

QUANTITATIVE EVALUATION OF TRANSMISSION OF MEANING

Hiroya Fujisaki
Department of Electrical Engineering
Faculty of Engineering
University of Tokyo
Bunkyo-ku, Tokyo, 113 Japan

Yasuhiro Katagiri
Department of Electrical Engineering
Faculty of Engineering
University of Tokyo
Bunkyo-ku, Tokyo, 113 Japan

As a step toward quantifying the processes whereby meaning is transmitted by language, an experiment was designed to facilitate observation of the process of verbalization by the sender as well as the process of comprehension by the receiver. Both sender and receiver characteristics were measured experimentally by confining the subject matter to ages and by restricting the available vocabulary to a small set of nouns. A model was then constructed for the entire process through which meaning is transmitted by a word, clarifying various causes of ambiguity and inaccuracy of transmission. The whole process was evaluated in terms of the amount of information and the r.m.s. error of transmission.

1. INTRODUCTION

Language is obviously an important medium of human thought and communication, and elucidation of its essential characteristics is indispensable not only for understanding the human processes of thought and communication, but also for machine processing of linguistic information. Most of the studies on language, however, have been concentrated on the nature of the language itself as a system of codes, and little attention has been paid to the relationship between the linguistic code and the information conveyed by the code [1].

The present study deals with the communicative roles of language between a sender and a receiver, and experimentally analyzes the sender's process of conversion from information to word as well as the receiver's process of conversion from word to information. The process of transmission of meaning is then formulated and quantitatively evaluated in terms of the amount of transmitted information and the accuracy of transmission [2].

2. PROCESSES IN TRANSMISSION OF MEANING

We deal here with a situation in which a physically definable stimulus is converted by a subject (*sender*) into a linguistic expression (a noun in this case) and is received by another subject (*receiver*). The input stimulus is first perceived by the sender and then expressed as a noun. The combined characteristics of perception and verbalization by the sender can be measured as the input-output characteristics, and will be referred to as the *coding* characteristics. On the other hand, the communication is generally terminated by the receiver's process of comprehension. In order to render the receiver's percept into a directly observable form, however, we add here another process in which the receiver reproduces the stimulus which he may find most appropriate for the received linguistic message. Thus the combined characteristics of comprehension and reproduction by the receiver can be measured as the input-output characteristics, and will be referred to as the *decoding* characteristics. The processes involved in this verbal communication situation is schematically shown in Fig. 1.

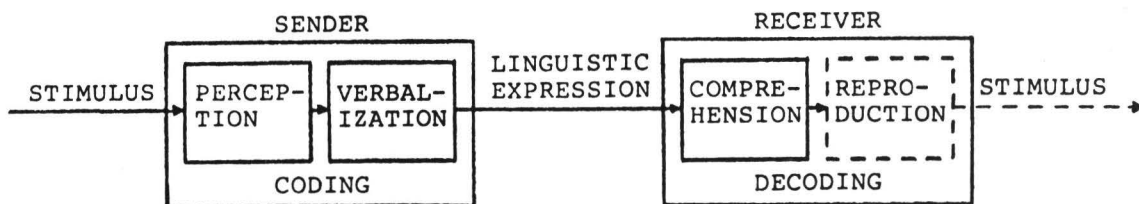


Fig. 1. Processes in transmission of meaning by language.

3. CODING AND DECODING CHARACTERISTICS

As an example of experimental determination of the coding and decoding characteristics, an investigation was conducted on the relationship between the age of a person and the nouns of Japanese commonly used to designate the age. These nouns were selected such that they were nearly complementary with each other in designating age and uniform in other aspects. The major part of the following experiments was conducted by using the five-word vocabulary: "yo-nen", "sho-nen", "sei-nen", "so-nen", and "ro-nen", corresponding roughly to the English "childhood", "boyhood", "youth", "manhood", and "old age", respectively. The vocabulary size was varied, however, in studying the effect of its change on transmission accuracy. The method of constant stimuli was used to measure the coding and decoding characteristics separately on each of the nine subjects. Because of individual differences, the data from each subject had to be analyzed separately.

Figure 2 shows the coding characteristics of one subject as the probability with which he assigns a noun against a given age. In the vicinity of a category boundary, each of the probability distribution functions can be approximated by a Gaussian distribution function with a mean θ and a standard deviation ρ , to be defined here respectively as the coding boundary and the coding accuracy, suggesting that the selection of a particular noun is based on a categorical judgment whose threshold fluctuates due to a number of psychological factors. Thus the coding characteristics of a subject can be represented by the set of θ 's and ρ 's for all the category boundaries.

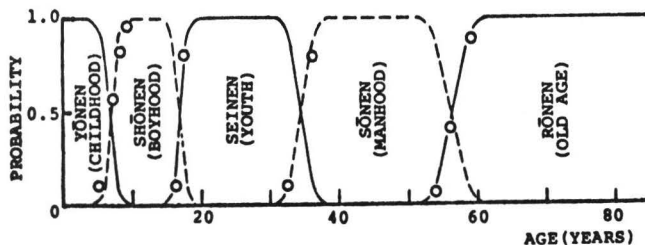


Fig. 2. An example of coding characteristics.

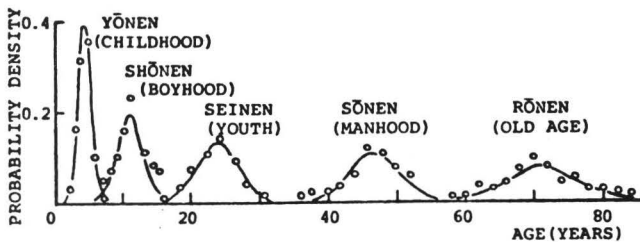


Fig. 3. An example of decoding characteristics.

Figure 3 shows the decoding characteristics of the same subject as a set of five probability density functions, each representing the probability density that a certain age is reproduced against a given noun. In this case, each of the density functions can be approximated by a Gaussian density function with a mean μ and a standard deviation σ , to be defined here respectively as the decoding reference and the decoding accuracy, suggesting that the reproduction of a particular age is based on a reference which is perturbed by a number of psychological factors. Thus the decoding characteristics of a subject can be represented by the set of μ 's and σ 's for all the categories (nouns).

4. A MODEL OF SEMANTIC INFORMATION TRANSMISSION

The results of the foregoing experiments lead to a model for the processes of semantic information transmission shown in Fig. 4. The coding process at the sender can be considered as quantization followed by code assignment. The quantization is probabilistic since each threshold θ_i is perturbed by a Gaussian noise m_i . On the other hand, the decoding process at the receiver can be considered as biased reproduction of the quantization levels. The bias β_i represents the discrepancy between the center of the i th coding step $(\theta_i + \theta_{i-1})/2$ of the sender and the i th decoding reference μ_i of the receiver. The reproduction is also probabilistic since each decoding reference is perturbed by a Gaussian noise n_i . The three variables x , w , and y in Fig. 4 respectively represent the meaning implied by the sender, the code word adopted for transmission, and the meaning recovered by the receiver. The entire process as a communication channel can be characterized by the conditional probability $q(y|x)$.

These formulations allow one to evaluate the mutual information R (in bit/word) and the r.m.s. transmission error E (in units of age) for a given probability density $p(x)$ of the stimulus x

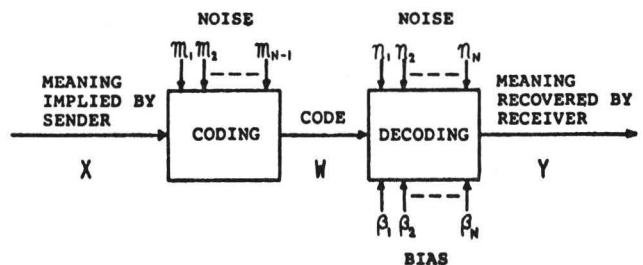


Fig. 4. Formulation of the process of semantic information transmission.

presented to a sender. They are given respectively by

$$R = \int \int p(x) q(y|x) \log[q(y|x)/q(y)] dx dy,$$

where

$$q(y) = \int p(x) q(y|x) dx,$$

and

$$E = \left[\int \int p(x) q(y|x) (y-x)^2 dx dy \right]^{1/2}.$$

The mutual information R indicates the requirement on the lower bound of the channel capacity for semantic information transmission necessary to achieve the accuracy of transmission indicated by the r.m.s. error.

These quantities were calculated from the measured data of coding and decoding characteristics of three subjects, obtained for the vocabulary size of two, five, and eight. The probability density $p(x)$ of the stimuli was assumed to be uniform over the range $[0, 80]$, and the calculation was made for all possible sender-receiver pairs. The results are shown in Fig. 5, where a small circle indicates the mean of all sender-receiver combinations, and a vertical line segment indicates the range of distribution. The results indicate that both the accuracy and the transmitted information increase with the vocabulary size, but also indicate that there is a limit to the accuracy attainable by increasing the vocabulary size. Thus there seems to exist an optimum size of vocabulary from the point of view of transmission efficiency.

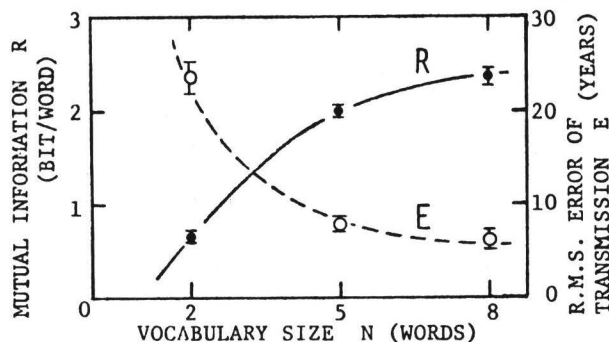


Fig. 5. Mutual information and r.m.s. error of transmission versus vocabulary size.

5. DISCUSSION AND CONCLUSIONS

The probabilistic nature of language use was earlier pointed out by Lenneberg and his co-workers [3,4]. Due to the use of unrestricted vocabulary and pooling of data from many subjects, however, these authors did not attain a clear insight into the underlying process of coding in a language user. The problem is also treated by Zadeh [5] in connection with the concept of

"fuzziness." Despite the conceptual distinction made by Zadeh between probability and the membership function, the present authors consider that the membership function is equivalent to the probability distribution in the coding characteristics. It should also be noted that these earlier studies did not separate the two kinds of indeterminacy in language use, i.e., the indeterminacy of coding and that of decoding.

The present study was based on the clear separation of the two processes in semantic information transmission, and developed experimental techniques to observe both the process of verbalization by the sender and the process of language comprehension by the receiver. The characteristics of these processes were measured and formulated. Based on these results, a model was constructed for the process of semantic information transmission by means of a word, and a method was also shown to evaluate the process in terms of the ambiguity and the accuracy of transmission.

Although the results shown in this paper are quite limited, the techniques described here are applicable to a wide range of problems of language use, such as the quantitative analysis of individual, contextual, dialectal, and language differences in semantic information transmission.

ACKNOWLEDGMENTS

The research reported here was supported by a Grant-in-Aid for Scientific Research (No. 310708) from Japanese Ministry of Education.

REFERENCES

- [1] Shannon, C. E. and W. Weaver: The Mathematical Theory of Communication, Univ. of Illinois Press, 1949.
- [2] Fujisaki, H., K. Hirose and Y. Katagiri: Transmission of Meaning by Language, Annual Bulletin, Research Institute of Logopedics and Phoniatrics, Univ. of Tokyo, 12, 127-140, 1978.
- [3] Brown, R. W. and E. H. Lenneberg: A Study in Language and Cognition, J. Abnormal and Social Psychology, 49, 454-462, 1954.
- [4] Lenneberg, E. H.: Biological Foundations of Language, John Wiley & Sons, 1967.
- [5] Zadeh, L. A.: Fuzzy Sets, Information and Control, 8, 338-353, 1965.