HUMAN CHANNEL CAPACITY IN PRODUCING ONE-DIMENSIONAL VOCAL SOUNDS

MICHEL HUPET

There is in the literature quite a number of studies which have explored the laws and the limits of the absolute or categorial judgment in the identification of various signals (Miller 1956). Stimuli are presented in such a way that they are continuously or uncontinuously varying in oniy one dimension; the subject's task is to classify them into a number of categories N, generally numbered from 1 to N. After each presentation of a stimulus, the subject has to respond by a number from 1 to N. By varying N, the maximum number of categories which can be used without any confusion can be determined; this number corresponds to the cologarithm of the amount of transmitted information, and expresses the 'channel capacity'. Measures taken on several perceptive dimensions had led to state rather small and equivalent capacities ranging from 2 to 3 bits, which corresponds to 4-8 correctly utilizable categories. Examples can be found for pitch (Pollack 1952), loudness (Garner 1953) and duration of sounds (Murphy 1966). It is with reference to such results, among others, that Jakobson and Halle (1956) try to warrant their well-known binary hypothesis.

Although the experiments on absolute judgment are numerous, only very few of them make use of the inverse procedure: instead of presenting a stimulus that the subject has to identify by a number, one would give him a number and he would then have to generate a corresponding motor response. Using this new procedure, it would be possible to measure the capacity of the human channel in selecting behaviors, i.e., to measure the maximum number of different categories of behaviors which can be accurately used. In other words, this new procedure states the question of knowing among how many types of behaviors, varying in only one or in several dimensions, the subject is able to choose without any risk of error. We have inquired into this problem about verbal behavior, starting with the most simple situation in which the signals to be emitted vary by only one dimension, here their intensity or their duration. Our hypothesis is that the number of uni-dimensional vocal signals the subject is able to emit without any confusion is very small; according to this hypothesis, the channel capacity in producing signals would be the same as that in identifying analogous signals.

Concretely, we have asked several subjects to emit vocal sounds — actually, the french vowel 'a' — pursuing to the following instruction for the intensity:

> The vowel 'a' can be pronounced with more or less intensity; divide mentally the scale of vocal intensity you can use into 3 (4 or 6) categories that you number from 1 to 3 (4 or 6), from the least to the most powerful 'a'. Everytime I give you a number from 1 to 3 (4 or 6), you will have to pronounce an "a" with the corresponding intensity. It is not necessary to pronounce an 'a' of a well fixed intensity in response to each number, but if I give you a higher number you will have to pronounce a more intense 'a' than the 'a' emitted in response to smaller numbers.

The analysis of the results first consists in ranging the subject's responses in their ascending order, indicating for each of them the stimulusnumber which evoked it. After that, we draw a line under the n_1 first responses (n_i = number of times that the stimulus i has been presented), then a line under the n_i following responses, and so on until the n_i last responses. Therefore we find the same number of categories for the responses and for the stimuli, each category of responses containing some correct responses and some incorrect ones which are to be interversions. An interversion between the stimuli 3 and 5 for example will be noted as a response 3 to the stimulus 5, and as a response 5 to the stimulus 3. By counting these interversions, it is finally possible to establish confusion matrices from which the transmitted information can be calculated.

When the subjects have to generate vocal sounds varying in only their intensity, the channel capacity is surprisingly very small, ranging about 1 bit, which corresponds to the correct selection of only two categories of vocal intensity. The locutor is therefore severely limited when he has to choose the intensity of vocal sounds; this severe limitation is very surprising if we compare it with that one obtained by Garner in the identification of analogous acoustic signals. It appears from that comparison that the listener is able to identify accurately four or five levels of intensity while the locutor can only produce two of them. The difficulty of setting up solid anchors must be taken in account, but this doesn't explain why the vocal generation of intensity is submitted to greater limits that the unvocal generation of similar signals. In a previous experiment indeed, we asked several subjects to adjust the intensity of acoustic signals electronically produced; in that case, we found a capacity of 2,3 bits. Nevertheless we find possible to explain the very small capacity if we agree with the hypothesis that in vocal generation of intensity, the categorisation of the sounds is made on a kinesthesic criterion (the capacity is very small in that field) which would be the 'vocal effort'. An outstanding experiment will have to prove this hypothesis.

When the subjects have to generate vocal sounds varying in only their duration, the channel capacity is ranging about 2,4 bits, which corresponds to six categories of duration correctly utilizable. For this dimension, it still must be said that the capacity is increased by some important anchor effects; it can be estimated that these anchor effects — evaluated in terms of transmitted information — increase the capacity of nearly 0,7 bit.

Just as in the assimilation of information, the human channel presents a limited capacity in the generation of information. The selection of an element in a repertory is submitted to really constraining limitations. The listener is submitted to these limitations while identifying the signals reaching him, but the locutor is also limited when he has to choose in his repertory the signals he will emit. It must be said, too, that these limitations in the categorial generation of vocal sounds cannot be explained by sensorial difficulties; it is well known how large is our ability to discriminate relative differences among signals or motor responses. These limitations proceed from the Central Nervous System itself which is accountable of the encoding of messages. In other words, in the two experiments reported here, the difficulty is not for the subject to pronounce different sounds, but to keep, in his repertory of vocal sounds, someones which remain different from others. Finally, the limits observed in the generation of vocal sounds must be able to account for the structure of the phonological systems which is nearly the same in all languages.

National Science Foundation of Belgium and
Centre de psychologie expérimentale
Université de Louvain

REFERENCES

Garner, W.R.
1953 "An Informational Analysis of Absolute Judgments of Loudness", *Journal of Experimental Psychology* 46:373-380.
Jakobson, R. and M. Halle
1956 *Fundamentals of Language* (Den Haag, Mouton).
Miller, G.A.
1956 "The Magical Number Seven, Plus or Minus Two", *Psychological Review* (1956):63.
Murphy, L.E.
1966 "Absolute Judgments of Duration", *Journal of Experimental Psychology* 71:260-263.
Pollack, I.
1952 "The Information of Elementary Auditory Displays", *Journal of the Acoustical Society of America* 24:745-749.

DISCUSSION

LÉON M. (Toronto)
Est-ce que vos expériences vous ont amené à considérer le problème de l'intensité spécifique des voyelles sur les plans génétique, acoustique et perceptif?

HUPET
Nous n'avons pas abordé directement ce problème; néanmoins, si l'on se rappelle que les tenants de la théorie motrice mettent en relation le phénomène de l'intensité spécifique avec l'effort articulatoire fourni par le locuteur, et si l'on se rappelle les

travaux du Prof. M. Rossi selon lesquels l'intensité spécifique résulterait principalement de la distribution de l'énergie sur le spectre (notamment au niveau du 2e formant), nous pouvons dire ceci: les résultats de nos expériences nous conduisent à penser que, dans la génération vocale, la catégorisation d'intensité se fait sur base d'un critère kinesthétique qui serait l' 'effort vocal à fournir'. En analysant les spectrogrammes des *a* prononcés par les sujets, nous avons constaté qu'il n'y avait toujours en fait que deux types de spectres: soit un spectre étroit (obtenu pour les *a* peu intenses), soit un spectre large (obtenu pour les *a* très intenses). On peut faire l'hypothèse que ces variations du spectre expriment la quantité d'effort vocal fourni; dans ce cas, la catégorisation des émissions se ferait, pour l'intensité du moins, sur base du critère suivant: peu d'effort — beaucoup d'effort. Pour vérifier ce point de vue, nous avons mis au point l'expérience que voici: il s'agit de la même tâche, mais les sujets sont tous casqués et reçoivent un bruit blanc d'environ 90 db (re). Ils n'ont plus recours dès lors aux contrôles auditifs interne et externe; ils ne disposent plus que d'un contrôle kinesthétique. Si notre hypothèse est correcte, nous devrions trouver dans ce cas une capacité sensiblement égale à celle que nous avions précédemment obtenue.

En ce qui concerne l'aspect génétique, nous avons entamé quelques travaux qui ont pour objet de déterminer chez des enfants de trois à six ans la 'capacité' du canal. Nous n'avons jusqu'ici mené ces travaux que dans le domaine visuel; si les 'capacités' obtenues sont assez faibles (plus faibles que pour l'adulte), c'est probablement en raison de ce que les enfants utilisent moins ou établissent moins d'ancrages perceptifs.