THE ALPHA-PHONEME-PHONOID

Towards experiments on the "Ultimate 'Unit' of Music and Speech"

C. R. Sankaran and K. S. Sampat

A proposed attempt within our theoretical framework to investigate the alpha-phonoid — the "Ultimate 'Unit' of Music and Speech" — is presented in this paper.

The alpha-phoneme is a Dedekind—'cut' which points to the *inter-phenomenon* which is neither space nor time separately but being both, is especially the "punctiform origin of time".[1]

The vistas claimed to have been opened up by the Alpha-Phonoid theory have been already discussed by B. Chaitanya Deva with particular reference to the *tambura* and music.[2] It has also been pointed out that the alpha-phoneme is a symbol of experience.[3]

In music and speech, we meet with both the non-observed and non-observeable.[4] From a so-to-say, dimensionless experience (the alpha-phoneme)[5] we are descending, as it were, towards an attempt to define an ultimate signal a 'Unit' of environment or 'Unit' of communication (the alpha-phonoid). Both in Music and Speech, we meet with structures or organisations at various 'levels'. Psychoacoustically, one could order structures from physicochemical undulations of acoustics to the psychological (i.e. mathematical) construction of phonemes (or notes) and to the ultimate alpha-phoneme whence we may derive the alpha-phonoid.

Now we have also discussed already alpha-phonoid of first order (which is of order of time or time itself) and the alpha-phonoid of secondary

order (as an expedient unit of temporal measurement). In the wake of this discussion, we pointed out both the 'motivational' and the 'activational' levels in the process of speech,[6] which may be true of music also. The motivational level represents the 'inner speech (or music, if you will) and its predicative function.'[7]

The first assumption behind the ultimate 'Unit' of music and speech is that a change in the time-constant T is necessary for any discrimination of duration, frequency or intensity. Now a detailed mathematical demonstration of this assumption is a mater of urgency.

At the very out set, it must also be pointed out that this assumption rests on the work of Mol and Uhlenbeck.[8] The simultaneous presence of change in the time-constant of the inner process and a change in duration, frequency or intensity in the outer process, that is respectively both in the neurolophysiological and acoustical media, is the very basis for the inductive postulation of the existence of the continuum due to Veronese in perceptual processes. We have already said enough of the deductive postulation of such a continuum in the wake of the alpha-phoneme theory.[9]

The first condition in all our investigations, concerning the "Ultimate 'Unit' of Music and Speech" is that duration, frequency and intensity should be reducible as a result of their bi-unique correspondences, through transformation equations in terms of T_t, T_a and T_f.

The time-constant T should be taken as being equivalent to an observer and if description of speech or music are made in terms of times t ($= T_t + T_a \eta$) + $T_f \varsigma$ . . . . . .) and T by an equation, such an equation should not differ in form from a description of the same phenomenon in terms of t' and T'[10].

It should be possible to establish transformational equations between t, T and t', T'. There will be a quantity X which will be an invariant for all observers T, T' etc.,[11] This invariant X may be taken as "the (ultra) elementary constituent of perception, which extends in the time-series over what may be called a *duration*, which since it is *sui generis* cannot be defined in terms of anything else."[12]

Adopting the words of A. N. Whitehead,[13] the 'pure interval' (which is the alpha-phoneme) is the "primitive experience 'vector feeling', that is to say, feeling from a beyond which is determinate and pointing to a beyond (the alpha-phonoid) which is to be determined."

There is no obtainable evidence from a psychological study of the passage of time in individual experience to prove that the notion of an 'instant' (a duration) corresponds to anything in the world of sense-data.[14]

II

Before giving a detailed account of acoustical structure of music and speech it is necessary to explain some basic ideas in the theory of analysis of speech oscillations.

In order to investigate music and speech processes thoroughly, we endeavour to experiment into the structure and to resolve the effect into individual elements or quanta. In seeking the appropriate quantal entity which is to appear in the ultimate analysis without eliminating completely the raster limits we find that music and speech dynamics presents substantially a more difficult task than does e.g. static structure of crystals. Music and Speech are continuous entities and if they are cut into individual elements of sound, the isolated elements give a distorted and uncharacteristic impression of flow. The choice of quantum is therefore empirical and it must depend on varying characteristics of individual audible sounds. There are no definite borders of the sound elements to be chosen from the flow of music and speech since each is influenced by neighbouring elements; one is thus obliged to choose a fictitious unit determined by musicological or linguistic properties, but this has then little correspondence with physical properties of the sound. An elementary quantity of this kind will be called a note in music or a phoneme in speech.[15]

On the other hand it has been found possible in Phonetics to analyse imperceptibly different sounds; the process can be carried to ultimate physical limits. This subdivision is necessary when the full-time and frequency range must be covered in the description of the spectrum of natural sounds; without such a complete description, the aural phenomena connected with the initiation and cessation of sounds cannot be understood.

We now briefly review the analysis of stationary (i.e. non-varying) sound phenomena; strictly, of course such phenomena do not happen in nature and they are therefore irrelevant to the discussions on music and speech, but this will lead us to the methods of making acoustical measurements on the vocal tract; finally we come to foundations of the formant theory in its present-day form.

Fourier's theorem allows us to understand a complex tone as a synthesis of sounds from harmonic partials. The frequencies of the harmonics are integral multiples of the fundamental. The intensities of the fundamental along with their partials is termed 'Spectrum of a complex tone'. The spectrum gives no indication of the relative phases of the harmonics. However, different phase-relationships between the same harmonic components gives an altogether different wave-shape of a complex tone on an oscilloscope. Hence oscilloscope presentation is not very suitable for the analytic assessment of sounds, apart from certain special purposes.

It is certainly true that phase relations have no significance in the study of music and speech-structures.

Of course, the wide-spread view that the ear carries out a Fourier analysis in the cochlea is now disputed; it is considered to be valid for at most a limited range of frequencies, as shown by Mol and others (1963). Mol further points out that Fourier analysis also might be true only in limited range of frequencies.

Naturally we are led to enquire what the other methods of analysis of a complex tone are possible.

In this connection the auto-correlation function $\varphi(\tau)$ and the power density spectrum $\varphi(\omega)$ are of particular interest; they are Fourier transforms of each other, according to the Wiener-Khintvhine theorem:[16]

The music or speech signal x (t) is delayed by the interval and the product x (t) x $(t + \tau)$ is formed and integrated.

K. Stevens was one of the pioneers to work out a simple instrumentation for this purpose.[17]

In auto-correlation analysis of speech, we can give a "short time" auto-correlation function by integrating over a finite time interval only. If this integration process on the product say f (t). f $(t + \tau)$ is allowed to operate continuously we obtain a running $\varphi(\tau)$ that varies with time, just as the speech spectrum does. The auto-correlation function obtained in this way will be very nearly equal the Fourier transform of the spectrum.

Measurement of this short time auto-correlation has been achieved in several ways. The accuracy of frequency determination by the auto-correlation function is a better method for visual display than other methods demonstrated so far (particularly for fricatives in speech). As early as 1955, Taskar had also calculated power spectra and auto-correlation functions of vowels in the wake of theoretical work of the senior author of the present paper.

Licklider[18] has made yet another approach to throw some more light on structure in Speech. He distorted the wave amplitude using the technique of infinite peak clippings; even though the result is a square wave

form, the articulation is maintained at 70% (Pollack and Licklider, 1948). The comprehensibility can be further increased by differentiating the curve.

In this experiment only the rhythmical succession of zeroes crossing the abscissa is retained after abridgment, of all the information contained in the original speech waves.

Since infinite clipping dichotomizes the amplitude dimensions and no further reduction can be achieved in amplitude, it is necessary to operate upon temporal pattern if the speech-wave is to be further simplified. The simplifying operation had been done by quantization of the time scale. In quantized time, a rectangular wave can switch only at predetermined instants. Auto-correlation functions of these amplitude dichotomized, time quantized speech-waves were calculated; also articulation test of these speech waves was carried out. Consequently, a theory of Hearing process might be constructed on quite a different basis from the purely physical method of spectral analyses.

III

Conclusions

Such are the few technical limitations in the study of music and speech that we have discussed. What is necessary now is to examine these and several other techniques for analyses of both music and speech in the light of the alpha-phoneme-phonoid theory. The alpha-phonoid has already been defined as the minimum common duration of a 'unit information cell' in the physical stimulus as well as in the neurological and psychological responses. In other words, the minimum common duration of all these three will serve as the key 'interval' for the basic representation of both the structure in speech and *sruti* which is the microtonal 'interval' in music.

The conceptual and technical difficulties facing an attempt to determine the "ultimate Unit" of music or speech as the minimum common duration of the physical stimulus, neuro-physiological and psychological components of the total event seem formidable in the present state of the brain sciences, especially as they bear on such high level functions as human music and speech. The lack of functional-structural clarity in this field is reflected in the sporadic and fragmentary attempts at studying the musical and speech processes as in terms of brain function.[19]

The following are some attempts in recent times:

So called "command potentials" (averaged evoked potential of surface electro-encephalograms i.e. EEG preceding various voluntary activities) was shown to differ in shape for phoneme O, T, and P (but not for numerals

2 and 10 or for the words 'yes' or 'no') when recorded over the left temporal speech area.

The peripheral nervous system involvement in the speech process (the orienting reflex, recorded by a new plethysmographic technique distinguishing orienting from defensive reaction by a vaso-motor criterion) led two Russian workers to postulate a semantic model of word nuclei surrounded by a semantic field of words linked to the nucleus by experimental meaning relations rather than logical content.[20]

Analysing human behaviour including verbal behaviour from sound motion pictures allowed two American workers to study 'motor phonetics' with the help of a 'basic form of "unit-in-change" or "process unit" by which they were able to describe the 'on-going flow', of "moments-of-sustaining-together of the body parts" in continual sequences of change.

These are the widely different techniques and instrumental approaches that are being employed, as Bjorn Merker rightly observes in the objectification of the semantic process and these varied techniques and experimental investigations might be relevant to the investigations of the alpha-phonoid which is qualitatively defined as the minimum common duration of the physical stimulus, as well as the neurophysiological and psychological response-components of a total event.

Bjorn Merker also remarks that the above-mentioned objectifications of the semantic process are the surface or overt event of the music and speech processes while the definition of the alpha-phonoid goes further, as indeed one must, in order to distinguish human speech and music from animal calls so as to include the psychological component at high levels of abstraction and integration. In terms of the brain, this points to the convergent structures of the limbic system which is subcortical and accessible only by implanted depth electrodes. H. Lesse and R. G. Heath have recorded electrical activity in limbic structures in humans directly correlated with emotional thought and recall, but intracranial recording is not a generally applicable technique in humans.

These technical problems are further complicated by the theoretical consideration that the information content of a stimulus event cannot be defined independently of the nature of the information processing and storage capacity, which in man is shaped by accumulating ontogenetic experience to an unusual degree, resulting in great individual and temporal variation.

The very magnitude of this problem makes it the more challenging a task and therefore, the theoretical formulation of the alpha-phonoid seems to provide, as Bjorn Merker thinks, a provocative perspective within which

one may strive to approach a solution. The results of any deep experimental investigation within this perspective, might find too an application within phylobiology when that baby science, to quote once again Bjorn Merker, "reaches the maturity of speech and concrete manipulations."

Notes and References

(1) Chaitanya Deva, B., *Psycho-acoustics of Music and Speech,* The Music Academy, Madras 14, p. 296

(2) *ibid* pp. 54

(3) *ibid* p. 120

(4) *ibid* p. 198

(5) *ibid* p. 272

(6) Sankaran, C. R. and Patil, N. K.: Alpha-Phonoid and Alpha-Phoneme, *Phonetica,* Vol. 17. 15-23 (1967), Chaitanya Deva, B., *Tonal Structure of Tambura,* p. 63, Sankaran, C. R. *Process of Speech,* pp. 24-25.

(7) Luria, A. R. and Tsvetkova, L. S. : The Mechanisms of "Dynamic Aphasia", *Foundations of Language* 4; 1968, pp. 302-4.

(8) Mol, H and Uhlenbeck, E. M.: The analysis of the phoneme indistinctive features and the process of hearing, *Lingua* 4: 167-83(1954-55).

(9) Sankaran, C. R., *Process of Speech,* Deccan College Monograph Series 27 (1963) p. 29.

(10) Cf. "The listener 'knows' that the way in which the speaker manipulates *his* code elements is the same way in which the other speakers manipulate their *code* elements" — Mol, H. The Relation between Phonetics and Phonemics, *Linguistics* 1 (1963) p. 61.

(11) Sankaran C. R., A Note on the quantitative and qualitative definitions of the alpha-phonoid, *Phonetica* 20 : p. 75 (169).

(12) Whittaker, Edmund: *From Euclid to Edington,* Cambridge. At the University Press, 1949; p. 41. Sankaran, C. R., A Philosophical analysis of the Alpha-Phoneme theory in relation to the problem of Speech-structure, *Bull. of Deccan College Research Institute* 14 : p. 94, 1955. Cf. also a discussion on durations on page 59 of C. R. Sankaran's *Process of Speech,* Deccan College Monograph Series 27 : 1963.

(13) White-head, A.N., *Process and Reality,* New York, 1941, p. 247. Cf. also Paul Ushenko, *Power and Events,* Princeton, 1946, p. 153.

(14) We have already said (C. R. Sankaran and P. C. Ganeshasundaram. Time and Speech-structure, *Bulletin of Deccan College Res. Instt.,* 12: 1954, p. 407 and footnotes 28 and 28a) that following Fechner himself, A. Maack, contends that Webner-Fechner Law does not hold good for duration phenomena, since the law is applicable only to intensive sense impressions, whereas duration is an 'extensive abstraction' in a slightly Whiteheadean sense". Cf. also Whitehead, A.N., *Process and Reality,* Social Science Publishers, New York, 1929, reprinted 1941, pp. 148 and 454-508; Cf. also White-head. A. N., *The Concept of Nature,* Turner Lectures delivered in Trinity College, Cambridge, 1919. At the Univ. Press, 1926, pp. 74-78 (for the method of extensive abstraction). Cf. also Andhre M. Weitzenhoffer, Mathematical structures and Psychological measurements, *Psychometrica,* Vol. 16, 1951, pp: 387, 406 for further remarks on intensive and extensive dimensions, wherein it is demonstrated that even in regard to intensive sense impression, Webner-Fechner's Law breaks.

(15) Winckel, F. Acoustical Foundation of Phonetics, *Manual of Phonetics* (ed) Malmberg, pp. 39 ff.

(16) Brillouin, Leon, *Science and Information Theory,* p. 101.

(17) Stevens, K. M. Autocorrelation analysis of Speech, *JASA* Vol. 22, 6, pp. 769 — 71.

(18) Licklider, J. C. F., *JASA,* Vol. 22 No. 6, pp. 820-23.

(19) Sankaran, C. R. (1969). A Note on the Quantitative and Qualitative definitions of the alpha-phonoid, *Phonetica.* 20 : 74 - 80.

(20) It is obvious that such a Semantic field operates to 'note'-nuclei also — B.C.D.