INSTITUUT VOOR FONETISCHE WETENSCHAPPEN

UNIVERSITEIT VAN AMSTERDAM

ON THE ABSTRACT NATURE OF THE PHONEME

by

Hendrik Mol

- AMSTERDAM - 1970 - PUBLIKATIE no. 27 -

ON THE ABSTRACT NATURE OF THE PHONEME.

The following contribution, a paper written in 1961 and not hitherto published, seems at the present time to have gained in topicality in view of the renewed interest in the abstract nature of the phoneme.

1. Introduction.

When a linguist makes a phonemic analysis of a language he uses an impressionistic method that is based on meaning on the one hand and on articulatory and perceptive similarity on the other hand. [1)]

He knows by experience that, for instance, the words

tip, dip, lip, nip,

have different meanings. This fact is evidently caused, he argues, by the circumstance that those words have a different beginning and that the listener is able to perceive those beginnings. In this environment the initial sounds are commutable.

The same series can be found in final position :

bit, bid, bill, bin.

Here the final sounds determine the meaning of the word and have, therefore, a phonemic function.

It is not only a matter of elegance or economy to say that initial t in tip belongs to the same phoneme as final t in bit. Neither is it astonishing that initial l in lip is grouped together with final ll in bill. Here are obvious articulatory and perceptive similarities.

Two sounds display articulatory similarity when their pronounciation requires virtually the same articulatory positions or movements as judged by a speaker. It is even imaginable that only part of these articulatory activities are common for both sounds.

Two sounds show perceptive similarity when they induce essentially the same patterns of nervous activity in the acoustic nerve of the listener.

Here also similarity may exist even for part of the nervous pattern.

---

1) See e.g. E. Fischer-Jørgensen, The Phonetic Basis for Identification of Phonemic Elements, JASA, vol. 24, no 6, 611 - 617, Nov. 1952.

The phonemic system as the linguist creates it is the result of an interesting interaction between functional motives and articulatory and perceptual arguments. For example, in the Dutch language there exist the following three words

doos	[d] [o] [s] ,	meaning	box
door	[d] [o_r] [r] ,	meaning	through
dor	[d] [ɔ] [r] ,	meaning	arid

When a Dutchman carefully pronounces these three words one after another a Dutch listener will find the three vowels decidedly different although he will admit that there is a certain resemblance , a mitigated sort of similarity among them. Now the famous question is: do [o_r] and [o] belong to different phonemes or not ? It is possible to argue as follows.

In Dutch [o_r] always appears before [r] .[2)] As a matter of fact [o] never appears before [r] . Therefore [o_r] and [o] are not commutable and might as well be considered as belonging to one phoneme. This reasoning is reflected in normal writing where both sounds are symbolized by the graphic symbol oo .

Here we let a functional motive override a perceptual difference.

On the other hand, [o_r] and [ɔ] are commutable so that we should regard them as belonging to two different phonemes from a functional point of view.

All these things seem self-evident but the decision to regard [o] and [o_r] as belonging to one phoneme has a dangerous consequence. As in Dutch the sounds [h] and [ŋ] can in no position be interchanged we might declare that they belong to one phoneme as for instance C.C. Berg [3)] does. In my opinion this goes too far as

---

2) In logopedics it is tradition to consider [o_r] as the result of a regressive assimilation caused by the succeeding [r]. It is held that the talker really wants to pronounce [o] but that the succeeding [r] does not allow that. Nevertheless , it is not impossible to say [d] [o] [r] or even [d] [o_r] [s] .

3) C.C. Berg, Poezie der herscheppende wetenschap, paper read on the 374th birthday of Leyden University , p. 9 .

a Dutchman perceives [h] and [ŋ] as totally different sounds. Though [h] and [ŋ] are not commutable a Dutchman when pronouncing for instance

hoop [h] [o] [p] , meaning hope and

lang [l] [α] [ŋ] , meaning long

will no doubt feel he does something quite different with his articulators while realising [h] or [ŋ] .

In phonemic analysis we should never loose contact with the realities of articulation and perception. The process of speech communication is based on the possibility of interpreting perceptible accoustic cues that root in intentional articulatory actions. The perceptible cues are the data the ear is able to extract from the sound waves in the form of patterns of nerve impulses suitable for processing and interpretation by the central nervous system.

Each language operates with its own perceptible cues. Cues that function in one language may be ignored in another language to the extent that phonetically untrained mono-lingual listeners cannot even hear the difference between some speech sounds of a foreign language. For instance, normally a Javanese does not hear the difference between the Dutch words vier , [v] [i] [r] , fier, [f][i][r] , and pier , [p] [i] [r] , because those differences do not function in his own language. He simply is not on the alert for those cues.

But though he cannot distinguish aurally [f] , [v] and [p] he will alway pronounce [p] and never [f] or [v] because he has learned his native language like a parrot by unconsciously imitating the [p] sound used by adult speakers.

A speaker of a language where there is only an [i] and no [I] cannot hear the difference between [i] and [I] . Nevertheless he will always pronounce [i] and not [I] .

We see here an overlap in perception not accompanied by a corresponding overlap in articulation. One can argue that this overlap occurs in an unnatural situation and that it disappears gradually as the person in question learns the foreign language.

But, as we shall point out in the next paragraph, the articulation and perception of people speaking the same native language is characterized as well by severe overlap, this time not cured by a process of learning.

Though the definition of a phoneme is a delicate subject on which there is no common opinion the term "phoneme" is loosely used by speech therapists, audiologists, engineers and phoneticians. They do not treat a phoneme as an abstraction resulting from linguistic contemplations involving arguments like " as it were " and " one might as well " .

On the contrary , many investigators see the phoneme as something tangible that can be transported from the brain of a speaker to the brain of a listener. It is conveniently assumed that on its way the phoneme is clothed in several disguises.

The application of information theory to speech is inspired on the technique of the teleprinter. In a teleprinter system two electrical typewriters are connected by a line. What is typed on the sending machine is printed on the receiving machine. In this case the coded signals transmitted by the line are indeed exact descriptions of the letters the receiving machine is expected to print. Unfortunately in speech communication the experimental findings deny the existence of such an "exact" link. Therefore the idea of the so-called voice typewriter (also named sonotype or phonetic typewriter) rests on the sandiest of foundations. A voice type-writer is a not yet existing machine which is expected to produce a typewritten text of what is spoken into it.

It is the object of this paper to discuss the abstract nature of the phonemic system.

2. Intention and result.

Current phonemic analysis is based on the paradigmatic comparison of isolated words. It is taken for granted that the words in question are pronounced with the intention of giving the listener 100 % intelligibility. It is tacitly assumed that this is possible. We shall now undermine that belief.

Peterson and Barney [4] plotted second formant frequency against first formant frequency for 10 vowel sounds uttered by 76 men, women and children .[5]

---

4) Peterson,G.E. and H.L.Barney, JASA, 24 (1952) 175, quoted in Manual of Phonetics, Amsterdam 1957,Edited by L.Kaiser,p. 198-201.

5) This method of representing the formants of a vowel we find as early as in 1930 in a paper by A.W. de Groot: Phonologie und Phonetik als Funktionswissenschaften, Travaux du cercle linguistique de Prague 4, 1931, p. 141 - 147.

All of the vowels were spoken [6] between [h] and [d], i.e. heed, hid, head, had, hod, hawed, hood, who'd , hud and heard. The results have been depicted in fig. 1.

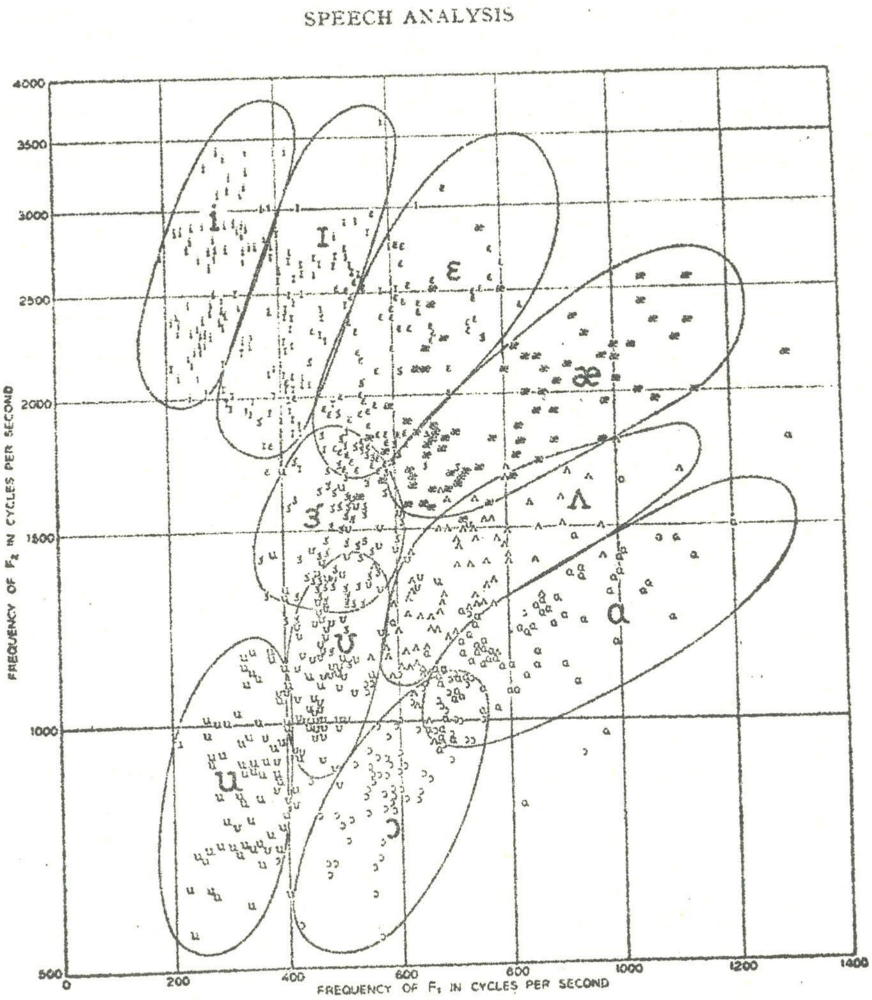

Figure 1

Frequency of second formant versus frequency of first formant for 10 vowels by 76 speakers (Peterson and Barney)

The plot shows 10 targets around which the results of the talkers spread.

---

6) In Manual of Phonetics p. 198 is stated, less fortunately,: all of the vowels were spoken between the <u>letters</u> h and d .

A close scrutiny of the figure reveals that in spite of the contours that have been drawn in an attempt to find a demarcation between the vowels there is considerable overlap between the targets.

It is interesting to see that, as the sounds [ü] and [ö] do not exist in English, there is a corresponding "vacuum" in the plot.

In spite of the above overlap the 10 targets betray at least the outlines of a system. The value of that system for the listener is greatly reduced, however, by the background of fig. 2.

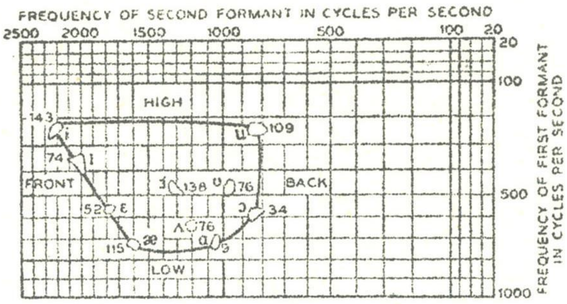

Figure 2

Vowel loop with numbers of sounds unanimously classified by listeners; each sound was heard 152 times (Peterson and Barney) .

As all words were recorded on a magnetic tape they could be presented to a jury , in this case consisting of 70 persons.

As every talker pronounced every vowel twice (at different sittings) there was a collection of 10 x 152 = 1520 vowel sounds. So a listener heard 152 vowels intended as [i] , 152 vowels intended as [I] etc. The numbers in the vowel loop of fig. 2 indicate how many of the 152 sounds intended by the talkers as a certain vowel were actually correctly identified by <u>all</u> members of the jury . The scores of [ɑ] and [ɔ] are very low, whereas those of, for instance [ɛ] and [I] are not too good either. In total only roughly one half of all vowels was unanimously correctly identified as the vowels the talkers intended to pronounce. In other words: The listener, when totally bereft of any cue that might be derived from the context or the situation, and being unable to adapt himself to the articulatory habits of the speakers, cannot with certainty identify the monosyllabes intended by the talkers. It is obvious that the listeners are not directly aware of the existence of the 10 articulatory targets.

Similar pronouncing and identification experiments were performed by Fairbanks and Grubb [7]. They reduced the overlap between the targets by selecting trained talkers who were allowed to adjust their pronunciation by listening to their own results, and by defining a vowel as succesfully identified when correctly identified by 75 % or more of the listeners.

For running speech there are not even 10 targets. In 1957 Blom and Mol [8] plotted the formants of the 12 vowels appearing in the following Dutch sentence :

"Wim kon de leunstoel van schoonmama niet tegen de muur zetten "[9] freely pronounced by 100 persons. The underlined vowels are [I],[ɔ],[ə],[ö],[u],[α],[o],[a],[i],[e],[ü] and [ɛ].

---

[7] G. Fairbanks and P. Grubb: A psychophysical investigation of vowel formants, Journal of Speech and Hearing Research, Sept. 1961, vol. 4, no 3, p. 203.

[8] H. Mol, Belief and superstition in phonetics (in Dutch), Inaugural address Amsterdam 1960.

[9] English translation: Bill could not put the easy-chair of his mother-in-law against the wall.

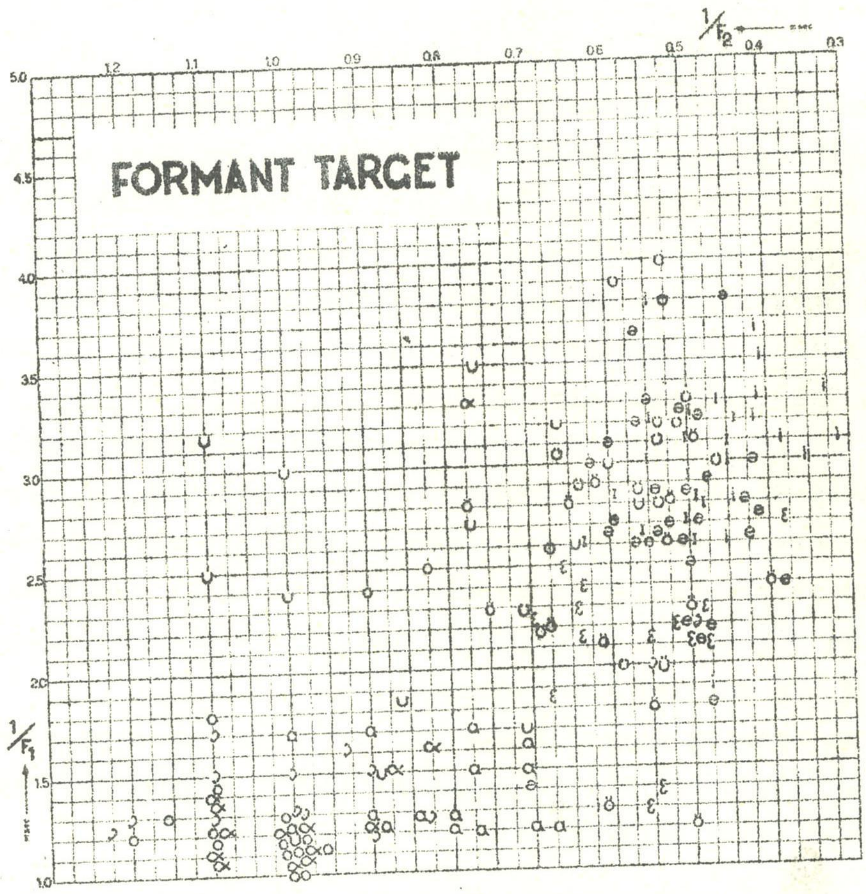

Figure 3

Period of second formant versus period of first formant for 12 Dutch vowels appearing in a simple, freely pronounced sentence. The data of this pilot investigation refer to 15 talkers taken at random from a group of 100 talkers. The numbers on the axes represent periods, in harmony with the fact, that the formants were indeed measured as periods and not as frequencies.

By way of a pilot investigation 15 talkers were selected in a random fashion from the total group. The first two formants of the 12 vowels of every talker were determined and plotted in fig. 3. There appeared to be only 2 targets around which the realisations spread randomly.

One target embraced [i],[I],[ü],[ə],[ö],[e] and [ɛ]; the other target [a],[α],[o],[ɔ] and [u].

It is evident that in running speech the listener is, as regards vowel identification, on very weak phonetic ground and that he needs extra-phonic cues in order to identify the sentences.

In view of the generally admitted success of voice communication one is led to believe that there will be less overlap between some consonants as their pronunciation suggests less freedom than that of the vowels. We must, however, thoroughly check this hunch by experiments before we content ourselves with such an explanation.

Nevertheless, there is the supporting fact that in Hebrew only the consonants are depicted in writing.

3. The description of phonemes.

The logical way to inventorize the phonemes of a language is to enumerate the keywords in which they function. After all, that is the way in which they are "discovered" and defined.

As there are only 26 letters and in general more than 26 phonemes one must take refuge in forming combinations of letters in order to write down the keywords. This need not be a handicap as long as this procedure is applied consistently. In English the same combination of letters may be used to indicate different phonemes. For instance,

through [θ] [r] [u]
dough [d] [ou]
rough [r] [ʌ] [f]

The same inconsistency holds for single letters.[10] For instance,

be [b] [i]
bet [b] [ɛ] [t]

As it is now, writing is not a system for showing foreigners how to pronounce the words of the language in question. Writing is a method for bringing into the mind of a native reader the words he already knows and can pronounce.

---

[10] Also in Dutch there exists this single letter ambiguity, for instance kat, [k][α][t] , versus kater, [k][a][t][ə][r] .

The so-called phonetic alphabet , in its international form or in its local varieties , is not more phonetic than the normal alphabet , it is merely more consistent, all be it at the cost of more letter types or the addition of diacritic signs. The international alphabet is unpractical because of the trouble of finding a printer who has those apocryphal letter types at his disposal.

As a warning to the reader the phonetic letters are often put between square brackets.

An inventory of phonemes may look like the following table

[I]	as in	pit
[ɛ]	as in	pet
[æ]	as in	pat
[ʌ]	as in	putt
[ɔ]	as in	pot
[u]	as in	put
etc.		
etc.		

Table I

Example of an inventory of phonemes

Now the investigator wants more than a simple enumeration of the keywords. He wants some sort of description that is more attractive than merely selecting "linearly" one keyword out of a big number N of keywords. Without perhaps realising it, he wants, as we shall show now, a <u>coding.</u>

items	abstract code-elements
No 1	A
No 2	B
No 3	A and B
No 4	C
No 5	A and C
No 6	B and C
No 7	A, B and C
No 8	D
No 9	A and D
No 10	B and D
etc.	etc.
.	
.	
.	
No N	

Table II

Coding N items als combinations of a (smaller number of n code-elements A, B, C, D, E, F etc. These code-elements are abstractions , products of contemplation. The biggest number one can code in this way is $N = 2^n$

Table II shows how a collection of N items can be considered as the result of combining a smaller number of n so-salled code-elements , the nature of which is, for the time being, arbitrary. This method is, clearly so , an abstraction. The reasoning is that, in order to select or point out a certain item, one might as well state the corresponding code-elements. For example, when the combination AD is given, one knows it is item No 9.

Mathematically speaking , this method comes down to the translation of a decimal number into the binary system. This can be seen by reading $2^0 = 1$ for A , $2^1 = 2$ for B , $2^2 = 4$ for C , $2^3 = 8$ for D etc. In that way for instance $9 = A + D = 1 + 2^3 = 1 + 8$.

In principle this system of combinations can be applied to a phonemic inventory by interpreting the phonemes as items. The problem is how to define the code-elements A, B, C, etc. so that the system becomes more than an abstract structure one can learn by heart as an aid for memorizing the inventory.

Jakobson et al [11] call the code-elements the ' distinctive features' . Nowadays one is interested in the nature of the process of identifying speech waves. If one supposes that the said process has a phonemic basis (which , after all is merely a dogma) one should choose distinctive features that root in the perceptual limitations of the organ of hearing and the nervous system. The distinctive features of the binary system are for the main part abstractions. For instance, the opposition [12] consonantal/non-consonantal is of a highly arbitrary nature.

Especially for the vowels a binary system of description seems inadequate and superfluous. The ten targets in fig. 1, displaying the ghost of an articulatory framework of reference are characterized by the bull's eye of each target. Each bull's eye is given by its two coordinates F_1 and F_2 . There is every reason to suppose the ear can more or less directly measure these two frequencies and use them as indications. There is no need for oppositions like tense/lax or compact/diffuse etc. in this process.

Summarizing, we think to have thrown light on the twofold abstract nature of the phonemic system. To begin with the phonemic inventory is the result of an abstraction. Secondly a further description of the phonemes is an abstraction too.

---

[11] R. Jakobson, C.G.M.Fant and M.Halle, Preliminaries to speech analysis. The distinctive features and their correlates. Technical report no 13. May 1952, Acoustic Laboratory Mass.Inst. of Technology.

[12] R. Jakobson and M. Halle, Fundamentals of language, Janua Linguarum Nr. 1, 's Gravenhage 1956 .

4. Conclusions.

When the unsophisticated listener identifies speech utterances he does not, step by step, identify every speech sound [13]. He has at his disposal a sequence of patterns of nervous activity. These patterns carry perceptible cues that are "keyed" by the articulatory activities of the speaker. The perceptible cues, in so far the listener has learned to use them, allude to the notion or idea behind what we call a word. We know by now, that the cues are not unambiguous and that the listener must often use his knowledge of the context, the situation and the articulatory habits of the speaker in order to discover the intention of the speaker. For the vowels the most probable perceptible cues are, among other possibilities , the frequencies (or their reciprocals the periods) of the first two formants F_1 and F_2 , the fundamental frequency F_0 (or the period $T_0 = 1/F_0$) of the vocal cords and the duration. At the moment we know that in situations where the talkers do their utmost to be unambiguous in their pronunciation of the vowels the overlap between the various phonemic catagories as indicated by the perceptible cues is minimal though still present. As A.W. de Groot pointed [14] out, apart from the pronunciation of isolated words an interesting precision-provoking method of uttering words consists in inviting a person to articulate contrasting words in this following carrier-sentence:

I do not say 'pit' , but 'put' .

It will certainly be enlightening to study the way in which precision and overlap are influenced by the character of the situation in which the talker pronounces his speech sounds. It will be equally interesting to determine how a listener reacts to the same perceptible cues in various situations. One expects intermediate situations between the extremes of isolated words and running speech.

Still, we must not think that this type of experiments, all be it very useful, will show us the mechanism we are locking for.

---

13) In view of the speed of talking the brain would be too slow for that anyhow. In reading every letter is not individually identified either.

14) Private communication.

One gets the impression that the listener, when interpreting ambiguous , often misleading or even totally missing perceptible cues uses a 'frame of reference' , a kind of system of pigeon-holes into which he can throw the incoming data. In our opinion in normal speech this frame of reference roots in the ideas the perceptible cues are able to call up, rather than in the phonemic system.

The situation seems quite different for experiments in which subjects are asked to pronounce or to identify isolated speech sounds. We then tie them down to an imposed system of pigeon-holes in the intentional or the perceptual domain. We know that the discrepancy between those two systems of pigeon-holes is considerable.

Just like information-theory and Fourier-analysis , phonemic analysis is a method and as such an abstraction. It is not a mechanism at work in the process of speaking and hearing.

One is tempted to postulate that there is much truth in the subtitle of a well-known book of Kenneth L. Pike : Phonemics, A technique for reducing languages to writing [15].

H. MOL
Institute of Phonetic Sciences
of the University of Amsterdam.

---

15) Ann Arbor, University of Michigan Press 1947.