

ENGINEERING RESEARCH INSTITUTE
THE UNIVERSITY OF MICHIGAN
ANN ARBOR

THE EFFECT OF VOCABULARY SIZE ON ARTICULATION SCORE

Technical Report No. 81

Electronic Defense Group
Department of Electrical Engineering

By: D. M. Green
T. G. Birdsall

Approved by:



A. B. Macnee

AFCRC-TR-57-58
ASTIA Document
No. AD 146 759

Contract No. AF19(604)-2277

Operational Applications Laboratory
Air Force Cambridge Research Center
Air Research and Development Command

January 1958

ERRATA SHEET

Technical Report No. 81.
Electronic Defense Group

Page 12 Paragraph 4, Line 1 should read:

"If one is not satisfied with a graphic interpolation for the values of a_n , a second method may be suggested which is perhaps somewhat more accurate."

Page 13, Line 8: Change Reference 4 to Reference 5..

Page 14, Last line should read:
(Equation 7).

Page 15, Add reference 5:
"Probability Tables for the analysis of Extreme-Value Data", U.S. Dept. of Commerce, National Bureau of Standards, Applied Mathematics Series 22.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	iv
ABSTRACT	v
1. INTRODUCTION	1
2. THE MODEL	1
3. APPLICATION OF THE MODEL	3
4. CONCLUSION	6
APPENDIX 1 PROBABILITY OF BEING CORRECT IN A CHOICE OF M ORTHOGONAL SIGNALS	10
REFERENCES	15
DISTRIBUTION LIST	16

LIST OF ILLUSTRATIONS

Figure 1 Maximum Probability of a Correct Forced Choice Among M Orthogonal Alternatives	4
Figure 2 S/N in db	5
Figure 3 Predicted versus Obtained Data	8

ACKNOWLEDGEMENT

The use of the one-of-m-orthogonal signal computations and their application to the problem of vocabulary size has been made earlier by W. P. Tanner, Jr. The particular form expressed in this paper is the responsibility of the author.

ABSTRACT

A statistical decision model is applied to the recognition of voice signals in noise. Certain strong simplifying assumptions are made to make the mathematics of the model manageable. The model is compared with the data of Miller, et. al. (Reference 1). The main problem dealt with is how the size of the vocabulary affects the articulation score. A discussion is included of the physical parameters involved in such tests. An appendix presents various approximations to the problem involved in predicting the percent correct recognition for the conditions considered by the model.

THE EFFECT OF VOCABULARY SIZE ON ARTICULATION SCORE

1. INTRODUCTION

The dependence of the articulation score upon vocabulary size has been studied empirically by Miller, Heise and Lichten (Reference 1). This paper will attempt to account for the data obtained in terms of a statistical decision model. The main virtue of the application will be to show how a single set of transformations of the data yield a single function relating an inferred variable, d' , to the physical measure employed in the study.

The model will not be developed in full: rather, one plausible manner of interpreting the model will be explained. The results derived, while encouraging, need confirmation from other studies. The problems involved in checking the model with other data will be discussed.

2. THE MODEL

Let the set of words be denoted W , and a particular word of the set $W_i(t)$. $W_i(t)$ may be interpreted as the voltage wave form of the word. Suppose the receiver is a cross correlation type where the received input $S_i(t)$ (the i^{th} stimulus wave form) is correlated with every expected word. Now $S_i(t)$ may be considered as composed of two parts:

$$S_i(t) = W_i(t) + n(t)$$

where $W_i(t)$ is the i^{th} word and $n(t)$ is random noise. Let us further assume that all words in set W are orthogonal with equal energy; that is:

$$\int_0^T W_i(t) W_j(t) dt = 0 \quad \text{for } i \neq j$$

$$\int_0^T W_i(t) W_j(t) dt = E \quad \text{for } i = j$$

where T is the duration of the word and E is the energy of the word. Note that the E is independent of i . All words are assumed to have the same energy. Now if a stimulus word $S_k(t)$ is presented, the receiver will cross correlate every stored word $W(t)$ with the received input. Suppose that n such words may be presented. There will be n correlations, $(n-1)$ will be of the type

$$C_{k \neq j} = \int_0^T S_k(t) W_j(t) dt = \int_0^T (W_k(t) + n(t)) W_j(t) dt = \int_0^T n(t) W_j(t) dt$$

and one cross correlation of the type:

$$C_j = \int_0^T n(t) W_j(t) dt + E \quad \text{for } j = k.$$

These correlations can be transformed so that $(n-1)$ correlations of the type $C_{k \neq j}$ will be normally distributed with zero mean and unit variance, while the one correlation C_j will be normally distributed with a non-zero mean and unit variance.* This normalized mean is called d' , and increases as E increases.

Let us assume that the receiver selects the largest correlation value and reports the corresponding word as the stimulus received. Then the probability of this response being correct is the probability that the correlation $C_{k=j}$ is larger than the largest of the $(n-1)$ correlations of the type $C_{k \neq j}$.

Using the transform of these correlations this is the probability that the largest of $(n-1)$ drawing from a normal deviate with mean zero and unit variance is smaller than the drawing from a single normal deviate with mean d' and unit variance. Birdsall and Peterson (Reference 3) have calculated a graphic answer for this problem to provide the relation between d' and percent correct

* C_k is normally distributed. This can most easily be seen by considering this integral as a sum using the sampling theorem. It then becomes a linear sum of normal variables. The sum therefore is a normal variable (Reference 4).

responses for various size vocabularies. Appendix 1 discusses these calculations in detail.

Figure 1 shows the results of these computations. The graph shows that if there are 32 possible words, the mean of the "word-sent" distribution must be 3 times greater than the variance in order to be correctly chosen 80 percent of the time.

For the receiver discussed previously $d' = \sqrt{2E/N_0}$, where E is the energy in each word and N_0 is the noise power in a one cycle per second band.

3. APPLICATION OF THE MODEL

In Miller, Heise and Lichten's study (Reference 1) we find the articulation score plotted as a function of signal-to-noise ratio with vocabulary size as a parameter.

From the raw data each percent correct identification was used to enter Figure 1. Using the line appropriate for this vocabulary size, a d' value was obtained. These values for d' were then plotted against the signal-to-noise ratio which was used in the study. Figure 2 shows the result of this work.

The following examples illustrate this procedure: for a vocabulary size of two, -12 db gave 87 percent correct; using Figure 1, 87 percent corresponds to a $d' = 1.60$ for a 2 word vocabulary. For a vocabulary of size 32, -12 db gives 39 percent correct; using Figure 1, 39 percent yields d' of 1.73 for a 32 word vocabulary. The actual data obtained in the experiment, not the smoothed curves, was used to construct Figure 2.

On Figure 2 a smooth line was drawn by eye through the cluster of points. The line drawn is presented below in table form.

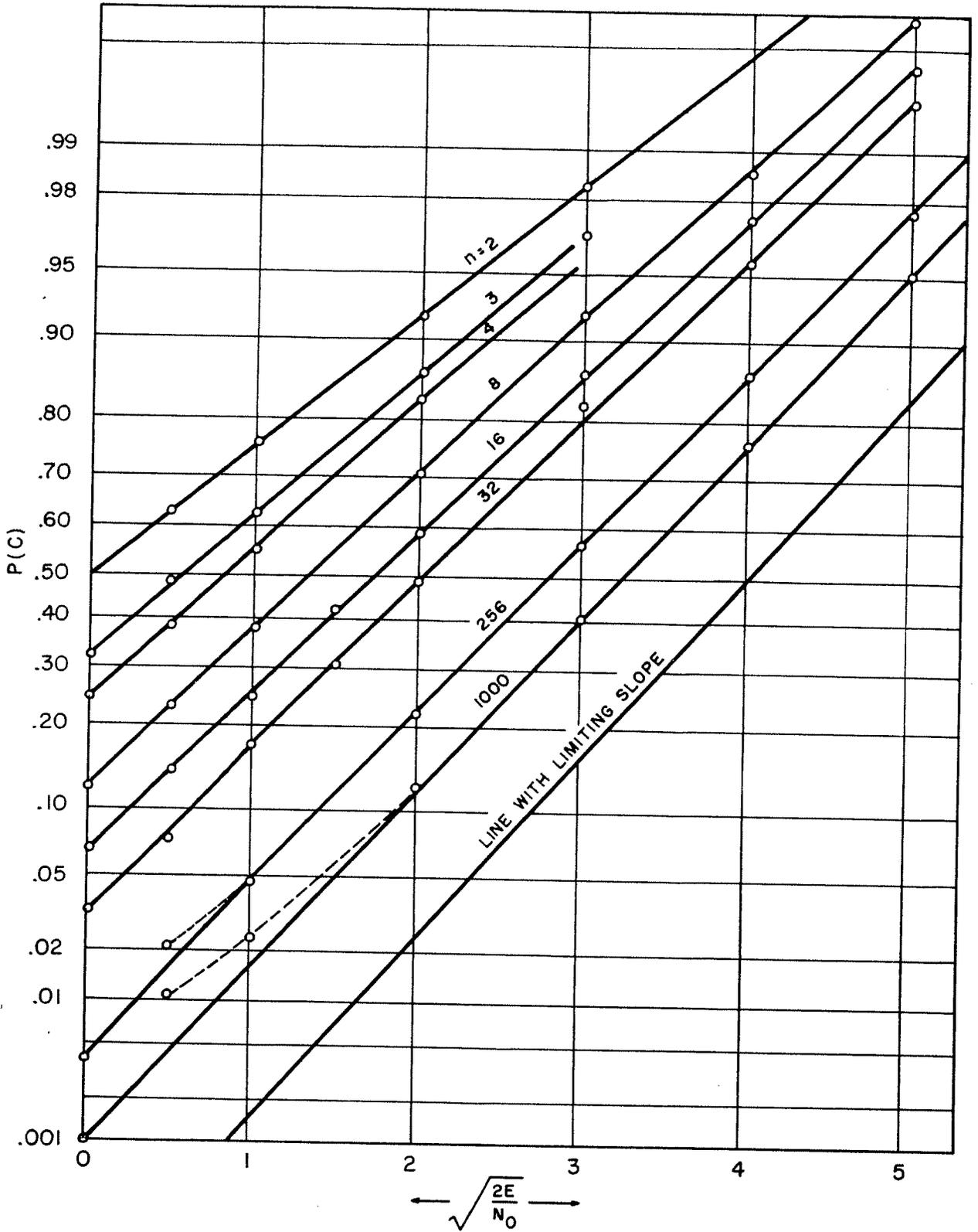


FIG. 1. MAXIMUM PROBABILITY OF A CORRECT FORCED CHOICE AMONG n ORTHOGONAL ALTERNATIVES.

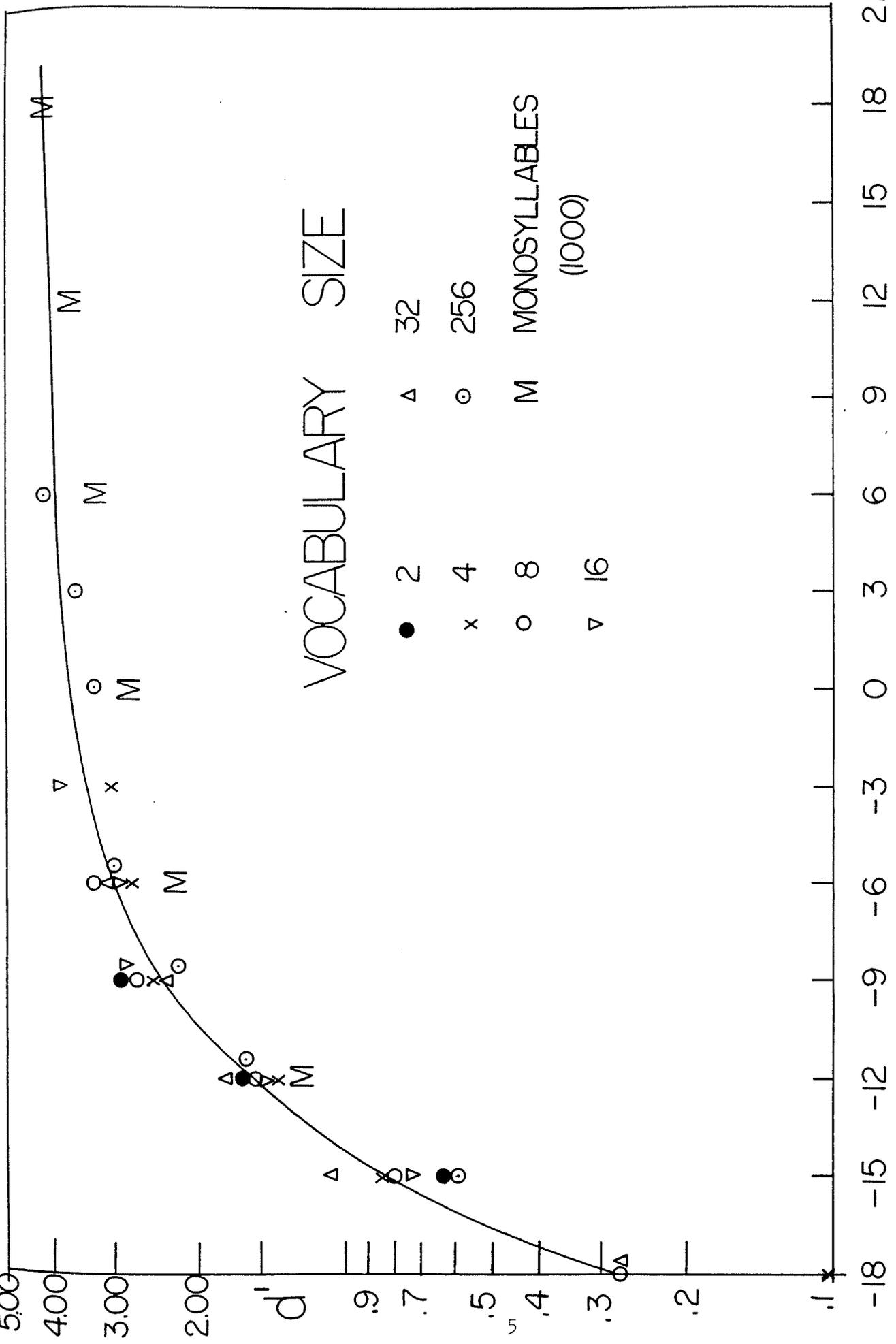


FIG.2 S/N in db

ENGINEERING RESEARCH INSTITUTE • UNIVERSITY OF MICHIGAN

S/N (in db)	-18	-15	-12	-9	-6	-3	0	3	6
d'	.27	.80	1.55	2.40	3.00	3.4	3.6	3.8	3.9

S/N (in db)	9	12	15	18	21
	4.0	4.1	4.2	4.3	4.4

In order to evaluate the fit of the line to the data, the following procedure was employed. For each value of S/N the d' given in the table was assumed. This was used to reenter Figure 1 and predict the articulation score. One therefore obtains a set of percent correct values as inferred from the line and those obtained in the experiment. These two values (of predicted and obtained percent correct) are displayed in Figure 3. As can be seen, the line drawn in Figure 2 appears to fit the data fairly well. Several data points do not appear on the graph. For a vocabulary of size two, the points obtained with the smallest signal-to-noise ratio were 49 percent and 51 percent. These led to estimates of d' which are too small to appear on Figure 2. Only one is displayed in Figure 3 since the other would have about the same value.

The points obtained with the monosyllables (M) depart considerably from the predicted value. One reason for this divergence is that the condition under which this data was collected is different from the procedure used with the other vocabulary size. The monosyllables were selected from a list of 1000 words, but the subjects had no list of the 1000 words. Therefore, he could not possibly perform the operations assumed for the model. If one assumes the apparent vocabulary size is between 2000 to 4000 words a better prediction of this data is obtained.

We now arrive at the knotty problem of what the measure (S/N in db) means. According to the article (Reference 1, page 330) "A S/N of zero db means, therefore, that the electrical measurements indicated the two voltages, speech and noise, were equal in magnitude." A consideration of the time wave form of speech and noise

VOCABULARY SIZE

● 2 Δ 32

x 4 ○ 256

○ 8 M (MONO)

▽ 16

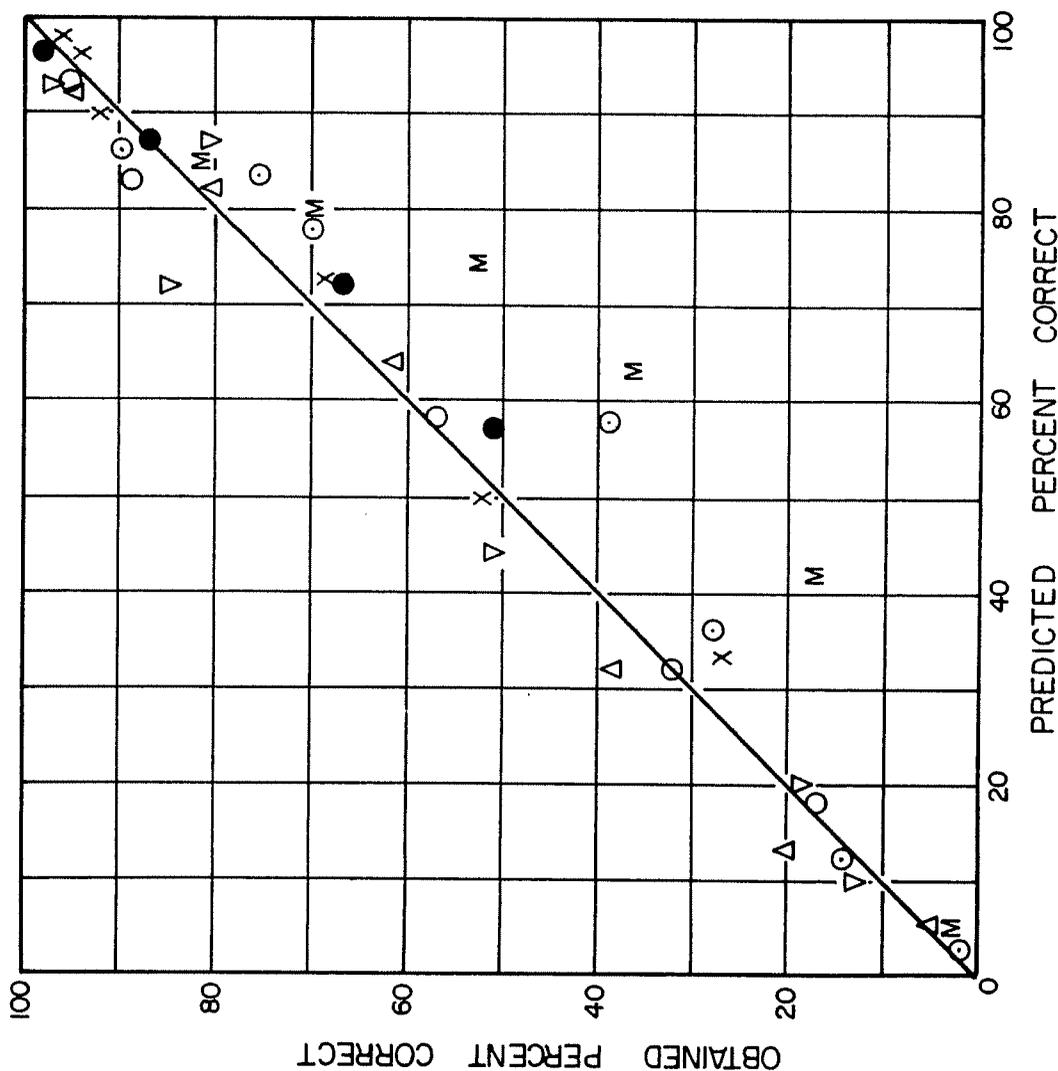


FIG. 3. PREDICTED VERSUS OBTAINED DATA.

ENGINEERING RESEARCH INSTITUTE • UNIVERSITY OF MICHIGAN

leads us to believe that these voltages were not equal in magnitude for any period of time. What was done was the familiar monitoring of the speech by a volume indicator (VU). The carrier phrase "You will write" was held constant by this method and the words spoken in their natural manner. The peak deflection of the meter were used to measure the individual words and the average, therefore, gave us a definition of S in db. The noise was then measured on the same meter and this gives us a definition of N in db.

If one interprets the S/N in db as varying linearly with $\log (2E/N_0)$ then one would expect $\log d'$ to vary linearly with S/N in db. Hence, in Figure 2, if the observer acts as an ideal receiver, one would expect a single straight line to fit the data. It must be remembered, however, that two assumptions were made in deriving the equation for the ideal receiver. First, the words were assumed to be orthogonal, and, secondly, they were assumed to be equal in energy. Both assumptions appear unlikely, especially for the larger vocabulary sizes.

The exact manner in which a violation of these assumptions affects the expected percent correct detections is difficult to evaluate. One can say with some certainty that they will decrease the expected percent correct. Beyond this trivial statement little of a concrete nature can be stated. The equations for the percent correct answers with correlated words or words of quite different energies are rather complex and no simplifying forms have been discovered as yet.

4. CONCLUSION

While the assumptions made by the model are obviously too strong, the present analysis provides a logical way of transforming the articulation score with different vocabulary size to a single function. The single function is derived empirically from the data and hence the main question remains, will the function

ENGINEERING RESEARCH INSTITUTE • UNIVERSITY OF MICHIGAN

work in other experiments? It is rather difficult to apply the functions to other experiments when the measure of the noise is dependent on the bandwidth of the system employed in the experiment. Also, until some relation is determined between "the peak VU meter deflection" and the energy in a word it remains impossible to test the model in any detailed way.

APPENDIX 1

PROBABILITY OF BEING CORRECT IN A CHOICE OF N ORTHOGONAL SIGNALS

Figure 1 of the report displays the probability of a correct choice among N orthogonal alternatives. Mathematically this is equivalent to the probability that a single drawing from a normal deviate, with mean d' and variance one, will be larger than the largest of (n-1) drawings from normal deviate with mean zero and variance one. Let this probability be denoted $P_n(d')$.

The original computation for this problem were done by Birdsall and Peterson (Reference 3). They used an approximate integration technique which had an error no greater than about 2%.

Table 1 shows the results of these computations.

Their results indicated that $P_n(d')$ could be approximated by a rather simple form; i.e.,

$$P_n(d') = \bar{\Phi}(a_n d' - b_n) \tag{1}$$

where $\bar{\Phi}$ is the (area) normal distribution function, i.e.,

$$\bar{\Phi}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

The value b_n is obtained by setting $d' = 0$ then $P_n(0) = \bar{\Phi}(-b_n) = 1/n$. The value a_n was computed for several values of n and is listed in Table 2.

TABLE 2

n	2	4	8	16	32	256	1000
a_n	.707	.827	.855	.884	.890	.916	.964

For values of n between those listed, a_n may be obtained by graphic interpolation.

Recently some work has been in progress which allows us to check partially the accuracy of these computations and extend the range on n.

TABLE 1

Computed Probability of Being Correct¹

Forced Choice Among N Orthogonal Alternatives

d'	N	0	0.5	1.0	1.5	2	3	4	5	6
Exact ²	2	.5000	.6381	.7602	-----	.9198	.9830	-----	-----	-----
	2	.5000	.6385	.7601	-----	.9213	.9831	-----	-----	-----
	3	.3333	.4827	.6336	-----	.8658	.9689	-----	-----	-----
	4	.2500	.3893	.5522	-----	.8321	.9573	-----	-----	-----
	8	.1250	.2375	.3853	-----	.7110	.9220	.9865	.9988	-----
	16	.0625	.1417	.2585	.4218	.5952	.8660	.9750	.9974	-----
	32	.03125	.0844	.1747	.3126	.4840	.8029	.9571	.9954	-----
	256	.00391	.0206	.0488	-----	.2249	.5691	.8658	.9787	.9984
	1000	.00010	.0106	.0224	-----	.1247	.4118	.7645	.9509	.9950

1. When $d' = 0$, all the variables have zero mean and unit variance. The probability that any one will be the greatest is $1/N$, and hence $P(c) = 1/N$. For the sake of clarity it must be pointed out that these P(Correct) curves are not "corrected for chance" or normalized in any manner except for that occurring from the definition of d' .
2. These computations are exact since the greatest of one drawing from a normal deviate is simply the normal deviate. The computation listed in the line below are those obtained using the approximation formula.

ENGINEERING RESEARCH INSTITUTE • UNIVERSITY OF MICHIGAN

Tippett's tables (Reference 2) have the distribution function for the largest of n drawings from a normal deviate with mean zero and unit variance. By convoluting this distribution function with the normal density function having mean d' and unit variance it is possible to obtain $P_n(d')$.

That is

$$P_n(d') = \int_{-\infty}^{+\infty} T_{n-1}(x) \phi(x-d') dx \quad (2)$$

where $T_{n-1}(x)$ is Tippett's distribution and $\phi(x-d')$ is the normal density function with mean d' and unit variance.

Table 3 shows the values obtained by convolution and those obtained from the report of Birdsall and Peterson.

TABLE 3

	a'	0	1	3.5		
$P_{31}(d')$.03225	.174974	.903456	Convolution with Tippett's Function	
$P_{32}(d')$.03125	.1747	.89480	Birdsall and Peterson	
a'	0	2	3.5	5	6	
$P_{1001}(d')$.000999	.120231	.59818	.950594	.99484	Convolution with Tippett's Function
$P_{1000}(d')$.001	.1247	.61179	.9506	.9948	Birdsall and Peterson

If one is not satisfied with a graphic interpolation for the values of a_n , a second method may be suggested which perhaps is somewhat interpolation for the values of a_n , a second method may be suggested which perhaps is somewhat more accurate. To obtain $P_n(d')$ for $n \geq 30$, Tippett's distribution may be approximated by a normal distribution function, for values of $.01 \leq T_n(x) \leq .99$. That is:

$$T_n(x) \approx \Phi\left(\frac{x-m_n}{\sigma_n}\right) \quad (3)$$

where

ENGINEERING RESEARCH INSTITUTE • UNIVERSITY OF MICHIGAN

$$m_n = \lim_{b \rightarrow \infty} \left[b - \int_0^b T_n(x) dx \right] \quad n > 30 \quad (4)$$

$$\sigma_n^2 = \lim_{b \rightarrow \infty} \left[b^2 - 2 \int_0^b x T_n(x) dx \right] - m_n^2 \quad n > 30 \quad (5)$$

Then,

$$\begin{aligned} P_{n+1}(d') &= \int_{-\infty}^{\infty} \bar{\Phi} \left(\frac{x-m_n}{\sigma_n} \right) \phi(x-d') dx \\ &= \bar{\Phi} \left(\frac{d'-m_n}{\sqrt{1+\sigma_n^2}} \right) \end{aligned} \quad (6)$$

Note that $\frac{m_n}{\sqrt{1+\sigma_n^2}}$ plays a role like b_n in equation 1 and $\frac{1}{\sqrt{1+\sigma_n^2}}$ is like a_n in Equation 1.

For larger n ($n > 1000$) the approximations suggested by the Bureau of Standards (Reference 4) are useful. They suggest:

$$T_n(x) = \bar{\Phi} \left(\frac{x-m_n}{\sigma_n} \right) \quad (7)$$

where $m_n = U_n + \frac{.57722}{\alpha_n}$

$$\sigma_n = \frac{\pi}{\sqrt{b} \alpha_n} = \frac{1.28255}{\alpha_n}$$

and U_n is such that $\bar{\Phi}(U_n) = 1 - \frac{1}{n}$

$$\alpha_n = n \phi(U_n)$$

Since for large x ($x \geq 4$)

$$\bar{\Phi}(x) = 1 - \bar{\Phi}(-x) \doteq 1 - \frac{1}{x} \phi(x)$$

$$\frac{1}{U_n} \phi(U_n) \doteq \frac{1}{n}$$

$$n \phi(U_n) \doteq U_n$$

$$\alpha_n \doteq U_n \quad \text{for } n > 10^3$$

Using the Bureau of Standards approximations as the constants in Equation 1

ENGINEERING RESEARCH INSTITUTE • UNIVERSITY OF MICHIGAN

$$n = 10^6 \quad P_n(d') = \bar{\Phi} (.96796 d' - 4.71416)$$

$$n = 10^9 \quad P_n(d') = \bar{\Phi} (.97885 d' - 5.96303)$$

$$n = 10^{12} \quad P_n(d') = \bar{\Phi} (.98434 d' - 7.00358)$$

The term $a_n \rightarrow 1$ for very large n , that is, the variance of the extreme value distribution (Tippett's distribution) approaches zero for very large n .

To summarize, Birdsall and Peterson's calculations provide good approximations to the probability $P_n(d')$ for moderate values of n , $n \leq 1000$, and for practical ranges of d' , (Equation 1). Better accuracy is guaranteed using convolutions with Tippett's tables, (Equation 2). Another method similar to Birdsall and Peterson's, but employing Tippett's table is demonstrated (Equation 6). Finally, for large n ($n > 1000$), the Bureau of Standards approximations are useful (Equations 3, 4, and 5).

REFERENCES

1. Miller, Heise and Lichten, "The Intelligibility of Speech as a Function of the Context of the Test Materials," J. Expt. Psy. 41, 329-335, 1951
2. Tippett, L. H. C., "On the Extreme Individuals and the Range of Samples Taken from a Normal Population," Biometrika 17, p. 364, 1925.
3. Birdsall, T. G., Peterson, W. W., "Probability of Correct Decision; In a Forced Choice Among M Alternatives", Electronic Defense Group Quarterly Progress Report No. 10, The University of Michigan, April 1954.
4. Peterson, W. W., Birdsall, T. G., and Fox., W. C., "Theory of Signal Detectability," Transactions of the I.R.E., Professional Group on Information Theory, 1954 Symposium on Information Theory, Sept. 15-17, 1954.

DISTRIBUTION LIST

1 Copy	Document Room Willow Run Laboratories University of Michigan Willow Run, Michigan
12 Copies	Electronic Defense Group Project File University of Michigan Ann Arbor, Michigan
1 Copy	Engineering Research Institute Project File University of Michigan Ann Arbor, Michigan
250 Copies	Operational Applications Laboratory Air Force Cambridge Research Center Air Research and Development Command Bolling Air Force Base 25, D. C. Contract No. AF19(604)-2277