The Journal of the Acoustical Society of America

MAY 01 1977

Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability

D. N. Kalikow; K. N. Stevens; L. L. Elliott

J. Acoust. Soc. Am. 61, 1337–1351 (1977)
https://doi.org/10.1121/1.381436

Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability*

D. N. Kalikow and K. N. Stevens†

Bolt Beranek and Newman Incorporated, Cambridge, Massachusetts 02138

L. L. Elliott

Auditory Research Laboratory, Northwestern University, Evanston, Illinois 60201
(Received 20 September 1976; revised 18 January 1977)

This paper describes a test of everyday speech reception, in which a listener's utilization of the linguistic-situational information of speech is assessed, and is compared with the utilization of acoustic-phonetic information. The test items are sentences which are presented in babble-type noise, and the listener response is the final word in the sentence (the key word) which is always a monosyllabic noun. Two types of sentences are used: high-predictability items for which the key word is somewhat predictable from the context, and low-predictability items for which the final word cannot be predicted from the context. Both types are included in several 50-item forms of the test, which are balanced for intelligibility, key-word familiarity and predictability, phonetic content, and length. Performance of normally hearing listeners for various signal-to-noise ratios shows significantly different functions for low- and high-predictability items. The potential applications of this test, particularly in the assessment of speech reception in the hearing impaired, are discussed.

PACS numbers: 43.70.Ep, 43.50.Qp, 43.70.Ve

INTRODUCTION

This paper discusses some of the issues related to the problem of designing a test of everyday speech reception, and describes a new sentence test that attempts to overcome some of the inadequacies of tests currently in use. While the proposed test has not yet received adequate calibration and validation with a clinical population, and hence cannot at this time be recommended for clinical use, the performance of a normally hearing population with the test has been evaluated, and some data have been collected from hearing-impaired individuals. Performance of normally hearing listeners is described in this paper, and results from a group of listeners with sensorineural losses are presented in a companion paper.[1]

Since everyday speech communication covers a wide range of spoken material and takes place in a variety of contexts, it is not possible in a single test to sample all types of speech events that might occur in practice. It is appropriate, however, to determine the kinds of sensory and cognitive processes that are involved in the reception of these kinds of speech material, and to devise tests that will assess the degree to which an individual shows impairment in the utilization of these processes.

Basically two kinds of operations are involved in the understanding of sentences. One is the reception and initial processing of acoustic information through the auditory system, and the other is the utilization of linguistic information that is stored in memory. As stated succinctly by Fry,[2] "[speech reception] depends upon both the condition of the peripheral hearing mechanism and the efficiency of the central decoding mechanism, the speech centers of the brain." One component in the decoding of a sentence by a listener is the extraction of a partial set of phonetic features from the acoustic signal. These phonetic features are placed in short-term memory, where they are available for further processing. The linguistic information available in the long-term memory of a listener includes knowledge of the phonological, lexical, syntactic, and semantic constraints that occur in language. The more these kinds of information provide a context for a particular utterance, the less it is necessary for the listener to depend on the detailed properties of the acoustic signal in order to understand the utterance. A test of a listener's ability to understand everyday speech must, therefore, assess both the acoustic-phonetic and the linguistic-situational components of the process.

I. BACKGROUND

We examine first the experimental evidence for the various factors that influence the performance of a listener in a sentence-understanding task, and that must be taken into account in the design of a test for listeners with hearing and other impairments.

A. Phonetic and prosodic factors

It is well known that the intelligibility of a word in noise depends upon the sequence of sounds that constitute the word. Some classes of sounds are more susceptible to masking by noise than others,[3] and consequently words containing these sounds are likely to be less intelligible than words containing sounds that are resistant to masking. In developing any test of speech intelligibility, therefore, care must be taken to select speech materials in which the phonetic content is properly balanced to reflect the distribution of speech-sound classes that occur in the language.

The acoustic attributes of sentences include not only the properties of phonetic units, but also prosodic parameters that signal the characteristics of larger units within the sentences. These consist of variations in the durations of sounds and in the fundamental frequency of voiced sounds. There is evidence that these prosodic parameters are used by a listener as cues for the understanding of sentences, since they contribute information about stress and the grouping of words.[4,5]

B. Effect of sentence context

The fact that in a noisy environment words in a sentence context are more intelligible than words spoken in isolation or without the benefit of sentence context has been demonstrated by Miller, Heise, and Lichten,[6] and by Miller.[7] These investigators argued that the sentence context imposes constraints on the set of alternative words that are available as responses at a particular location in a sentence, and noted that the intelligibility of words increases when the number of response alternatives decreases.

This conclusion was supported and quantified further by Duffy and Giolas,[8] who examined the intelligibility of words in sentences in which the words had various degrees of predictability. Their experiments showed directly that the predictability of a word has an influence on its intelligibility. Bransford and Nitsch[9] further showed that, when a sentence is presented to a listener in a particular context that is established by earlier utterances and by the situation in which the utterance is presented, then the comprehension of the sentence is greater than when it is presented in a more neutral context.

C. Word familiarity

The intelligibility of words in noise is influenced not only by the predictability of the words, but also by their familiarity. A number of investigators have shown the effect of word familiarity[10–14] using as a measure of familiarity the relative frequency of occurrence of a word in specified kinds of spoken or written language.[15]

D. Noise interference

An everyday listening situation is usually characterized by noise that interferes with understanding of the speech. The principal effect of the noise is to mask some of the sounds, so that the listener has less acoustic information on which to base his interpretation of the acoustic signal. Furthermore, there is evidence that increased noise can require increased effort by a listener in identifying words,[16,17] and it is not unlikely that such interference can affect the cognitive and memory processes that are involved in understanding sentences.

One kind of noise that is often encountered is a babble of voices produced by several speakers. It has been shown that this type of noise interferes with speech intelligibility more than a (stationary) random nonspeech noise, and that the amount of masking depends upon the number of different voices that are mixed to produce the noise.[18,19] This finding suggests that a babble of a few voices can produce interference that exceeds the interference due solely to masking of individual sounds. This enhanced interference arises both because the babble contains false speech cues and because it increases the load on the attentional and memory processes that are involved in understanding sentences.

E. Listener-related factors

In addition to acoustic, linguistic, and situational influences on a listener's understanding of spoken language, there are listener-related factors that can cause a deterioration in performance. Hearing impairment can obviously distort and reduce the acoustic information available to a listener, with a consequent reduction in the understanding of speech in all environments. When noise is mixed with the speech signal, the masking effect of the noise is often greater than that for normally hearing listeners.[18,20] One might also expect substantial individual differences in the degree to which a listener is able to make use of linguistic and contextual constraints in understanding sentences. For example, it is known that people have different vocabulary sizes, one source of these differences being the fact that an individual's vocabulary increases with increasing age.[21] The effect on word familiarity resulting from differences in vocabulary size would presumably tend to influence listener performance with spoken language. On the other hand, there is a decline in response speed and in short-term memory performance with increasing age,[22,21] and this decline could tend to have a negative effect on performance in a sentence-understanding task.

F. Problems with existing tests of speech perception

There has been a long history of attempts to develop speech tests, particularly those that examine the communicative capabilities of the hearing impaired. The objective of these tests is to provide a measure or measures of the ability of a hearing-impaired listener to understand speech in an everyday listening situation. The general form of such tests is a series of utterances—usually lists of monosyllabic words—which are presented to a listener, who is asked to write down or repeat back what he hears. Performance is scored in terms of the number of words or the number of sounds correctly identified.

There have been a number of objections to these tests, primarily because they do not provide sufficiently close approximations to everyday communicative situations, and do not adequately assess or effectively control the various components of the speech-understanding process, as discussed above. An especially difficult aspect of the design of sentence tests is the generation of sentences that are reasonably natural and that are at the same time well controlled for word predictability, phonetic content, and other attributes. It is also desirable that the audiometric test be easily administered as well as reliable and valid. If a test requires the hearer to respond by repeating the entire utterance, clinicians may not accept it because the "whole-response" mode (1) is time consuming, (2) limits the range of hearers capable of dealing easily with the task, and (3) complicates the scoring of responses. Progress has been made toward developing sentence tests that satisfy some of the objections just noted, including the control of word familiarity, grammatical structure, and phonetic content,[2,23] the use of everyday sentences of various lengths and types,[24,25] the construction of nonsense sentences with various degrees of approximation to real sentences,[26] and the generation of questionlike sentences that require a one-word answer.[27,28] However, no test has yet had a sufficient range of positive characteristics to enable it to achieve wide clinical acceptance.

II. TEST DEVELOPMENT

A major objective in developing this test was to produce a measure that would assess utilization of the linguistic–situational information of speech in comparison with utilization of acoustic–phonetic information. In contrast to most tests of speech intelligibility which examine only processing of acoustic–phonetic information, we wanted to place equal emphasis on examining the contribution of "cognitive" variables of memory and language competence. As used here, "language competence" refers to receptive knowledge of the lexical, syntactic, and semantic constraints of standard American English. We anticipated that a test that measured utilization of linguistic-situational information as well as acoustic–phonetic information would prove to be a more useful index of everyday speech reception than a measure that assessed only the latter quantity.

Our basic strategy in preparing a sentence test was (1) to make certain initial decisions concerning some of the properties of the sentences to be used in the test; (2) to generate a number of sentences with these properties; (3) to subject these sentences to various kinds of testing and analysis; (4) on the basis of appropriate selection criteria derived in (3), to prepare from the larger list in (2) a smaller, more homogeneous ensemble of sentences; (5) to assemble these sentences into a set of equivalent forms for a proposed test; and (6) to test for equivalence of performance scores for the different forms.

A. Initial selection of sentence characteristics

One of the first questions that must be addressed in formulating a sentence test concerns the type of response to be elicited from the subject. In order to simplify the task of the listener while maintaining a reasonable degree of flexibility in the design of the sentences, it was decided to require a single-word response, and to make the response word the last word in the sentence. The subject's task is simply to write down or to repeat the last word in the sentence. In order to further control the types of sentences, an additional restriction was imposed that this final word (called the key word) be always a monosyllabic noun. The requirement that the last word be a noun is not, in fact, a severe restriction, since, at least in written text, many more than half of all sentences end in a noun. By limiting the word to a monosyllable, it is easy to formulate test sentences in which the key word receives main stress. In this way, a certain degree of acoustic control over the prosodic aspects of the sentences can be achieved.

An additional requirement that was imposed on the key words was that they be neither little used nor very frequently used words in English. All key words in the test materials were selected from items in the Thorndike–Lorge[15] list with frequency counts in the range 5 to 150 per million words.

A further decision that was made prior to generating sentence materials was that there be a reasonable degree of homogeneity in sentence length. All sentences were constrained to contain five to eight words and six to eight syllables.

At the outset, it was decided to attempt to generate sentence lists in which the various key words had different degrees of predictability from the sentence context. If the predictability of the last word from the remainder of the sentence is *high*, the listener is aided in this identification of that word by the syntactic, semantic, and prosodic cues available in the sentence, as well as by any perceived acoustic characteristics of the word itself. A stored lexicon of English words is presumably involved as well. When the key word has *low* predictability, on the other hand, the listener must depend primarily on acoustic properties and lexical information concerning the key word itself. Higher functions and cues, while they may be invoked initially, are ultimately of far less utility to the listener in making the correct response. Sentences with low-predictability key words were constructed by utilizing a set of neutral sentence contexts which placed essentially no restriction on the semantic features of the final word. An example is *John was talking about the bay*. The rule for generating sentences with high-predictability key words was to use within each sentence a set of "pointer" words (usually two or three) that provide semantic links to the key word. An example is *The boat sailed across the bay*, in which the three content words *boat*, *sailed*, and *across*, all provide semantic links to the key word *bay*. Finally, sentences in which the key word has an intermediate degree of predictability were generated according to the rule that one content word (with one or two syllables) in the sentence provided a relatively strong semantic link to the key word. An example of such a sentence is *They tried to follow the trail*.

B. Formulation of an initial corpus of sentences

The initial step in formulating a corpus of sentences was to select from the Thorndike–Lorge list a set of 479 monosyllabic nouns with word frequencies in the specified range of 5 to 150 per million words. Beginning with this list of words, and given the other constraints just noted with regard to sentence length, position of final stress, and predictability, several of the professional staff at Bolt Beranek and Newman generated a set of 574 high-predictability (PH) sentences and 95 medium-predictability (PM) sentences. A conscious attempt was made to introduce some variety in the syntactic structure of the PH and PM sentences, subject to the various constraints noted above. For 95 words, two PH sentences were constructed. Since there was no obvious way to decide in advance which of these two competing sentences to retain, these items were included in the group of PH sentences that were to undergo detailed testing and culling procedures. In addition, a set of 479 low-predictability (PL) sentences (one for each word) was produced by using various combinations of constructions like "John was discussing the ..., I'm glad you spoke about the ..., We are considering the ...," etc.

The sentences resulting from this procedure were entered into computer data files, and were subjected to two kinds of testing: a paper-and-pencil test that yielded a measure of the predictability of the key word for each PH sentence, and a series of listening tests (in speech babble noise) that provided a measure of the intelligibility of the key word for each sentence. In addition, an analysis of the phonetic content of the key words and of the content words in the PH sentences was made.

C. Tests of key-word predictability for PH sentences

Two different paper-and-pencil tests were conducted to determine the predictability of the key words in the 574 PH sentences. In both tests, the sentences were listed on answer sheets with the final word deleted. Subjects were instructed to "fill in the word that you think is most likely to occur at the end," and they were told that the target words were all one-syllable nouns. In one of the tests, no further hints were given. In the other test, the vowel nucleus in the target word was identified for the subjects by giving a prompt word (a proper name) having the same vowel (e.g., Joe, Mike, Mark). The prompt was included to simulate the fact that the vowel nucleus of the target word is often preserved when the entire sentence is presented in noise. A different group of twelve high-school students served as subjects in each test. The mean age was 17 for the unprompted test and 16 for the prompted test.

For purposes of data presentation, we shall define predictability as the number of subjects (out of 12) who wrote down the intended target word. The number of sentences (out of 574) that were in each predictability class is given in Fig. 1. The figure shows that the unprompted predictability was less than the prompted predictability, on the average, as expected, but the difference was not large. The sentences are biased toward the high-predictability end of the scale.

The data from the paper-and-pencil tests were used as a basis for eliminating sentences in which the target words were either highly predictable or minimally predictable. There were 72 sentences for which both prompted and unprompted predictability scores were 12, and these were deleted from the corpus, since the key words in such sentences were judged to be too predictable. An example of a sentence with such a highly-predictable key word is: *Hit the baseball with the bat.* Likewise, sentences which achieved a prompted or an unprompted predictability score of 0 or 1 were also deleted, on the grounds that the key words were not sufficiently predictable. There were 65 such sentences. A typical sentence in this class, with zero for both unprompted and prompted predictability scores is: *He quit the race with a bad cramp.* Thus 137 sentences were eliminated, and each of the remaining 437 sentences was characterized by a key-word predictability in the range 2–12, with the provision that the sentences with a key-word predictability of 12 achieved this score on only the unprompted *or* the prompted test.

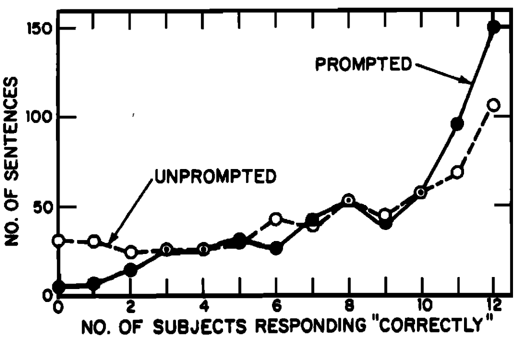

FIG. 1. Summary of performance of 12 subjects in paper-and-pencil tests of predictability of key words in 574 PH sentences, for the unprompted and prompted situations.

TABLE I. Comparison of intelligibility of different key words in a higher predictability (PH) context, a medium-predictability (PM) context, a low-predictability (PL) context, and an artificial sentencelike context (P0). Numbers represent percent of key words correctly identified. Data for PM context are based on 95 words; data for other contexts are based on 479 different words. A score for PH words at +10 dB S/N ratio was not obtained; it would be close to 100 percent.

S/N ratio, (dB)	PH	PM	PL	P0
−2	79	62	42	50
+10	···	93	88	86

D. Tests of intelligibility in noise

An intelligibility testing program with the PH, PM, and PL sentences in noise was carried out using normally hearing subjects. The intelligibility tests with these sentence materials were supplemented by additional tests of the more conventional type, in which the key words were embedded in the neutral context *"I will now say the word ..... ."* These were called the P0 sentences. The purpose of this initial series of tests was to use the intelligibility results for individual sentences as a basis for rejection of sentences with deviant key-word intelligibility, and thus to arrive at sets of sentences with relatively homogeneous intelligibility scores that would be suitable for use in an ultimate test instrument.

The materials were recorded by a male talker with experience in making such recordings. A babble noise was generated on a separate tape by mixing 12 voices reading continuous text. A total of 1722 sentences was recorded; 574 PH sentences, 95 PM sentences, 479 PL sentences, and 574 P0 sentences (95 of which were duplicates). The sentences were divided into lists of about 40 items each, and were presented to a total of 81 college-age subjects, who were instructed to write down the last word of each sentence. Several signal-to-noise ratios were used in a prescribed protocol, with suitable randomization of conditions.

Detailed results for these tests, showing the influence on intelligibility of word familiarity, word predictability, phonetic context, and other factors are presented elsewhere.[29] Some typical data for two different signal-to-noise ratios are summarized in Table I. The results indicate that scores for PH sentences were higher than scores for PM sentences, which are in turn higher than those for PL sentences. Table I compares scores for words in PL context and words in the artificial P0 context, at the two S/N ratios for which both sets of sentences were administered. These data suggest that intelligibility scores for words in the PL context are essentially comparable to word intelligibility determined by more conventional procedures, which usually involve a neutral context similar to the P0 carrier phrase used here. The advantage of the PL sentence over the P0 carrier phrase is that it combines minimum predictability with variable rhythmic structure. It is a pseudo-normal English utterance which can be intermixed with PH sentences without being unduly conspicuous.

On the basis of results of this experiment, a first step in the selection of sentences to be used in an ultimate test instrument was to eliminate PM sentences from consideration. It was felt that a test that included items with the two predictability levels PL and PH would permit an adequate evaluation of the influence of word predictability on listener performance. The intelligibility scores of sentences with intermediate key-word predictability are probably not sufficiently different from either the PH or PL types to permit reliable assessment of a difference in performance traceable to level of predictability, especially when testing with a relatively small number of items.

The intelligibility of individual items in the PH sentences and in the PL and P0 sentences was examined in order to select items with more homogeneous intelligibility scores and to maximize the difference between the scores for PH and PL sentences. Since one aim of an ultimate sentence test is to provide a means for evaluating the extent to which hearing-impaired listeners can take advantage of context in listening to sentences in noise, it is desirable to develop an instrument that makes this difference as large as possible and hence permits the difference to be assessed with reasonable reliability.

Two steps were taken to achieve these goals. First, a measure of the intelligibility of each key word in its PH sentence was determined by computing an average of seven PH scores at various S/N ratios and babble levels. This average score, called PH-AVG, for each word is based on responses of about 40 different listeners under various conditions. A histogram of these scores was generated for the 437 PH items that survived the cut based on predictability scores, and the 55 items with lowest PH scores were then eliminated. For the remaining items, five scores for each of the words in the PL context (at five different S/N ratios) and the two scores in the P0 context (two S/N ratios) were averaged, yielding a quantity called P0L-AVG. A histogram of these scores was created, and the 55 items with the highest scores were eliminated. Of the surviving 327 items, there were 42 pairs of PH sentences based on the same target word (out of the original 95 PH sentences with the same key word). One member of each pair was arbitrarily eliminated, leaving a total of 285 items.

E. Phonetic coding

An analysis was made of the phonetic content of the 285 PH sentences (and key words) that survived the initial selection process. The purposes of this type of analysis were threefold: (1) to determine whether the phonetic content of the words and sentences is representative of utterances used by speakers of American English (to the extent that data on American English utterances are available); (2) to determine whether particular sentences are heavily weighted toward certain phonetic classes, in which case there might be grounds for rejecting these sentences; (3) to use the phonetic analysis as a basis for selecting 25-item sets to be included in different test forms, each of which should have a similar phonetic profile.

For each PH sentence, the content words that were judged to be related semantically to the final key word were marked. Thus, for example, in the sentence "The *king wore* a *golden crown*," the italicized words provide semantic pointers to the key word. The phonetic content of each of these pointer words and of the key word was assessed by counting the number of occurrences of segments in each of several phonetic categories that were considered to be the primary phonetic determinants of intelligibility. The phonetic categories that were used are listed in Table II. Counts were made both for the pointer-plus-key words and for the key words alone. In addition, the number of syllables and the number of words in each PH sentence were recorded, as shown at the bottom of each part of the table.

The phonetic data for each sentence and for each key word were entered into computer data files. From the phonetic counts for the PH sentences, distributions of the number of occurrences of different phonetic categories were prepared. An example of such a distribu-

TABLE II. Example of phonetic counts. The symbols in parentheses under the consonant classes are mnemonics that are used in Fig. 3.

Sentence: The *king wore* a *golden crown.*
Pointer words: king, wore, golden
Key word: crown

Consonants			Vowels		
Voiceless plosives (NVP)	sent. word	2 1	Front	sent. word	1 0
Voiced plosives (VP)	sent. word	2 0	Back	sent. word	2 0
Voiceless fricatives (NVF)	sent. word	0 0	High	sent. word	1 0
Voiced fricatives (VF)	sent. word	0 0	Nonhigh	sent. word	2 0
Nasals (NAS)	sent. word	3 1	Reduced	sent. word	1 0
Liquids, semivowels	sent. word	4 1	Diphthongs	sent. word	1 1
No. of syllables:	7				
No. of vowels:	6				

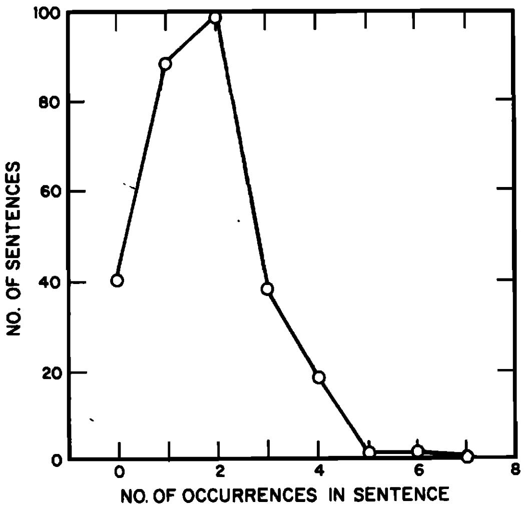

FIG. 2. Distribution of sentences according to number of occurrences of voiceless fricative consonants in PH pointer words and key words. $N = 285$ sentences.

tion for voiceless fricative consonants is shown in Fig. 2. This figure indicates that most of the sentences contain one or two voiceless fricatives, but there is one sentence with five such consonants and one with six. (An example of a sentence of this kind is *The sandwich was Swiss cheese and ham.*) Based on this analysis, these two sentences were eliminated from the corpus. Several other sentences were also eliminated on the basis of similar kinds of observations for other phonetic categories.

A final scanning of the sentences that survived this process revealed some PH sentences that, despite previous culls, seemed too highly predictable or "sloganlike," and some PL sentences which, because of the abstract nature of the target word, were semantically somewhat abnormal. These sentences were eliminated.

IV. DEVELOPMENT OF EQUIVALENT TEST FORMS

A. Initial selection and recording of equivalent test forms

The result of the various culling procedures described above was a set of 250 PH sentences, together with a corresponding set of 250 PL sentences based on the same set of key words. These sentences were next divided into a series of lists of 50 sentences each. The aim was to generate several lists of forms of a test with the following characteristics: each form contains 50 items, consisting of 25 different PH sentences and 25 different PL sentences; the PH and PL items are interleaved in a random fashion, but with different key words for the different sentence types (i.e., a given key word appears only once in each form); and each form is equivalent to each other form with respect to phonetic content, average intelligibility of PH and PL items, predictability of key words in the PH sentences, and word familiarity. The objective was to develop a set of test forms such that testing of a hearing-impaired individual with just *one* form would yield reliable scores for words in PH sentences and for words in PL sentences, and that these scores would be independent of the test form used.

The 250 words were assigned into 10 sets of 25 each, such that each set had roughly the same mean value of PH-AVG and POL-AVG. These sets of 25 words were then scrutinized for average phonetic content, word predictability (PH sentences), and word familiarity. When particular 25-word sets showed appreciable deviation from the mean for these characteristics, items were exchanged between sets in order to achieve a better balance, while maintaining similar average intelligibility scores. The result of this process was 10 sets of 25 words, each of which identified a PH and a PL sentence. These sets were then paired to produce ten 50-item test forms, each of which contained an equal number of PH and PL sentences, randomly interleaved. Each odd-numbered test form had an even-numbered counterpart with the same key words, but the words were embedded in the opposite type of sentence context. A further constraint was that each half of each test form contained 12 or 13 PH sentences and 12 or 13 PL sentences. Within this constraint, the ordering of the sentences was pseudo-random (runs of no more than three sentences of a given predictability class being proscribed), except that a final informal scanning of the lists revealed some adjacent sentences that were semantically related. Some reordering was done to eliminate such sequences.

These ten forms were then re-recorded by the same male speaker who had made the original recordings. The recordings were copied onto one track of a magnetic tape, and a 12-speaker babble at the same rms level (measured according to procedures described in Pearsons, Bennett, and Fidell[30]) was recorded on the second channel. In order to compensate for possible long-term variations in speaking level, adjustments in level were made in the dubbing process so that the average level (rms) of the test sentences in each half-list (25 items) was the same, within 0.1 dB. Appropriate calibration tones were inserted on both speech and babble tracks before the recording of each test form.

B. Tests of form equivalence

Although the various test forms were constructed from items that had given approximately equal PH scores and equal PL scores (averaged over a variety of signal-to-noise ratios), it was considered desirable to conduct further testing to determine directly whether the newly recorded forms gave equivalent scores. These tests were carried out using 80 listeners who were screened for normal hearing (less than 20 dB hearing level in the range 500–4000 Hz) using a Békésy audiometer. Their native language was English. The forms were heard at a 0-dB S/N ratio with monaural earphone presentation. The SPL of both signal and noise was 80 dB *re* 0.0002 dyn/cm^2. The S/N ratio of 0 dB was selected because previous data had indicated that the maximum difference between PH and PL scores was to be found in that region. (In clinical testing with hearing-impaired individuals, it will be desirable in most instances to select an SPL and an S/N ratio appropriate to the patient's individual sensitivity; in this way the possibility of obtaining good separation between PH and PL scores is maximized.)

The experimental design was aimed at evaluating the equivalence of the 10 test forms, but the issue was complicated by the presumed influence on form performance of the babble track that was time-locked to the test sentences. We decided to conduct a two-part test. In the first part, all ten test forms were presented (each with a given babble time-lock associated with it) to a group of listeners, with each listener responding only to two forms of the test. The forms so tested were designated by a form number identifying the sentence (from 1.1 through 1.10). Since each form could be paired with only eight others (excluding the form which contained the same key words in opposite-predictability sentences), and since both orders of presentation of each pair were to be tested (to compensate for possible learning effects), there were 80 possible pairs and hence 80 listeners. Their ages ranged from 28 to 50 years (mean = 38), and there were 34 males and 46 females. For this set of data on the ten forms with the original babble (designated as babble A) time-locked to the sentences, therefore, there are 16 different individuals responding to each form.

A subset of 25 individuals served in a second part of the study directed at measuring the possible influence of the babble version. Two of the test forms (1.1 and 1.10) were prepared with babble version B—simply an arbitrarily selected portion of the master babble recording different from Version A. The listeners participating in this subexperiment thus heard two test forms in addition to the two forms with babble A. Listeners providing data for the babble-B versions of forms 1.1 or 1.10 had not previously heard those same sentences in their initial pair of forms.

Prior to hearing the first actual test form, listeners were instructed and allowed to practice the task using a standardized test form. They heard from 15 to 25 practice sentences at a + 10 dB S/N ratio.

Data are shown in Table III for the 10 forms in their babble-A version, based on 16 listeners per form, and for the babble-B versions of forms 1.1 and 1.10, with 25 listeners. The standard deviations of the distributions are in the range normally expected for speech discrimination tests (see, for example, Rintelmann *et al.*[31]). The dispersions of the difference scores are generally in the same range as those of the PH and PL scores, suggesting that PH and PL scores are correlated, and that the reliability of the difference score is about the same as that of the two scores forming that difference.

TABLE III. Mean percent of key words correct, and standard errors of the mean, for each of the original 10 forms of the sentence test, some with two versions of the babble track. The talker is male and the S/N ratio is 0 dB. Scores for "A" versions of the babble are based on data from 16 listeners; for "B" versions, on data from 25 listeners. Forms showing a "-" under "New Form No." were deleted from the final set of test forms.

Orig. Form No.	Babble Version	New Form No.	PH Mean	PH s.e.	PL Mean	PL s.e.	PH–PL Mean	PH–PL s.e.
1.1	A	-	90.5	2.0	25.3	3.1	65.3	3.1
1.1	B	2.1	88.2	1.7	38.2	2.4	49.9	2.6
1.2	A	2.2	91.0	1.7	43.7	2.9	47.2	2.1
1.3	A	-	80.3	1.8	28.0	1.7	52.2	1.9
1.4	A	-	87.0	1.8	46.0	2.3	41.0	2.0
1.5	A	2.3	82.0	3.4	33.5	3.0	48.5	3.4
1.6	A	2.4	91.8	1.2	44.0	3.1	47.7	2.4
1.7	A	2.5	87.3	1.6	45.0	1.7	42.3	1.3
1.8	A	2.6	89.0	2.4	44.5	2.1	44.5	2.8
1.9	A	2.7	82.0	1.6	33.0	1.9	49.0	1.8
1.10	A	-	92.0	1.8	31.5	1.6	60.5	2.0
1.10	B	2.8	90.4	1.2	40.5	1.9	49.9	2.1
Mean:			87.6		37.8		49.8	
Range:			80.3		25.3		41.0	
			92.0		46.0		60.5	

Three separate one-way analyses of variance were performed on these data, for the PH scores, the PL scores, and the difference scores. Each analysis involved 12 "treatments" (i.e., a different set of sentences and/or babble version), and each assumed independence between the data in the various treatments. To the extent that the same listeners provided data for more than one treatment, that assumption will lead to artifactually high sensitivity to differences between treatments, and therefore to a conservative test. The three analyses showed significant differences between the treatments (F ratios of 4.4, 8.6, and 7.7, respectively, df $= \frac{11}{198}$, $p < 0.001$ in all cases). We concentrated our attention on the treatments resulting in outliers: Form 1.1, babble A (low PL, high difference); Form 1.10, babble A (high PH, high difference); Form 1.3, babble A (low PH, low PL); and form 1.4, babble A (high PL, low difference). We reasoned that if we substituted the babble-B versions of Forms 1.1 and 1.10, we could reduce the range of several performance parameters across the remaining forms; and if a case could be made for eliminating either Form 1.3 or 1.4, the other would perforce be eliminated because they are a complementary pair. We carried out *post hoc* comparisons[32] on the performance data of the four above treatments in an effort to evaluate the probability that a given treatment's data had been drawn from the population defined by the data of the remaining set of 8 treatments we wished to retain. We found that there was reason to reject the hypothesis of equivalence, at the 5% level of significance, for all but Form 1.4, which was on the borderline.

Table IV summarizes the data for the remaining 8 forms and associates a new form number, ranging from 2.1 through 2.8, with each set of data. The three analyses of variance, on PH, PL, and difference score, were redone for the reduced set. The null hypothesis of form equivalence was rejected for the PH and PL analysis, and accepted for the analysis of the difference scores (F ratios of 3.6, 3.7, and 1.2 respectively, df $= \frac{7}{136}$, reject at $p < 0.05$ for the first two). Since the statistical model is conservative (in that it overstates differences when listeners are shared between two treatments), the apparent lack of equivalence in PH and PL scores across the forms should not be regarded as a serious problem, particularly in view of the equivalence of the difference scores.

The current version of this test, as it exists in these eight equivalent forms, has been called the SPIN test (Speech Perception in Noise). For normally hearing listeners, at 0 dB S/N and for the various other stimu-

TABLE IV. Performance properties of the current set of eight forms of the SPIN test that meet the specified criteria of homogeneity for normally-hearing subjects at 0 dB S/N ratio. Scores are in percent.

New Form No.	PH Mean	PH s.e.	PL Mean	PL s.e.	PH–PL Mean	PH–PL s.e.
2.1	88.2	1.7	38.2	2.4	49.9	2.6
2.2	91.0	1.7	43.7	2.9	47.2	2.1
2.3	82.0	3.4	33.5	3.0	48.5	3.4
2.4	91.8	1.2	44.0	3.1	47.7	2.4
2.5	87.3	1.6	45.0	1.7	42.3	1.3
2.6	89.0	2.4	44.5	2.1	44.5	2.8
2.7	82.0	1.6	33.0	1.9	49.0	1.8
2.8	90.4	1.2	40.5	1.9	49.9	2.1
Mean:	87.7		40.3		47.4	
Range:	82.0 91.8		33.0 45.0		42.3 49.9	

lus and experimental parameters noted above, the range of scores shown in Table IV for different test forms corresponds roughly to a shift of 1 dB in S/N ratio for PH sentences and 2 dB for PL sentences. This was considered to be an adequate degree of homogeneity for a clinical test instrument.

The eight test forms that survived this intensive screening process are listed in Appendix A.

V. PHONETIC, LEXICAL, AND OTHER LINGUISTIC PROPERTIES OF THE SENTENCES IN THE FINAL TEST INSTRUMENT

We summarize now some of the properties of the finally selected set of test forms listed in Appendix A. These forms exist as a set of two-track recordings with the sentences on one track and babble on the second track. In Appendix B we describe the acoustic characteristics of this particular implementation of the forms by a male speaker, and of the babble-type masking noise that accompanies the recordings.

As noted earlier, an analysis was made of the phonetic content of the key words in all the sentences and of the content words in the 285 PH sentences that survived the initial culling process. For the sentences that survived the final culling process, i.e., the sentences listed in Appendix A, these phonetic data are summarized in Fig. 3. Two phonetic counts are shown: one for the 200 key words, and one for the total set of content words (including key words) in the PH sentences. Phonetic counts based primarily on written English and reported by Dewey[33] are also displayed in the figure. These counts for the different phonetic categories are not substantially different from the data from telephone conversations reported by French, Carter, and Koenig,[34] as retranscribed by Tobias.[35]

There is reasonable agreement between the Dewey phonetic profile and the two profiles for the test sentences, except in certain categories. For example, because of the constraint that key words be monosyllabic nouns, they obviously contain no unstressed syllables. Furthermore, because function words were not included in the phonetic counts for the words in the sentences, the occurrence of certain classes of segments is less frequent than that indicated by Dewey (e.g., the segment [ð] in the frequent word *the* inflates the voiced-fricative count in the Dewey data). With these minor and explainable exceptions, it can be concluded that the sentences in the proposed test instrument have phonetic profiles that are representative of English.

Similar phonetic counts were made for the sentences within individual forms of the test. These data are summarized in Table V. No significant deviations from the average are evident in particular forms of the test. Attention was, of course, paid to phonetic balance in generating the different forms, and hence Table V simply indicates that degree to which this goal was achieved. It should be noted that phonetic balance in this set of forms means a phonetic profile comparable to that of the language, and not an equal representation of different phonetic classes.

Also given in Table V are the number of syllables and the number of words in the PH sentences, and the average number of syllables in the semantically related pointer words. Again, the average lengths are comparable across forms.

The average key-word predictability and key-word familiarity for each form are listed in Table VI. As noted earlier, predictability is defined here as the proportion of subjects (out of 12) who were able to fill in the intended key word for the PH sentences in a paper-and-pencil test, averaged over an unprompted condition and a prompted condition in which the vowel nucleus of the

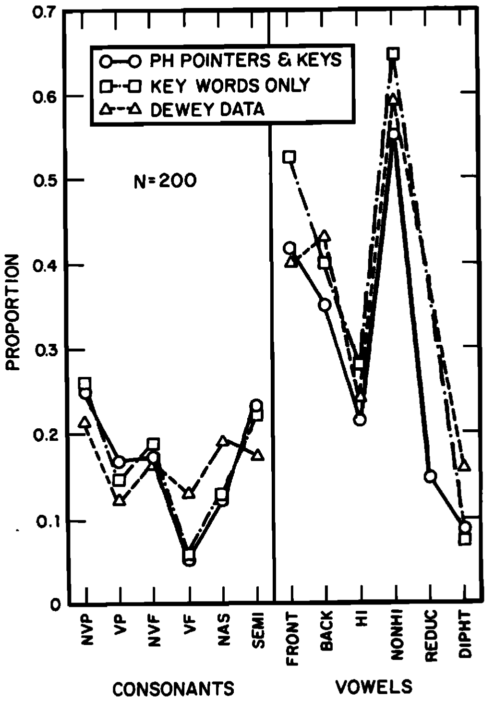

FIG. 3. Distribution of occurrences of different phonetic classes in the final corpus of 200 items, compared with data from Dewey[33]. See Table II for interpretation of phonetic classes.

TABLE V. Summary of phonetic content of words in the 8 forms of the SPIN tests listed in Appendix A. The entries represent the percent of phonetic segments in the 25 PH sentences that are in the phonetic class indicated. The table gives characteristics of key words alone (upper section) and characteristics of all content words, including key words (lower section). The number of "pointer" syllables designates the number of syllables in the content words that are semantically related to the key words. The column labeled "Dewey" gives the phonetic counts for American English (in percent) from Dewey.[33] As noted in the text, the key words for PL sentences in odd-numbered forms are the same as the key words for PH sentences in the next higher even-numbered forms, and vice versa.

	Key Words									
	Test Form									
Phonetic Class	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	Mean	S. D.
voiceless plosives	32.4	28.5	23.2	21.8	25.0	23.1	29.6	23.5	25.9	3.8
voiced plosives	10.8	17.8	16.0	10.9	15.0	12.3	15.6	19.1	14.7	3.1
voiceless fricatives	24.3	14.2	17.4	21.8	20.0	13.9	18.7	19.1	18.7	3.5
voiced fricatives	2.7	7.1	5.8	6.2	3.3	7.7	4.7	8.8	5.8	2.1
nasals	10.8	16.0	11.6	17.2	11.7	16.9	12.5	4.4	12.6	4.2
semivowels	18.9	16.0	26.1	21.8	25.0	26.2	18.7	25.0	22.2	3.9
consonant sum	74	56	69	64	60	65	64	68	65	5.5
front vowels	60	60	52	44	60	52	48	44	52.5	6.9
back vowels	36	24	44	44	36	40	48	48	40.0	8.0
high vowels	32	16	36	32	32	28	28	20	28.0	6.8
nonhigh vowels	64	68	60	56	64	64	68	72	64.5	5.0
diphthongs	4	16	4	12	4	8	4	8	7.5	4.5
vowel sum	25	25	25	25	25	25	25	25	25	25

	All Content Words										
	Test Form										
Phonetic Class	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	Mean	S. D.	Dewey
voiceless plosives	22.0	24.3	26.1	28.6	24.2	23.3	27.3	24.6	25.0	2.2	19.1
voiced plosives	15.9	20.2	15.4	13.1	19.5	14.3	16.1	18.8	16.7	2.6	11.0
voiceless fricatives	16.3	13.6	19.0	20.0	15.7	17.6	19.0	17.9	17.4	2.1	16.0
voiced fricatives	6.4	3.7	5.5	3.7	5.5	5.3	5.8	6.2	5.3	1.0	14.8
nasals	13.6	14.8	10.7	14.3	10.6	12.7	12.8	10.0	12.4	1.8	17.8
semivowels	25.7	23.4	23.3	20.4	24.6	26.6	19.0	22.5	23.2	2.6	21.4
consonant sum	264	243	253	245	236	244	242	240	246	8.8	
front vowels	40.0	49.1	43.9	38.5	44.6	39.4	37.2	42.0	41.8	3.90	52.0
back vowels	33.9	30.2	36.0	38.5	33.9	34.6	37.2	33.6	34.7	2.54	41.2
high vowels	20.0	19.0	29.8	22.1	22.3	19.2	20.4	19.6	21.6	3.56	33.4
nonhigh vowels	53.9	60.3	50.0	54.8	56.2	54.8	54.0	56.1	55.0	2.88	59.8
diphthongs	8.7	7.8	3.5	9.6	8.0	12.5	9.7	8.4	8.6	2.51	6.8
reduced vowels	17.4	12.9	16.7	13.5	13.4	13.5	15.9	15.9	14.9	1.76	···
vowel sum	115	116	114	104	112	104	113	107	111	4.90	
No. "pointer" syllables (per sentence)	3.60	3.64	3.56	3.16	3.48	3.16	3.52	3.28	3.42	0.20	
No. syllables	187	192	190	182	187	186	188	183	187	3.3	
No. words	161	168	162	165	164	166	162	158	163	3.1	

intended word was identified for the subject. Data on word familiarity were taken from the Thorndike–Lorge counts, and the variation of these data from form to form is small in comparison with the range (5 to 150 per million) from which the original selection of words was made.

VI. LISTENER PERFORMANCE WITH TEST INSTRUMENT

A. Performance versus S/N ratio

After the culling had been completed to the point where 250 PH sentences and 250 PL sentences had been assembled into ten equivalent test forms, an experiment was carried out with 20 normally hearing subjects to determine the performance for this reduced set of sentences at various S/N ratios. This experiment was done with the ten forms recorded with babble A, i.e., before two of the ten forms were eliminated and before babble B was selected as the babble track to be used for another two of the forms. The slight error in mean performance caused by averaging over the ten forms used in this experiment rather than the 8 forms finally selected is very small (cf. Tables III and IV), and the results of this experiment are taken to be representative of the final test instrument listed in Appendix A.

TABLE VI. Mean predictability and median familiarity of key words in PH sentences, final set of eight test forms. Predictability is defined as the proportion of subjects who correctly guessed the key word in a paper-and-pencil test. Familiarity data are from Thorndike–Lorge counts.

Form	Predictability (mean of prompted and unprompted tests)	Familiarity (median, per million words)
2.1	0.67	44
2.2	0.68	30
2.3	0.67	32
2.4	0.78	40
2.5	0.65	37
2.6	0.72	33
2.7	0.76	45
2.8	0.72	34

While the principal objective of this experiment was verification of the expected separation between PH and PL functions for normally hearing listeners, a subsidiary aim was to investigate the effects of age on performance, for normally hearing subjects. Two groups of subjects were used: ten young subjects ranging in age from 18 through 25 years, and ten subjects between 60 and 75 years of age. Both groups listened to ten different test forms presented at various S/N ratios from − 5 to + 10 dB. The overall speech level was always 80 dB *re* 0.0002 dyn/cm^2. All of the subjects had better than 20 dB hearing level for pure tones at all frequencies through 4 kHz, as determined by Békésy audiometry. The order of presentation of test forms and S/N ratios was partially counterbalanced [based on Graeco-Latin Square principles (cf. Fisher and Yates, [36])] so that form equivalence, learning effects, and effects of age could be assessed.

Functions relating PH and PL scores to S/N ratio for young and old subjects are shown in Fig. 4. The data show the expected separation between the two functions, the PH functions rising much more rapidly than the PL functions. That is, the intelligibility of the words in PH context is more homogeneous than the intelligibility of the words in PL context. Both functions for the older subjects are slightly below those for the younger subjects. This difference can be ascribed either to the presumed greater hearing loss for the older Ss at higher frequencies (even though the pure-tone loss for both groups is less than 20 dB up to 4 kHz) or to a loss in cognitive abilities for the older group. There is no evidence, however, that the difference between PH and PL scores becomes smaller for the older Ss. That is, the older individuals are as adept as the younger Ss at taking advantage of the sentence context in order to gain improved performance for the PH sentences.

Comparison of PH and PL scores is shown in a different way in Fig. 5, in which the algebraic difference between PH and PL score is plotted against PL performance. This difference can be regarded as a measure of the contribution of context. The negative diagonal indicates the limits of performance: it is impossible for a datum point to appear above that line. The region of principal interest is where the PH–PL differences reach their maxima. For both groups, this maximum difference appears to be at around 40% or 50%. Interestingly, the older Ss' maximum appears to be shifted leftwards with respect to the younger group. This apparent shift may, however, be a consequence of the fact that testing was not done at S/N ratios between − 5 and 0 dB, where the younger subjects may show a peak in the PH–PL curve.

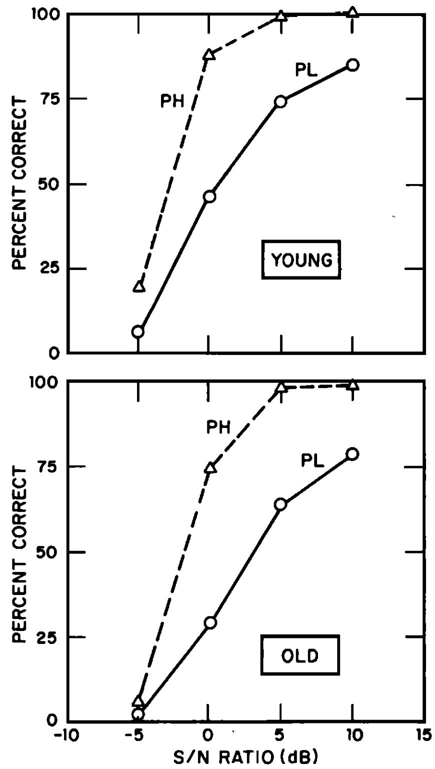

FIG. 4. PH and PL scores versus S/N ratio for the original 10 test forms (250 items of each type), for normally hearing young and old listeners.

B. Effects of learning on performance

Since the test instrument is designed for clinical use in which efficient and rapid testing is a prerequisite, it

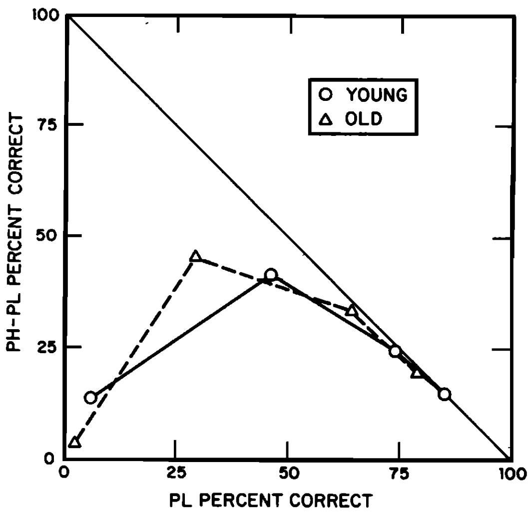

FIG. 5. PH minus PL scores versus PL score, for old and young listeners. The diagonal line corresponds to a PH score of 100 percent.

TABLE VII. Comparison of average scores (in percent words correct) for the test forms administered first and second, showing the effects of learning. Each score is based on data from 80 listeners. Signal-to-noise ratio is 0 dB.

	PH score	PL score
First test	86.0	36.8
Second test	88.5	38.0

is important to determine the effects of learning on performance on the SPIN test. If there are substantial effects of learning when successive test forms are presented, or even within a given test form, then the value of the test as a diagnostic instrument diminishes. The results of the experiment that was designed to test for form equivalence can be used to provide some indication of learning. In that experiment, subjects heard two test forms in sequence, always at 0 dB S/N ratio. All forms were represented equally in first and second positions. Thus a comparison of average performance on the first and second tests gives a measure of the effect of learning. The data, given in Table VII, show that there was a very slight effect of learning on the administration of a second form. This learning effect amounted to only 2.5% on the PH score and 1.2% on the PL score, on the average. This effect is much smaller than the expected variability across test forms, and is equivalent to the change in performance that would occur for a shift in S/N ratio of only 0.2 dB. It should be noted, however, that formal testing in this experiment was preceded by 15–25 practice items, presented at a S/N ratio of +10 dB, so that any possible initial transient in performance due to lack of orientation to the test situation was presumably minimized.

VII. DISCUSSION

The sentence tests that have evolved from this work have characteristics that make them potentially useful in several applications. An obvious application, noted earlier, is in speech testing of the hearing impaired. One form of the test, at one signal-to-noise ratio, can be administered in about 10 min. Each of 50 items requires a simple one-word response from the subject. The test leads to two scores—a score for low-predictability sentences (similar to more conventional monosyllabic word tests), and a score for high-predictability sentences. Some combination of these two scores has the potential of predicting the ability of a hearing-impaired listener to perform in everyday communicative situations, and thus may help to estimate the benefit that the individual is likely to gain from a hearing aid. The predictive value of the test would presumably be enhanced if it were administered at more than one signal-to-noise ratio, representative of a range of conditions encountered in everyday situations. Suggested protocols for administering the test to hearing-impaired listeners in a clinical setting are discussed elsewhere.[1]

Tests of this type could have application beyond use in evaluating auditory speech-processing capabilities of the hearing-impaired. To the extent that the test provides a measure of the involvement of cognitive and memory processes in speech perception, it might be assumed that it can assess these capabilities in individuals suspected of deficiencies in these aspects of speech comprehension. If a small difference is found between PH and PL scores, then some deficiencies in cognitive and memory processes might be suspected. The validity of this suggestion must, however, be determined through testing of appropriate population of subjects.

A related application of the sentence material is in testing of comprehension of English for those who are learning it as a second language. Differences between PH and PL scores at an appropriate (possibly high) S/N ratio would presumably indicate the degree to which the listener has mastered the ability to profit from the semantic, syntactic, and prosodic information provided by the sentence context and to conduct the rapid lexical searches necessary for sentence comprehension.

This sentence test could also be useful as a component of a battery of tests to be used for evaluating the benefit derived from nonauditory aids for those with severe or profound hearing loss. It is reasonable to assume that, for such a tactile or visual aid to be effective, possibly in conjunction with speech reading, a user should be able to profit from sentence context, and should be able to perform more effectively with sentences of the PH type than with PL sentences. Testing with these sentences should indicate the extent to which such a person is able to make use of this context.

Still another potential use of sentence material of the kind described here is in the evaluation of speech processing devices of various kinds. For example, devices that attempt to synthesize speech by rule from a phonetic string are known to produce reasonable approximations to most segmental aspects of speech, but to generate rather unnatural prosodic characteristics in sentences.[37] A measure of the degree to which the sentences produced by the synthesizer approximate natural sentences would be provided by the PH and PL scores when sentences like those in Appendix A are generated by rule. A similar type of evaluation can be made of the speech produced by individuals who are learning English as a second language, or for whom there is evidence of speech pathology, such as profoundly deaf children, or those with cleft palate.

In addition to providing a practical means for assessing potential deficits in an individual's ability to process speech, work with sentences of the type discussed here could lead to insights into the process of sentence decoding by normal listeners—a process that is only poorly understood at present. The procedure would be to manipulate various aspects of sentences and to determine the effects of these manipulations on listener performance. One such factor—the semantic links between words in the sentences—has been shown in this and in other studies to exert a strong influence on performance. This factor needs to be examined in more detail, with a more precise specification of the semantic and syntactic constraints that link the words. In this experimental situation it would be appropriate to obtain responses to the entire sentence rather than just to the key word. The effect of sentence length on performance should also be assessed, in order to indicate the constraints imposed by short-term memory in sentence decoding. The role played by fundamental-frequency (F_0) contours in the decoding process could be examined through manipulation of these contours by means of vocoder-type devices. Reducing the F_0 contour to a monotone, for example, may not influence substantially the performance for PL sentences, but could have a significant effect on scores for PH sentences, since the function of the F_0 contour in grouping the words into larger units would be missing.

Experimental studies such as these could help to lead to a more adequate model of sentence processing. Such a model would then provide a much sounder basis on which to construct sentence materials that can be used to assess deficits in everyday speech communication.

ACKNOWLEDGMENTS

Dr. Lois Elliott originated the project and served as the Technical Monitor of this work, during her tenure as Head of Directed Research in Communicative Disorders, NINCDS.

The contributions of the following members of the BBN technical staff are gratefully acknowledged: B. Freeman, for invaluable assistance in compiling and analyzing the data; A. W. F. Huggins, for many contributions to the experimental design; the two preceding people, plus J. Berliner and W. Hawkins, for their assistance in production and discussion of the corpus of sentence materials; R. S. Nickerson, for helpful comments and guidance; and K. S. Pearsons, S. Fidell, R. Bennett, B. Curtis, and S. Tomooka for their painstaking production of the recorded stimulus materials, and for their assistance in gathering the data reported herein.

We also wish to thank H. L. Gerstman of the Tufts-New England Medical Center Hospital for contributions to the experimental design, and to give special thanks to R. Morrison and H. Schuknecht of the Massachusetts Eye and Ear Infirmary for making their patients and facilities available to this project.

APPENDIX A. SPIN TEST SENTENCE LISTS

The H or L at the left of each sentence indicates whether the key word has high or low predictability.

Form 2.1

(H) 1. The watchdog gave a warning growl.
(H) 2. She made the bed with clean sheets.
(L) 3. The old man discussed the dive.
(L) 4. Bob heard Paul called about the strips.
(L) 5. I should have considered the map.
(H) 6. The old train was powered by steam.
(H) 7. He caught the fish in his net.
(L) 8. Miss Brown shouldn't discuss the sand.
(H) 9. Close the window to stop the draft.
(H) 10. My T.V. has a twelve-inch screen.
(L) 11. They might have considered the hive.
(L) 12. David has discussed the dent.
(H) 13. The sandal has a broken strap.
(H) 14. The boat sailed along the coast.
(H) 15. Crocodiles live in muddy swamps.
(L) 16. He can't consider the crib.
(H) 17. The farmer harvested his crop.
(H) 18. All the flowers were in bloom.
(L) 19. I am thinking about the knife.
(L) 20. David does not discuss the hug.
(H) 21. She wore a feather in her cap.
(L) 22. We've been discussing the crates.
(L) 23. Miss Black knew about the doll.
(H) 24. The Admiral commands the fleet.
(L) 25. She couldn't discuss the pine.
(L) 26. Miss Black thought about the lap.
(H) 27. The beer drinkers raised their mugs.
(H) 28. He was hit by a poisoned dart.
(H) 29. The bread was made from whole wheat.
(L) 30. Mr. Black knew about the pad.
(L) 31. You heard Jane called about the van.
(H) 32. I made the phone call from a booth.
(L) 33. Tom wants to know about the cake.
(L) 34. She's spoken about the bomb.
(H) 35. The cut on his knee formed a scab.
(L) 36. We hear you called about the lock.
(L) 37. The old man discussed the yell.
(H) 38. His boss made him work like a slave.
(H) 39. The farmer baled the hay.
(L) 40. They're glad we heard about the track.
(H) 41. A termite looks like an ant.
(H) 42. Air mail requires a special stamp.
(H) 43. Football is a dangerous sport.
(L) 44. Sue was interested in the bruise.
(L) 45. Ruth will consider the herd.
(H) 46. We saw a flock of wild geese.
(L) 47. The girl talked about the gin.
(L) 48. Paul can't discuss the wax.
(H) 49. Drop the coin through the slot.
(L) 50. I hope Paul asked about the mate.

Form 2.2

(L) 1. You're glad they heard about the slave.
(L) 2. The girl knows about the swamps.
(H) 3. Hold the baby on your lap.
(H) 4. For your birthday I baked a cake.
(H) 5. The railroad train ran off the track.
(L) 6. They did not discuss the screen.
(L) 7. They were interested in the strap.
(H) 8. Tear off some paper from the pad.
(L) 9. I had a problem with the bloom.
(L) 10. Peter should speak about the mugs.
(H) 11. The fruit was shipped in wooden crates.
(H) 12. The rancher rounded up his herd.
(L) 13. She wants to speak about the ant.
(L) 14. We're discussing the sheets.
(L) 15. The boy would discuss the scab.
(H) 16. The lonely bird searched for its mate.
(L) 17. Tom could have thought about the sport.
(L) 18. You'd been considering the geese.
(H) 19. They drank a whole bottle of gin.
(H) 20. On the beach we play in the sand.
(L) 21. Mr. Black considered the fleet.
(H) 22. The airplane went into a dive.
(H) 23. We're lost so let's look at the map.
(L) 24. I want to know about the crop.
(H) 25. Household goods are moved in a van.
(H) 26. The honey bees swarmed round the hive.
(L) 27. Betty has talked about the draft.
(L) 28. Tom discussed the hay.
(L) 29. Jane was interested in the stamp.
(H) 30. The airplane dropped a bomb.
(H) 31. Cut the bacon into strips.
(L) 32. I had not thought about the growl.
(H) 33. The drowning man let out a yell.
(H) 34. I gave her a kiss and a hug.
(L) 35. Paul should know about the net.
(H) 36. I cut my finger with a knife.
(H) 37. The candle flame melted the wax.
(L) 38. Tom heard Jane called about the booth.
(L) 39. We can't consider the wheat.
(H) 40. This key won't fit in the lock.
(L) 41. We have not discussed the steam.
(L) 42. Miss Brown might consider the coast.
(L) 43. Mr. Brown can't discuss the slot.
(H) 44. The little girl cuddled her doll.
(H) 45. Tom fell down and got a bad bruise.
(L) 46. He hasn't considered the dart.
(H) 47. The furniture was made of pine.
(H) 48. How did your car get that dent?
(L) 49. Mr. Smith thinks about the cap.
(H) 50. The baby slept in his crib.

Form 2.3

(H) 1. A rose bush has prickly thorns.
(L) 2. We should have considered the juice.
(H) 3. The shipwrecked sailors built a raft.
(L) 4. Bob could have known about the spoon.
(H) 5. Ruth poured the water down the drain.
(H) 6. The boy gave the football a kick.
(L) 7. Bill might discuss the foam.
(H) 8. The cop wore a bullet-proof vest.
(L) 9. Tom could not discuss the barn.
(L) 10. You were considering the gang.
(H) 11. After his bath he wore a robe.
(L) 12. Nancy should consider the fist.
(H) 13. I can't guess so give me a hint.
(H) 14. The soup was served in a bowl.
(L) 15. I've spoken about the pile.
(L) 16. Jane has a problem with the coin.
(H) 17. The bomb exploded with a blast.
(L) 18. Mary could not discuss the tack.
(L) 19. They have a problem with the limb.
(L) 20. Nancy had considered the sleeves.
(H) 21. Lubricate the car with grease.
(H) 22. The workers are digging a ditch.
(L) 23. Bill heard Tom called about the coach.
(H) 24. They marched to the beat of the drum.
(H) 25. No one was injured in the crash.
(L) 26. The old man thinks about the mast.
(H) 27. The sailor swabbed the deck.
(L) 28. Tom will discuss the swan.
(L) 29. Ann was interested in the breath.
(H) 30. This nozzle sprays a fine mist.
(L) 31. Ruth hopes he heard about the hips.
(L) 32. Tom is talking about the fee.
(L) 33. Miss Smith considered the scare.
(H) 34. The ship's Captain summoned his crew.
(H) 35. They fished in the babbling brook.
(H) 36. The hockey player scored a goal.
(L) 37. David should consider the blame.
(H) 38. They played a game of cat and mouse.
(L) 39. He's glad you called about the jar.
(L) 40. Tom will discuss the cot.
(H) 41. The steamship left on a cruise.
(H) 42. She faced them with a foolish grin.
(L) 43. He hopes Tom asked about the bar.
(L) 44. Miss Black could have discussed the rope.
(H) 45. A chimpanzee is an ape.
(H) 46. He wiped the sink with a sponge.
(H) 47. We shipped the furniture by truck.
(L) 48. Ruth's Grandmother discussed the broom.
(L) 49. I've been considering the crown.
(H) 50. A bear has a thick coat of fur.

Form 2.4

(L) 1. I want to speak about the crash.
(H) 2. Harry slept on the folding cot.
(L) 3. She's glad Jane asked about the drain.
(H) 4. The doctor charged a low fee.
(L) 5. He had considered the robe.
(L) 6. I haven't discussed the sponge.
(H) 7. The guilty one should take the blame.
(L) 8. You cannot have discussed the grease.
(H) 9. The cookies were kept in a jar.
(H) 10. Let's invite the whole gang.
(L) 11. Mr. White discussed the cruise.
(H) 12. The sport shirt has short sleeves.
(L) 13. They knew about the fur.
(L) 14. We've spoken about the truck.
(H) 15. The cushion was filled with foam.
(H) 16. How long can you hold your breath?
(L) 17. She wants to talk about the crew.
(H) 18. The cow was milked in the barn.
(H) 19. That accident gave me a scare.
(H) 20. The kitten climbed out on a limb.
(L) 21. You're glad she called about the bowl.
(L) 22. The man could not discuss the mouse.
(H) 23. He tossed the drowning man a rope.
(L) 24. You hope they asked about the vest.
(L) 25. You want to talk about the ditch.
(H) 26. Stir your coffee with a spoon.
(L) 27. We hear she called about the drum.
(H) 28. Bob stood with his hands on his hips.
(H) 29. The teacher sat on a sharp tack.
(L) 30. She might have discussed the ape.
(H) 31. The storm broke the sailboat's mast.
(H) 32. At breakfast he drank some juice.
(H) 33. He hit me with a clenched fist.
(L) 34. Peter knows about the raft.
(L) 35. The old man considered the kick.
(L) 36. We have not thought about the hint.
(H) 37. The team was trained by their coach.
(L) 38. Bill hopes Paul heard about the mist.
(H) 39. The king wore a golden crown.
(H) 40. The sand was heaped in a pile.
(L) 41. The boy can't talk about the thorns.
(L) 42. Miss Brown will speak about the grin.
(H) 43. The duck swam with the white swan.
(H) 44. Let's decide by tossing a coin.
(L) 45. She has a problem with the goal.
(L) 46. Jane didn't think about the brook.
(L) 47. He hears she asked about the deck.
(H) 48. He got drunk in the local bar.
(H) 49. The girl swept the floor with a broom.
(L) 50. The class will consider the blast.

Form 2.5

(L) 1. Miss White would consider the mold.
(L) 2. Ruth has a problem with the joints.
(L) 3. The boy might consider the trap.
(H) 4. To store his wood he built a shed.
(H) 5. The lion gave an angry roar.
(L) 6. He is considering the throat.
(L) 7. They hope he heard about the rent.
(H) 8. The car was parked at the curb.
(L) 9. Peter should consider the bow. (as in "no")
(L) 10. The old woman discussed the thief.
(H) 11. A round hole won't take a square peg.
(L) 12. You're discussing the plot.
(L) 13. The woman knew about the lid.
(H) 14. Peter dropped in for a brief chat.
(L) 15. You were interested in the scream.
(H) 16. The gambler lost the bet.
(H) 17. The burglar escaped with the loot.
(L) 18. He could discuss the bread.
(H) 19. He was scared out of his wits.
(L) 20. He doesn't discuss the mop.
(H) 21. Eve was made from Adam's rib.
(H) 22. Get the bread and cut me a slice.
(L) 23. Bill won't consider the brat.
(H) 24. We heard the ticking of the clock.
(H) 25. Greet the heroes with loud cheers.
(H) 26. This camera is out of film.
(L) 27. Ruth wants to speak about the sling.
(H) 28. My jaw aches when I chew gum.
(L) 29. The man could consider the spool.
(H) 30. The bloodhound followed the trail.
(H) 31. The doctor prescribed the drug.
(H) 32. He rode off in a cloud of dust.
(L) 33. He was interested in the hedge.
(L) 34. Ruth hopes she called about the junk.
(H) 35. Playing checkers can be fun.
(L) 36. We're glad Ann asked about the fudge.
(H) 37. The super highway has six lanes.
(H) 38. Unlock the door and turn the knob.
(L) 39. Ruth is speaking about the meal.
(H) 40. Maple syrup is made from sap.
(L) 41. Bill cannot consider the den.
(L) 42. We are speaking about the prize.
(H) 43. The car drove off the steep cliff.
(L) 44. Miss Smith couldn't discuss the row. (as in "no")
(H) 45. The glass had a chip on the rim.
(H) 46. Old metal cans were made with tin.
(L) 47. Miss White thinks about the tea.
(L) 48. Miss White doesn't discuss the cramp.
(H) 49. That job was an easy task.
(L) 50. Mr. White spoke about the firm.

Form 2.6

(H) 1. Throw out all this useless junk.
(H) 2. She cooked him a hearty meal.
(H) 3. Her entry should win first prize.
(L) 4. Ruth could have discussed the wits.
(L) 5. We could discuss the dust.
(H) 6. The stale bread was covered with mold.
(H) 7. The firemen heard her frightened scream.
(L) 8. We spoke about the knob.
(H) 9. Your knees and your elbows are joints.
(H) 10. I ate a piece of chocolate fudge.
(L) 11. Paul hopes we heard about the loot.
(H) 12. Instead of a fence, plant a hedge.
(H) 13. The story had a clever plot.
(L) 14. David might consider the fun.
(H) 15. The landlord raised the rent.
(L) 16. Paul could not consider the rim.
(L) 17. He heard they called about the lanes.
(H) 18. Her hair was tied with a blue bow. (as in "no")
(L) 19. They had a problem with the cliff.
(H) 20. He's employed by a large firm.
(L) 21. Harry will consider the trail.
(L) 22. We are considering the cheers.
(H) 23. To open the jar, twist the lid.
(L) 24. She has known about the drug.
(L) 25. Bill had a problem with the chat.
(L) 26. We hear they asked about the shed.
(H) 27. The swimmer's leg got a bad cramp.
(L) 28. Jane had not considered the film.
(H) 29. Our seats were in the second row. (as in "no")
(L) 30. Jane did not speak about the slice.
(L) 31. Paul was interested in the sap.
(L) 32. I am discussing the task.
(H) 33. The thread was wound on a spool.
(H) 34. They tracked the lion to his den.
(L) 35. Ruth has discussed the peg.
(H) 36. Spread some butter on your bread.
(L) 37. Tom is considering the clock.
(L) 38. He's thinking about the roar.
(H) 39. A spoiled child is a brat.
(L) 40. I should have known about the gum.
(H) 41. Keep your broken arm in a sling.
(H) 42. The mouse was caught in the trap.
(L) 43. They heard I asked about the bet.
(H) 44. I've got a cold and a sore throat.
(L) 45. Betty doesn't discuss the curb.
(L) 46. He had a problem with the tin.
(H) 47. Ruth poured herself a cup of tea.
(H) 48. The house was robbed by a thief.
(L) 49. He wants to know about the rib.
(H) 50. Wash the floor with a mop.

Form 2.7

(L) 1. I did not know about the chunks.
(H) 2. The chicken pecked corn with its beak.
(L) 3. Bob could consider the pole.
(H) 4. The judge is sitting on the bench.
(L) 5. Mr. Smith knew about the bay.
(L) 6. You've considered the seeds.
(H) 7. The heavy rains caused a flood.
(H) 8. For dessert he had apple pie.
(L) 9. She hopes Jane called about the calf.
(H) 10. The detectives searched for a clue.
(L) 11. Mary hasn't discussed the blade.
(H) 12. The chicks followed the mother hen.
(L) 13. Mr. Brown thinks about the vault.
(L) 14. Bob was considering the clerk.
(H) 15. We camped out in our tent.
(H) 16. Paul took a bath in the tub.
(L) 17. Mary can't consider the tide.
(L) 18. The old man talked about the lungs.
(H) 19. The candle burned with a bright flame.
(H) 20. My son has a dog for a pet.
(L) 21. Bob has discussed the splash.
(H) 22. The plow was pulled by an ox.
(H) 23. The flood took a heavy toll.
(L) 24. Mr. Smith spoke about the aid.
(L) 25. Mary had considered the spray.
(H) 26. The pond was full of croaking frogs.
(L) 27. The girl should not discuss the gown.
(H) 28. Please wipe your feet on the mat.
(L) 29. Ruth hopes Bill called about the cop.
(L) 30. We will consider the debt.
(L) 31. Peter could consider the dove.
(H) 32. She shortened the hem of her skirt.
(H) 33. The cabin was made of logs.
(L) 34. Bill can't have considered the wheels.
(L) 35. He has a problem with the oath.
(H) 36. The dealer shuffled the cards.
(H) 37. The shepherd watched his flock of sheep.
(H) 38. The flashlight casts a bright beam.
(L) 39. We could consider the feast.
(H) 40. The scarf was made of shiny silk.
(H) 41. The guests were welcomed by the host.
(L) 42. Betty has considered the bark.
(H) 43. The sick child swallowed the pill.
(L) 44. Paul should have discussed the flock.
(H) 45. Tighten the belt by a notch.
(L) 46. She might discuss the crumbs.
(L) 47. Tom has not considered the glue.
(H) 48. The swimmer dove into the pool.
(L) 49. Tom has been discussing the beads.
(H) 50. Follow this road around the bend.

Form 2.8

(H) 1. The bird of peace is the dove.
(L) 2. Tom had spoken about the pill.
(H) 3. The cigarette smoke filled his lungs.
(L) 4. They've considered the sheep.
(H) 5. Cut the meat into small chunks.
(H) 6. Watermelons have lots of seeds.
(L) 7. The man should discuss the ox.
(L) 8. Miss Smith knows about the tub.
(H) 9. Raise the flag up the pole.
(L) 10. Peter has considered the mat.
(H) 11. The bride wore a white gown.
(L) 12. She might consider the pool.
(H) 13. We swam at the beach at high tide.
(H) 14. The poor man was deeply in debt.
(L) 15. She's glad Bill called about the beak.
(L) 16. Harry had thought about the logs.
(H) 17. Banks keep their money in a vault.
(H) 18. The witness took a solemn oath.
(L) 19. Bill didn't discuss the hen.
(L) 20. Ruth must have known about the pie.
(H) 21. The shepherds guarded their flock.
(L) 22. Bob has considered the tent.
(L) 23. We're speaking about the toll.
(H) 24. A bicycle has two wheels.
(H) 25. Ann works in the bank as a clerk.
(L) 26. Tom won't consider the silk.
(H) 27. Ruth had a necklace of glass beads.
(L) 28. She's discussing the beam.
(H) 29. Paul hit the water with a splash.
(H) 30. The nurse gave him first aid.
(H) 31. The wedding banquet was a feast.
(L) 32. Nancy didn't discuss the skirt.
(L) 33. The girl should consider the flame.
(H) 34. Tree trunks are covered with bark.
(H) 35. Break the dry bead into crumbs.
(L) 36. Mr. Black has discussed the cards.
(L) 37. The woman considered the notch.
(L) 38. The man spoke about the clue.
(H) 39. The boat sailed across the bay.
(L) 40. I'm talking about the bench.
(L) 41. They heard I called about the pet.
(H) 42. The cow gave birth to a calf.
(L) 43. I'm glad you heard about the bend.
(H) 44. It was stuck together with glue.
(L) 45. The woman talked about the frogs.
(H) 46. Bob was cut by the jacknife's blade.
(H) 47. Paul was arrested by the cops.
(L) 48. Bill heard we asked about the host.
(H) 49. Kill the bugs with this spray.
(L) 50. The class should consider the flood.

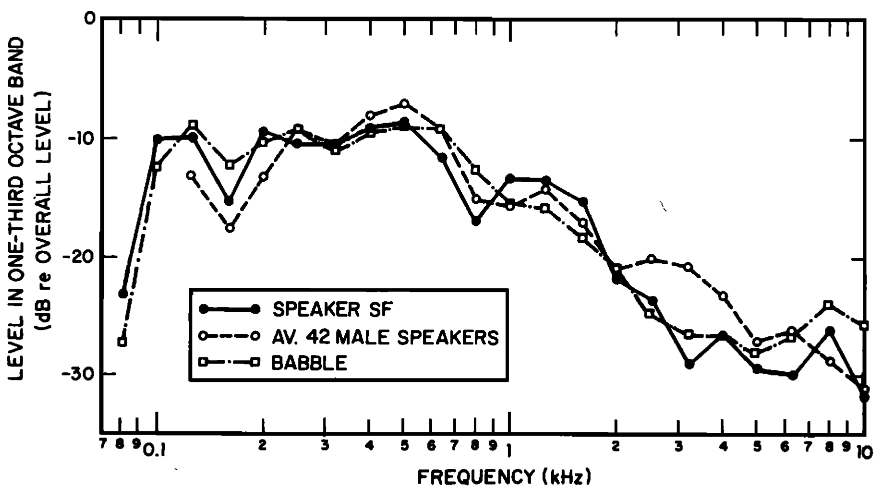

FIG. B1. One-third octave band spectra of (1) speaker SF who recorded test forms; (2) average of 42 male speakers, from Ref. 30; (3) 12-voice babble used in test instrument.

APPENDIX B. ACOUSTIC CHARACTERISTICS OF A PARTICULAR IMPLEMENTATION OF THE TEST INSTRUMENT

It is well known that the characteristics of a particular talker's speech can influence the intelligibility of a recorded test (Ref. 38 and references cited therein). Thus we include here a summary of certain relevant acoustic properties of the recording of the test forms that were prepared at the Bolt Beranek and Newman Los Angeles Laboratories.

The talker (SF) is an adult male speaker of the General American dialect. He has had extensive previous experience as a radio announcer, and in reading similar speech materials. He read the test sentences in an anechoic chamber with a condenser microphone located 12 in. from his lips.

The average speech spectrum for the talker is shown in Fig. B-1. Comparison of this spectrum with the average spectrum for a large number of male speakers shows that this talker has a spectrum that is close to the average.

The average fundamental frequency (F_0) for the talker is about 120 Hz. He followed the usual style of producing the sentences with a gradual baseline fall in F_0, with superimposed rises and falls indicating phrasing and stress. The initial F_0 prominence, which usually occurred on the first stressed vowel in the sentence, has a peak frequency of 156 Hz on the average, and F_0 on the final key word falls from 131 to 72 Hz, on the average. This speaker could be regarded as showing slightly more F_0 fluctuation within a sentence than an average speaker.

The average length of the PH sentences is 1.74 sec and of the PL sentences is 1.70 sec. The average speaking rate, therefore, is about 4.3 syllables per second. An identifying number precedes each sentence, beginning 2 sec before the sentence onset. The recorded sentences are spaced 10 sec apart from onset to onset, and thus the mean time available for making a response is about 6 sec.

The babble was generated by recording each of six adult talkers (three male, three female) reading the same section from a child's storybook in an anechoic chamber, mixing these six recordings (each at the same rms level), and then combining two repetitions of this six-voice babble to produce the final 12-voice babble. The average spectrum of the babble (in one-third octave bands) is shown in Fig. B-1. Measurements of the fluctuation in the overall level of the babble show that it is within ± 4 dB of the mean 90% of the time, when sampled at $\frac{1}{10}$-sec intervals.

The number cues and the test sentences are on track 1 of the standard $\frac{1}{4}$-in. magnetic tape, using a half-track recording format; the babble is on track 2. The overall rms level of the babble is constant during the test items, recorded at the same level as that of the test speech and number cues. Just prior to the onset of the number cue for each test sentence, the overall level of the babble is smoothly diminished 10 dB, held at that level through the cue, and smoothly restored to its original level prior to the onset of the test sentence. This "scoop" in the babble level provides a better chance for the listener to keep his place in the list, even under severe noise conditions.

*This work was supported by the National Institute of Neurological and Communicative Disorders and Stroke under contract N01-NS-4-2322. This paper was presented in part at the 92nd meeting of the Acoustical Society of America, in San Diego [J. Acoust. Soc. Am. 60, S28(A)(1976)].

†Also at Massachusetts Institute of Technology, Cambridge, MA 02139.

[1]D. N. Kalikow, K. N. Stevens, H. L. Gerstman, and R. K. Morrison, "The Speech Perception in Noise Test: Description and Clinical Data" (unpublished).

[2]D. B. Fry, "Word and Sentence Tests for Use in Speech Audiometry," The Lancet, 197–199 (22 July 1961).

[3]G. A. Miller and P. E. Nicely. "An Analysis of Perceptual Confusions Among Some English Consonants," J. Acoust. Soc. Am. 27, 338–352 (1955).

[4]V. A. Kozhevnikov and L. A. Chistovich, *Speech: Articulation and Perception* (Moscow-Leningrad, 1955). English translation J. P. R. S., Washington, DC, No. JPRS 30543.

[5]C. Speaks, "Intelligibility of Filtered Synthetic Sentences," J. Speech Hear. Res. 10, 289–298 (1967).

[6]G. A. Miller, G. A. Heise, and W. Lichten, "The Intelligi-

bility of Speech as a Function of the Context of the Test Materials," J. Exp. Psychol. 41, 329–335 (1951).
[7]G. A. Miller, "Decision Units in the Perception of Speech," IRE Trans. Inf. Theory IT-8, 81–83 (1962).
[8]J. R. Duffy and T. G. Giolas, "Sentence Intelligibility as a Function of Key Word Selection," J. Speech Hear. Res. 17, 631–637 (1974).
[9]J. D. Bransford and R. E. Nitsch, "How can we come to understand things that we did not previously understand? in *Speech and language in the laboratory, school, and clinic*, edited by J. F. Kavanagh and W. Strange (MIT, Cambridge, in press).
[10]D. Howes, "On the Relation Between the Intelligibility and Frequency of Occurrences of English Words," J. Acoust. Soc. Am. 29, 296–307 (1957).
[11]M. R. Rosenzweig, and L. Postman, "Intelligibility as a Function of Frequency of Usage," J. Exp. Psychol. 54, 412–422 (1957).
[12]I. Pollack, H. Rubenstein, and L. Decker, "Intelligibility of Known and Unknown Message Sets," J. Acoust. Soc. Am. 31, 273–279 (1959).
[13]E. Owens, "Intelligibility of Words Varying in Familiarity," J. Speech Hear. Res. 4, 113–129 (1961).
[14]A. Epstein, T. G. Giolas, and E. Owens, "Familiarity and Intelligibility of Monosyllabic Word Lists," J. Speech Hear. Res. 11, 435–438 (1968).
[15]E. L. Thorndike and I. Lorge, *The Teacher's Word Book of 30,000 Words* (Teachers College, Columbia University, New York, 1952).
[16]D. E. Broadbent, *Perception and Communication* (Pergamon, London, 1958).
[17]P. Rabbitt, "Recognition Memory for Words Correctly Heard in Noise," Psychon. Sci. 6, 383–384 (1966).
[18]R. Carhart and T. W. Tillman, "Interaction of Competing Speech Signals with Hearing Losses," Arch. Otolaryngol. 91, 273–279 (1970).
[19]R. Carhart, C. Johnson, and J. Goodman, "Perceptual Masking of Speakers by Combinations of Talkers," J. Acoust. Soc. Am. 58, S35(A) (1975).
[20]J. C. Cooper and B. P. Cutts, "Speech Discrimination in Noise," J. Speech Hear. Res. 14, 332–337 (1971).
[21]J. E. Birren and W. Spieth, "Age, Response Speed, and Cardiovascular Functions," J. Gerontol. 17, 390–391 (1962).
[22]D. E. Broadbent and A. Heron, "Effects of a Subsidiary Task Upon Performance Involving Immediate Memory by Younger and Older Subjects," Br. J. Psychol. 53, 189–198 (1962).
[23]D. B. Fry and P. M. T. Kerridge, "Tests for the Hearing of Speech by Deaf People," The Lancet (14 January 1939), 106–113.
[24]S. R. Silverman and I. J. Hirsh, "Problems Related to the Use of Speech in Clinical Audiometry," Ann. Otol. Rhinol. Laryngol. 64, 1234–1244 (1955).
[25]H. Davis and S. R. Silverman (Eds.), *Hearing and Deafness* (3rd ed.) (Holt, Rinehart, and Winston, New York, 1970).
[26]C. Speaks and J. Jerger, "Method for Measurement of Speech Identification," J. Speech Hear. Res. 8, 185–194 (1965).
[27]C. V. Hudgins, J. E. Hawkins, J. E. Karlin, and S. S. Stevens, "The Development of Recorded Auditory Tests for Measuring Hearing Loss for Speech," Laryngoscope 47, 57–89 (1947).
[28]D. H. Irvine, "A New Type of Speech Intelligibility Test," Ergonomics 17, 783–788 (1974).
[29]D. N. Kalikow, Quarterly Progress Report No. 3, on a project entitled, "Development of a Test Instrument for Speech Discrimination," Bolt Beranek and Newman Inc., Cambridge, MA (15 April 1975).
[30]K. S. Pearsons, R. L. Bennett, and S. Fidell, "Conversational Speech Levels in Various Environments," Bolt Beranek and Newman Report No. 3281 (1976).
[31]W. F. Rintelmann, *et al.*, "Six Experiments on Speech Discrimination Utilizing CNC Monosyllables," J. Aud. Res. Suppl. 2 (1974).
[32]W. L. Hays, *Statistics for Psychologists* (Holt, Rinehart and Winston, New York, 1964), pp. 483–487.
[33]G. Dewey, *Relativ Frequency of English Speech Sounds* (Harvard University Press, Cambridge, MA, 1923).
[34]N. R. French, C. W. Carter, and W. Koenig, "The Words and Sounds of Telephone Conversations," Bell System Tech. J. 9, 290–324 (1930).
[35]J. V. Tobias, "Relative Occurrence of Phonemes in American English," J. Acoust. Soc. Am. 31, 631 (1959).
[36]R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research* (Oliver and Boyd, London, 1953).
[37]D. H. Klatt, personal communication.
[38]M. H. L. Hecker, "A Study of the Relationship between Consonant-Vowel Ratios and Speaker Intelligibility," Ph.D. thesis (Stanford University, December 1973).