

FOR AN UP-TO-DATE VISUAL REPRESENTATION OF SPEECH

Raymond THOMAS, Laboratoire de Langues, Fac.Sciences de Luminy, 13288 MARSEILLE Cedex9, France
 Serge BAGNOLI, Fac. Sciences St-Charles, U. d'Aix-Marseille I, 13331 MARSEILLE Cedex 3, France
 Jacques GENIN, DAII-SEDE, 38-40, rue du Général Leclerc, 92131 ISSY LES MOULINEAUX, France
 Robert H. GREEN, Dpt of Psychology, Ann Arbor University, Michigan 48100, USA
 Hubert GREVEN, Fac. de Lettres et Sc. H., Université de Rouen, 76130 MONT SAINT AIGNAN, France
 Nils-Olof JÖNSSON, Språkpedagogiska Forsknings AB, Box 63, S-230 10 SKANÖR, Sweden
 Antony McKENNA, Laboratoire de Langues, Fac.Sciences de Luminy, 13288 MARSEILLE Cedex 9, France
 William WEISS, Dpt of Theatre, University of Ottawa, OTTAWA, Ontario, CANADA K1N 6N5

RESUME

Une meilleure information sur les connaissances acquises en acoustique de la parole pourrait être immédiatement disponible pour tous grâce aux techniques informatiques. Nous donnons un exemple de la forme qu'elle pourrait adopter et prions membres et organisateurs du Congrès de préparer l'établissement d'un organisme représentatif qui poursuivrait la réalisation d'une image visuelle, probablement syllabique, de la parole, informant ainsi d'une façon satisfaisante de l'acquis scientifique en ce domaine.

AIMS

At recent meetings of acousticians, I have pointed to the predicament in which we, who teach foreign languages, find ourselves: we must know what speech habits the student should change in order to speak another language well, but the scientific information about pronunciation that we need for that purpose, is too often lacking, or practically unavailable, or uncertain. At the 1986 International Congress on Acoustics at Toronto, I was asked to give a model of what we precisely want. As may be expected, that question can be answered only in part, and that's why the drawings we'll show to give a clear answer, will be somewhat like dreams, made of both true and doubtful elements. Nevertheless, as will be seen, they should help in stating more precisely what elementary problems still plague our general knowledge of speech, and suggest steps to be taken in order that better information on this subject be made readily available to all.

WHAT PARTS ARE BEST DESCRIBED?

Though speech statements are unpredictable, their parts are limited in diversity, and thus susceptible of scientific description. The most obvious elementary parts in speech chains are the syllables, but these have long been divided into vowels and consonants, and lately these smaller parts, in their turn, were analysed into smaller elements variously described as phonemes, features, indices, etc. The first problem therefore is: what elementary parts of pronunciation should be selected for a practical, realistic, and up-to-date visual description of speech?

Three Reasons for Choosing Syllables.

Tentatively in these drawings, we have chosen syllables as basic units, and for a long time past, the three following reasons could have been given for this choice. First, they are the only divisions that are immediately perceptible to all language users. Second, we may therefore say that syllables are the basic units used by our minds to process speech if, for the sake of brevity, you forgive us for using such simple terms. Thirdly, the finer divisions vary according to the syllables in which they are to be found, and, for a scientific description, these variations nullify the fundamental advantage of such smaller elements, i.e. that of being less numerous than syllables.

Why Finer Divisions Are Unsatisfactory.

If vowels and consonants, and all the more phonemes, features, etc., vary according to the other components with which they are to be found, two attitudes may be adopted. First, if we want to give a realistic visual representation of speech with vowels and consonants, we must distinguish between the various sounds that each of these may represent. This recently led to the International Phonetic Alphabet, and we know that such notations cannot be satisfactory because too many diacritic signs make them impracticable, even though only the more important variations are noted.

A second attitude is to forgo writing as an unequivocal system of signs; this is what we generally do and, particularly for the study of our Western languages, it is one of the basic obstacles students have to overcome. But, more fundamentally, such an equivocal use of vowels and consonants cannot be a satisfactory basis for a scientific description of language sounds since what is then spoken of, in one instance, may not be the same object as what is described in another, without our being aware of this change in the object. Many problems of present day acoustical research should be ascribed to this fundamental cause.

The Decisive Reason for Choosing the Syllables.

But the practical reason which motivated our choice of the syllables as the basic units to be described in the speech chain, is pedagogical. For over twenty years now, teaching experience has consistently shown that a predominant attention to syllables,

rather than to the individual vowels and consonants, is a decisive factor in helping English and French speakers to use the other tongue without those irrepressible defects that mark our two communities as particularly unsuccessful among European students of foreign languages. This pedagogical advantage of stressing attention to syllables does not seem to be restricted to French and English, but to apply it to other languages requires special studies for each of these, and this was why we first asked acousticians for an improved general description of language sounds. The benefits for the millions of language students the world over, suggest the importance of the pursuit, and its interest is not limited to pedagogy.

A New Reason for this Choice.

Lastly, finer analyses of speech sounds have recently shown that vowels and consonants are not always separated in the time sequence as we formerly thought, and as is recorded in our spellings. On the contrary, syllables follow strictly the sequential character of speech. At bottom, this may be but another aspect of the first two reasons we mentioned, but it is a most compelling argument when considering visual representation.

HOW COULD SUCH DESCRIPTIONS BE ACCOMPLISHED?

This question requires two different answers. As we were invited to do, we must give a precise view of what we would find in such new descriptions, and this we shall do with our drawings. But a more practical answer is also needed: showing what should, or could, be done in itself ineffectual; we must also say how the necessary work can be accomplished. This need for a better description of languages is not just theoretical; congresses such as this are occasions to start a process leading to some organization for a cooperation among acousticians who can enlighten us in their various fields. The first step we mentioned: switching more scientists' attention to syllables, is too heavy a task for anyone, and therefore it requires the action of such a representative body. And in the course of our description of a desirable new visual representation of speech, the need for such a research organization will again be found in several convincing ways.

A FEASIBLE SYLLABIC VISUAL REPRESENTATION

Though it cannot be done in this black and white paper, we would take advantage of the widening use of color and information processing. We would have three separate drawings for production, transmission, and decoding of syllables. In these drawings, the common parameter of duration will be shown on the horizontal or "x" axes of our two-dimensional models. We have assigned different colors to each of the drawings by specifying the lengths of the corresponding light-waves, as will be seen in the fourth view that can be shown only with colors. In black and white, the superposition of lines and areas would be confusing.

Light-wave lengths between 7200 and 5600 Angstroms may mark the range of colors to be used in figure 1; for Fig. 2, the choice would be more limited (5600 to 5000 A.) since the diversity in the airwaves pressures is not so large as for the phenomena described in Fig. 1 and 3. For this third figure, the corresponding lengths would extend from 5000 to the end of the specter.

Using light-wave lengths as practical units may be deemed artificial, and since these drawings will be interpreted with the help of computer processing, the specifications of colors that computers manufacturers provide, could be more conveniently used. Hopefully, our research organization could work for standardization of these color specifications.

Further Notes on the Proposed Representation of Syllables.

The various intensities, or degrees, in the recorded phenomena, could be shown by the intensity of the colors used. Within each of the three groups of recorded phenomena, related facts or events could be pictured in the same colors. Some type of color correspondance could also be specified for related parts in the three drawings, whenever such relations are generally recognized.

Our drawings show individual production of precisely determined syllables, and not more or less general models of syllables, since "norms" and "averages" are not yet established for a sufficient classification among them. Only an important research organization could usefully set standards or "averages" to obtain such classifications.

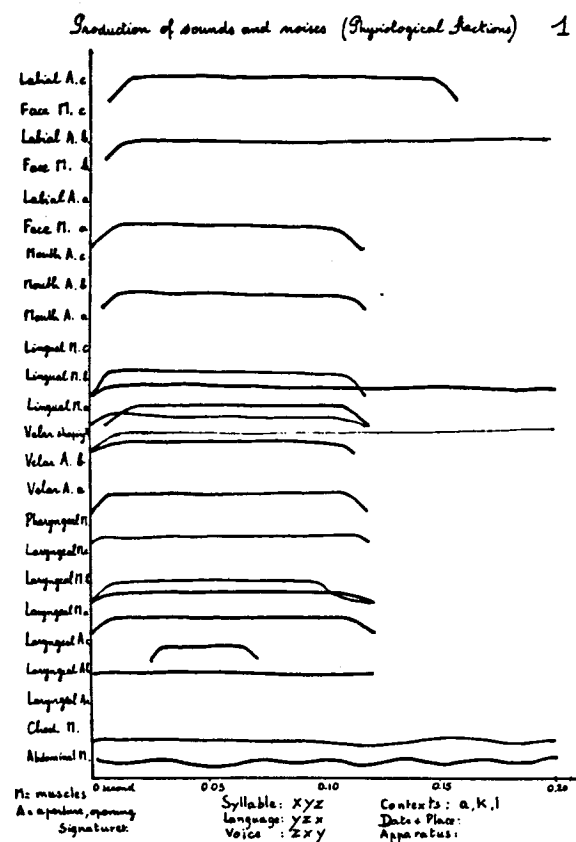
"Model" or "average" syllables could thus be determined, but several types of occurrences must be separately represented. To give just one instance of this unavoidable multiplicity, the duration of accented syllables cannot usually be assimilated to that of unaccented syllables. In our use of this complex information, a limited number of properly selected syllables would be sufficient for an efficient study of the sound characteristics of a given foreign language. On the other hand, storing the large volume of the pertinent information on all the syllables in any language, is not a major problem for computers. Similarly, the calculations needed to obtain "average" or "normal" models from a sufficient number of individual performances does not seem to be unusually large either.

Since such "models" are not available now, our present drawings show the performances by a "normal" speaker or hearer in a given language for what we deem a representative occurrence of the chosen syllables.

We must also note that, similarly to the establishment of "norms" of syllable occurrences, the decisions on many points in our drawings require a knowledge and an authority that can rest only with a representative research organization.

In our drawings, we have purposely included some obviously unusual opinions in order to underline both that these pictures are intended just to show one possible kind of up-to-date visual representation and that our generally accepted opinions may have to be modified, for instance as a result of a wider study of speech in the various communities.

FIGURE 1 : SPEECH PRODUCTION



F.2 "TRANSMISSION OF SPEECH" OR "AIR PRESSURE WAVES"

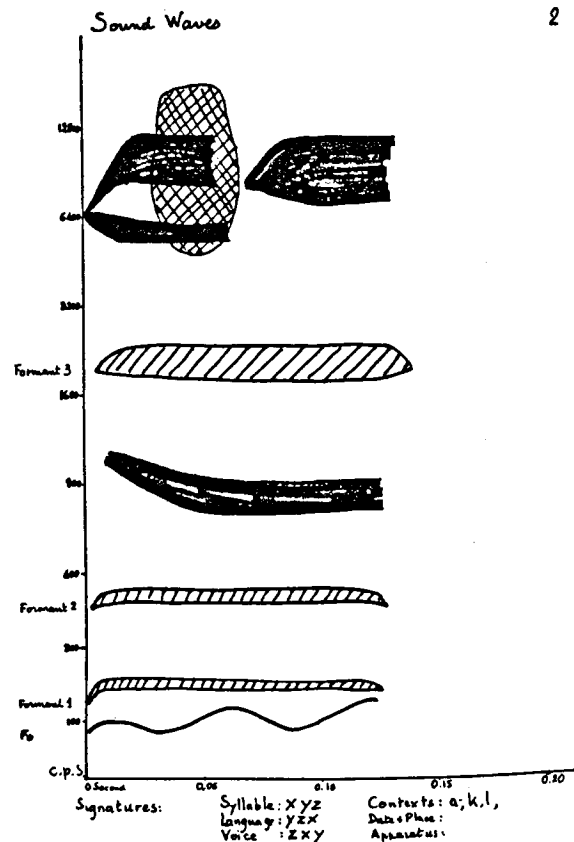
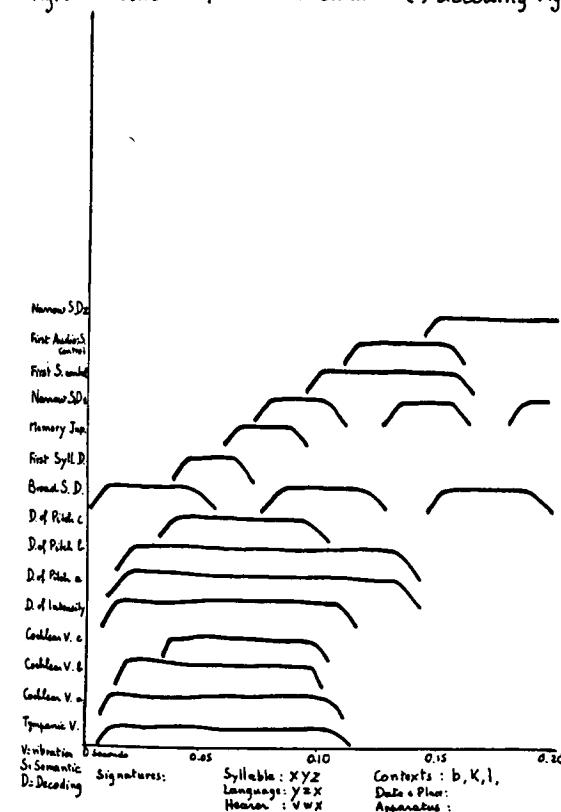


Fig.3 Aural reception and semantic(?) decoding Fig.3



In this, and in the third pictures, we should imagine as many "x", or horizontal, axes, as there are phenomena, or groups of phenomena, indicated in the left margins. In these drawings, intensities or degrees are often marked by curves, and therefore various intensities of colors are not needed here; so these could be used for information that is not now available about these phenomena. Though we have hinted at much of present day knowledge, we beg to be forgiven for insisting that these drawings are not meant to be complete, but merely to show how knowledge about language sounds may be accurately recorded in visual representations. Perhaps we should have grouped together muscular actions in the higher or lower part, and the positions of the various organs of speech in the other part of the figure. The aspect of the picture would change much according to such options. There are at least two often conflicting rules to be followed in all our drawings: 1/in the conventions we adopt, we should try to keep as near to actual facts as we can; 2/the drawings should be as expressive as possible. In this, and still more in the fourth totalizing picture, various factors will be mixed that cannot readily be isolated by an untrained user. However, though in a different way, this is also true for the sounds of real speech. So long as the same conventions are strictly adhered to for all representations, their artificial character is not to be unduly regretted.

Since neither the voice, nor the language, and not even the syllable are specified, this drawing is most arbitrary, and the sounds of speech it partly represents, may cause more surprise than in the first figure. In the fundamental frequency, we purposely marked a variation which is unusual in our European languages, in order to stress the fact that most of the five thousand, or so, languages have not yet been scientifically studied, not even briefly. Therefore, what we consider as basic and general rules of the production, transmission, and decoding of speech, could be less universal and necessary than we hold.

Fig. 3 DECODING THE NATURE OF THE TOTAL SOUND.

Picture on top of next column. For a long time, this third view should remain the least complete of the three. Recognition of pitch is not achieved in one way only, as could be inferred from our drawing; but that is not the only point for which more knowledge could already be noted. Still, in our picture we have shown our opinion that it is not realistic to describe syllables produced outside meaningful speech, without due caution. To be sure, not everybody would agree that, meaning being essential to speech, an analysis that unguardedly excludes the meaningful aspect of language, runs a risk of being pointless or wrong. Perhaps will it still take many years before this opinion is examined by influential persons, and yet we have noted it in this third figure,

our attitude about semantics was marked, not with a view to have it accepted, but especially to show the inclusive character that these new visual representations of speech could have. A limited organization, not to speak of an isolated individual, cannot be responsible for such a vast endeavor.

CONCLUSION

Especially in their totalizing pictures, which cannot be given without colors, our successors may bring a measure of beauty that must now remain a conjecture. But, as we have beautiful scripts and prints, perhaps our conventions will be modified in the future with this preoccupation in mind. For now, let's limit ourselves to voicing a wish that is probably in each of us: to have an easy and accurate means of access to all the findings which our colleagues will have gathered on our subject. But, to be practical in this conclusion, our duty now is to beg you, members and organizers of this meeting, not to leave it to a hypothetical future to take steps towards establishing a representative authority that can change our present dream into a not too far distant reality.

BIBLIOGRAPHY

BERNARD J. Comprendre et organiser le traitement automatique de l'information Dunod Paris 1967.

CHOMSKY N. Language and Mind Harcourt NY 1968.

DELATTRE P. "De la hiérarchie des indices acoustiques pour la perception de la parole" Proc. 5th IC Phon. Sciences 1962 244-251

" Principes de phonétique française, Middlebury College, Middlebury Vermont 2è éd. 1951

EIMAS P.D. "The Perception of Speech in Early Infancy" Scientific American Jan. 1985

FOUCHE P. Traité de prononciation française Paris, Klincksieck 1956

GENIN J. Contribution à l'étude des systèmes de prothèse auditive sensorielle Ms thèse BU Grenoble '87

GREVEN H. Elements of English Phonology Paris Puf '72

HAGEGE C La structure des langues Paris Puf 1982

LAVER J. The Phonetic Description of Voice Quality Cambridge UP 1980

LENAT D. B. "Computer Software for Intelligent systems" Scientific American September 1984

POLANYI M. "Sense-Giving and Sense-Reading" Philosophy Macmillan Oct. 1967

SAUSSURE F. (de) Cours de linguistique générale 1916

THOMAS R. Evaluation de l'information en langues naturelles Manusc. Thèse BU Rouen 1983

" "Teaching Foreign Pronunciations" Abstract in JASA Suppl. 1 Vol 76 Fall 1984 Minneapolis

" "Recherches pour l'enseignement des prononciations étrangères" JEP ENST Paris 1985

" Pour décrire efficacement les phonétismes" 12 ICA Toronto 1986.

Time-Life Les images électroniques coll. Le Monde des ordinateurs, Amsterdam Déc. 1986

WINOGRAD T. "Computer Software for working with language" Scientific American September 1984.

to show that such representations of speech could be used to other ends than mere information. To defend the validity of this opinion however, perhaps we might note that, even in a language like French, in which syllables are relatively unaffected by the meaning they convey, there are still important variations due to their specific role in a sentence. For instance, the syllable "in" which often means "negation" in some way or other, sounds quite differently in words such as "inexact", "innombrable" and "impossible", and still more unexpectedly in "innovation" and "innové". Even more directly in keeping with our third drawing, the syllable "ma" may be more easily identified as a part of a word like "matériel" than if it is pronounced separately, as in contrast with "pa" or "ba", "fa", etc.

Semantic decoding cannot be overlooked therefore, without unpredictable consequences for an objective study of speech. We have even noted that, for the same speech fact, contexts generally differ for the speaker and the hearer. We have suggested only minor differences in this case, but in our view, this point cannot be overlooked.

THE NEED FOR A REPRESENTATIVE BODY

These facts are not usually taken into account, and more generally speaking, points that require discussion are many. Only a representative body can decide on these controversial issues. Realistic drawings should be aimed at, and the relative serenity of a representative institution is most desirable to achieve this. And perhaps, for the third drawing, should we have stressed even more than we did, that