p_w = p_p^n, \quad (5)$$

where  $p_w$  is the probability of recognition of a whole,  $p_p$  is the probability of recognition of a part, and  $n$  is the number of parts in a whole. We allow for violation of the assumption of statistical independence by assuming that the exponent of Eq. (5) is lowered (Fletcher, 1953). Thus

$$p_w = p_p^j, \quad (6)$$

where  $1 \leq j \leq n$ . A value of  $j = n$  implies that the recognition of any part is unchanged by recognition of other parts. A value of  $j = 1$  implies that recognition of any part is sufficient to ensure recognition of the other parts and, therefore, of the whole. From Eq. (6), we obtain

$$j = \log(p_w) / \log(p_p). \quad (7)$$

Application of Eq. (7) to the intelligibility scores for sam-

ples of the speech of deaf children gave a value of  $j = 2.4$  for the recognition of phonemes in real CVC words. This value was constant across a wide range of intelligibilities (Boothroyd, 1985). If similar findings are obtained for other types of material and other forms of stimulus degradation, the factor  $j$  would provide another dimensionless quantity reflecting the magnitude of context effects.

To summarize: It is suggested that the effect of linguistic context on the recognition of speech units can be expressed by two factors. The  $k$  factor is derived from measurements of recognition probability for units presented with and without context. Here,  $k$  is the ratio of the logarithms of the two error probabilities. The value of  $k$  should be unity when context has no effect. It should increase monotonically as the contribution of context increases. The  $j$  factor is derived from measurements of recognition probabilities for wholes and for the parts (units) within the wholes. Here,  $j$  is the ratio of the logarithms of the two recognition probabilities. The value of  $j$  should equal  $n$  (where  $n$  is the number of parts in a whole) when context has no effect. It should decrease monotonically as the contribution of context increases, approaching a limiting value of unity. In the derivations of the  $k$  and  $j$  factors it was assumed that they are constant, i.e., that they are not affected by the underlying recognition probabilities for the no-context situation. This assumption is open to empirical test.

The goals of the studies to be described were: (a) to measure and compare the effects of lexical, syntactic, and semantic context on the perception of speech in noise; (b) to obtain values of the  $k$  and  $j$  factors for these effects; (c) to test the hypothesis that these factors are constant across a range of recognition probabilities; and (d) to test the hypothesis that, for nonsense syllables and nonsense sentences,  $j$  equals the number of parts in a whole.

## I. EXPERIMENT 1: PHONEME AND WORD RECOGNITION IN WORDS AND NONSENSE SYLLABLES

### A. Purpose

The purpose of the first experiment was to measure the percent correct recognition of phonemes and whole syllables in consonant-vowel-consonant (CVC) syllables at several S/N ratios, and to compare the scores obtained from meaningful CVCs (i.e., monosyllabic words) with those obtained from nonsense syllables.

### B. Subjects

Thirty-two normally hearing young adults served as listeners. All were native talkers of English and had no known history of speech or hearing problems. Twenty-four of the subjects were students in Masters programs in speech and hearing. The remaining eight subjects were staff and students in the Speech and Hearing Sciences Program of the City University of New York Graduate Center.

### C. Test material

Phonetically balanced lists of CVC syllables were developed for this study. Their structure follows that of the short,

isophonemic word lists developed by Boothroyd (1968a, 1984) for clinical speech audiometry. That is, each list consists of ten syllables constructed from the same pool of ten initial consonants, ten vowels, and ten final consonants. Because of differences of acoustic structure and phonotactic constraints, choices for initial and final consonants were made separately. Twelve of the lists were of meaningful words that appeared in the Thorndike-Lorge (1944) word count. The words covered a wide range of frequency of occurrence: 40% occurred 100 or more times per million; 40% occurred between 11 and 99 times per million; 17% occurred between 2 and 10 times per million; and 3% occurred 1 or less times per million. The other 12 lists were of nonsense syllables. They did not appear in the Thorndike-Lorge count and were not known to be meaningful words in British or American English by either experimenter. These lists are presented in Appendix A.

### D. Test recordings

The 24 CVC lists were recorded by an adult male talker with pronunciation typical of the Middle Atlantic region. They were spoken in the carrier phrase "you will write ... please." The talker monitored levels with a VU meter and was asked to maintain a constant level.

The master tape recording was analyzed using a graphic level recorder and the level of each carrier phrase, with its test word, was adjusted during preparation of a submaster so that peak levels during the word "write" varied by no more than  $\pm 1$  dB from item to item. In this way, any variations of average effort were corrected, but the test words, and their constituent phonemes, preserved their natural intensity relationships. Six copies were prepared in which the orders of the lists were randomized.

An FFT analyzer was used to prepare a long-term spectrum of the speech material, which was then matched using white noise passed through a  $\frac{1}{3}$ -oct graphic equalizer. The spectrally shaped noise, which was intended to have an equal masking effect at all frequencies, was mixed with each of the six randomizations of the test material, but at a different S/N ratio for each randomization. The S/N ratios, measured as the difference between the average level of the speech peaks and the average level of the noise, were  $-9$ ,  $-6$ ,  $-3$ ,  $0$ ,  $+3$ , and  $+6$  dB. On the basis of the results of a pilot study, the four ratios from  $-6$  through  $+3$  dB were selected for the experiment.

### E. Procedure

The test tapes were played on a UHER portable reel-to-reel tape recorder and groups of subjects listened diotically under Sennheiser headphones connected to the tape recorder via an eight station distribution box. Each subject was instructed to select a comfortable listening level during presentation of practice material.

Subjects were divided into four groups of eight. Each group listened to the test material at only one S/N ratio. Subjects were told in writing that they would be listening to lists of real words and nonsense syllables, presented in a carrier phrase, and they were instructed to write down as much

TABLE I. Proportion of phonemes and whole syllables correctly recognized in the contexts of nonsense syllables and real words by groups of eight subjects listening at four different S/N ratios. Each phoneme score is based on 360 items. Each syllable score is based on 120 items. Note that a different group of subjects listened at each S/N ratio.

| S/N ratio | Context  | Scoring unit | Subject |       |       |       |       |       |       |       | Mean         | s.d.  |
|-----------|----------|--------------|---------|-------|-------|-------|-------|-------|-------|-------|--------------|-------|
|           |          |              | A       | B     | C     | D     | E     | F     | G     | H     |              |       |
| - 6 dB    | Nonsense | Phoneme      | 0.414   | 0.492 | 0.495 | 0.443 | 0.406 | 0.486 | 0.420 | 0.341 | <b>0.437</b> | 0.053 |
|           |          | Syllable     | 0.075   | 0.117 | 0.100 | 0.042 | 0.042 | 0.092 | 0.100 | 0.075 | <b>0.080</b> | 0.027 |
|           | Word     | Phoneme      | 0.533   | 0.551 | 0.553 | 0.522 | 0.542 | 0.575 | 0.528 | 0.440 | <b>0.531</b> | 0.040 |
|           |          | Syllable     | 0.200   | 0.275 | 0.192 | 0.217 | 0.233 | 0.250 | 0.217 | 0.167 | <b>0.219</b> | 0.034 |
| - 3 dB    | Nonsense | Phoneme      | 0.647   | 0.571 | 0.603 | 0.608 | 0.556 | 0.585 | 0.573 | 0.587 | <b>0.591</b> | 0.028 |
|           |          | Syllable     | 0.233   | 0.125 | 0.183 | 0.242 | 0.142 | 0.200 | 0.192 | 0.175 | <b>0.187</b> | 0.040 |
|           | Word     | Phoneme      | 0.670   | 0.658 | 0.682 | 0.673 | 0.673 | 0.710 | 0.699 | 0.647 | <b>0.677</b> | 0.020 |
|           |          | Syllable     | 0.358   | 0.367 | 0.392 | 0.358 | 0.392 | 0.425 | 0.412 | 0.342 | <b>0.381</b> | 0.029 |
| 0 dB      | Nonsense | Phoneme      | 0.696   | 0.718 | 0.753 | 0.647 | 0.675 | 0.617 | 0.688 | 0.705 | <b>0.687</b> | 0.042 |
|           |          | Syllable     | 0.308   | 0.392 | 0.458 | 0.300 | 0.258 | 0.125 | 0.317 | 0.317 | <b>0.309</b> | 0.097 |
|           | Word     | Phoneme      | 0.886   | 0.772 | 0.806 | 0.803 | 0.756 | 0.709 | 0.803 | 0.799 | <b>0.789</b> | 0.045 |
|           |          | Syllable     | 0.692   | 0.550 | 0.600 | 0.625 | 0.542 | 0.467 | 0.533 | 0.608 | <b>0.577</b> | 0.069 |
| + 3 dB    | Nonsense | Phoneme      | 0.865   | 0.852 | 0.888 | 0.821 | 0.851 | 0.799 | 0.868 | 0.863 | <b>0.851</b> | 0.028 |
|           |          | Syllable     | 0.692   | 0.633 | 0.717 | 0.608 | 0.608 | 0.492 | 0.658 | 0.625 | <b>0.629</b> | 0.068 |
|           | Word     | Phoneme      | 0.870   | 0.907 | 0.927 | 0.925 | 0.930 | 0.933 | 0.908 | 0.944 | <b>0.918</b> | 0.023 |
|           |          | Syllable     | 0.708   | 0.767 | 0.825 | 0.850 | 0.825 | 0.825 | 0.792 | 0.858 | <b>0.806</b> | 0.049 |

of each word as they could, guessing if necessary. Those listeners who were not familiar with a phonetic notation were asked to spell nonsense syllables as well as they could, using traditional orthography. Subjects were informed at the beginning of each list whether it would consist of real words or nonsense syllables.

## F. Results

### 1. Percent correct

Each subject provided four scores: (i) the proportion of phonemes correctly recognized in nonsense syllables; (ii) the proportion of phonemes correctly recognized in real words; (iii) the proportion of whole nonsense syllables correctly recognized; and (iv) the proportion of whole words

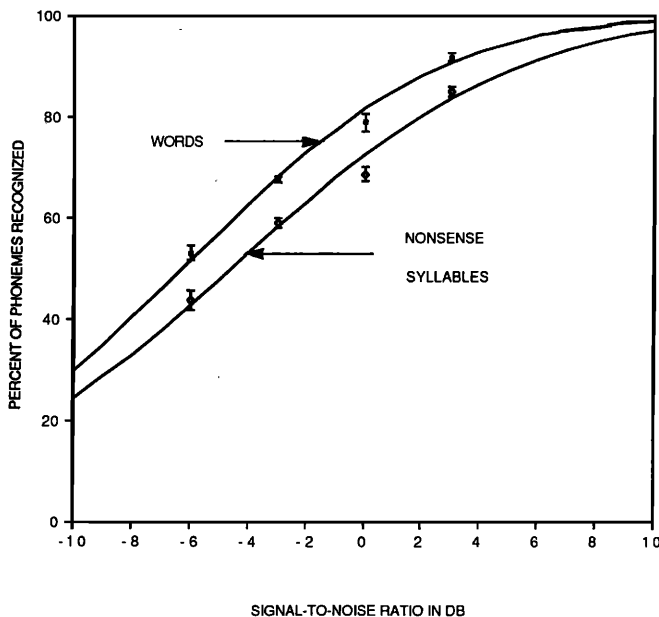


FIG. 1. Percent phonemes correctly recognized by normal listeners in meaningful CVC words and nonsense CVC syllables, as a function of S/N ratio. Data points are means ( $\pm 1$  standard error) for groups of eight subjects. Curves are best fitting normal ogives (i.e., linear least-squares fits to z transforms of recognition probabilities).

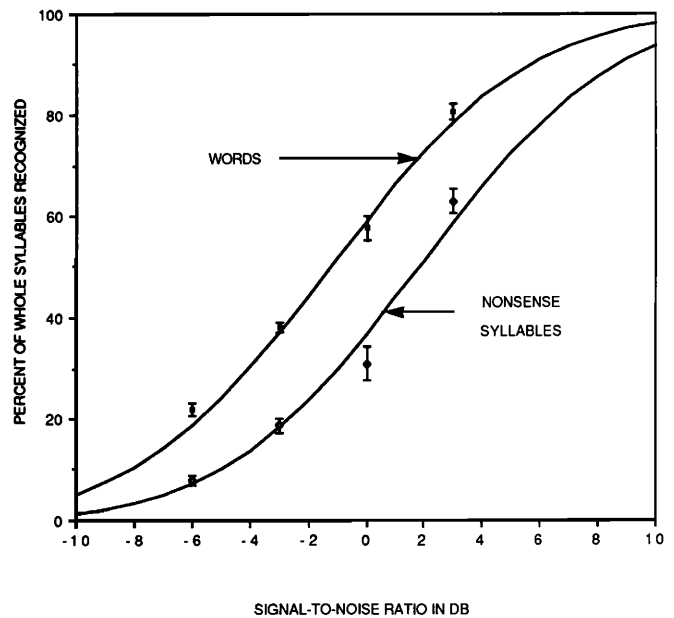


FIG. 2. Percent whole CVC words and whole CVC nonsense syllables recognized by normal listeners as a function of S/N ratio (see Fig. 1 for details).

correctly recognized. Each phoneme score was based on 360 items. Each syllable or word score was based on 120 items. Table I shows the resulting scores together with the group means. The mean data are illustrated in Figs. 1 and 2. The curves in these figures show the best fitting normal ogives, i.e., they are the least-squares linear fits to the  $z$  transforms of the group means of the percent recognition scores. It will be seen from Fig. 1 that the probability of phoneme recognition in words is about 10 percentage points higher than it is in nonsense syllables over the range of S/N ratios used here. It will be seen from Fig. 2 that the probability of recognition of whole syllables is some 20–30 percentage points higher for words than for nonsense syllables.

## 2. Recognition of phonemes with and without lexical context

Individual scores for the recognition of phonemes in nonsense syllables and in real words were converted to  $k$  values using Eq. (4). No significant correlation was found between the value of  $k$  and the probability of recognition for phonemes in nonsense syllables ( $r[30] = -0.058$ ,  $p > 0.05$ ). The mean value of  $k$  is 1.32 with a standard error of 0.03 and 95% confidence limits of  $\pm 0.06$ . By substituting this value in Eq. (3), we obtain the predicted relationship between recognition probability for phonemes in nonsense syllables and that for phonemes in meaningful words. This relationship is illustrated in Fig. 3, together with individual data points from Table I. The coefficient of correlation between measured scores for words and those predicted from the scores for nonsense syllables is 0.969. The standard deviation of the difference between measured and predicted scores is 3.7 percentage points.

## 3. Recognition of wholes and of parts within wholes

Individual scores for the recognition of phonemes and whole syllables, in both nonsense syllables and CVC words, were converted to  $j$  values using Eq. (7). Because the estimates of  $j$  and  $k$  are highly sensitive to measurement errors at

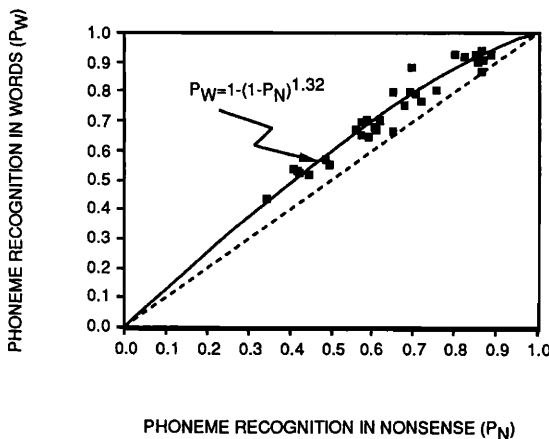


FIG. 3. Recognition probability for phonemes in CVC words as a function of recognition probability for phonemes in CVC nonsense syllables. Data points are individual scores from Table I. The curve shows the prediction of Eq. (3).

very low and very high probabilities, subjects were excluded from this and subsequent analyses if either of the two probabilities was less than 0.05 or greater than 0.95. Using these criteria, 30 subjects provided values of  $j$  for nonsense syllables and 32 subjects provided values of  $j$  for meaningful words.

a. *Nonsense syllables.* No significant correlation was found between the value of  $j$  and phoneme recognition probability for nonsense syllables ( $r[28] = -0.102$ ,  $p > 0.05$ ). The mean value of  $j$  is 3.07, with 95% confidence limits of  $\pm 0.14$ . This value is not significantly different from the value of 3.0 that is predicted if the recognition of a whole syllable requires independent recognition of the three constituent phonemes. By substituting the mean value in Eq. (6), we obtain the predicted relationship between recognition probability for phonemes and that for whole nonsense syllables. This relationship is illustrated in Fig. 4, together with individual data points from Table I. The coefficients of correlation between measured scores for whole nonsense syllables and those predicted from the phoneme scores is 0.989. The standard deviation of the difference between measured and predicted scores is 3.4 percentage points.

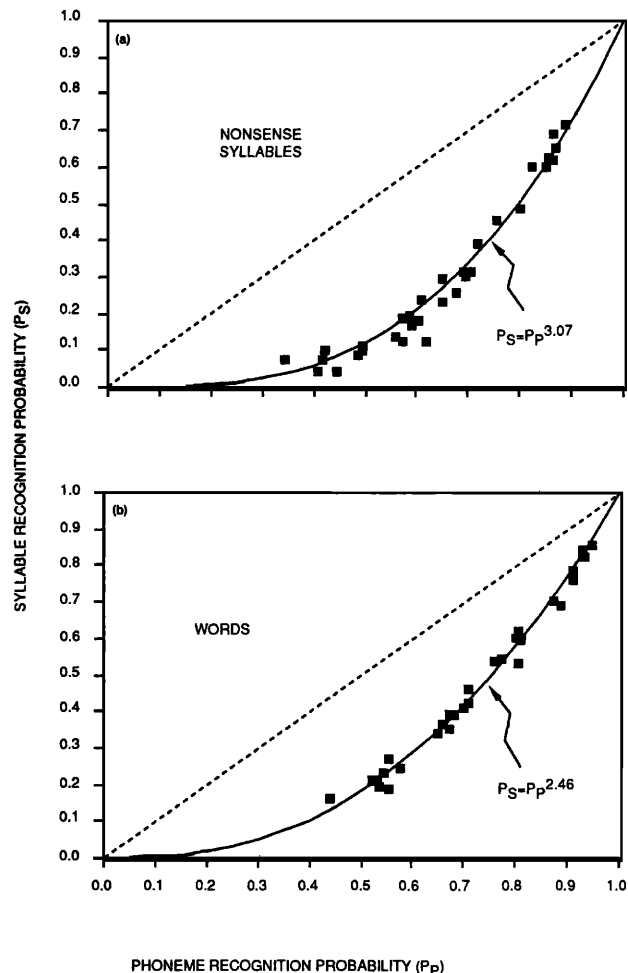


FIG. 4. Recognition probability for whole syllables ( $p_s$ ) as a function of recognition probability for constituent phonemes ( $p_p$ ) in: (a) CVC nonsense syllables and (b) CVC words. Data points are individual scores from Table I. Curves show the predictions of Eq. (6).

*b. Meaningful words.* No significant correlation was found between the value of  $j$  and phoneme recognition probability for real words ( $r[30] = 0.278, p > 0.05$ ). The mean value of  $j$  is 2.46, with 95% confidence limits of  $\pm 0.08$ . This mean value is significantly less than that obtained for nonsense syllables. By substituting the mean value in Eq. (6), we obtain the predicted relationship between recognition probability for phonemes and that for whole words. This relationship is illustrated in Fig. 4, together with individual data points from Table I. The coefficient of correlation between measured scores for whole words and those predicted from the phoneme scores is 0.994. The standard deviation of the difference between measured and predicted scores is 2.5 percentage points.

## II. EXPERIMENT 2: WORD AND SENTENCE RECOGNITION IN NONSENSE AND MEANINGFUL SENTENCES

### A. Purpose

The purpose of the second experiment was to measure, at several S/N ratios, the percent correct recognition of words and of whole sentences, as a function of the syntactic and semantic constraints within the sentences.

### B. Subjects

Forty normally hearing undergraduate students served as listeners. All were native talkers of English and had no known history of speech or hearing problems.

### C. Test material

The test material consisted of sets of four-word sentences. This length was selected as being enough to permit syntax to play a role, but not so large as to tax the listeners' ability to retain and recall strings of unrelated words. All meaningful sentences were simple declaratives, and all the words were monosyllabic. No constraints were placed on the frequency of occurrence of the words.

Three types of sentence were developed: (a) zero predictability (ZP) sentences, consisting of random sequences of words (e.g., "girls white car blink"); (b) low predictability (LP) sentences, that were syntactically appropriate but semantically anomalous (e.g., "ducks eat old tape"); and (c) high predictability (HP) sentences, that were both syntactically and semantically appropriate (e.g., "most birds can fly").

Identical sentence frames were used for the HP and LP sentences. The ZP sentences were constructed from the vocabulary used in the HP and LP sentences. A total of 20 HP and 20 LP sentences was developed, their constituent words providing 40 ZP sentences.

As a check on the construction of the HP and LP sentences, they were written on cards, shuffled, and presented to each of five adults who were native speakers of American English and had no known history of speech or language disorder. The adults were asked to sort the sentences into the two categories. All five were able to do this without error. The sentence materials are presented in Appendix B.

## D. Test recordings

The 80 sentences were recorded by the talker used in experiment 1. Using the same noise as that used in experiment 1 (i.e., matched to the long term-spectrum of the speech), four test tapes were prepared, the S/N ratio varying in 3-dB steps from  $-9$  to  $0$  dB, these being selected from a wider range on the basis of a pilot study.

## E. Procedure

The equipment and procedure were basically identical to those used in experiment 1. Subjects were given 15 practice sentences, five of each type, and an additional nine sentences for the purpose of setting listening level. Test sentences were then presented in blocks of five. The orders of presentation of the blocks were randomized across groups, and subjects were told at the beginning of each block whether they would be hearing nonsense sentences or real sentences. Subjects were tested in groups of eight. Two groups listened at the  $-3$ -dB S/N ratio. One group listened at each of the other ratios.

## F. Results

### 1. Percent correct scores

Each subject provided six scores: the proportion of words correctly recognized in ZP, LP, and HP sentences, and the proportion of whole ZP, LP, and HP sentences correctly recognized. The word scores for the ZP materials were based on 160 items, while those for LP and HP materials were based on 80 items each. The sentence scores for ZP materials were based on 40 items, while those for LP and HP materials were based on only 20 items each. Table II shows these scores together with the group means for the four S/N ratios. These data are further illustrated in Figs. 5 and 6. The curves in these figures show the best fitting normal ogives, i.e., they are the least-squares linear fits to the  $z$  transforms of group means of percent recognition scores. It will be seen from Fig. 5 that the probability of word recognition in HP sentences is about 20–35 percentage points higher than in ZP sentences, over the range of S/N ratios used here. This finding is similar to that of Miller *et al.* (1951) for the difference between words in isolation and words in sentence context. It will be seen from Fig. 6 that the recognition probability for whole sentences was as much as 60 percentage points higher for HP sentences than for ZP sentences.

### 2. Recognition of words with and without sentence context

Individual scores for the recognition of words in ZP, LP, and HP sentences were converted to  $k$  values using Eq. (4). Three values were obtained: (a)  $k_S$ , being the ratio of the logarithms of the error probabilities for words in LP and ZP sentences and reflecting the contribution of syntactic constraints alone; (b)  $k_M$ , being the ratio of the logarithms of the error probabilities for words in HP and LP sentences and reflecting the contribution of meaning (i.e., semantic constraints) alone; and (c)  $k_{S+M}$ , being the ratio of the logarithms of the error probabilities for words in HP and ZP

TABLE II. Proportion of words and whole "sentences" correctly recognized in three types of sentence context by groups of eight subjects listening at four S/N ratios. ZP sentences have neither syntactic nor semantic constraints, LP sentences have only syntactic constraints, and HP sentences have both syntactic and semantic constraints. Note that a different group of subjects listened at each S/N ratio, and two groups listened at  $-3$  dB.

| S/N ratio | Context | Scoring unit | Subject |       |       |       |       |       |       |       | Mean         | s.d.  |
|-----------|---------|--------------|---------|-------|-------|-------|-------|-------|-------|-------|--------------|-------|
|           |         |              | A       | B     | C     | D     | E     | F     | G     | H     |              |       |
| -9 dB     | ZP      | Word         | 0.269   | 0.219 | 0.144 | 0.156 | 0.138 | 0.175 | 0.075 | 0.175 | <b>0.169</b> | 0.058 |
|           |         | Sentence     | 0       | 0     | 0     | 0     | 0     | 0     | 0     | 0     | <b>0.000</b> | 0.000 |
|           | LP      | Word         | 0.388   | 0.313 | 0.188 | 0.175 | 0.213 | 0.250 | 0.213 | 0.250 | <b>0.249</b> | 0.071 |
|           |         | Sentence     | 0       | 0     | 0     | 0     | 0     | 0     | 0.050 | 0.050 | <b>0.013</b> | 0.023 |
|           | HP      | Word         | 0.450   | 0.450 | 0.463 | 0.288 | 0.300 | 0.350 | 0.275 | 0.388 | <b>0.371</b> | 0.078 |
|           |         | Sentence     | 0.150   | 0.150 | 0.250 | 0.150 | 0.200 | 0.150 | 0.100 | 0.150 | <b>0.163</b> | 0.044 |
| -6 dB     | ZP      | Word         | 0.394   | 0.425 | 0.350 | 0.388 | 0.394 | 0.244 | 0.300 | 0.281 | <b>0.347</b> | 0.065 |
|           |         | Sentence     | 0.050   | 0.050 | 0.025 | 0.025 | 0     | 0     | 0.025 | 0     | <b>0.022</b> | 0.021 |
|           | LP      | Word         | 0.500   | 0.613 | 0.500 | 0.513 | 0.525 | 0.313 | 0.388 | 0.438 | <b>0.474</b> | 0.092 |
|           |         | Sentence     | 0.100   | 0.150 | 0.150 | 0.150 | 0.150 | 0.050 | 0.050 | 0.100 | <b>0.113</b> | 0.044 |
|           | HP      | Word         | 0.713   | 0.825 | 0.675 | 0.788 | 0.888 | 0.550 | 0.575 | 0.600 | <b>0.702</b> | 0.124 |
|           |         | Sentence     | 0.500   | 0.700 | 0.400 | 0.550 | 0.750 | 0.250 | 0.300 | 0.250 | <b>0.463</b> | 0.196 |
| -3 dB (a) | ZP      | Word         | 0.519   | 0.450 | 0.481 | 0.481 | 0.569 | 0.594 | 0.494 | 0.563 | <b>0.519</b> | 0.051 |
|           |         | Sentence     | 0.050   | 0.025 | 0.050 | 0.025 | 0.100 | 0.150 | 0.075 | 0.050 | <b>0.066</b> | 0.042 |
|           | LP      | Word         | 0.700   | 0.450 | 0.588 | 0.588 | 0.638 | 0.750 | 0.550 | 0.525 | <b>0.599</b> | 0.096 |
|           |         | Sentence     | 0.350   | 0.050 | 0.150 | 0.150 | 0.150 | 0.300 | 0.100 | 0.150 | <b>0.175</b> | 0.100 |
|           | HP      | Word         | 0.875   | 0.775 | 0.775 | 0.800 | 0.988 | 0.900 | 0.850 | 0.938 | <b>0.863</b> | 0.078 |
|           |         | Sentence     | 0.650   | 0.500 | 0.550 | 0.600 | 0.950 | 0.800 | 0.750 | 0.850 | <b>0.706</b> | 0.157 |
| -3 dB (b) | ZP      | Word         | 0.525   | 0.350 | 0.438 | 0.450 | 0.456 | 0.556 | 0.413 | 0.419 | <b>0.451</b> | 0.065 |
|           |         | Sentence     | 0.025   | 0     | 0     | 0.025 | 0.025 | 0.075 | 0.025 | 0.025 | <b>0.025</b> | 0.023 |
|           | LP      | Word         | 0.625   | 0.400 | 0.513 | 0.513 | 0.413 | 0.650 | 0.513 | 0.575 | <b>0.525</b> | 0.090 |
|           |         | Sentence     | 0.150   | 0.100 | 0.250 | 0.150 | 0.100 | 0.300 | 0.150 | 0.100 | <b>0.163</b> | 0.074 |
|           | HP      | Word         | 0.888   | 0.600 | 0.988 | 0.788 | 0.800 | 0.913 | 0.763 | 0.850 | <b>0.824</b> | 0.117 |
|           |         | Sentence     | 0.750   | 0.250 | 0.950 | 0.700 | 0.600 | 0.800 | 0.600 | 0.700 | <b>0.669</b> | 0.203 |
| 0 dB      | ZP      | Word         | 0.725   | 0.669 | 0.681 | 0.475 | 0.644 | 0.825 | 0.763 | 0.706 | <b>0.686</b> | 0.103 |
|           |         | Sentence     | 0.250   | 0.250 | 0.175 | 0.025 | 0.200 | 0.375 | 0.375 | 0.175 | <b>0.228</b> | 0.115 |
|           | LP      | Word         | 0.813   | 0.825 | 0.825 | 0.613 | 0.813 | 0.850 | 0.825 | 0.850 | <b>0.802</b> | 0.078 |
|           |         | Sentence     | 0.450   | 0.450 | 0.450 | 0.200 | 0.450 | 0.500 | 0.550 | 0.650 | <b>0.463</b> | 0.127 |
|           | HP      | Word         | 0.975   | 0.988 | 0.938 | 0.750 | 0.963 | 1     | 0.975 | 0.925 | <b>0.939</b> | 0.080 |
|           |         | Sentence     | 0.900   | 0.950 | 0.850 | 0.500 | 0.850 | 1     | 0.900 | 0.850 | <b>0.850</b> | 0.151 |

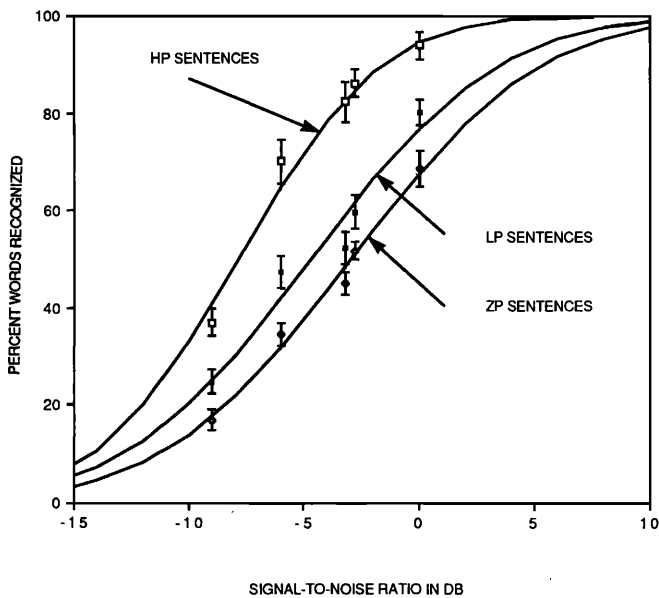


FIG. 5. Percent words correctly recognized by normal listeners in three types of sentence, as a function of S/N ratio. HP sentences provide semantic and syntactic context. LP sentences provide only syntactic context. ZP sentences provide neither. Data points are means ( $\pm 1$  standard error) for groups of eight subjects. Curves are best fitting normal ogives (i.e., linear least-squares fits to z transforms of recognition probability).

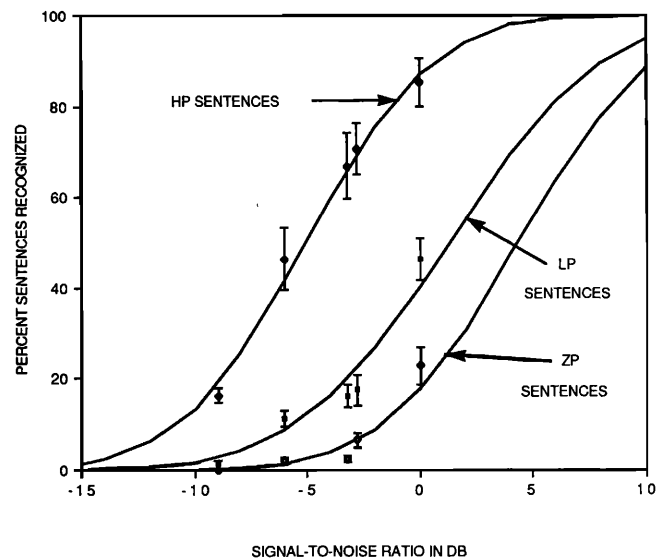


FIG. 6. Percent whole sentences correctly recognized by normal listeners in three types of sentence, as a function of S/N ratio (see Fig. 5 for details).

sentences and reflecting the combined effects of syntactic and semantic constraints. Those subjects for whom either of the two scores being compared was less than 0.05 or greater than 0.95 were excluded from the analysis. One additional subject was excluded from two of the analyses because his score on the ZP sentences and the resulting  $k$  factors were completely out of line with those of the other subjects in the  $-9$ -dB S/N ratio group. Application of these criteria resulted in 39 values of  $k_S$ , 34 values of  $k_M$ , and 33 values of  $k_{S+M}$ . No significant correlations were obtained between the  $k$  values and the corresponding word recognition probabilities ( $r[37] = -0.188$ ,  $r[32] = -0.014$ , and  $r[31] = 0.082$ , for  $k_S$ ,  $k_M$ , and  $k_{S+M}$ , respectively,  $p > 0.05$ ). Mean values and 95% confidence limits were  $k_S = 1.38 \pm 0.07$ ;  $k_M = 1.97 \pm 0.19$ ; and  $k_{S+M} = 2.72 \pm 0.19$ . By substituting the mean values in Eq. (3), we obtain the pre-

dicted relationships among recognition probabilities for words in the three types of sentence. These relationships are illustrated in Fig. 7, together with individual data points from Table II. The coefficients of correlation between measured and predicted scores are 0.961 for the prediction of LP from ZP scores, 0.936 for the prediction of HP from LP scores, and 0.962 for the prediction of HP from ZP scores. The standard deviations of the differences between measured and predicted scores are 5.5, 7.6, and 5.6 percentage points, respectively.

### 3. Recognition of wholes and parts within wholes

Individual scores for the recognition of words and whole sentences in ZP, LP, and HP sentences were converted to  $j$  values using Eq. (7). Applying the criteria that scores should not be less than 0.05 or greater than 0.95 resulted in 16, 32, and 33  $j$  values for ZP, LP, and HP sentences, respectively.

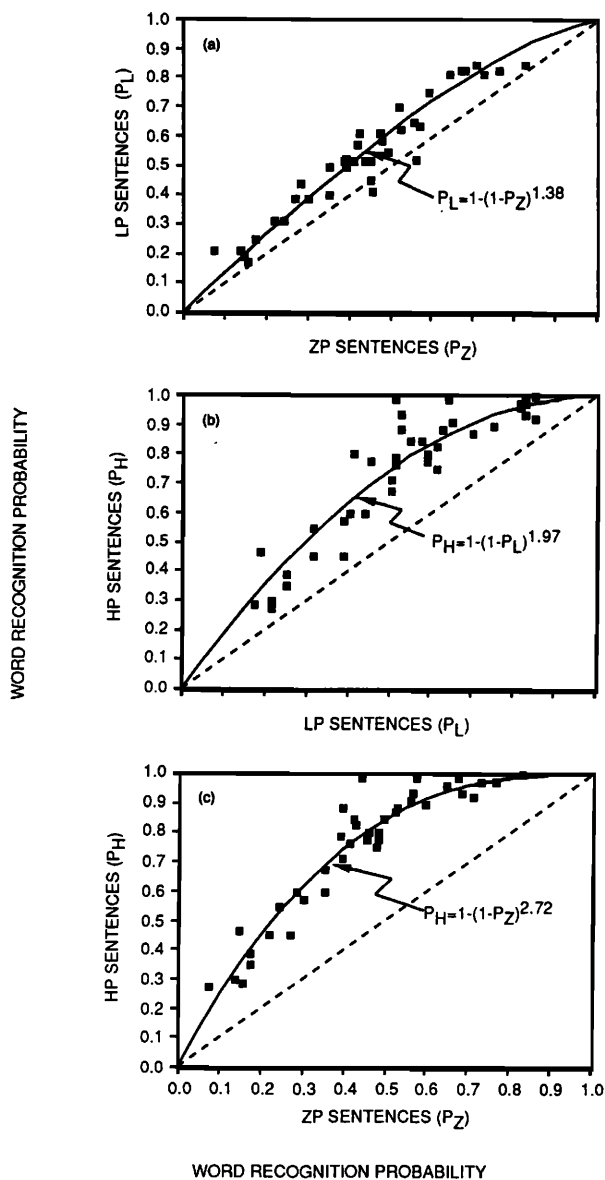


FIG. 7. Relationships among recognition probabilities for words in (a) LP and ZP sentences; (b) HP and LP sentences; and (c) HP and ZP sentences. Data points are individual scores from Table II. Curves show predictions of Eq. (3).

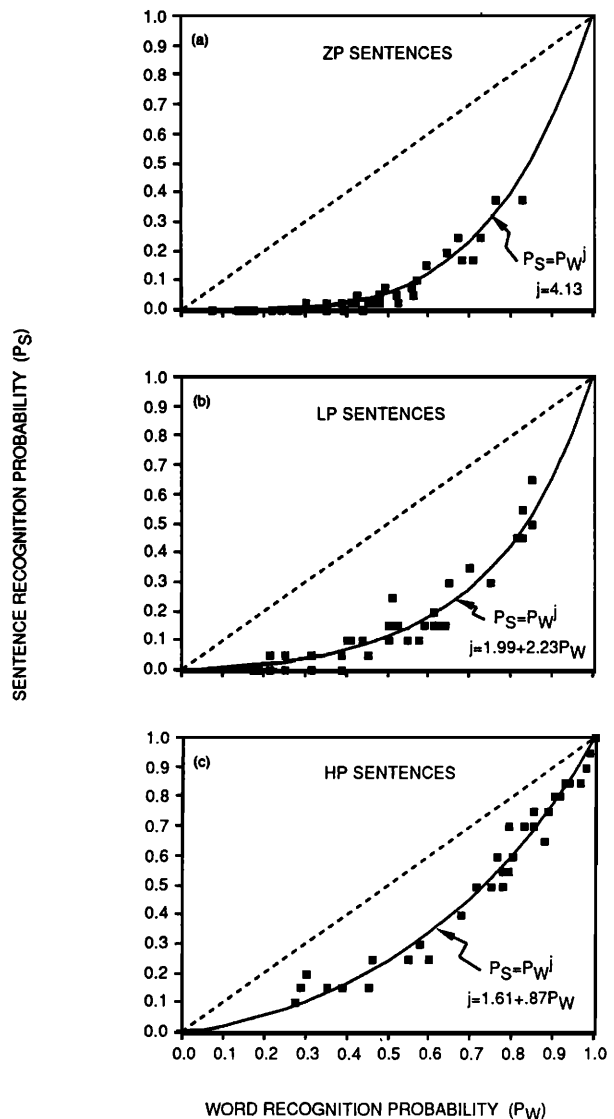


FIG. 8. Relationships between recognition probabilities for words and whole sentences in (a) ZP sentences; (b) LP sentences; and (c) HP sentences. Data points are individual scores from Table II. Curves show predictions of Eqs. (6), (8), and (9) (see text).



a. *ZP sentences.* The coefficient of correlation between values of the  $j$  factor and the recognition probability for words in ZP sentences failed to reach the 5% level of significance ( $r[14] = 0.416$ ). The mean value of  $j$  was 4.13, with 95% confidence limits of  $\pm 0.32$ . This value is not significantly different from the value of 4.0 that is predicted if the recognition of a whole sentence requires independent recognition of the four constituent words. By substituting the mean value in Eq. (6), we obtain the predicted relationship between recognition probability for words and that for whole ZP sentences. This relationship is illustrated in Fig. 8, together with individual data points from Table II. The coefficient of correlation between measured scores for whole ZP sentences and those predicted from the word scores is 0.945. The standard deviation of the difference between measured and predicted scores is 3.9 percentage points.

b. *LP sentences.* In contrast to the earlier analyses, the data for LP sentences showed a significant correlation between the value of  $j$  and the probability of recognition of the constituent words ( $r[30] = 0.538, p < 0.01$ ). The mean value of  $j$  is 3.33, with 95% confidence limits, based on the standard error of regression of  $\pm 0.19$ . The linear regression function for  $j$  on word recognition probability ( $p_p$ ) has an intercept of 1.99 and a gradient of 2.23. Thus, rather than being a constant,  $j$  for LP sentences is more accurately represented by

$$j = 1.99 + 2.23 * p_p. \quad (8)$$

Substituting in Eq. (6) gives the predicted relationship between word and sentence recognition probabilities for LP sentences. This relationship is illustrated in Fig. 8, together with data points from Table II. The coefficient of correlation between measured and predicted scores is 0.955. The standard deviation of the difference between the two is 4.9 percentage points.

c. *HP sentences.* As with the LP sentences, there was a significant correlation between the value of  $j$  in HP sentences and the probability of recognition of the constituent words ( $r[31] = 0.455, p < 0.01$ ). The mean value of  $j$  is 2.21, with 95% confidence limits, based on the standard error of regression of  $\pm 0.14$ . The linear regression function for  $j$  on word recognition probability ( $p_p$ ) has an intercept of 1.61 and a gradient of 0.87. Thus

$$j = 1.61 + 0.87 * p_p. \quad (9)$$

Substituting in Eq. (6) gives the predicted relationship between word and sentence scores for HP sentences. This relationship is illustrated in Fig. 8 together with data points from Table II. The coefficient of correlation between measured and predicted scores is 0.982. The standard deviation of the difference between the two is 5.1 percentage points.

### III. DISCUSSION

#### A. Findings re original goals

In terms of the original goals of this study, these data confirm that lexical, syntactic, and semantic constraints serve to increase the recognition probabilities for phonemes and words presented in noise. Moreover, the derived values of  $k$  appear to be constant across a wide range of recognition

probabilities, lending support to the notion that this factor provides a quantitative estimate of the effects of linguistic context that is independent of the specific conditions under which the two underlying recognition probabilities are measured. The derived values of  $j$  are similarly independent of recognition probability for the CVC data and for the ZP sentence data. The expectation of independence is not met, however, for the LP and the ZP sentence data. This finding may reflect the fact that linguistic constraints become increasingly important for the recognition of whole sentences as S/N ratio deteriorates. It was predicted that, for CVC nonsense syllables, and the ZP nonsense sentences, the value of  $j$  would be equal to the number of parts in a whole. The results support that prediction.

#### B. Interpretation of the $k$ factor

The  $k$  factor may be thought of as a proportional increase in the number of channels of statistically independent information available in the stimulus. Consider, for example, the speech spectrum divided into  $k$  frequency bands, each giving the same recognition probability  $p_i$  for words in isolation. The recognition probability can be increased to  $p_c$  either by combining the  $k$  bands and presenting the words in isolation, or by listening to a single band and presenting the words in sentence context. In other words, by using the  $k$  factor, we are simply expressing the context effect in terms of the increase of channels of information that **would be required** to produce the same change of percent correct recognition in the absence of context.

With this interpretation in mind, we can now examine the  $k$  values obtained in these studies. For phonemes in words, we obtained a  $k$  of 1.3, indicating that the addition of lexical context is equivalent to multiplying by 1.3 (or increasing by 30%) the channels of statistically independent information available in the nonsense syllables. The corresponding values for the four-word sentences were 1.4 for the effects of syntax (a 40% increase), 2.0 for the effects of semantics (a 100% increase), and 2.7 for the combined effects of syntactic and semantic constraints (a 170% increase).

These results suggest that sentence context is much more important than word context and that semantic constraints are the single most important contextual factor. They should, however, be interpreted cautiously. The effect of lexical constraints was measured in CVC words only. It is probable that the  $k$  factor for polysyllabic words will be found to be much higher. (Even in the present data, we found a word-frequency effect in that the value of  $k$  obtained from those CVC lists consisting mainly of common words was higher than that obtained from the other lists.) Note also that the assertion that the LP sentences provided only syntactic constraints is open to question. "Ducks eat old tape," for example, is by no means devoid of semantic information and the juxtaposition of the first two words is not semantically anomalous. The removal of additional semantic information from the LP sentences, however, would reduce the value of the derived  $k$  factor for syntax and increase it for semantics, lending even stronger support to the conclusion that semantic constraints produce the largest context effects.

### C. Interpretation of the $j$ factor

The  $j$  factor may be interpreted as the effective number of statistically independent parts within a whole, that is, as a measure of the perceiver's tendency to perceive the stimulus in "chunks," each chunk extending over more than one of the obvious units (i.e., phonemes in experiment 1 and words in experiment 2). Thus the data suggest that, on average, subjects respond to CVC words as though they consist of 2.5 independent parts (each part containing the equivalent of 1.2 phonemes), and they respond to meaningful four-word sentences as though they consist of 2.2 independent parts (each part containing the equivalent of 1.8 words). A  $j$  factor of 1 in the sentence data would have indicated that subjects were recognizing either complete sentences, or nothing at all. In fact, for the HP sentences, the  $j$  factor approaches 1.6 as the probability of recognition of parts approaches zero, indicating a close approach to the condition in which a sentence is recognized either as a unit or not at all. The fact that  $j$  for HP sentences rises under more favorable listening conditions presumably reflects the fact that the listener becomes less dependent on internal linguistic constraints as the information in the stimulus becomes more accessible.

As predicted, the nonsense syllables and sentences were perceived as though they consisted of three and four independent parts, respectively. The fact that  $j$  for nonsense sentences was not significantly different from 4 supports the validity of using nonsense sentences to obtain measures of word recognition probability under the no-context condition. A second test of the validity of this approach is available from the comparison of data from the two experiments. It will be seen from Figs. 2 and 5 that the scores for CVC words in a carrier phrase were, on average, some 10 percentage points higher than those for words in ZP sentences. This difference, however, was not statistically significant at any of the three S/N ratios for which comparative data were available.

### D. Alternative methods of quantifying context effects

Alternative methods of treating the differences between scores obtained under the context and no-context situations were mentioned in the Introduction. From Eq. (3), we can predict the relative outcomes of these approaches for a given value of  $k$ . Illustrative results are shown in Fig. 9 for  $k = 2.72$ , the empirically determined value for the combined effects of syntactic and semantic context. It will be seen that the arithmetic difference between the two scores reaches a maximum of 35 percentage points for a no-context probability of 45% but approaches zero at very low and very high probabilities. The difference between the  $z$  transforms of the two probabilities rises with increasing probability, the gradient increasing rapidly as the no-context probability approaches 100%. Also shown is the context effect expressed as a ratio of probabilities for the context and no-context conditions. This curve shows that the relative contribution of context to the percentage score falls with increasing probability. The ratio between the two scores approaches the value of  $k$  as the no-context probability approaches 0%, and it approaches unity as the no-context probability approaches

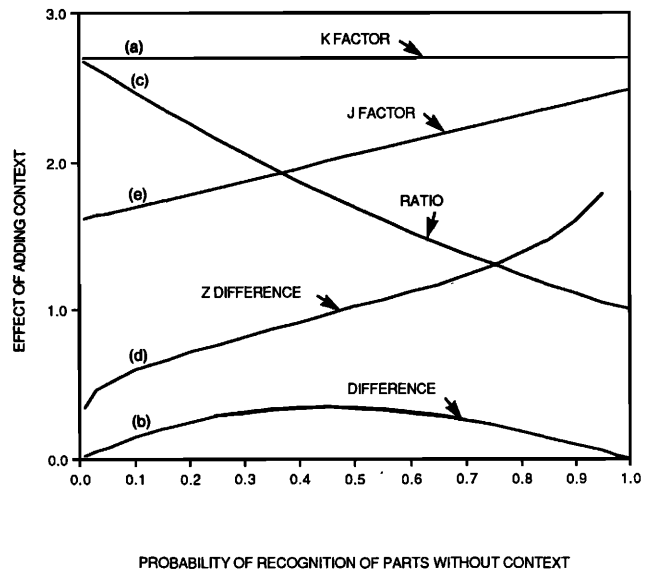


FIG. 9. Comparison of several methods of quantifying context effects. The relationship between recognition probabilities for items with context ( $p_c$ ) and items without context ( $p_i$ ) is taken to be  $p_c = 1 - (1 - p_i)^k$ , with  $k = 2.7$  (the experimentally determined value for the effects of sentence context on word recognition). The measures shown are (a) the  $k$  factor; (b) the arithmetic difference between the two recognition probabilities ( $p_c - p_i$ ); (c) the ratio of the two recognition probabilities ( $p_c/p_i$ ); (d) the difference between the  $z$  transforms of the two recognition probabilities; and (e) the  $j$  factor for HP sentences [Eq. (9)].

100%. The limiting values of  $k$  and unity can be shown to result directly from the underlying mathematics by expanding the binomial term on the right of Eq. (3). The implication of this relationship is that, at very low probabilities, the sensory data and the context tend to provide complementary information and the percent correct scores for the two information sources are additive. At high probabilities, however, the two sources tend to provide duplicate information and the contribution of context to percent correct score is small. This is not to say that the presence of contextual constraints has no effect on speech perception under favorable listening conditions. Other measures, such as reaction time and spare channel capacity, would be expected to respond to contextual constraints even though the change in percent correct is, of necessity, small.

Also illustrated in Fig. 9 are the empirically determined  $j$  factor values for HP sentences [Eq. (9)]. The fact that  $j$  increases under more favorable listening conditions probably reflects the diminishing role of context as a determinant of the percent correct score. Thus the  $j$  factor is, in a sense, the mirror image of the ratio of percent correct scores. Because  $j$  is not independent of the underlying probability, at least in sentences, it is less valuable than the  $k$  factor as a quantitative measure of context effects. The advantage of the  $j$  factor, however, is that it can be calculated from the responses to only one kind of test material. Unlike the  $k$  factor, it does not require the collection of recognition scores under the no-context condition.

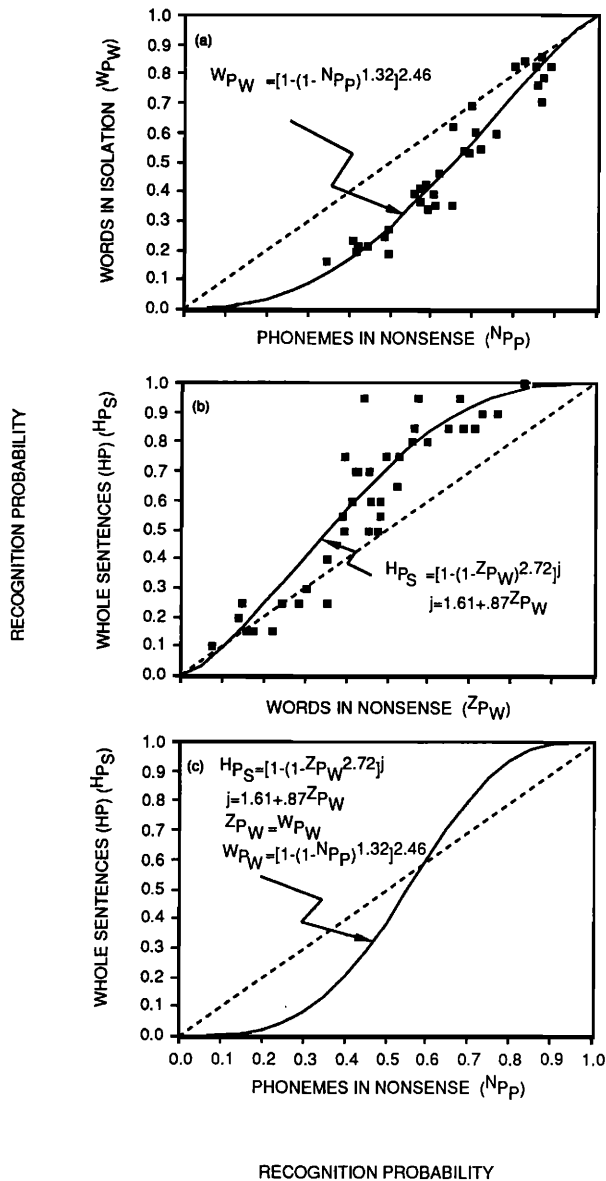


FIG. 10. Predicted relationships between recognition probabilities for (a) phonemes in nonsense syllables and whole CVC words; (b) words in nonsense sentences (or isolation) and whole HP sentences; and (c) phonemes in nonsense syllables and whole HP sentences. The data points for (a) and (b) are taken from Tables I and II. There are no data points for (c) because different subjects were used in the two experiments.

### E. Derived relationships

Equations (3) and (5) can be combined to provide predicted relationships among all of the measures that were obtained in these studies. For example, the predicted relationship between the probabilities for recognition of phonemes in nonsense syllables and of whole CVC words is

$${}^w p_w = [1 - (1 - {}^n p_p)^{1.32}]^{2.46}, \quad (10)$$

where  ${}^w p_w$  is the probability of recognition of words in CVC words,  ${}^n p_p$  is the probability of recognition of phonemes in nonsense, 1.32 is the  $k$  factor for lexical context in CVC words, and 2.46 is the  $j$  factor for CVC words. This relation-

ship is illustrated in Fig. 10(a) together with experimental data from Table I.

Similarly, the relationship between the probabilities for recognition of words in ZP sentences and for whole HP sentences is

$${}^h p_s = [1 - (1 - {}^z p_w)^{2.72}]^j, \quad (11)$$

where  ${}^h p_s$  is the probability of recognition of sentences in HP material,  ${}^z p_w$  is the probability of recognition of words in ZP material, and  $j$  is the  $j$  factor for HP sentences =  $1.61 + 0.87 \cdot {}^z p_w$ . This relationship is illustrated in Fig. 10(b) together with experimental data from Table II.

Accepting that  ${}^w p_w$  and  ${}^z p_w$  are equal, Eqs. (10) and (11) can be combined to give the predicted relationship between recognition probabilities of phonemes in nonsense syllables and whole HP sentences, as illustrated in Fig. 10(c). In this case, there are no data points since the two experiments reported here used different subjects.

This last analysis demonstrates how the combined effects of lexical, syntactic, and semantic constraints can produce a large increase in the probability of sentence recognition with a small increase in the phonetic detail available from the sensory data. Such analyses are in keeping with the observation that certain sensory aids for the deaf function as highly effective aids to lipreading, even though the phonetic detail they provide by themselves may be very limited.

### F. Clinical application

Throughout this article, we have been assuming that the listener has sufficient language and world knowledge to be able to take full advantage of contextual redundancy. This assumption may not be justified for children, geriatrics, persons who are not operating in their native language, individuals with prelingually acquired hearing loss, and persons with disorders of linguistic and/or cognitive function. In follow-up studies, to be reported later, we have already shown that the  $k$  factor for semantic context is lower in young children than in young adults. This finding presumably reflects the young child's incomplete world knowledge. We have also shown that the  $j$  factor for nonsense syllables is higher in geriatrics than in young adults, presumably reflecting an increased tendency to impose meaning on nonsensical stimulus patterns. These considerations lead to the possible use of the materials and methods reported here as evaluative and diagnostic tools in clinical populations.

Another clinical application might be in the evaluation of responses to speech perception training within a rehabilitative program. The techniques may also assist in planning individual programs of intervention by separating problems of access to sensory data from those of effective use of linguistic context. Yet another application might be in the prediction of probable benefits from cochlear implants or other forms of sensory assistance.

Even in the present study, we found individual differences of speech recognition in noise and in the use of contextual constraints. Our goal was to investigate the changes of both measured and derived quantities as the subjects' access

to information in the acoustical stimulus was varied. We introduced the variations in two ways. One was to change S/N ratio. The other was to take advantage of intersubject differences in the ability to extract information from speech in noise. It is quite possible that these two effects are not equivalent. In other words, a subject who has inherent difficulties in processing speech in noise may use context differently from one who obtains the same no-context score, but at a less favorable S/N ratio. The design of these experiments does not permit these two effects to be examined separately. Examination of the data, however, suggests that there *were* marked intersubject differences in both the ability to deal with speech in noise and the ability to take advantage of contextual constraints, even in a group that was audiometrically and linguistically "normal." The nature, causes, and consequences of these differences could well be a topic for further research.

Although the mathematical treatment of context effects reported here has potential clinical applications, there is a serious limitation to the use of these methods with individuals. The estimates of recognition probability from which values of  $j$  and  $k$  would be derived have high test-retest variability unless obtained with a very large number of items (Boothroyd, 1968a,b; Thornton and Raffin, 1978). Both  $j$  and  $k$  are essentially difference scores, and, therefore, have even greater test-retest variability than the scores from which they are derived. The consequences of this limitation can be seen in the scatter of individual data points from experiment 2, in which sentence recognition scores were based on only 20 HP and 20 LP sentences (see Figs. 7, 8, and 10). At the present time, it is not clear how best to address the problem of test-retest reliability within the time constraints of clinical evaluation.

### G. Generalization

The estimates of  $j$  and  $k$  obtained in these studies relate to the specific materials and subjects with which they were obtained. Although they provide a general indication of the magnitude of the effects of lexical, syntactic, and semantic constraints on phoneme and word recognition, they can by no means be taken as norms. Changes of test material, type of stimulus degradation, and the subject pool will affect the magnitude of context effects and, hence, the values of the derived parameters.

Note, also, that there is nothing in the foregoing analyses to suggest the processes by which an individual extracts or applies linguistic information from the sensory data. We are not proposing a theory of speech perception. We are, however, proposing a mathematical approach to the quantification, comparison, and manipulation of the known effects of contextual constraints as they affect the probabilities of phoneme and word recognition across large stimulus aggregates.

## IV. CONCLUSIONS

(1) The effects of contextual constraints on the recognition of linguistic units may be expressed as the ratio of the logarithms of the error probabilities measured with and without the source of context present. The resulting  $k$  factor provides a measure of the context effect that is independent of recognition probability under the no-context condition. It may be interpreted as the amount by which one would need to multiply the number of statistically independent channels of information in order to produce an effect equal to that of the context.

(2) The data reported here indicate that the most effective source of contextual constraint is sentence meaning. The combined effects of syntax and meaning gave a  $k$  factor of 2.7 for the recognition of words in four-word sentences. In contrast, the value of  $k$  for the effects of word meaning on phoneme recognition in CVC syllables was found to be only 1.3.

(3) The effects of internal linguistic constraints may be expressed as the ratio of the logarithms of the recognition probabilities for wholes and for parts within wholes. The resulting  $j$  factor was found to be equal to the number of parts within the whole for materials that are free of internal constraints (nonsense materials), as predicted from the underlying mathematics. The value of  $j$  falls as the effects of linguistic constraints increase. The  $j$  factor may be thought of as the number of statistically independent parts within a whole — that is, as a measure of the perceiver's tendency to chunk the parts into larger perceptual units.

(4) From the data reported here, it appears that CVC words are perceived as though they consist of 2.4, rather than 3, independent units. Meaningful four-word sentences appear to be perceived as though they consist of around 1.6 independent units at very poor signal-to-noise ratios, but 2.5 independent units at highly favorable signal-to-noise ratios.

(5) By combining the equations for  $k$  and  $j$ , one can predict relationships among many measures of speech perception.

(6) In addition to their theoretical interest, we feel that the  $k$  and  $j$  factors have potential clinical application.

## ACKNOWLEDGMENTS

The help of the following people is gratefully acknowledged: Irving Hochberg of the City University of New York; Theresa Hnath-Chisolm, Liat Kishon-Rabin, and Eddy Yeung of the Center for Speech and Hearing Sciences Research of the City University of New York's Graduate Center; Daphne Fox and students of Marymount Manhattan College; Nancy McGarr and students of Molloy College; Robert Domingo and students of Queens College; Mindy Sirlin and students of Iona College; Adrienne Rubenstein and students of Brooklyn College; and staff and students of the City University of New York Graduate Center. This work was funded by PSC CUNY Awards #6-63137 and #6-64363. Additional support was provided by NIH Program Project Grant #17764.

## APPENDIX A: CVC WORD AND NONSENSE SYLLABLE LISTS

TABLE AI. Meaningful CVC words.

| 1 <sup>a</sup> | 2 <sup>a</sup> | 3      | 4      | 5      | 6    |
|----------------|----------------|--------|--------|--------|------|
| pass           | hall           | pies   | boss   | time   | make |
| rule           | come           | mock   | sip    | caught | laws |
| cause          | bag            | room   | pal    | beg    | rice |
| time           | rose           | dad    | coat   | rid    | bell |
| log            | suit           | loan   | rod    | loon   | tote |
| sick           | made           | beg    | moon   | mop    | cod  |
| mean           | like           | tell   | hem    | doze   | ham  |
| bed            | peace          | keep   | take   | says   | deep |
| hope           | dip            | hiss   | league | pack   | pig  |
| date           | ten            | sought | dies   | heel   | soon |
| 7              | 8              | 9      | 10     | 11     | 12   |
| seal           | hide           | call   | lean   | lice   | dike |
| dawn           | tame           | buys   | hag    | mall   | ball |
| boom           | rule           | same   | bed    | tomb   | mace |
| hogs           | cause          | miss   | sews   | bag    | rig  |
| toes           | big            | rot    | cop    | soap   | lose |
| mid            | sass           | hoop   | root   | rake   | sop  |
| cat            | pope           | load   | pick   | pen    | comb |
| like           | don            | peck   | maim   | keys   | ten  |
| pep            | meek           | tag    | toss   | hid    | pad  |
| race           | let            | dean   | dial   | dot    | heat |

<sup>a</sup> Lists 1 and 2 consist almost entirely of words selected from the 1000 most commonly used words of the Thorndike and Lorge (1944) count. The remaining lists each cover a wide range of frequencies.

TABLE AII. CVC nonsense syllables.

| 1       | 2       | 3       | 4       | 5       | 6       |
|---------|---------|---------|---------|---------|---------|
| m a I g | b e I p | k e z   | b i g   | r u k   | p u k   |
| k o s   | r a I k | p I d   | t a t   | t e I t | k i z   |
| h e I m | l i m   | s u m   | p e m   | l a I l | t e I d |
| d I t   | m I n   | r a I g | r o U s | b o p   | d o s   |
| t u p   | p u t   | t e I n | k a l   | p I m   | h a I m |
| b a k   | s o U g | l o U l | h u n   | k a z   | l i t   |
| p o U d | t a z   | m a t   | d I z   | d o U n | s a n   |
| s e m   | d a s   | d o k   | l e I p | h a d   | b e p   |
| l a l   | h o d   | b i s   | s o k   | m i s   | m o U g |
| r i z   | k e l   | h a p   | m a I d | s e g   | r a l   |
| 7       | 8       | 9       | 10      | 11      | 12      |
| r I t   | r e I g | l e n   | p a z   | r e I m | k I g   |
| s e l z | t u k   | d a p   | h i s   | l o k   | p e I k |
| b o d   | m o t   | p a I t | m e k   | m u l   | t a I s |
| k u m   | k I p   | b I m   | s o U m | d a s   | d o d   |
| l e p   | b o U d | s i g   | t a l   | h e z   | m i z   |
| p o U n | h a s   | h a l   | l o t   | t i d   | r e m   |
| h a I s | l a n   | k e I z | k e I d | p o U t | h a p   |
| d a k   | s a I l | r u s   | b a I p | b a n   | l a n   |
| t i g   | d e z   | t o d   | d u g   | k a I p | s o U t |
| m a l   | p i m   | m o U k | r I n   | s I g   | b u l   |

## APPENDIX B: HP, LP, AND ZP SENTENCE LISTS

### A. High predictability (HP)

Tough guys sound mean  
Warm sun feels good

Birds like long worms  
Dad buys new shirts  
Cows give sweet milk  
Cooks make hot food  
Feed your dog meat  
Take that trash out  
Dull paint won't shine  
Most boys play ball  
Tall men jump high  
Red wine can stain  
Tell Mom those jokes  
Sharp nails might hurt  
Fall nights turn cool  
Apes swing from trees  
Boats sail at sea  
Small kids need help  
Pour me more tea  
Fresh bread smells great

### B. Low predictability (LP)

Thin books look bright  
Wide pens swim fast  
Cats get good ears  
Feet catch slow thieves  
Ducks eat old tape  
Cups kill fat leaves  
Throw his rat ice  
Lead this coat home  
Green hands don't fall  
Pink chalk bakes phones  
Blue chairs speak well  
Dumb shoes will sing  
Find girls these clouds  
Tin hats may laugh  
Late forks hit low  
Knees talk with mice  
Teeth sleep on fields  
Sad cars want chills  
Lend them less joy  
Soft rocks taste strange

### C. Zero predictability (ZP)

Sing his get throw  
Tin bread blue more  
Pour this kids lead  
Jokes boats speak fast  
Trees nights from might  
Rock Mom sea strange  
Apes home ducks clouds  
Soft these ice give  
Birds warm great fields  
Knees on sleep feels  
Sad slow cars bakes  
Chairs pink well wine  
Milk can swim girls  
Play mean help less  
Stain tape dull mice  
Old ball paint swing

Shoes trash feet want  
 Find worms coat tail  
 Leaves joy high talk  
 Guys late phones chalk  
 Hats ears rat will  
 Cups small at fresh  
 With chills me red  
 Sharp thin books look  
 Hurt tell laugh low  
 Taste long cats thieves  
 Lend men feet those  
 Bright tea green catch  
 Dumb good fall new  
 Forks may shirts jump  
 Good take sail sun  
 Shine them won't wide  
 Make meat teeth don't  
 Hands cooks cows nails  
 That like fat smells  
 Dog you sweet Dad  
 Out cool pens most  
 Hot wear hit need  
 Food boys kill fall  
 Sound tough turn buys

Bilger, R. C. (1984). "Speech recognition test development," in *Speech Recognition by the Hearing Impaired*, ASHA Rep. No. 14, 2-7.

Boothroyd, A. (1968a). "Developments in speech audiometry," *Br. J. Audiol.* **2**, 3-10.

Boothroyd, A. (1968b). "Statistical theory of the speech discrimination score," *J. Acoust. Soc. Am.* **43**, 362-367.

Boothroyd, A. (1978). "Speech perception and sensorineural hearing loss," in *Auditory Management of Hearing-Impaired Children*, edited by M. Ross and T. G. Giolas (University Park, Baltimore, MD).

Boothroyd, A. (1984). "Auditory perception of speech contrasts by subjects with sensorineural hearing loss," *J. Speech Hear. Res.* **27**, 134-144.

Boothroyd, A. (1985). "Measurement of speech production in hearing-im-

paired children: Some benefits of forced-choice testing," *J. Speech Hear. Res.* **28**, 185-196.

Broadbent, D. E. (1967). "Word-frequency effect and response bias," *Psychol. Rev.* **74**, 1-15.

Duffy, J., and Giolas, T. G. (1974). "Sentence intelligibility as a function of key word selection," *J. Speech Hear. Res.* **17**, 631-637.

Fletcher, H. (1953). *Speech and Hearing in Communication* (Van Nostrand, New York).

Giolas, T. G., Cooker, H., and Duffy, J. (1970). "The predictability of words in sentences," *J. Aud. Res.* **10**, 328-334.

Giolas, T. G., and Epstein, A. (1963). "Comparative intelligibility of word lists and continuous discourse," *J. Speech Hear. Res.* **6**, 349-358.

Hirsch, I., Reynolds, E. G., and Joseph, M. (1954). "Intelligibility of different speech materials," *J. Acoust. Soc. Am.* **26**, 530-538.

Hnath, T., Hanin, L., and Boothroyd, A. (1985). "Effect of knowledge of topic on the lipreading of sentences," paper presented at the 1985 Convention of the American Speech, Language, and Hearing Association, Unpublished Report, City University of New York.

Howes, D. (1957). "On the relation between the intelligibility and frequency of occurrence of English words," *J. Acoust. Soc. Am.* **29**, 296-307.

Kalikow, D. H., Stevens, J. N., and Elliott, L. L. (1977). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," *J. Acoust. Soc. Am.* **61**, 1337-1351.

Massaro, D. W., and Oden, G. C. (1980). "Evaluation and integration of acoustic features in speech perception," *J. Acoust. Soc. Am.* **67**, 996-1013.

Miller, G. A., Heise, G. A., and Lichten, W. (1951). "The intelligibility of speech as a function of the context of the test material," *J. Exp. Psychol.* **41**, 329-335.

Owens, E. (1961). "Intelligibility of words varying in frequency," *J. Speech Hear. Res.* **4**, 113-129.

Pollack, I., Rubenstein, H., and Decker, L. (1959). "Intelligibility of known and unknown message sets," *J. Acoust. Soc. Am.* **31**, 273-279.

Rosenzweig, M., and Postman, L. (1957). "Intelligibility as a function of frequency of usage," *J. Exp. Psychol.* **54**, 412-422.

Savin, H. (1963). "Word-frequency effects and errors in the perception of speech," *J. Acoust. Soc. Am.* **35**, 200-206.

Schiavetti, N., Sitler, R. W., Metz, D. E., and Houde, R. A. (1984). "Prediction of contextual speech intelligibility from isolated word intelligibility measures," *J. Speech Hear. Res.* **27**, 623-626.

Schultz, M. (1964). "Word familiarity influences in speech discrimination," *J. Speech Hear. Res.* **7**, 395-400.

Thorndike, E., and Lorge, I. (1944). *The Teacher's Word Book of 30 000 Words* (Columbia University, New York).

Thornton, A. R., and Raffin, M. J. M. (1978). "Speech discrimination scores modeled as a binomial variable," *J. Speech Hear. Res.* **21**, 507-518.