PREDICTING THE INTELLIGIBILITY OF WORDS II

JOHN W. BLACK

The central issue of one series of studies in the Phonetics Laboratories of the Ohio State University has been to determine a relationship between the intelligibility of words and the intelligibility of the constituent phonemes.[1] If a word is an ordered sequence of phonemes, and if a phoneme is – at least in part – a differentiating feature among words, then it would seem that a word would represent a compounding of the phonemes.

The foregoing relationship is straightforward in treatments of nonsense syllables. Although a particular sound may be more or less intelligible when it follows or precedes other particular sounds the ultimate outcome is that the intelligibility of nonsense syllables is well anticipated by the joint probability, i.e., intelligibility, of the member phonemes. The intelligibility of each phoneme is an average, determined from the transmissions of a number of nonsense syllables.

The words of a natural language, as English, are not nonsense syllables; and the constituent sounds are not randomly assembled. Rather, the order is biased and is amenable to a considerable degree of prediction. Assemblages of speech sounds have been learned as words, and presumably this has been accompanied by the learning of many of the probabilities of one's own language. This latter feat may not be consciously achieved; however, it manifests itself in an individual's orthography, syntax, pronunciation, reading speed, errors (flubs) in oral reading, and pointedly in evaluation, as *right-wrong*, of the usages of the language that he hears about him.

The criterion measure in the present series of studies is the intelligibility of a word, that is the identification of a heard word as indicated by a written response.

Preliminary to the present study, closely related work has been reported with the following salient features in methodology[2]: a) two sets of approximately 600 words each were responded to by 200 listeners; b) word- and phoneme-intelligibility values were determined (these were based on all responses in which at least two listeners were in agreement); c) the intelligibility of words was, on the average, equivalent to the

[1] This study was conducted at The Ohio State University, Columbus, Ohio, under a contract between the Offices of Naval Research and the Ohio State University Research Foundation. (Contract No. Nonr-495(18) NR 145–993).

[2] John W. Black, "Predicting the Intelligibility of Words," *Folia Phoniatrica*, 12:260–272 (1960).

joint probability (intelligibility) of the relative intelligibility values of the first two phonemes of the word. The author suggested,

> The listener is a vocabulary....Apart from context, a phoneme or two, a transition, a pattern of duration, another of intensity – and the listening vocabulary has been tapped. Instead of phonetic "recognizability" (intelligibility) a better concept might be "word suggestion". The amount of suggestion that is needed to elicit a correct word is equivalent to the joint-recognition value of two phonemes.

A disturbing result in the foregoing study arose from one comparison of the obtained intelligibility value of a word and the predicted value of the word, each word separately. The standard deviation of the distribution of discrepancies between these two sets of values was of the same order of magnitude as the standard deviation of the set of obtained word intelligibility values. An assumption followed that if a limited number of words, of speakers, and of listeners were employed experimental error would be at a minimum and the standard deviation of the intelligibility values of a list of words would exceed the standard deviation of the distribution of values representing the discrepancies, word-by-word, between obtained intelligibility scores and predicted values. This assumption inherently attributed some of the imprecision in the earlier reported predictions of intelligibility values to experimental error in the relatively large-scale approach that was used. The projected plan was more economical than the earlier work, which, in turn, was a by-product of the construction of intelligibility tests extending over fifteen years.

PROCEDURE

Five individuals, knowledgeable in phonetics, served as speakers and as listeners. The material was 300 English words, all among the 1,000 most-frequently-used words in American English by University students in formal public address. One-half of the words were of one syllable, and one half of two syllables. Daily, Monday through Friday, for four weeks each participant selected unsystematically twenty-five words from the master list. He recorded these on a tape recorder, attempting to maintain a level of approximately 85 db (re .0002 dyne/cm^2) at a condenser microphone that was positioned along the side of the face and mouth, out of the breath stream. He read a numerical identifying phrase with each word, and read the words at ten-second intervals.

The speaker wrote in IPA symbols the words he recorded.

The participant turned from speaker to listener. He activated a masking noise (white noise) that he heard through a headset along with recorded words at approximately O-db signal-to-noise ratio. As he listened to the diurnal lists of his four colleagues he held scripts of the words they had recorded, each word written orthographically and in phonetic notation. He kept the list shielded from his view until

he had consciously determined the word that he had heard. He then exposed the corresponding word on the list of stimuli. If the word he uncovered confirmed the judgment he had made, he wrote nothing; otherwise he wrote his error response, and either immediately or later transcribed it phonetically. Although the five participants knew that the signals they heard were English words, they were permitted to respond with a nonsense syllable or a pair of syllables, phonetically transcribed if the stimulus did not elicit an English word.

The listener was responsible for comparing the stimulus and the response and enumerating phonetic substitutions, deletions, and additions. These, as well as the retentions of the sounds of the stimulus words were tallied by two of the five participants.

Recapitulating, each of five participants daily selected and recorded acoustically groups of twenty-five words from a 300-word list. Over a fourweek period he read 500 words; in the same span of time he heard 2,000 words. The participants, following a single set of rules, indicated the phonetic discrepancies and agreements between the stimuli and the error-responses. These were summarized in the following sets of values:

a) relative intelligibility-values of the words;

b) relative intelligibility-values of the phonemes, based upon the intelligibility of the words;

c) relative intelligibility-values of the phonemes, based upon their retention in incorrect responses – called *preservation-in-error* (P-I-E);

d) relative intelligibility values of the phonemes, based upon the sums of the two foregoing values for phonemes;

e) a matrix of phonetic substitutions.

In these compilations, values were determined separately for the sounds in initial, medial, and terminal positions within the word.

RESULTS

Tables 1 shows the relative intelligibility scores of the phonemes – consonants and vowels separately – and gives separate values for the sounds in various positions; it also shows that part of the score of a phoneme that was derived from correct responses and the part that arose from the sound's being preserved in error responses. By way of illustration, [t] appeared in the initial position of 335 of the 10,000 stimuli. These 335 words were heard correctly in 50.1 per cent of the responses; the sound was preserved in 13 of the 167 incorrect responses. Thus, the intelligibility of the sound in the initial position was 57.6 per cent. Moreover, it should be observed, especially in the instance of the vowels, that a high intelligibility score based on correct responses precluded the possibility of a high intelligibility score based on the sound's being preserved in error responses.

Table 1. Relative intelligibility, in per cent, of the consonants and vowels of 300 English words of one and two syllables. Partial scores, derived from "words correctly heard" and from "phoneme preserved in an error response" (P-I-E), are shown; also separate scores for each of the positions in which the sound occurred, as **initial, medial,** and **final** (I, M, F).

	Intelligibility From Correct Words			Intelligibility From P.I.E.			N, Occurrences			Phoneme Intelligibility		
						CONSONANTS						
	I	M	F	I	M	F	I	M	F	I	M	F
t	50.1	63.3	61.9	7.5	*13.3*	*18.2*[a]	335	994	1548	57.6	76.6	79.1
d	61.2	68.2	65.9	*12.2*	*12.0*	*17.3*	237	233	805	73.4	80.2	83.2
p	68.7	70.7	62.7	11.6	*13.2*	7.0	588	431	172	80.3	83.9	69.7
b	70.2	90.1	68.0	11.8	1.5	0.0	363	132	25	82.0	91.6	68.0
k	58.0	68.0	63.2	*12.1*	*12.4*	8.4	591	473	332	70.1	80.4	71.6
g	60.6	82.1	—	3.8	5.7	—	206	140	—	64.4	87.8	—
s	54.6	62.6	49.6	*14.6*	*17.2*	6.7	828	638	596	69.2	79.8	56.3
z	—	69.3	56.2	—	6.1	7.0	—	114	128	—	75.4	63.2
f	62.2	77.3	68.6	*14.4*	4.7	2.2	514	212	134	76.6	82.0	70.8
v	65.3	78.0	65.1	0.0	7.4	6.4	26	228	109	65.3	85.4	71.5
θ	65.0	79.4	76.9	0.0	2.7	0.0	40	73	26	65.0	82.1	76.9
ð	59.0	73.3	66.6	3.6	7.6	0.0	166	169	3	62.6	80.9	66.6
ʃ	43.9	56.9	—	6.5	2.2	—	107	93	—	50.4	59.1	—
tʃ	47.6	70.2	46.2	3.1	8.5	5.2	63	47	134	50.7	78.7	51.4
dʒ	58.8	70.5	47.2	1.9	4.4	2.7	51	68	36	60.7	74.9	49.9
h	68.1	93.7	—	*14.1*	4.5	—	298	44	—	82.2	98.2	—
hw	50.0	—	—	1.8	—	—	110	—	—	51.8	—	—
w	82.3	64.1	—	6.7	12.8	—	488	78	—	89.0	76.9	—
m	73.0	75.4	58.7	*15.2*	6.6	12.5	342	374	216	87.2	82.0	71.2
n	66.2	66.7	72.9	4.4	*20.6*	*17.4*	246	1516	876	70.6	87.3	90.3
ŋ	—	50.9	67.3	—	*14.5*	7.3	—	55	205	—	65.4	74.6
r	56.3	64.5	71.1	9.1	*17.7*	*15.3*	142	1489	1039	65.4	82.2	87.0
l	63.8	71.9	77.1	*26.5*	*14.5*	9.3	362	787	323	90.3	86.4	86.4
j	67.2	76.9	—	3.4	*15.4*	—	58	13	—	70.6	92.3	—
						VOWELS						
ə	86.0	74.6	50.0	7.0	*16.6*	—	315	1863	10	93.0	91.2	50.0
ʌ	85.9	70.3	—	4.0	*15.9*	—	99	1074	—	89.9	86.2	—
i	77.2	60.9	41.4	13.9	*25.8*	*26.8*	79	562	41	91.1	86.4	68.2
ɪ	71.7	65.6	75.5	*20.7*	*17.0*	*14.7*	251	1178	442	92.4	81.6	90.2
e	90.0	58.5	85.7	6.6	*25.7*	11.9	30	415	42	96.6	84.2	97.6
ɛ	84.8	67.0	—	6.8	*14.2*	—	191	973	—	91.6	81.2	—
æ	66.6	59.0	—	19.0	*13.5*	—	63	481	—	85.6	72.5	—
ɑ	95.9	68.3	—	0.0	*13.2*	—	148	452	—	95.9	81.5	—
ɔ	88.9	70.7	—	7.7	*13.6*	—	154	475	—	96.6	84.3	—
o	91.8	71.5	89.7	5.8	*17.0*	6.0	86	341	166	97.6	88.5	95.7
ʊ	—	72.3	—	—	11.8	—	—	152	—	—	84.1	—
u	—	59.6	44.9	—	*13.5*	*28.9*	—	185	69	—	73.1	73.8
aʊ	74.1	71.1	93.3	9.6	*14.4*	0.0	31	180	45	83.7	85.5	93.3
ɔɪ	—	88.0	—	—	0.0	—	—	25	—	—	88.0	—
aɪ	89.6	56.0	81.3	3.4	*26.7*	14.6	29	321	75	93.0	82.7	95.9
ɪu	—	58.0	—	—	4.6	—	—	50	—	—	62.6	—
eɪ	—	—	94.8	—	—	5.2	—	—	19	—	—	100.0

[a] Italicized values indicate a statistically significant deviation from chance expectancy; computations based on a minimum of 36 error responses.

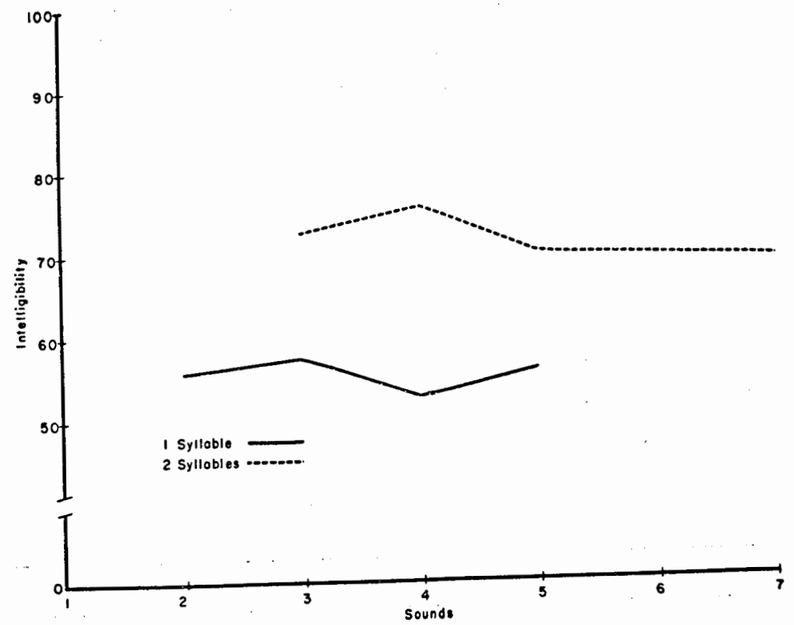

Fig. 1. Mean obtained intelligibility of words of one and two syllables and of different lengths.

One fact that can be extracted from Table 1 is that within categories of sounds, as plosive consonants, fricative consonants – wherever the comparison is applicable – voiced sounds tended to be more intelligible that their voiceless counterparts.

Figure 1 (also Table 2) shows the mean relative intelligibility of the different classes of words. A striking feature of the data is the increment in intelligibility that accompanies two-syllable words. In the light of these data two artifacts in Table 1 became apparent. It would appear from Table 1 that an initial vowel is inherently more intelligible than a medial vowel. An investigation of the 300 words showed that the initial vowel occurred in two-syllable words six times as frequently as in one-syllable words. Thus, the higher intelligibility values associated with initial vowels is, at least in part, attributable to the relatively high intelligibility values associated with two-syllable words in comparison with one-syllable words. A similar artifact relates to the apparent advantage of medial consonants over initial consonants in intelligibility. Again, medial consonants were more common in two-syllable words than in one-syllable words.

An optimal method for predicting the intelligibility of the words under treatment from the values of Table 1 would arrange the words in terms of intelligibility in the

Table 2. Means and standard deviations of the obtained relative intelligibility values of 300 words; also similar values for intelligibility scores predicted by five methods: 1) initial sound squared; 2) joint probability of the first two sounds, retaining the values peculiar to each position; 3) joint probability of the first two sounds, using the value for the initial position only; 4–5) joint probability of the two and three most intelligible sounds in the word.

ONE-SYLLABLE WORDS

Sounds N, words		2 10	3 73	4 59	5 11		
Obtained Value	Mean S.D.	57.3 26.3	58.5 18.9	52.7 17.4	56.1 17.2		
Predicted Values							
Method 1	Mean	62.0	57.4	57.9	54.8		
	S.D.	22.7	17.4	16.3	18.8		
Method 2	Mean	65.4	62.1	62.3	60.4		
	S.D.	11.1	10.2	12.9	8.5		
Method 3	Mean	65.0	67.8	66.0	54.0		
	S.D.	9.8	11.9	13.8	14.8		
Method 4	Mean	63.1	66.7	65.4	59.2		
	S.D.	7.0	8.1	9.3	9.8		
Method 5	Mean	63.4	68.2	65.7	56.8		
	S.D.	11.2	10.1	12.1	11.6		

TWO-SYLLABLE WORDS

Sounds N, words			3 4	4 37	5 61	6 35	7 10
Obtained Value	Mean S.D.		72.7 16.1	75.8 17.9	70.6 17.2	68.9 16.7	69.6 19.5
Predicted Values							
Method 1	Mean		86.0	77.1	67.2	63.9	59.0
	S.D.		1.0	17.3	15.2	17.4	12.0
Method 2	Mean		81.8	73.0	66.8	67.1	65.4
	S.D.		11.7	9.5	9.8	9.8	7.2
Method 3	Mean		64.2	71.0	70.4	64.3	63.9
	S.D.		14.8	9.2	9.8	8.7	10.3
Method 4	Mean		67.5	65.9	68.3	65.3	66.0
	S.D.		8.8	7.3	7.2	6.3	8.3
Method 5	Mean		55.4	65.3	68.1	64.0	55.8
	S.D.		16.6	9.9	10.1	9.4	22.3

order of their empirically derived values, would re-establish the dispersion of values present in the empirically derived ones, and would, as a consequence of these two criteria, show no discrepancy between empirically derived and predicted values. Earlier efforts at predicting word intelligibility have pointed towards the formula of using the intelligibility values of the first two phonemes of a word in the manner

Table 3. The standard deviations of the distributions of discrepancies, word by word, between obtained intelligibility values and ones predicted by five methods: 1) initial sound squared; 2) joint probability of the first two sounds, retaining the values peculiar to each position; 3) joint probability of the first two sounds, using the value for the initial position only; 4–5 joint probability of the two and three most intelligible sounds in the word.

ONE-SYLLABLE WORDS

Sounds N, words	2 10	3 73	4 59	5 11		
Obtained	26.3	18.9	17.4	17.2		
Predictions:						
Method 1	8.6	13.3	3.4	7.6		
Method 2	14.3	12.7	10.7	5.9		
Method 3	8.8	13.2	12.9	2.7		
Method 4	7.0	12.5	12.3	7.6		
Method 5	14.8	13.1	15.5	10.2		

TWO SYLLABLE WORDS

Sounds N, words	3 4	4 37	5 61	6 35	7 10	
Obtained	16.1	17.9	17.2	16.7	19.5	
Predictions:						
Method 1	16.1	11.8	9.9	14.4	15.7	
Method 2	14.3	10.5	8.8	10.5	9.8	
Method 3	9.9	11.0	9.8	9.1	5.9	
Method 4	7.9	10.3	10.9	9.8	5.7	
Method 5	14.1	11.1	15.1	9.0	10.1	

of joint probability. This method remains plausible – although other possibilities are not yet excluded – and is referred to in Tables 2–3 as *Method 2*. Other methods that are included in Tables 2–3 for comparison are as follows:

Method 1: the intelligibility of the initial sound squared;

Method 3: joint probability (intelligibility) of the first two sounds, using the value of the sound in the initial position only;

Method 4: joint probability (intelligibility) of the two most intelligible sounds in the word;

Method 5: joint probability (intelligibility) of the three most intelligible sounds in the word.

The first statistical procedure applied to the data was a test of independence between the predicted and the obtained values. This was applied separately to the several groups of words as determined by length in syllables and sounds, and again to the one-syllable words pooled and to the two-syllable words pooled. The hypothesis of independence was rejected in all instances except the separate category of two-syllable, five-sound words.

The second statistical procedure involved computing the discrepancies between

empirical and predicted values, word by word. These data are summarized in Table 2. The diminished variance in the distribution of disparities relative to the variance in the original dispersion of obtained intelligibility scores is a hopeful sign that the product of the intelligibility values of the first two phonemes of a word provides a helpful estimate of the relative intelligibility value of a word.

The preceding results are in keeping with an assumption that the member phonemes of a word contribute to the distinctive character of the word. The concept and data relative to "preservation-in-error values" relate further to this matter. The underlined values in Table 1 indicate the sounds that were apparently guiding stimuli in the selection of the error response. Thus all of the medial vowels appeared in the error responses with a frequency that exceeded chance; also particular consonants, especially the plosives and [r], [l] and [s].

SUMMARY

This study represented an economical approach to the task of determining the relative intelligibility of phonemes and then predicting the relative intelligibility of words from these results. A small number of common English words was used; a single listener replaced the usual listening panel; the listeners described their own errors in listening; every response contributed to the outcome, i.e. no response was treated as bizarre or anomalous. The results seem to confirm and strengthen the possibility that the intelligibility of a word relies upon an amount of intelligibility-information equivalent to the joint probability (intelligibility) of the first two phonemes of a word.

The Ohio State University