DOCUMENT RESUME

ED 103 920 | CS 501 014

AUTHOR	Black, John W.
TITLE	Information Theory: A Method for Human Communication Research.
PUB DATE	Dec 74
NOTE	8p.; Paper presented at the Annual Meeting of the Speech Communication Association (Chicago, December, 1974)
EDRS PRICE	MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS	Communication (Thought Transfer); *Educational Research; *Information Theory; Language Research; *Listening Comprehension; *Phonetic Analysis; Phonetics; Syllables

ABSTRACT

This paper describes seven experiments related to human communication research. The first two experiments discuss studies treating the aural responses of listeners. The third experiment was undertaken to estimate the information of sounds and diagrams which might lead to an estimate of the redundancy ascribed to the phonetic structure of words. A total of 1549 words of one syllable and 2151 of two syllables were transcribed phonetically using 41 phonetic symbols. From this experiment it was concluded that the intelligibility of a word is most closely associated with the intelligibility of the initial sounds of a word. The fourth experiment was undertaken with contextual material and short phrases from two sources: flight patter and newspapers. The fifth experiment was related to letter prediction. The sixth experiment used single words in different contexts. The seventh experiment was concerned with the nature of the distribution of responses. (TS)

INFORMATION THEORY: A METHOD FOR HUMAN COMMUNICATION RESEARCH

John W. Black

The Ohio State University

Presented At

Speech Communication Association Convention

Chicago, IL, 1974

INFORMATION THEORY: A METHOD FOR HUMAN COMMUNICATION RESEARCH

John W. Black, The Ohio State University

SCA Chicago 1974

I. George Miller and Patricia Nicely reported a study in 1955, that has become a standard among works on speech perception. Sixteen consonants were spoken initially before the vowel /a/ in a list of 200 nonsense syllables. This list was recorded by each of 5 female speakers. Four listeners wrote the syllables they believed they heard, providing 800 syllable-response events per talker for which confusions could be studied. Pooling the five talkers, gave 4,000 observations at each condition that was tested. The conditions were 6 signal-to-noise ratios, -18, -12, -6, 0, +6, +12 dB. Furthermore, the bandwidths were varied to provide low-pass cutoffs at 300, 400, 600. 1200, 2500, and 5000 Hz and high-pass cutoffs at 200, 1000, 2000, 2500, 300-, and 4500 Hz. The responses were published in 18 confusion matrices. Supposedly this work dates the origin of the terms <u>confusion matrix</u>, <u>error matrix</u>, <u>response matrix</u>, and the like. The scope of the work and the reporting of the results are only two of several reasons why the study is considered a classic within the area of concern. Another reason is that the consonants were studied in terms of a system of distinctive features, specifically a system that included voicing, nasality, affrication, duration, and place of articulation. When I say "studied" I refer more specifically to the fact that each distinctive feature of a consonant was considered as a separate channel, and the ratio between the input and output of each channel was determined for each of the 18 experimental conditions. The most satisfactory experimental condition was bandwidth 200 - 6500 Hz and signal-to-noise ratio 12 dB. Under this condition voicing was conveyed properly 97% of the time, nasality 100%, affrication 85%, duration 93%, and place of articulation 71%. The fact that these channels summated to more than the theoretical whole was attributed to interaction or "crosstalk". The presence of one feature made the presence of another feature more likely. This crosstalk is what we are experiencing today as we inadvertently slip into the area assigned to one of the other speakers. In the parlance of the day we are calling this redundancy.

II. The Miller and Nicely study illustrates a procedure that is mandatory as far as I know in working with information theory: a process of absolute judgment or of identifying a stimulus. My high regard for the work of Miller and Nicely is shown by the fact that Sadanand Singh and I used it as a model in a study of speech production and speech perception by four language groups. Twenty-six intervocalic consonants were recorded by three speakers of each of four languages, Hindi, English, Arabic, and Japanese and heard by 24 speakers of each of them. The data were treated in two ways, 1) an analysis of variance of correct responses (all speakers spoke better and all listeners listened better when saying and hearing sounds of their native language)} and 2) the quantitative procedure employed singularly by Miller and Nicely -- our topic today. This was adapted to ascertain which features were retained by the listeners in their error responses and whether or not these were similar from one language group to another. We used seven channels instead of five and, of course, had four language groups instead of one. The striking outcome lay in the rank order of the seven channels in the relative amounts of information per channel, or the relative importance of the channels. A single rank order in this regard obtained for all of the listening groups: 1) nasality, 2) place, 3) liquidness, 4) voicing, 5) duration, 6) friction, and 7) aspiration.

This outcome varied in one important detail from the related result of Miller and Nicely. With them, place of articulation was the hardest to hear correctly. In our instance, it was the second easiest.

III. The two preceding studies treated the aural responses of listeners. I come now to an approach to our code of communication. This piece of work I'm delighted to have behind me. First, I am pleased with it, and second, I would never want to tackle it again. The study was a finger exercise in applying elementary techniques of information theory to phonetic probability. One- and two-syllable isolated words were sampled with respect to (a) the relative frequency of speech sounds, $p(i)$ ('the probability of sound *i*'), (b) the relative frequency of occurrence of the sounds at different positions in the word, (c) the probability of two sounds occurring in succession, $p(i, j)$ ('the joint probability of a sequence of sounds designated *i* and *j*'), (d) the probability that one sound follows another, $p_i(j)$ (('the conditional probability that sound *j* follows sound *i*'), and (e) the probability that a sound will precede another, $p_j(i)$ ('the conditional probability that sound *i* precedes sound *j*'). Words of one- and two-syllables and of differing numbers of sounds were treated separately. The principal objective was to estimate the information of sounds and digrams. This would lead to an estimate of the redundancy that might be ascribed to the phonetic structure of words. A secondary objective was enumerative, to find the probability of the sounds and digrams represented by the sample.

The narrowness of the problem is emphasized. The sample was not continuous language and it might be regarded as zero-order approximation to the language of speech, i.e., an array of words not weighted by their probability. There were 1549 words of one syllable and 2151 of two syllables. They had been selected from words of Thorndike ratings 1 - 10 though excluding homonyms, homographs, and words of greatest and least intelligibility. The words were transcribed phonetically using 41 phonetic symbols. Thus there were 5.35 bits per phoneme. Let me summarize a few results. 1) The average information per sound in bits was:

	One-syllable words	
	i-sounds	*j*-sounds
3-sound	5.04	4.65
4-sound	4.15	4.40

	Two-syllable words	
	i-sounds	*j*-sounds
4-sound	4.68	4.31
5-sound	4.48	4.15
6-sound	4.40	4.33

2) Computation of the information of the digrams yielded the following information per symbol:

	One-syllablewwords	Two-syllable words
3-sound	4.21	....
4-sound	3.35	3.89
5-sound	....	3.89
6-sound	....	3.75

3) The decrements in information from a) maximum (5.35 bits) to b) observed average bits per symbol to c) observed bits per digram symbol follow:

3-sound, 1 syllable	5.35	4.85	4.21
4-sound, 1 syllable	5.35	4.28	3.35
4-sound, 2 syllable	5.35	4.50	3.89
5-sound, 2 syllable	5.35	4.32	3.89
6-sound, 2 syllable	5.35	4.37	3.75

The fact that the relative uncertainty of successive sounds decreases in the course of a word provides an interesting parallel -- perhaps a coincidence -- to some results I obtained in a different line of research. Briefly, the intelligibility of a word was found to be most closely associated with the intelligibility of the initial sounds of a word. The parallel is striking in that in the present instance the greatest uncertainty in the phonetic complex of a word is at the outset.

IV. A somewhat complicated approach to the code of communication was undertaken with contectual material, short phrases from two sources: flight patter and newspapers. Approximately 400 new students in a naval flight training program served as experimental subjects, each responding to 4, 5, or 6 phrases. One group was told the source of each phrase, newspaper or flight patter. The second group heard the phrases described as "newspaper language". In keeping with Shannon's original procedure, a subject was encouraged to continue guessing successive letters in a phrase through ten trials if necessary and then was told the letter if he had not guessed it. He kept a record of his guesses and of the correct letters and was told when the phrase ended. The corpus had the following dimensions:

	Flight phrases	Newspaper phrases
Total phrases and total syllables	64; 320	64; 320
Length in typewriter spaces and median length in spaces	13-25; 19	14-23; 19
Total words and number of different words	260; 140	240; 148
Total letters	1252	1201

On the average uncertainty decreased throughout a phrase from approximately 4 bits for the first space to approximately 1.2 bits for the 21st space. The correct predictions increased from approximately 25% in the first word to 65% in the fifth word. These results are in keeping with ones obtained by Shannon and with our intuitive notions of the power of context. Let me note that these results were obtained from the averages of several phrases superimposed, one upon another. I have recently been somewhat disillusioned in this regard through the results of 10 predictors working with the same paragraphs. I have not extended the results to measures of information. That will come later, perhaps much later. However, the first predictions of the letters of words throughout the paragraph -- and this was carried through to 10 paragraphs from varied sources, each paragraph of approximately 400 spaces -- showed little effect of context. Using a signal-detection model, I found no effect of context beyond 15 spaces. There is still the possibility that measures of information will show the rising function that I have described with newspaper and flight patter phrases and with which we are familiar.

V. One more study that is related to letter prediction. This was one component in a comparative study of the styles of Rufus Choate, his contemporaries, and our contemporaries -- all orators. Again the corpus was phrases, 50 from each source. The phrases included at least 30 typewriter spaces, were randomly selected, and were not necessarily at the outset of a sentence. Each phrase was predicted, letter-by-letter by 10 university students. In contrast to the preceding procedure a predictor was given the correct response after each prediction. The undertainty conveyed by the predicted first letter of words 1, 2, 3, and 4 of the materials provide an interesting contrast.

	Choate	Choate's Contemporaries	20th Century Speakers
(1st word)	(2.674)	(2.674)	(2.674)
2nd word	2.526	2.168	2.269
3rd word	2.371	1.878	2.304
4th word	2.516	2.314	2.158
Mean	2.471	2.120	2.244

My first observation would be that knowing one word of a Choate phrase provided fewer clues to the next word than knowing the first word of phrases of other speakers. The same was true when knowing the second word; and when knowing the third word. I sometimes view this process as the ability of a listener to track a speaker. In this parlance Choate was more surprising from word to word than either his contemporaries or our contemporaries. This surprise was measured by 20th Century university students.

VI. Many people are impatient with letter predictions. Even a word-to-word approach may be too atomistic to satisfy them. I am working in the next instance with 50 separate words and putting them in 5 contexts.

Context 1: The word in isolation.
Context 2: The word repeated in isolation.
Context 3: The word in conjunction with a warning that the word is not a closely identified word aurally.
Context 4: The word in 2nd order word approximation.
Context 5: The word in 3rd order word approximation.

(These last two were determined in the manner of Miller and Selfridge.) The criterion score was word intelligibility as determined by 25 listeners who heard the materials in a signal-to-noise ratio of 2 dB. Maximum uncertainty would be represented by 25 different word responses by the 25 listeners; minimum uncertainty by a single response on the part of all listeners. The range could be from 4.65 bits to 0 bits in a response set to one stimulus. The obtained results were: context one, isolated word 2.68 bits; context two, repeated word, 2.32 bits; context three, word followed by a warning not ______ (the word that had been written most frequently in error), 2.19 bits; context four, second order word approximation, 1.78 bits; context five, third order word approximation, 1.40 bits.

VII. I come now to a different kind of application of information theory, the information conveyed by a closed message set as illustrated by a multiple-choice test. I have a series of multiple-choice intelligibility tests. Pooled, these are comprised of 1124 items. Each listener's response set includes 4 words per item, the test word and the three words that were written most frequently in error when the test words were used in write-down intelligibility testing. Thus the available foils for grew are drew, crew, and groove. These tests were recorded by 4 speakers and heard by 32 listening panels of 12-15 members. Sixteen of the panels heard the list in quiet lines with the signals at 60, 50, 40, and 30 dB listening-level, and 16 heard the lists at 12, 8, 4, and 0 dB signal-to-noise ratio. With decreased level the ratio of the uncertainty of the response to maximum uncertainty increased from 17 to 24 to 58 to 89%; with deteriorating signal-to-noise ratio the percentages increased from 22 to 30 to 54 to 77%. Here, one is concerned with the nature of the distribution of the responses. As the four available responses become more and more equally probable, the uncertainty conveyed by them increases to approximately 2 bits. I anticipated on the basis of earlier work that the principle of a constant ratio among the three error responses would be demonstrated with changing listening circumstances. This was not the case. However, the prospect invites further investigation.

VIII. I have illustrated only a few applications of the use of information theory in human communication. These illustrations have included uses of Shannon's first method, that is allowing a predictor to continue predicting until all possibilities are exhausted, and his 2nd method, that is giving the experimental subject the correct response after each prediction. I have illustrated the use of information theory with word responses, letter responses, and phoneme responses. The focus of interest has varied from context to distributions of available responses and has included inferences about an ill-defined concept, style. The source of the materials investigated has been varied and the effect of knowing the source as opposed to not knowing it has been measured. By inference it seems clear that a prerequisite for the use of information theory is a closed message set. Whether this exhaustive enumeration is an alphabet or the lines of reasoning of Aristotle is inconsequential.

The limited examples have made no reference to one important aspect of information transmission, namely the rate of transmission. I shall not expand on this omission. It has not come within my ken.

REFERENCES

George A. Miller and Patricia E. Nicely, An Analysis of Perceptual Confusions Among Some English Consonants, The Journal of the Acoustical Society of America, 27 (1955), 338-352.

Sadanand Singh and John W. Black, Study of Twenty-Six Intervocalic Consonants as Spoken and Recognized by Four Language Groups, The Journal of Acoustical Society of America, 39 (1966), 372-387.

John W. Black, The Information of Sounds and Phonetic Digrams of One- and Two-Syllable Words, Journal of Speech and Hearing Disorders, 19 (1954) 397-411.

John W. Black, Predicting the Content of Short Phrases, Quarterly Journal of Speech, XLV (1959), 299-304.

John W. Black, Predictability as Related to Style, Speech Monographs, 40 (1973), 101-112.

Gail Nygaard Traul and John W. Black, The Effect of Context on Aural Perception of Words, Journal of Speech and Hearing Research, 8 (1965), 363-369.

John W. Black, Responses to Multiple-Choice Intelligibility Tests, Journal of Speech and Hearing Research, 11 (1968), 453-456.