7 February 1958, Volume 127, Number 3293 SCIENCE

Frequency of Usage and the Perception of Words

Mark R. Rosenzweig and Leo Postman

Telephoning in London with a poor connection, Mr. Sedgwick spells his name, identifying each letter with a word: "*S* as in 'sugar,' *E* as in 'Edward,' *D* as in 'David,' *G* as in 'George,' *W* as in 'William,' *I* as in 'Isaac,' *C* as in 'Charlie," *K* as in 'king.' " (The words he uses are suggested in "Aids toward clearness of speech" in the front of the London telephone directory.)

Monsieur Dumartin, telephoning under similar conditions in Paris, identifies the letters of his name in this way: "*D* comme 'Désiré,' *U* comme 'Ursule,' *M* comme 'Marcel,' *A* comme 'Anatole,' *R* comme 'Raoul,' *T* comme 'Thérèse,' *I* comme 'Irma,' *N* comme 'Nicolas'." (These alphabetic equivalents are in the list given in the Paris telephone directory.)

In both cases, the procedure works and the name is recognized. The alphabetic equivalents are more intelligible than the surnames or the letters. Such examples have intrigued psychologists who are interested in perception. What makes some words more intelligible than others? How should *groups* of words be chosen for maximum intelligibility? Within the past few years, several psychologists (most of them alumni of the Psycho-Acoustic Laboratory at Harvard University) have conducted studies to find what factors determine intelligibility. The results have general significance for our understanding of perception, not only in audition but in other sensory modalities as well. In fact, some of the same investigators have also studied the visual recognition of words, and comparison of results in audition and vision will help to show what is general in the perception of words and what is peculiar to one sensory modality. Studies of the perception of words have broad theoretical as well as practical interest because they illustrate some of the ways in which past experience influences perceptual discrimination.

The authors are on the staff of the department of psychology, University of California, Berkeley.

Since 1950, three main findings have been reported. We shall list them here and then consider how they were obtained and what they imply:

1) A listener can identify words more accurately if he knows the list from which the words will be chosen. The shorter the list, the more intelligible the words will be (*1*). Evidently, then, any list of alphabetic equivalents would be an aid to intelligibility, provided only that the speaker and listener agree upon them. (How much of an aid an arbitrarily chosen list would be, we will consider later).

2) The greater the relative frequency of usage of a word in the language, the more intelligible it tends to be (*2*). Thus, in writing and speaking we use the word *age* about a hundred times more often than the word *cud,* and tests show that *age* is recognized considerably more readily than *cud.*

3) The greater the length of a word, the more intelligible it tends to be, when frequency of usage is held constant (*3*). It had been pointed out earlier that words employed in a list will be more intelligible if they cannot readily be confused with one another.

Measurement of Intelligibility

All the studies from which these results come have used the same basic technique to measure the intelligibility of words: The words are presented one by one in the presence of a masking noise, and listeners attempt to identify each word. The relative intensity of the noise to the words is varied systematically—that is, the sound intensity of the words is held constant, but the intensity of the noise is changed from one part of the experiment to another. (In some experiments, the intensity of the noise is held constant, and the intensity of the speech is varied.) The relative intensity of the noise at which listeners can just hear a word gives a measure of intelligibility, or a recognition threshold for that word. To mask some words may require 20 decibels more noise than to mask others. The importance of this difference may be gauged by noting that the noise of a pneumatic drill (or of a noisy auto) is about 20 decibels more intense than the level of ordinary conversation; these two levels are respectively about 80 and 60 decibels more intense than those of the weakest sounds that normal ears can hear. The fact that one word can be heard over the noise of a pneumatic drill while another word is just intelligible over a background of conversation does not mean that the two words differ greatly in intelligibility in the quiet. The measure tells only how well a word can overcome the effects of masking noise. It is of course precisely where there is noise—in the office, on the street, over the telephone, in the presence of other conversations—that we are concerned about intelligibility, so the measure of resistance to masking noise is a useful one.

To select words with regard to their frequency of usage in English, the Lorge magazine count (*4*) has generally been used. This is a tabulation of 4.5 million words from 1929–1938 issues of five popular American magazines. No large-scale count of *oral* usage in English is available, but the magazine-count frequencies correlate fairly highly (about 0.75) with estimates of oral usage of college students. In French, a count of oral usage has recently been completed (*5*), and we have used it in a study of intelligibility as a function of frequency in French.

Frequency of Usage

Now let us consider the results obtained when intelligibility is measured as a function of frequency of usage. The masked thresholds show that, for each tenfold increase in frequency of usage, about 4 decibels more of masking noise can be overcome. Figure 1 shows this relationship for monosyllabic English words. The most frequent words used in the experiment, shown at the right of the graph, can be heard over about 12 decibels more noise than can the least frequent words used. A similar experiment was performed with French words and French listeners, and similar results were obtained; thus the effect of frequency on intelligibility seems to be a general one.

English listeners would of course be expected to do poorly on the French list, since the frequencies of usage of French words—pronounced correctly—are near zero for English listeners. Similarly, French listeners would perform poorly on the English list. We have even found that American university students whose maternal language was not English score significantly lower on our intelligibility test than do students whose original language was English. Mr. Sedgwick and M. Dumartin clearly require separate lists of alphabetic equivalents while telephoning within their own countries. For international telephonic communication a special list of alphabetic equivalents has been worked out, a list consisting of common geographical terms whose pronunciations are reasonably similar in Western European languages.

Restriction of Alternatives

The effects of changing the length of the list of test words have also been determined experimentally. For a tenfold increase in list length, intelligibility of the words falls about 4 decibels. This indicates that a word does not have an absolute intelligibility. The listener can recognize a word more readily when it is one of two alternatives than when it is one of twenty. Thus the word is perceived in terms of the alternatives available, as the mathematical theory of information would predict. The selection of any group of alphabetic equivalents brings the length of the list to 26. Estimates indicate that, with no list available, listeners respond in terms of a set of alternatives of the order of 30,000 words. Thus an arbitrary selection of alphabetic equivalents reduces the alternatives by 1000-fold and can be expected, on the average, to improve intelligibility by about 12 decibels.

Since the selection of an arbitrary group of alphabetic equivalents would be expected to improve intelligibility, it may be asked why the names of the letters themselves are not satisfactory enough. In favor of the letter names, it could be added that many of them are homonyms of frequently used words such as *a, be,* and *see.* Letter names have one distinct disadvantage, however: Many of them are readily confused with each other. The confusions occur because many of the letter names have the same vowel sound, the largest group sharing one vowel sound being the nine letters *B, C, D, E, G, P, T, V,* and *Z.* The sound power of a word is concentrated chiefly in the vowels, while the consonants are relatively weak. Consequently, it is difficult in the presence of noise to distinguish words that differ in their consonant sounds but not in their vowels.

Length of Word

The influence of word length on intelligibility can be interpreted as a special case of the effects of length of list. When frequency is held constant, intelligibility increases steadily with word length. This beneficial effect of length is found even up to the longest words tested—words of 21 letters. In interpreting this length effect, Howes (*3*) has suggested that the listener can discriminate the length of a word even if he does not recognize the word correctly. The population from which the test word could come would thus be reduced to words of the discriminated length. If this is so, then we have once more a case where the number of alternatives has been restricted. Since there are many short words but fewer long ones, the restriction of alternatives would be greater for, say, an 11-letter word than for a 3-letter word. On the basis of relative frequencies of words of each length in the Lorge magazine count, Howes has calculated the degree to which word length might be expected to aid intelligibility, and his calculated values fit his data well.

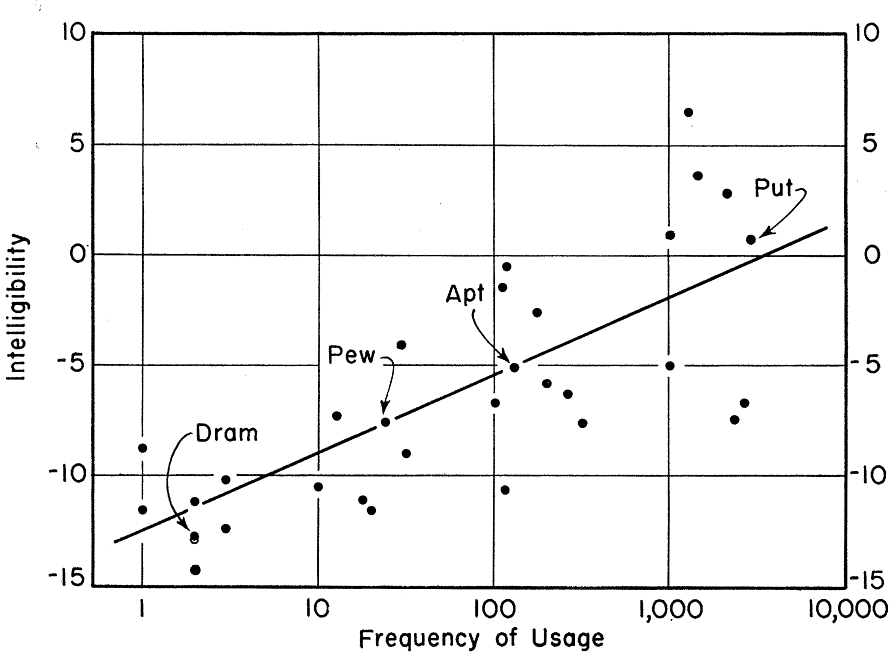

Fig. 1. Intelligibility of monosyllabic English words increases as a function of their frequency of usage. The intelligibility scale indicates the difference in decibels between the intensity of the masking noise and the intensity of the word when the word is just intelligible. Each dot represents one of the words used in the experiment, and four typical words are identified. The figure is modified from Rosenzweig (*2*).

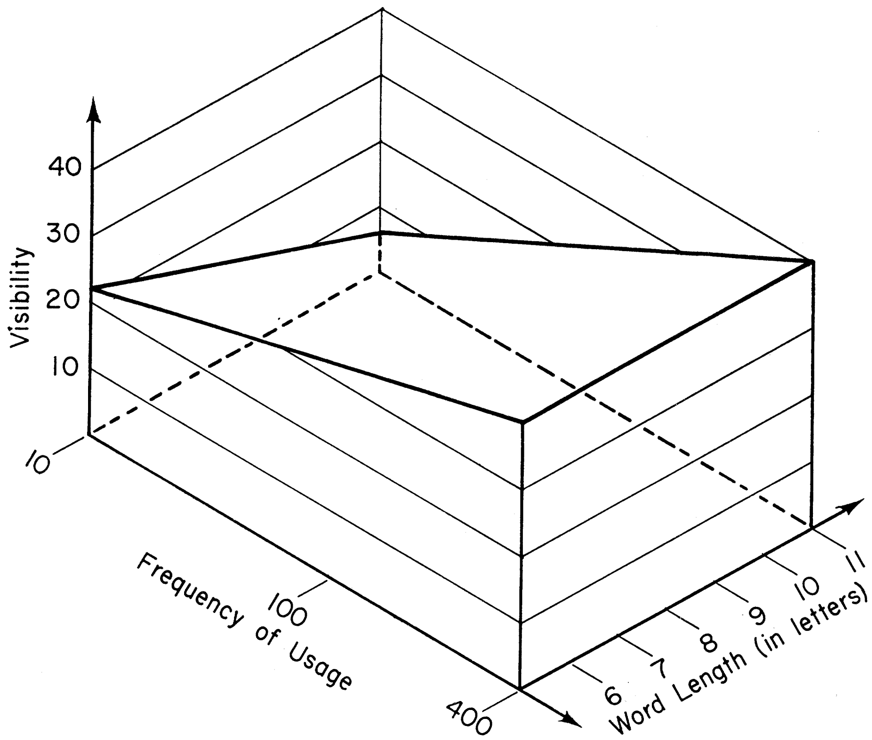

Fig. 2. A three-dimensional representation of visibility of words as a function of both frequency and word length. The sloping upper surface of the figure represents mean thresholds, greater height indicating greater visibility. The greater the frequency of the words, the less long they need be exposed to be seen, and thus the greater their visibility. Visibility decreases with word length, except for the most frequent words. (The values on the vertical scale equal 100 minus exposure duration in milliseconds.) The figure was prepared from data of Postman and Adis-Castro (*9*).

At this point a comparison of auditory and visual experiments is instructive. Experiments similar to the auditory ones have been performed with *tachistoscopic* perception of words. In this case the words are flashed on a screen for exposures of perhaps 1/100 second and the observer attempts to perceive and record them. Thresholds are determined by varying either the duration or the intensity of the flash exposure. Here, too, there is a word-frequency effect; in fact, it was discovered earlier in vision than it was in audition. The more frequently a word is used in the language, the more readily it can be seen (*6*). But in vision, contrary to the case in audition, the *longer* the word is, the *less* chance there is that the observer will report it correctly (*7*). The combined effects of frequency and word length on the visual threshold are shown in Fig. 2. The threshold is indicated by the level of the sloping surface: The higher the level, the more readily the items can be seen. The fact that greater frequency favors perception is shown by the upward slope of the surface from left to right. Length has negligible effects when word frequency is high. With words of lower frequency, greater length impairs ease of perception, as shown by the *upward* slope from back to front at the left of the figure. Figure 3 shows the effects of frequency and word length in audition. Here, too, a higher level indicates greater ease of perception. The frequency effect is similar to that in vision—the threshold surface slopes upward from left to right as frequency increases. But, contrary to the case in vision, intelligibility increases with word length—the surface slopes *downward* from back to front—for both frequent and infrequent words.

The fact that word length has opposite effects in audition and vision requires that we seek for differences between the auditory and visual cases. (Howes' formulation, as stated, would seem to be just as applicable to vision as to audition, but it clearly would not predict correctly for visual perception.) The divergent effects of word length in audition and vision may be explained partly in terms of differences in tendencies to give complete responses to auditory and visual stimuli. In two experiments, one auditory and the other visual, we employed the same list of three-letter words (*8*). The subjects had to write down words masked by noise or words flashed briefly on the screen. The thresholds for correct responses in the two sensory modalities could not, of course, be compared since one cannot measure masking noise and flash duration in the same terms. Taking only incorrect responses, we found that subjects tended to give complete, three-letter responses in the auditory case but that they tended to give incomplete one- or two-letter responses in the visual case. The responses to auditory stimuli tended to be made in units of a word or syllable; the responses to visual stimuli tended to be made in terms of letters or pairs of letters. If a perceived fragment of an auditory stimulus tends to be completed in the response, we would expect that, the longer the word, the greater the chance that the listener will perceive at least part of it and then complete the rest. In vision, however, the longer the word, the greater the chance that part of it will not be perceived and that the report will then be incomplete. It should also be noted that length is penalized by the fixed time of tachistoscopic presentation in a way that does not occur in auditory presentation. In tachistoscopic presentation, words of different length are compared at the same durations of exposure, and there is not time to shift the fixation of the eye as in ordinary reading. The greater the number of units to be discriminated during a single exposure, the more rapidly must each be discriminated. In auditory presentation, each word is pronounced at a normal rate,

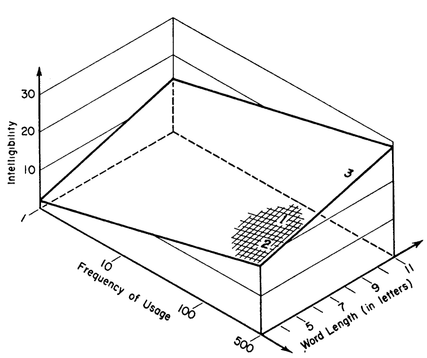

Fig. 3. Intelligibility as a function of frequency of usage and word length. The sloping upper surface represents mean thresholds. With increases in both frequency of usage and word length, words can be heard over greater intensities of masking noise. (The values on the vertical scale equal 20 plus intensity of masking noise in decibels minus intensity of the words in decibels.) The numbered locations on the surface are discussed in the final section of this article. The figure was prepared from data of Howes (*3*).

and longer words consequently are given proportionately longer presentations. The rate of discrimination then need not vary with the length of the word.

Habits and Performance Factors in the Perception of Words

Vision and audition share the beneficial effects of frequency of past usage and of restriction of alternatives upon ease of perception of words. Frequency of past usage may be considered to develop habits of responding to a particular stimulus with a particular response. The more frequently a word has been used in the past, the more readily it is recognized. The restriction of alternatives in a particular situation operates differently. It does not change the perceiver's habits, which are based on his history of past usage. Rather it is a "performance factor" which influences the use made of already formed habits. The smaller the number of alternatives included in the set of words expected by the perceiver, the more ready he can be to use each of the corresponding habits. Thus both frequency of usage and restriction of alternatives make it easier to give a complete response to a stimulus word on the basis of fragmentary discriminations.

A Practical Suggestion

Our survey of the factors governing intelligibility suggests a way in which more highly intelligible alphabetic equivalents might be selected. First let us see where the current alphabetic equivalents are located on Fig. 3. The words of the English telephone list (Andrew, Benjamin, Charlie, David, and so on) have an average length of 5.7 letters and an average frequency of 254 on the Lorge magazine count. These values determine the location labelled *1* near the lower right corner of the surface. Words of the joint U.S.–British radiotelephone list (the familiar Able, Baker, Charlie, Dog, and so on) have an average length of 4.1 letters and an average frequency of 355; this group centers around the 2 on the surface. (In both cases the individual words scatter rather widely around the average values, considerably more widely than the cross-hatching indicates.) Now these lists were not deliberately selected to be on the lower right corner of the diagram, because the studies on effects of frequency of usage and word length had not yet been done when the lists were prepared. The main criterion employed was that the words be common, familiar ones. In the case of the U.S.–British list, empirical tests were used to choose highly intelligible words from a pool of familiar items. The use of the criterion of familiarity leads to words of high frequency and low length, since familiarity implies high frequency and since, in turn, frequent words tend to be short. It should be noted, however, that although frequently used words tend to be short, there are many frequent words of considerable length. The figure indicates that by selecting longer words toward the upper right corner of the surface, the region labelled *3*, a gain of over 6 decibels in resistance to noise might be achieved. In fact, wartime tests of intelligibility demonstrated that many of the items of the U.S.–British list could be improved. Twenty-two changes were suggested on the basis of empirical tests; 19 of the changes substituted items of greater length for the original words. This improved list, which was never adopted, had an average length of 6.4 letters and an average frequency of 56. Length was gained, though with some loss in frequency.

Now that the effects of frequency and of length on intelligibility have been made explicit, it should be possible to choose items from a pool of words that are both frequent and long. The diagram, of course, presents only average thresholds. Within any region, some words are more intelligible and some are less intelligible than would be predicted on the basis of their frequency and length. Empirical tests will therefore be required to determine whether the most intelligible items from the pool of longer words are clearly superior to the most intelligible of the shorter words. If they are, we may find Mr. Sedgwick spelling his name on a future occasion in this way: "*S* as in 'student,' *E* as in 'examination,' *D* as in 'department,' *G* as in 'grandmother,' *W* as in 'welcome,' *I* as in 'industry,' *C* as in 'companion,' *K* as in 'kindness.' "

References

1. G. A. Miller, G. A. Heise, D. Lichten, *J. Exptl. Psychol.* 41, 329 (1951).
2. M. R. Rosenzweig, *Cahiers d'Etudes Radio-Télévision* 12, 283 (1956).
3. D. Howes, *J. Acoust. Soc. Am.* 29, 296 (1957).
4. E. L. Thorndike and I. Lorge, *The Teacher's Word Book of 30,000 Words* (Teachers College, Columbia Univ., New York, 1944).
5. G. Gougenheim, P. Rivenc, R. Michéa, A. Sauvageot, *L'élaboration du français élementaire* (Didier, Paris, 1956).
6. D. Howes and R. L. Solomon, *J. Exptl. Psychol.* 41, 401 (1951).
7. E. McGinnies, P. B. Comer, O. L. Lacey, *ibid.* 44, 65 (1952).
8. L. Postman and M. R. Rosenzweig, *Am. J. Psychol.* 69, 209 (1956).
9. L. Postman and G. Adis-Castro, *Science* 125, 193 (1957).

Strontium-90 in Man, II

Walter R. Eckelmann, J. Laurence Kulp, Arthur R. Schulert

The continuation of nuclear tests makes a thorough understanding of the movement and uptake of strontium-90 a necessity while the levels in man and his environment are still relatively small compared with natural background radiation. In a previous publication (*1*) the first data on a world-wide analysis of strontium-90 in human tissue were presented. The study of the geochemical distribution of strontium-90, of its transfer through the food chain, and of its variation in human populations has continued (*2*). This article (*3*) summarizes over a thousand analyses of human bone and interprets these data in terms of present concentrations of strontium-90 in the various critical phases of the geosphere and biosphere. The new data permit a closer definition of the average concentration of strontium-90 in a large part of the human race, the geographical and dietary variation, the increase in the concentration with time, and the distribution in urban populations. From the existing data, an attempt is made to predict future levels under specified conditions.

Dr. Kulp and Dr. Schulert are on the staff of the Lamont Geological Observatory (Columbia University), Palisades, N.Y. Dr. Eckelmann, formerly a member of the staff of the Lamont Geological Observatory, is now at the Carter Research Laboratories, Tulsa, Okla.