`

Concepts of *Prayatna and Svara* of *Shiksha Shastra* in light of Experimental Phonetics

Dr. T. V. Ananthapadmanabha

Voice and Speech Systems, 53 Temple Road, Malleswaram, Bangalore 560003

tva.blr@gmail.com

Preparation Date: 16-11-2022
Release Date: 07-05-2023

ABSTRACT

Vedic literature is the bed rock of ancient Indian philosophies, religions, culture, heritage and customs. Vedas were transmitted by oral tradition over a long period of time. In order to preserve the precise pronunciation of Vedas as well as for the sake of proper interpretation of the Vedas, there exist six support branches of knowledge called 'Vedangas'. Shiksha (Phonetics) is the first Vedanga. A knowledge of Shiksha is a pre-requisite to understand the rules of Vyakarana (Grammar), the second Vedanga. Sanskrit language was used not only in Vedas but also in day-to-day house-hold communication. Shiksha has six major components of which only the following Prayantna and Svara aspects are considered here.

An overview of speech production is presented from the viewpoint of both Shiksha Shastra and contemporary speech science. The contemporary knowledge helps us to truly appreciate the insights of the ancient Indian Phoneticians and to look-back in order to clarify or re-interpret or expand some of the aspects of ancient Shiksha Shastra. Specifically, the distinction between prayatna (effort) and mechanism of production of voiced sounds, fricatives and stops is elaborated. In ancient texts, there is a divergence of opinion on the description of svarita svara (High-Low tone) as well as the specification of relative durations of High/Low segments. By the application of tools of experimental phonetics on a chosen example, it is demonstrated how such divergence of opinions may be resolved.

`

I. Prelude

At the outset (*'atha'*) the author would like to make some opening remarks on his background and the scope of presentation. Learning Sanskrit Phonetics and Grammar in a traditional style demands twelve years of full-time rigorous study. The author has not been trained in *Shiksha* or *Vyakarana* in a traditional school. However, the author has a general knowledge of the subject matter of Sanskrit Phonetics and Grammar. In the present day educational system, Phonetics is taught as a subject under Linguistics that comes under the discipline of Humanities. The author is not a trained linguist or a phonetician. The author is a communication engineer with specialization in the area of speech signal processing. He has over forty years of research experience with publications in peer reviewed Journals. He has worked with pioneers in the area of acoustic-phonetics in world renowned labs. Further, as an entrepreneur, the author has developed software tools in the area of speech therapy and speech science (*www.vagmionline.com*). It is with such a technical backup, the author is attempting this presentation.

Pronunciation rules specific to a Veda are dealt with in the respective *Pratishakhyas*. Also, there are various schools of *Shiksha* dealing with spoken Sanskrit. We refer to the entire ancient literature by the term *'Shiksha Shastra'* and abbreviate it by SS. The material on SS has been collected from a few secondary sources [1-4] and not from the original texts. Some key references on the acoustic theory of speech production and experimental phonetics are given in the references [5-7].

The subject matter of SS as well as the modern 'Acoustic Theory of Speech Production' and 'Experimental Phonetics' is very vast. Hence, only an overview of speech production and some representative aspects of SS are selected for this presentation. The writing assumes a general readership and it doesn't demand a background of Sanskrit Phonetics or Sanskrit Grammar or modern acoustic theory or experimental phonetics. A basic background on the concept of sound waves is presented in Appendix-A for a general reader.

The main aims of this work are

(a) Appreciate the insights of the ancient Indian Phoneticians in light of the recent findings

(b) With the background of the modern theory of speech production, look-back into SS in order to clarify or re-interpret or enhance some of the concepts in SS.

(c) It is demonstrated that the subjective qualitative remarks in SS may be substantiated by quantitative objective measurement tools of experimental phonetics and any difference of opinion in the interpretation found in SS may be resolved.

`

Outline of the topics: The scope of SS is presented in the Introductory Section-II. An overview of speech production both from the viewpoint of SS as well as contemporary knowledge is presented in Sec. III. In Sec. IV, based on the contemporary knowledge in speech production, we look back critically into the concepts of external and internal efforts of SS. Sec. IV also illustrate the measurement of duration and tone height or pitch of a *svara*. In the Concluding Section (Sec. V), some suggestions for future work are presented.

II. Scope of *Shiksha Shastra*

II.A. Human communication and Sanskrit

Communication of thoughts via language, that too via spoken language, marks an epoch in the evolution of human culture and civilization. It is believed that experience of self-awareness arose because of speech communication. In animal communication, there are only a finite number of pre-programmed sequences of sounds that are automatically elicited based on the situation. In contrast, in a human language, an innumerable number of ever new sentences may be formed with different arrangements of words drawn from a finite vocabulary. A large number of meaningful words are formed as a sequence of a few speech sounds that are selected out of an inventory of about fifty speech sounds. Perhaps, the reductionistic view that innumerable words can be formed out of a small inventory of speech sounds might have motivated the philosophers of *Vaisheshika* and *Sankhya* to look for the building blocks of not only the physical but also the mental worlds.

Two main abilities of a human being have enabled the development of speech communication, viz., abstract thinking and deductive logic. Sanskrit happens to be an ancient spoken language. The holistic Vedic literature is in Sanskrit. Sanskrit was a house-hold language at one point in history. Usage of spoken Sanskrit was in vogue much prior to the advent of writing. The subject matter of *Shiksha* and *Vyakarana* has been developed based on spoken Sanskrit as the input material and not the other way around. That is, *Shiksha* and *Vyakarana* didn't impose rules on the native speaker of Sanskrit. SS is based mainly on abstract thinking ability while *Vyakarana* is based mainly on deductive logic.

`

II.B. Phonemics Vs Phonetics

If we ask an illiterate person what distinguishes the words 'cat' and 'bat', he/she wouldn't give the expected reply that these two words differ in terms of the initial sound. An illiterate person perceives and uses each word as one whole indivisible unit. Even children, before they are taught writing, will not be decomposing the words into basic sound units. It is only on introspective analysis and by a process of abstract thinking that the concept of basic sounds of a language can be conceptualized. SS has discovered the basic sounds of spoken Sanskrit even before the advent of writing. This requires a keen sense of auditory perception and abstract thinking ability.

II.B.1. Concept of phoneme from a listener's point of view

Process of abstraction works at various levels and the abstracted information is dictated by the purpose. A listener can distinguish speech sounds from all other sounds. A listener can identify a speaker irrespective of what is spoken. A listener can identify *svara* irrespective of the identity of vowel. The main purpose of a sequence of speech sounds is to convey a linguistic message. Spoken sounds invariably carry not only the linguistic message but also the speaker's voice quality, *svara* etc. That part of the abstracted sound quality which carries the linguistic message irrespective of the voice quality, duration, *svara* etc corresponds to the phoneme. Further, a listener can identify a phoneme from spoken utterances irrespective of the position where the phoneme occurs within a word, i.e., irrespective of whether the phoneme occurs in the initial or mid or final position and irrespective of the preceding and following sounds.

II.B.2. Concept of a phoneme from a speaker's point of view

A word is represented within a speaker's mind by a sequence of speech sounds. This mental representation of speech sounds shouldn't be confused with a sequence of graphemes or alphabets since Sanskrit was a spoken language prior to the evolution of writing. A speech sound within a word is an intended sound that has to be pronounced in order to be conveyed to a listener. An audible speech sound is the physical manifestation of the internal intended sound. During the pronunciation, there

`

are some unavoidable attributes that get associated with a sound such as duration, pitch, speaker's voice quality etc. The abstracted mental representation of an intended sound that serves to convey linguistic message is the phoneme of a language whereas its other extraneous physical attributes (duration, pitch etc) constitute the phonetic properties of the sound. In other words, a phoneme is a universal (or genus) and spoken sound is its particular physical manifestation. A clear distinction between phonemics and phonetics is made in SS and *Vyakarana*. Thus for example, vowel 'a' (the very first letter of Sanskrit language) may manifest phonetically in eighteen (3X3X2) different ways such as *hrasva* (short) or *dirga* (long) or *pluta* (prolated) and/or at different tone levels (low, high, mixed) and/or as non-nasalized or nasalized. Further, *hrasva* Vs *dirga* (or *pluta)* pronunciations of vowel 'a' are known to differ not only in terms of duration but also in their sound quality as *samvrutam* Vs *vivrutam* (tense Vs lax), which is considered as a phonetic feature. However, as far as the application of grammatical rules is concerned, the underlying phoneme 'a' is to be considered as one and the same irrespective of all these diverse phonetic manifestations. The presence of dialectical (regional) differences in pronunciation is recognized in SS (*Ashtadhyayi*) and these are considered as phonetic differences. Auditory speech signal also carries with it an individual's voice quality, identity of the gender and age. Such divergences in voice quality are considered as extraneous differences. A phoneme, being a universal ('genus'), is eternal ('*a-kshara*') whereas its physical or phonetic manifestation as a sound is ephemeral.

A philosophical question: “Is phoneme a sound?” If the answer is 'Yes', then it can be imagined in the mind with an associated sound quality and it becomes a particular manifestation. If the answer is 'No' then how come the concept of a phoneme is formed by listening to speech?

II.B.3. Phonology

In fluent speech, speech sounds are uttered continually in a sequence. During such a narration by a native speaker, the quality of speech sounds at the junction of two words (or two sub-units such as *pratipadika*, root of a noun, plus suffix or *dhatu*, root

`

of a verb, + suffix) gets altered. Such changes come about involuntarily or naturally and might have been instinctively learnt based on the principle of economy of effort.

Grammatical rules formalizing changes in speech sounds across word (or sub-word) boundaries are well known as *sandhi*. Grammatical rules, such as *sandhi* rules, are determined by the phonemes and are not influenced by the phonetic differences. This can be compared to chemical and physical properties of an element. Charcoal and diamond are one and the same chemical element, viz., carbon. They behave in the same manner in a chemical reaction though they differ drastically in terms of physical properties, while one is a black fuel the other is a dazzling jewel.

Phoneticians and Grammarians of ancient India have discovered that one and the same *sandhi* rule is applicable to a group of speech sounds, groups such as vowels Vs consonants, stops Vs nasals etc. In the absence of such a grouping, the entire list of sounds for which a rule is applicable has to be enumerated. By the use of the grouping of sounds, a rule can be stated succinctly. However, grammatical rules didn't dictate the grouping of speech sounds. As already said, grammatical rules are formed based on spoken Sanskrit and not the other way around. There has to be an underlying unification principle or a common mental process operating over a group of speech sounds since a native speaker makes use of such a grouping principle involuntarily in fluent speech.

On what basis do speech sounds belong to a common group? A group of sounds that behave in a similar manner must share a 'common property'. In the contemporary literature, the term 'Distinctive Feature (DF)' [8, 9] is used instead of the term 'common property'. This implies that a phoneme by itself is not an 'atom of a language' (not an ultimate indivisible wholesome unit) and that a phoneme is made up of a bundle of certain attributes (bundle of DFs). Identifying one or more common properties or DFs amongst a group of speech sounds is the subject matter of Phonology. There are diverse schools of Phonology [8, 9]. In Sanskrit language, speech sounds are divided into 14 (9+5=*nava pancha*) groups called Shiva Sutras. As per SS, Phonology is based on the production of speech sounds and hence it is referred to as 'articulatory phonology'.

SS states that there are two types of effort, internal and external, involved in the production of a speech sound. Details on the internal and external efforts of speech

`

production, both from the ancient and contemporary viewpoints, will be elaborated in Sec. III. Two speech sounds can be grouped together if they share at least one common property in either internal or external effort. Thus, for example, both 'p' and 'b' share two common attributes of internal effort *oshtau* + *sprustam* (labial+stop). Thus, a grammatical rule that is applicable to labial stops operates on both these sounds. Two speech sounds are considered distinct if they differ by at least one property or attribute. In the above example, 'p' and 'b' are distinct speech sounds since they differ in terms of the external effort attribute, viz., voicing; It is said that sound 'p' is voiceless (external effort: *vivara*) and 'b' is voiced (external effort: *samvara*).

II.B.4. Prosody

The study of duration of speech sounds and *svaras* (tones) comes under 'Prosody'. In experimental phonetics the term 'segmental' is used to denote the characteristics of an individual speech sound and the term 'supra-segmental' is used to denote the characteristics that extend over several speech sounds such as a word or an utterance.

Unlike a still picture where all objects in a scene are perceived almost at once, SS recognizes that a spoken utterance consists of a sequence of speech sounds that are temporally spread out. As per SS, the unit of timing is called *matra kala*. SS raises certain interesting questions as related to the duration: Is the duration of different speech sounds the same or different? For example, what is the difference in *matra kala* of a vowel Vs a consonant ('a' Vs 'k') or an aspirated Vs an un-aspirated stop sound (Ex. 'k' Vs 'kh' or 'p' Vs 'ph')? Etc.

Another aspect of prosody is related to the pitch or tone heights or *svaras* associated with the vowels as in chanting. The ancient experts of SS have raised some interesting questions such as for example: In the pronunciation of a syllable with *svarita svara* (mixed tone or High-to-Low tone), how long is the 'high tone' relative to the 'low tone'? Is it half a *matra* or one *matra*? Formulation of such questions implies that experts possessed a keen auditory sense to discriminate duration of sub-segments and ability to perceive variations in pitch.

`

Certain compound words (*samasa*) consist of one and the same sequence of basic speech sounds. As an example, the compound word *'sthulaprasati'* of Sanskrit describing a cow may denote [2] (a) 'having big spots' or (b) 'stout and spotted'. The ambiguity in the meaning is resolved based on the placement of the accent, whether it is on the first or the second member of the compound word. This reminds one of the words in English language like 'object' or 'judge' etc that may be used either as a noun or as a verb depending on the placement of stress. Emphasis has been laid on the proper pronunciation of accents in addition to the proper pronunciation of *varnas* (speech sounds).

Today's instruments, including, software tools, in the area of experimental phonetics may be used for visualization of a speech signal and to quantitatively study supra-segmental phenomena to a very fine temporal resolution. In Sec. IV we present the application of experimental phonetics on the measurement of m*atra kala* as well as relative pitch of *svaras*.

III. An Overview of Speech Production

A brief outline of speech production, specifically the internal and external efforts (*pryatna*) as given in SS are presented. These concepts are placed in the contemporary context.

III.A. Speech Production

Human speech production begins with a desire (*'ichcha'*) to express one's thoughts or ideas (*'jnana'*). This desire manifests by means of speech activity (*'kriya'*). Speech production is said to occur in four stages or levels, viz., *para*, *pashyanti*, m*adhyama* and v*aikhari*. Various interpretations are given to these four stages. It is said that the first three stages or levels are cognizable only to yogis. It is only the fourth stage that is accessible to a common man. In other words, the fourth stage corresponds to the physical level. We are concerned here only with the fourth stage.

Speaking is a learnt voluntary skilful motor activity. Brain and Nervous system issue well planned neuro-physiological commands (electrical current via efferent nerves) to the motor organs (*vak* - *indriyas*) involved in speech production. In SS, this activity is described briefly as *'the mind then excites the bodily fire,...*'. The motor activity finally results in a sequence of audible sounds. A listener perceives speech (via the *jnyanedriya* of the ears) and interprets speech with a meaning in his/her mind.

`

III.B. Organs of Speech Production

At the peripheral level, three major motor organs are involved in speech production, viz., the lungs, the larynx and the vocal organs (or articulators). See Fig.1. See also Fig.3. These organs have evolved primarily for the sake of survival; Lungs for respiration; Larynx acts as a valve preventing food and water from entering the wind pipe or trachea; Vocal organs are used while eating. These organs have been skilfully adapted for vocalization such as speaking or chanting or singing.

SS describes the functional role of these three organs in speech production. The coordinated functioning of lungs and larynx together has been termed as 'external effort'. On the other hand, the term 'internal effort' is assigned to the appropriate positioning of articulators like velum, tongue body, tongue tip, lower jaw and lips. In addition there is also a mention of *anupradana* (secondary effort). In contemporary literature, the term 'manner of articulation' and 'place of articulation' are used to denote 'external' and internal' efforts, respectively.

The terms 'external' and 'internal' are relative to the mouth (or buccal) cavity that extends from the lips to the top most point of the larynx; Mouth cavity comprises the pharynx (from larynx till uvula) and oral cavity (from uvula to lips). In contemporary literature mouth cavity is referred to as Vocal Tract (VT) and the space between the velum and the nostrils is referred to as 'Nasal Tract'. Velum can either completely or partially close the passage of air vibrations to nasal tract.

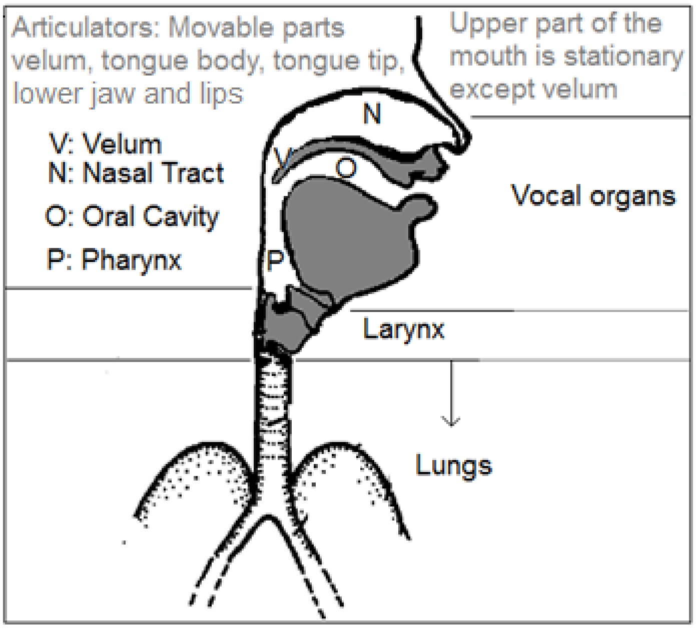

Fig.1. Organs of speech production

`

III.C. Role of Lungs in Speech Production

When a palm is held close to the lips, air flow may be felt during the pronunciation of some of the sounds like 'u'(oo) or 'kh' or 'p' etc. Speech sounds are vibrations in the air. Thus there has to be a supply of air into the mouth so that speech sounds may be uttered. Lungs act as a reservoir for the supply of air required during speech production. In SS, it is a well recognized fact that speech is produced during the exhalation phase of breathing. Recall the statement *'yasya nishvasitam vedah'* ('Vedas are verily the exhaled breath of the Cosmic Purusha').

Initially, we describe the role of lungs in speech production as per contemporary knowledge and later relate them to the statements in SS. In normal breathing, inhalation is active and exhalation is passive with the duration of inhalation and exhalation being approximately equal. In order to produce vocalization, as in speaking or chanting or singing, air is filled-up in the lungs with a deep inhalation. The natural involuntary recoil of muscles to exhale as in respiratory activity is inhibited with an effort. Air is released from the lungs in a controlled manner such that the exhalation prolongs as long as possible. If there are frequent shallow breaths, speech would sound as though the speaker has just completed running and has abruptly stopped for speaking.

There are various types of breathing. In abdominal breathing a larger quantity of air can be filled-up within the lungs. This enables one to speak for a longer duration with one inhalation. Hence, abdominal breathing is recommended. The maximum quantity of air stored in the lungs is called ' vital capacity'. It is typically of the order of 5 litres of which about 3 litres of air is utilized in normal speaking. A simple and an indirect method of determining the quantity of air available for speaking is to measure the maximum duration for which one can continuously utter vowel 'a' as in singing the same note. The measured duration is referred to as 'maximum phonation time (MPT) or maximum phonation duration (MPD)'. For a well trained singer or chanter, MPT (or MPD) is about 20 seconds. Typically, MPT is about 8 to 10 seconds. In contrast, in normal breathing exhalation duration is typically about 2 to 3 sec. By a practice of proper breathing exercises and by the use of abdominal breathing, MPT can be increased.

Spoken speech can be soft or loud. When you are conveying a secret to a friend, you speak very softly. When you are angry, you speak loudly. A commander of an army when issuing a command speaks at a very high volume or level. The level or the volume of speech is determined by the excess pressure with which air is stored within the lungs. This excess pressure in the lungs is referred to as 'lung pressure' or 'sub-glottal pressure'. Typically, lung pressure is of the order of 8 cm of water for conversational speech. Since this pressure is

`

very low the unit used for lung pressure is 'cm of water' instead of 'mm of mercury'. Thus, both the volume of air and the lung pressure play a role in speech production process.

Role of lungs as in SS: With the above background, we can now appreciate the insights of the phoneticians of ancient India. They describe the role of lungs succinctly as follows: "*The breath (air), circulating in the lungs (implies air filled-up within the lungs), is forced (implies excess pressure in lungs) upwards (implies exhalation) and, impinging upon the head (i.e., top of trachea, viz., the larynx), reaches the speech-organs (implying vocal folds and mouth cavity) and gives rise to speech-sounds*" [Ref.1, Sec.1.0]. It is understood that the exhaled air is modified by the larynx in order to produce audible speech sounds as described below.

III.D. Role of Larynx in Speech Production

Moving silently the articulators like tongue etc within the mouth would not produce audible sounds. Merely exhaling air from the lungs via the mouth doesn't produce audible speech sounds. Typical respiratory cycle is about 12 to 15 per minute or 4 to 5 per second. Such cycles of low frequency breathing doesn't lie in the audible range (See Appendix-A).

Larynx plays a versatile role in converting the slowly exhaled air from lungs either into periodically interrupted air flow or random disturbances in the air so as to produce sounds in the audible range of frequencies (20 to 20000 Hz) by means of aerodynamic processes. See Sec. IV.B on the mechanism of production of voiced sounds and Sec IV.C on the mechanism of production of fricative sounds.

III.D.1. Structure of larynx

A schematic structure of larynx is shown in Fig. 2 to highlight the most important parts [10-12]. The two vocal folds are bound to each other at the front towards the thyroid cartilage (or Adam's apple). At the other end (towards the back), each fold is connected to a separate muscle called 'arytenoid' that can bring the folds together (a movement called medial compression or adduction) or separate them apart (a movement called 'abduction') relative to each other. This action is like two windows mounted on hinges. The gap between the two folds is referred to as the glottis. The degree of opening of the glottis can be controlled. The two vocal folds can be made to be in contact with each other along their entire lengths or only partially along their lengths leaving a glottal opening over the rest of their lengths.

A contemporary knowledge of the anatomy of larynx has been gathered during dissection. Further, a live image of the larynx can be captured using the latest instrument, called fibre laryngoscope. A miniature camera is mounted on a thin fibre that can be inserted into the

`

mouth and the camera is held over the vocal folds. The subject is asked to say vowel 'ee' as the pharynx is wide open for this vowel. A video can be captured and transferred to computer memory. See Figs. 2d and 2e.

Larynx is well hidden within the mouth. Yet, SS identifies the important parts of the larynx. The top-most point of larynx, thyroid cartilage or 'Adam's Apple', is referred to as *'kakalaka'* in SS. Glottis is the narrow gap or orifice formed between the two vocal folds within the larynx. SS refers to the glottis by the term *bilam kan**th**asya*. The fact that glottis refers to an air gap is further recognized in SS by the usage of the term *kha:* (space). The two vocal folds are referred to as *ghoshatantryau* (literally strings of voice - note the usage of dual number). The author is not sure if this term '*ghoshatantryau*' is a later insertion or if it is found in the ancient texts.

Probably, knowledge of the structure of larynx might have been acquired in the ancient India during medical dissection indicating a good coordination between professionals of different fields. That such knowledge existed in the past is not as surprising as the fact that the functional role of larynx was so well understood.

III.D.2. Role of larynx as in SS

During normal respiration, the two vocal folds are widely separated (about 5 mm) forming a wedge like shape (Fig.2b or Fig 2d). There is a wide open glottis. In this resting position, the vocal folds are relaxed, i.e., there is no tension in the vocal folds. To change the vocal folds from their resting position in order to speak or sing or chant, a muscular effort has to be exercised.

In SS, the manner in which steady exhaled air from lungs is disturbed by the larynx to produce sound energy of audible frequencies is termed as *bahya prayatna* (external effort). There are two types of external efforts: vi*vara* (open) and *samvara* (closed) [3]. Speech sounds produced with these two efforts are referred to as 'voiceless' and 'voiced', respectively. Recent technical literature uses the term 'unvoiced' instead of the term 'voiceless'. In this Chapter we use the term voiceless.

The effort *vivara*: In the relaxed or resting position as in respiration, vocal folds are held apart and the glottis is open. Why then the term *vivara*, meaning 'open', has to be mentioned as an effort? It has to be understood that *vivara* is an effort to bring the vocal folds closer to each other from their relaxed position and at the same time to make sure that they don't make contact and close the glottis. In other words, this is a controlled degree of openness and hence it is considered as an effort.

`

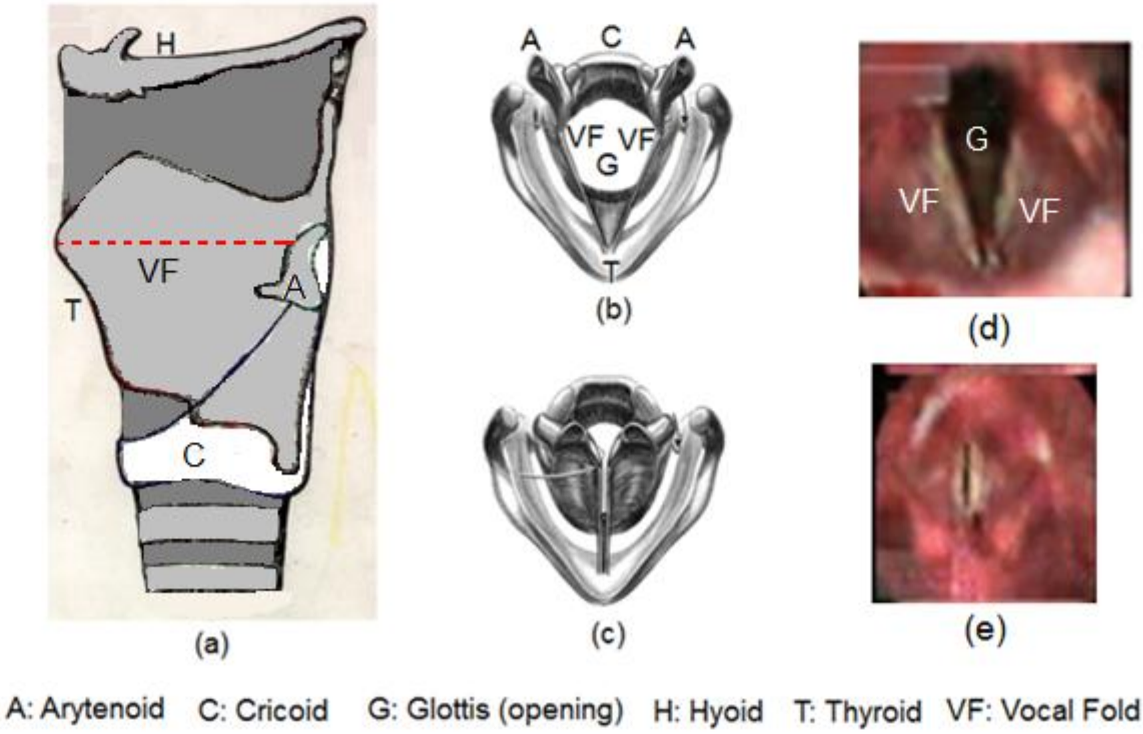

Fig.2. Structure of larynx. (a) Profile or Mid-sagittal view (b) Top view with vocal folds (VF) widely separated (exaggerated) as in normal breathing giving rise to a wide open glottis, G (c) Vocal folds are in contact and the glottis is almost closed as required for initiating voicing (d) Laryngoscope image of the larynx of a live subject in respiratory state (e) A frame of Laryngoscope image of the larynx of a live subject during voicing.

SS notes that in case of *vivara* there is a controlled exhalation of air via a small opening in the glottis. Voiceless sounds such as fricatives ('s', 'sh', "S**h**' and 'h') as well as voiceless stops ('k', 'kh', ... 'p', 'ph') are produced with this effort. Further, SS recognizes that different voiceless sounds have different auditory qualities of noise. The audio quality of noise is referred to as ***shvasa*** in SS and in some texts it has been translated as 'breathy', though not very appropriate. As seen later in case of *samvara* there is periodic disturbance (*ghosha*) in air flow. Such a periodicity is absent (***aghosha)*** in case of *vivara*. See Fig.5 for a graphic representation of a typical voiceless sound.

Note: Although, the effort *vivara* (open) is at the level of glottis, i.e., external to the mouth cavity, the actual mechanism of noise generation is NOT at the level of the glottis for all voiceless sounds. Noise is generated at the glottis only for 'voiceless h' (aspiration). For other voiceless sounds, noise is generated within the mouth at the *sthana*. See discussion on 'Mechanism of production of fricative sounds' given in Sec. IV.C and 'Mechanism of production of stop sounds' in Sec. IV.E.

`

The effort *samvara*: In this effort vocal folds are 'contracted', i.e., brought together (medial compression or adduction), so as to close the glottis. It should NOT be interpreted to mean that the glottis remains closed for the entire duration of voiced sounds. This effort is accompanied with two consequences: (i) ***ghosha*** - periodic vibrations of vocal folds and (ii) ***nada*** - a distinct resonant quality of the sound as perceived by the auditory system. It is an extraordinary insight that bringing together the vocal folds alone is sufficient to cause periodic vibrations (*ghosha*) of vocal folds. Specifically, note that there is no mention of any effort for every cycle of vibration. Instead, the **effort is only to initiate voicing**. Vowels and semi-vowels produced with the external effort of *samvara* are called voiced sounds. Also voicing is present for voiced stops ('g', gh',...'b', 'bh'). See Fig.5 for a graphic representation of a typical vowel sound. See Sec. IV.E.

The details on the mechanism of generation of periodic vibrations of vocal folds seem to be missing in SS. In the contemporary literature, this phenomenon of periodic vibrations associated with voicing is referred to as phonation or self-oscillations and the same is explained in Sec. IV.B.

Some texts ascribe the external effort only to the consonants and not to the vowels. This is an incorrect view. Three **additional** types of external effort are associated with the *svaras* of vowels, viz., *anudatta (*Low or Grave tone), *udatta* (High or Acute tone) and *svarita* (Mixed or High-to-Low tone). *Svara* or tone can only be associated with a periodic sound and hence it is obvious that vowels are also voiced sounds produced with the external effort *samvara*.

Note: SS also mentions 'mixed type of sounds' i.e., both voiced as well as voiceless [Ref. 1, Sec.1.120]. In a mixed type (simultaneous *samvara-vivara*), part of the vocal folds, along their length, are in vibration while the rest of the vocal folds are held apart resulting in a small glottal opening. This is a regular feature in the production of voiced stops. See Sec. IV.E. During fluent speech, the transition from one effort to another doesn't occur abruptly. Thus, during transitions from voiced to voiceless or vice-versa there exists an interval of mixed type. It is also possible to bring about such a mixed state voluntarily, as in the case of 'z' of English.

III.E. Internal Effort or Articulation

III.E.1. Sound quality and the geometry of vocal tract

When an empty bottle is blown at the opening of a narrow neck, a sound of distinct resonant quality is heard. Quality of the sound depends on the shape and size of the bottle, length and area of opening of the neck. In general, for any container, the sound quality depends on the geometry of the space occupied by the vibrating air particles. This basic principle is

`

utilized in speech production. Speech sounds of differing quality are produced by altering the geometry of the space within the mouth cavity by moving the articulators.

Silent articulation doesn't produce audible speech. Either a random noise or periodic pulses of air flow provides the sound energy as mentioned earlier. The sound energy (called the source) generated as a consequence of the external effort is modified by the vocal and/or nasal tracts. A device that modifies the energy distribution with respect to frequency, i.e., the quality of a sound, is referred to as a filter. In speech production, Vocal and Nasal Tracts act as filters.

III.E.2. Unmovable and movable parts

A profile view of vocal and nasal tracts is shown in Fig.3. The hollow part from the glottis to the root of the soft palate or velum is called pharyngeal cavity or simply pharynx. The geometry of pharynx is fixed for a given individual except for a minor detail that larynx as a whole can be raised or lowered thereby altering the length of the pharynx. There is no movable part within the pharynx. The space from the root of the tongue to the lips is called the oral cavity. The unmoving upper part of the oral cavity consists of upper jaw, alveolar ridge, hard palate. That part of the part palate towards the pharynx is called the soft palate and it is movable.

Karanas (Articulators)**:** The lower part of the oral cavity is dominated by the tongue. Tongue is divided broadly into three parts: the tongue root, mid-part of tongue body and the tongue tip. Tongue body can be moved up or down and forwards or backwards. Tongue tip can be raised to any desired angle to touch the alveolar ridge or the upper teeth. Tongue tip can be folded backwards to touch the roof of the mouth or the hard palate. Beyond the teeth, the lower jaw can be moved along an angular arc. Lips can be spread out or protruded with rounded shape. The movable parts within the mouth cavity, viz., velum, tongue, lower jaw and lips are called *karanas* ('articulators'), literally, instruments. By the movement of the articulators, the geometry of the space within the VT can be altered.

The part of soft palate near the pharyngeal wall is the velum. Velum can be pushed backwards that raises the uvula to act as a valve to either couple or de-couple the nasal tract. The degree of coupling can be controlled. For non-nasal sounds, passage to nasal tract is completely closed. Movement of velum is considered as a secondary feature. Sounds coming out of the mouth with partial opening to nasal tract are called 'nasalized sounds'. Sounds coming out of only the nostrils and not the mouth are called 'nasals'.

Sthanas **(place of articulation):** The curved path along the mid-line of pharynx and oral cavity from the glottis to the lips is called the **axis of VT**. By a controlled movement of the

`

articulators, the gap between the unmoving and moving parts within the mouth can be made narrower or broader at any chosen location along the axis of VT. **The location of the narrowest gap along the axis is referred to as *sthana* (place of articulation).** Although the narrowest gap could be formed anywhere along the axis, only FIVE places have been specified in SS. Sounds with the same *sthana* are said to be homogenous. The fact that there are only five places of articulation is universally true across all languages of the world. Reasons for choosing only five places for speech production are presented in Sec. IV.G.

The five places of articulation are (Refer Fig.3) (i) *kan**th**a* (literally 'throat' but referred to as 'velar') - region marked 1 (ii) *talu* (alveolar) - region marked 3 (iii) *murdhni* (literally 'head', here, roof of the mouth or hard palate) - region marked 2 (iv) *dantaha* (ridge of the upper row of teeth) - marked T (v) *ostau* (lips) - marked L.

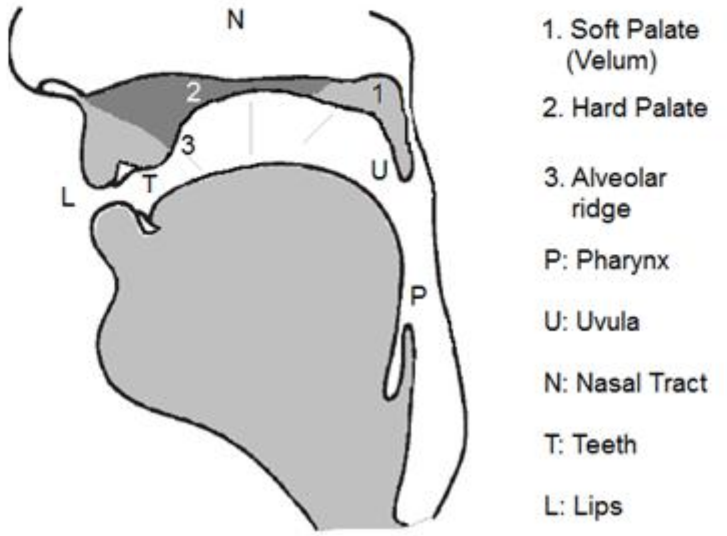

Fig.3. A profile view of vocal tract and nasal tract showing places of articulation

III.E.3. Acquiring knowledge about organs of speech production

In 1950-60s, X-ray imaging technique was used to acquire knowledge about articulation of vowels. To capture the movement of articulators during pronunciation of syllables, cine-radiography was used. Only a profile view is obtained by X-ray technique. Lateral view of the front part of the mouth is deduced using instrumentation called Palatography. At present advanced MRI technique is used to obtain 3-D view of the vocal tract. However, at present MRI images can be captured only for steady vowels since scanning time is large and the subject must not move the articulators during the scanning time.

`

The movement of parts like lower jaw and lips is clearly visible. The action of tongue tip can be sensed with a tactile feedback. However, the action of parts like vocal folds, velum, actual contour of the tongue body etc. are neither visible nor are they felt. Despite this, SS describes correctly the production of different types of sounds as confirmed by the modern day techniques. We can appreciate the insights of Phoneticians of ancient India in arriving at such an accurate description of these hidden processes. However, it is difficult to comprehend the methodology they might have used to arrive at such knowledge.

III.E.4. *Abhyantara Praytna*

Controlling the degree of gap between *karana* and *sthana* is called a***bhyantara prayanta*** or the internal effort. Although the size of a gap can be controlled in a fine graded manner (from full contact to widest possible opening), only five types of gap formations are used in speech production. The fact that there are only five types of gap formations is also universally true across all languages of the world. The reasons for using only five internal efforts for speech production are discussed in Sec.IV.G.

Speech sounds are grouped based on the five types of internal effort.

(i) ***Sprustam*** (literally 'contact'): This effort actually produces a complete closure within the VT, at one of the five places, preventing air vibrations to pass beyond the point of closure. If *sprustam* produces a complete closure for the passage of air, how come stop sounds are produced at all? This requires a proper interpretation of the production of stops. See Sec. IV.E for a detailed description on the mechanism of the production of stops based on the recent findings. Some additional aspects associated with the production of stops like micro-phonetic intervals, transition interval and co-articulation are also elaborated in Sec. IV.E.

A total of twenty five sounds are produced with this internal effort. These are the sounds 'ka' to 'ma' in the standard the vernacular Table of Sanskrit. There are two groups within these twenty five sounds, viz., nasals and stops.

With this internal effort, five voiced sounds corresponding to the five places along with a coupling only to the nasal tract are the five 'Nasals'. Recall that voiced sounds are produced with the external effort *samvara.* See Sec.IV.F for a discussion on the sound quality of nasals.

With this internal effort, sounds produced excluding the five nasals (no coupling to nasal tract) are called 'stops'. In the case of stops, for each *sthana*, the external effort can be one of the four different types thus resulting in twenty different stop sounds. The four external efforts are: *vivara* (open, voiceless) or *samvara* (closed, voiced) followed by an additional

`

sequential external effort called *maha-prana* (aspirated) or its absence (silence), *alpa-prana* (un-aspirated). See Sec.IV.E.

(ii) ***Isad-sprustam*** (slight or partial contact): Here, tongue tip or edges of the tongue make a partial contact with the upper part of the mouth. Voiced 'h' and semi-vowels ('y', 'r', 'l', 'v') are produced with this internal effort combined with the external effort of *samvara* ('voicing'). Hold the tongue in the position to say 'l' or 'r' but breath-in to feel the passage of air along the sides. This clarifies the meaning of 'partial contact' and shows that vibrations of air can pass beyond the place of partial contact.

(iii) ***Isad-vivrutam*** (slight or narrow opening like that of a constriction): Doesn't a partial contact or partial closure (*isad-sprustam*) result in a partially open state? What is the distinction between the internal efforts, partial contact and partial open? In the case of *isad-vivrutam* air is forced to pass via a narrow opening. To emphasize this aspect a separate internal effort has been designated. Fricatives (s, sh, Sh, h) are produced with this internal effort. The external effort for producing fricatives is *vivara* ('voiceless'). See Sec. IV.C on the mechanism of production of fricative sounds. The external effort is different for *'isad-sprustam'* and *'isad-vivrutam'*.

Note: There are two manifestations of sound 'h'. Firstly, voiced 'h' grouped under semi-vowels. Such a voiced 'h' occurs when preceding and following sounds are voiced, ex., 'aham'. Secondly, voiceless 'h' grouped under fricatives.

(iv) ***Vivrutam*** (open): Although this effort is called 'open' yet the narrowest gap is at the place of articulation (*sthana*) compared to elsewhere within the mouth cavity. This effort provides just enough gap between the articulator and *sthana* for the vibrating air particles to pass without being obstructed (without causing friction). All vowels are produced with this internal effort. Also, all vowels are produced with the external effort sam*vara*, ('voiced') as already noted. This would imply that there has to be only five vowels corresponding to the five places. Additional vowels are produced by a combination of internal efforts. Nasalized vowels are produced with a partial coupling to nasal tract (*anunasika*). *Visarga* is a vowel ending in aspiration.

(v) ***Samvrutam*** (contracted): In SS it has been noted that sound quality of *hrasva* (short) vowel 'a' is distinct from that of a long vowel 'a'. This is apart from a difference in duration. The internal effort of short 'a' is said to be *samvrutam.* In Sanskrit, the contrast *vivrutam* Vs *samvrutm* is considered to be a phonetic difference for the application of grammatical rules. In modern phonetics, such a distinction is referred to as 'lax' Vs 'tense' contrast. The presence of such a distinction has been noted in several languages as well and also for vowels other than 'a'. It is not clear why for Sanskrit such a distinction has been mentioned

`

only for vowel 'a'. Tools of experimental phonetics may be used to investigate if such a distinction exists as well between *hrasva* and *dirga* for other vowels of Sanskrit.

IV. In Retrospect

This section has two broad aims: (i) To emphasize the distinction between the concept of *prayatna* and the mechanism of speech sound production (ii) To illustrate the procedure for the measurement of duration and pitch of *svaras* using tools of experimental phonetics.

The physical process arising as a consequence of an external effort is independent of the language and hence recent knowledge acquired with respect to other languages may be applied to Sanskrit as well. In SS it is an implicit knowledge that the external effort *samvara* leads to periodic vibrations (*ghosha*). However, an explanation of the mechanism behind this is not seen in SS. Based on recent knowledge, the mechanism of periodic vibrations, called phonation, is given in this Section. Also, a discussion with respect to fricatives and stops is dealt with in this Section.

IV. A. *Prayatna* Vs Mechanism of Speech Production

Descriptions of external and internal efforts given in SS seem to be applicable at the physiological level. The author is not aware of ancient texts explaining the physical process (mechanism) that follows as a consequence of external and internal efforts. It might have been felt that such details are unwarranted for several reasons. Firstly, the classification of speech sounds based on efforts is adequate enough to apply grammatical rules. Secondly, in the Gurukula system of education, proper pronunciation of chanting or speaking was taught directly under the supervision of a teacher by oral tradition. Knowledge of the process of speech sound generation might have been considered as superfluous or of academic interest only. Thirdly, even if the process was presumably known, conveying the same without writing and graphics is a very difficult task.

In fluent speech, consisting of a sequence of speech sounds, the articulators are continually moving from one set of positions to another. In other words, there is a continuous change in *sthana* and *karana* during fluent speech. Even the transition from voiced to voiceless or vice-versa (manner of articulation) is gradual. SS shows an awareness of this aspect. See Sec.IV.E.4 on the contextual effect. But, a detailed discussion on the consequences of such dynamics seems to be missing in the traditional SS.

Based on the recent theoretical research into the acoustic theory of speech production and availability of technological devices and instruments of experimental phonetics, we can look

`

back into some of the aspects in SS and gain a better understanding, especially as related to the mechanism and dynamics of speech production etc.

IV.B. Mechanism of production of voiced sounds

We mentioned in Sec. II.D that during the production of voiced sounds, the external effort of *samvara (*contracted vocal folds or closed glottis) results in *ghosha*, i.e., periodic vibrations. This mechanism is referred to as phonation. The current understanding of the process of phonation by which the vocal folds are set into self-oscillations during voicing is described below.

IV.B.1 Phonation

See Fig.4. During normal respiration, vocal folds are in their rest or **relaxed position** and are wide apart (typically about 5 mm). When there is a desire to vocalize, the **two vocal folds are forcefully brought together with an effort** (*samvara*) by a process of adduction exerted by arytenoids. This introduces tension in the vocal folds.

(a) The tensed vocal folds come into contact and **close the glottis**.

(b) Air in the lungs below the closed glottis is at a pressure greater than the atmospheric pressure. Air above the closed glottis is nearly the same as the atmospheric pressure. There is a large **pressure difference** across the closed glottis. This is called 'trans-glottal pressure'. The pressure difference forcefully pushes aside the vocal folds creating a narrow glottal opening. A **jet of air gushes out** to equalize this pressure difference. This is like pinching the open end of a tube through which water is flowing from a height.

(c) The **vocal folds move away** from each other. The **area of the glottis increases** and the quantity of **air flowing through the glottis also increases**.

(d) The folds reach the maximum separation (of about 2 to 3 mm) with the volume of air flow through the glottis reaching a peak value.

(e) As the air gushes out through the glottis, the pressure just below and just above the glottis gets equalized. This is like the pressure difference between two close-by points on a tube through which water is freely flowing. Because of free air flow, the **force** that separated the vocal folds **is now absent**.

(f) At the beginning, for the initiation of voicing, tension (adduction or medial compression) was applied on the folds. The folds are still under this tension. This **tension brings the vocal folds together** in a manner similar to automatic closing action of a pushed spring door.

`

(g) As the **folds approach each other**, the **area of the glottis decreases**. The quantity of air flow decreases.

(h) Finally, **vocal folds make contact**.

(i) Due to momentum, the vocal folds continue to press against each other and thereby keeping the glottis closed for a while (closed glottis interval). **Air flow** from the lungs **ceases**. The state of the glottis and vocal folds now is the same as that of state (a). The sequence of actions (a) to (i) repeats itself cyclically.

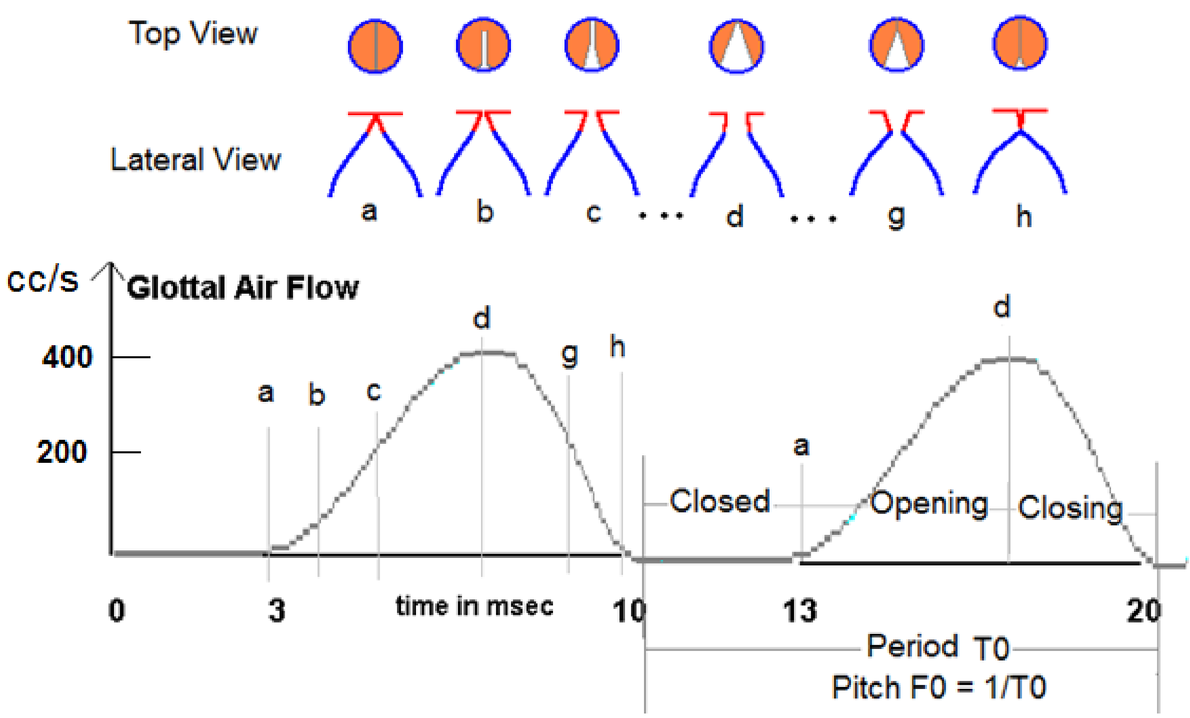

Fig.4. Top: Schematic representation of vocal fold movement for one cycle. Bottom: Typical air flow via glottis. Notations (a) to (h) correspond to the description given in the text. Note: Unit of x-axis 'msec' is milli-second or 1/1000th of a second.

IV.B.2 Glottal Cycle

There are three intervals in each glottal cycle: (i) Opening: Vocal folds separating from each other; glottal area and air flow increasing (ii) Closing: Vocal folds approaching each other; glottal area and air flow decreasing and (iii) Closed: Vocal folds in contact; glottis is closed; there is no air flow. Theoretically, the shape of glottal area Vs time and glottal flow Vs time are different. However, we will not go into these details here. These three intervals repeat periodically. The steady exhaled air flow from the lungs is thus interrupted producing periodic pulses of air flow called glottal pulses. The time interval of one full cycle is referred to as the period or pitch period denoted by T0. The number of glottal cycles per second, 1/T0, is

`

called the fundamental frequency denoted by F0. For the example illustrated in Fig.4, T0 is 10 msec and hence F0=1/(10*0.001) = 100 Hz (cycles per second). F0 determines the sensation of pitch (*svara*) associated with voiced sounds.

Note-1: Phonation is a phenomenon of self-oscillations. There is need to apply neuro-muscular command every cycle to separate or bring together the vocal folds. Instead, the neuro-muscular command (effort) is applied **only once** at the beginning in order to bring the vocal folds together with an intention to initiate voicing. This is unlike cyclic heart beats where distinct electrical commands are issued to the muscles of the heart during the different phases of pumping of blood.

Note-2: Sometimes the vibrations of vocal folds are compared with the vibrations of a string. This is an incorrect analogy. In case of a string, air surrounding the vibrating string is physically displaced by the coupling. The quantity of air that gets displaced because of physical contact with the vibrating vocal folds is negligibly small. The maximum horizontal excursion of vocal folds is only about 2 to 3 mm and there is hardly any vertical movement. It is the periodically interrupted air flow via the varying area glottis that produces voiced sounds. These glottal pulses have frequency components in the entire audible range, though they have dominant low frequency components.

IV.B.3. Control of *svara* or tone

Voicing is invariably associated with periodic vibrations (*ghosha*), which in turn is associated with a sensation of pitch. Typically, F0 is of the order of 70 to 130 Hz for adult male speakers, 130 to 180 Hz for adult female speakers and above 180 Hz for children. This natural frequency of vibration is determined by the mass and effective length of vibrating vocal folds as well as the tension applied on the vocal folds to initiate voicing. Only in this respect, the analogy of vocal fold vibrations to the vibrations of a string is meaningful.

It is possible to control F0 voluntarily by an additional neuro-muscular effort independent of the natural settings of vocal folds. Even during normal speaking, pitch varies over an utterance with emotions, emphasis and/or stress. Pitch typically rises towards the end of an interrogative utterance.

The additional voluntary control of pitch can be appreciated when one listens to Vedic chanting or vocal singing in contrast to casual speaking. In SS, broadly three tones, classified under external effort, have been mentioned and are called *anudatta* (Low), *udatta* (High) and *svarita* (mixed or High-Low). See Sec.IV.H.

`

IV.C. Mechanism of production of fricative sounds

The external effort *virvara* (open) is at the level of the glottis. As already mentioned, the term 'open' here means maintaining a controlled narrow opening of the glottis. It is used in the production of (voiceless) fricatives. The external effort *vivara* is necessary but not sufficient to produce a voiceless audible sound.

There are rigorous theories on the noise generation process. As per the recent findings, when a jet of air at a very high velocity (not slowly moving air) passes through a narrow constriction, noise like sound is produced. The fundamental requirement is that the velocity of air particles must be very high, which is achieved by forcing air at high pressure through a narrow constriction. The sound quality of noise is influenced by the shape of constriction, downstream obstruction etc [6].

Consider production of sound 's' whose *sthana* is *dantau* (teeth) and internal effort is *isad-vivrutam* (slight opening). High energy air passing through the narrow opening between the tongue tip and the upper ridge of teeth, produces the 's' sound. Note that the noise is generated **internally** within the mouth cavity at the *sthana*, viz., upper teeth ridges, though the **external effort** *vivara* for 's' is said to be at the level of **glottis**. This shows that the location of external effort and the location of the noise generation are distinctly different. Similar remarks are applicable to all the fricatives except voiceless 'h' sound for which noise is generated at the glottis itself. Noise generated at the glottis is referred to as aspiration noise. Also, see the mechanism of production of voiceless stops described below.

IV.D. Visualization of voiced and voiceless sounds

There are two types of speech sounds called voiceless (or unvoiced) and voiced sounds based on the two external efforts, *vivara* and *samvara*, respectively. Some broad features of these two classes of sounds are illustrated using the latest visualization tools.

When sound waves impact on the diaphragm of a microphone in a magnetic field an electrical voltage or current is generated that ideally mimics the pressure changes in the atmosphere produced by the sound waves. An electrical signal can be recorded or digitized and saved in computer memory. Software may then be used to display the signal.

A graphic representation of an electrical signal is called ‘waveform’. Figure 5 shows a typical waveform of a vowel followed by a fricative. Here, the value along the y-axis at any given instant of time represents the amplitude of the acoustic pressure (but for a scale factor). The x-axis represents time interval. In this example, the measured durations of vowel and fricative segments are 200 and 160 msec, respectively. In actual practice the duration of

`

segments in fluent speech varies widely. However, in Vedic chanting the duration of different segments may follow some rules.

Note the repetitive structure in the signal waveform (Fig.5) during the mid-part of vowel segment. The repetitive pattern seen in Fig.5 during the vowel segment is different from that seen in Fig.4 for the glottal air flow. This difference arises because the signal shown in Fig.4 is for the flow of air through the glottis as calculated theoretically whereas the signal of vowel segment shown in Fig.5 corresponds to the measured acoustic wave in the atmosphere. The glottal pulses are modified by the filtering action of vocal tract to produce a vowel segment. In fact the structure of the waveform is different for different vowels. Compare the structure of vowel seen in Fig.5 with the structure of vowel segments seen in Figs. 6 and 7.

Small vertical lines are drawn coinciding with the positive peaks. The temporal interval between two successive small vertical lines is nearly the pitch period. In this example, it is about 10 msec. The actual temporal spacing varies continually from one cycle to another. Note the increasing peak amplitude near the beginning and decreasing peak amplitude towards the ending of the vowel segment. Since the temporal spacing between two successive peaks as well as the peak amplitude is changing from cycle to cycle, a vowel segment in fluent speech is said to be quasi-periodic (approximately periodic).

Voiceless fricative seen in Fig.5 has random distribution of low amplitude peaks typical of a noisy signal. On an average, the maximum peak amplitude associated with a vowel is much higher than the maximum peak amplitude associated with a fricative. For the fricative in Fig.5, there are frequent changes in the signal from positive to negative polarity and vice-versa indicative of greater concentration of energy at higher frequencies.

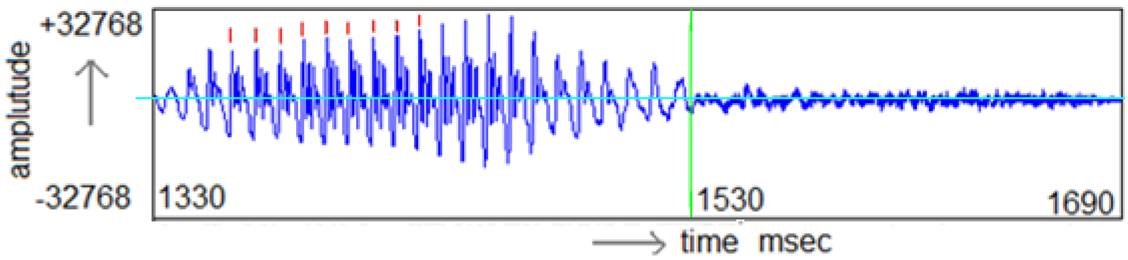

Fig.5. An example of a waveform of a vowel (between 1330 to 1530 msec) followed by a fricative (1530 to 1690 msec).

`

IV.E. Mechanism of production of stop sounds

In the production of a stop sound two different external efforts occur in immediate succession of one another. The first external effort corresponds to *aghosha* (voiceless) or *ghosha* (voiced) distinction. The second external effort corresponds to *alpa-prana* (un-aspirated) or *maha-prana* (aspirated) distinction. We refer to the physical process arising as a consequence of these efforts as sub-process-1 and sub-process-2, respectively. The four types of stop sounds based on these two external efforts are: (a) *Aghosha* + *Alpa-prana* i.e., Voiceless-Unaspirated ex. 'k' (b) *Aghosha* + *Maha-prana* i.e., Voiceless-Aspirated ex. 'kh' (c) *Ghosha* + *Alpa-pana*, i.e., Voiced-Unaspirated ex. 'g' (d) *Ghosha* + *Maha-prana*, i.e., Voiced-Aspirated, ex. 'gh'. Each of these four types along with the five *sthanas* result in twenty different stop sounds. Recall that aspiration refers to the noise generated at the glottis.

In English language, aspiration is optional. For example, the word 'pay' may be pronounced as 'phay' without a change in the meaning. Thus 'p' and 'ph' are phonetically distinct but phonemically identical. In Sanskrit, aspirated and un-aspirated stops are phonemically distinct.

The internal effort for stop sounds is *sprustam* or closure. How can a closure within the mouth produce a sound? The effort *sprustam* has to be considered only as an initial pre-requisite for the production of a stop sound. The complete process of stop production is described below.

Production of a stop sound involves rapidly changing sequence of events. The waveform of a typical voiceless aspirated stop preceded and followed by a vowel is shown in Fig.6. The waveform of a typical voiced un-aspirated stop preceded and followed by a vowel is shown in Fig.7. The significance of the various terms given in the Figs.6 and 7 are covered in the following description on the sub-processes.

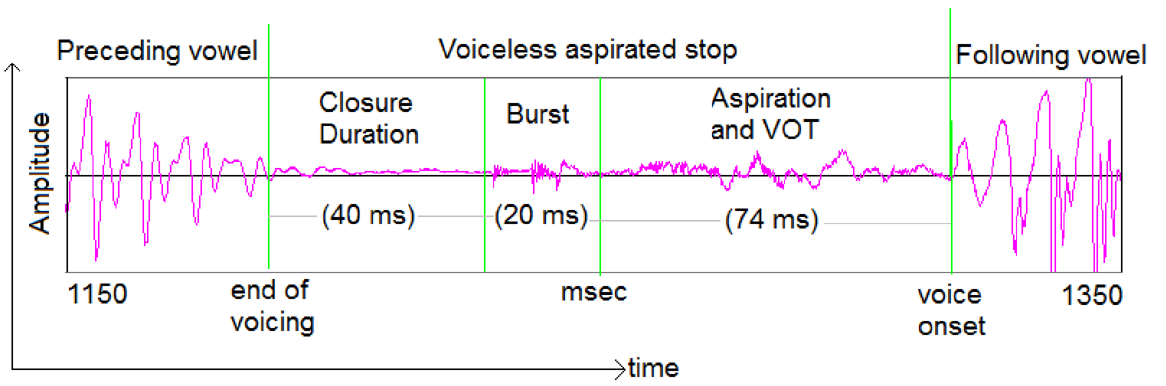

Fig.6. Typical waveform showing part of the preceding vowel, a voiceless-aspirated stop and part of the following vowel. Various events along with their typical durations are shown in the Figure.

`

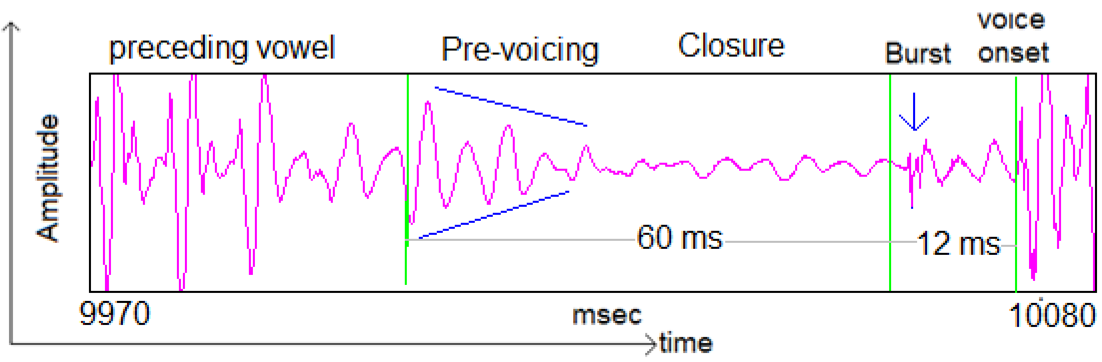

Fig.7. Typical waveform showing a part of the preceding vowel, a voiced un-aspirated stop and a part of the following vowel. Note the presence of voicing preceding the burst whose peak amplitude for each cycle decreases and then almost vanishes during closure.

IV.E.1. Sub-process 1: Pressure Build-up and Release

There are two intervals within sub-process-1: Closure duration and Burst duration.

(i) The **first step** in the production of a stop sound is to form a complete closure (*sprustam*) at the place of articulation (*sthana*).

(i.a) Consider voiceless stops. (Fig.6.) For the external effort *vivara* (open), air is exhaled via a narrow opening of the glottis. This steady flow of air is blocked by the closure at the *sthana*. Because of the stoppage of air flow there is a silence interval (note the closure duration in Fig.6). Like blowing a balloon, air accumulates behind the closure. The pressure above the glottis builds-up. For a voiceless stop, continue from step (i.c) below.

(i.b) Consider voiced stops (Fig.7). Periodic pulses of air produced at the glottis enters the mouth cavity. Since there is a closure at *sthana*, air accumulates behind the closure, like blowing a balloon.

Several possibilities arise in the case of a voiced stop. The pressure above the glottis builds-up. The pressure below the glottis is the lung pressure. The difference in pressure across the glottis (trans-glottal pressure) gradually decreases. This reduces the peak amplitude of voicing (see pre-voicing marked in Fig.7). As the air flow builds-up behind closure, vocal folds vibrate partially along their length and the remaining parts of vocal fold are held apart. This is a mixed state of *samvara-vivara* [Ref.1 Sec. 1.120].

In order that the vocal folds may sustain cyclic vibrations, a trans-glottal pressure difference of at least 3 cm of water is required. When the trans-glottal pressure falls below 3 cm of water, the periodic vibrations completely stop. Note this trend in Fig.7 (interval marked 'closure'). See Note-2 below for an exception. Subsequently, air flow via the glottis continues as in the case of a voiceless stop.

`

(i.c) Building-up of air pressure continues with the air coming out of the narrow open glottis till the pressure above the glottis is equal to the lung pressure. Then, the air flow via the glottis stops. But, see Note-2 below for an exception.

(i.d) When the pressure behind the closure is building-up, there is no sound output. This interval is referred to as 'closure duration'. During the closure duration there is silence (Fig.6) or very low amplitude voicing (Fig.7). Paradoxically, during the internal effort of *sprustam* (closure) there is no sound output.

(ii.a) The **second step** is to suddenly move the articulator away from the *sthana*, such as the lowering of the tongue tip or the separation of lips, resulting in a small opening at *sthana*. This is called the 'release' of the articulator.

(ii.b) Before the release, the pressure just behind the closure is almost equal to the lung pressure. The pressure in front of the closure is the atmospheric pressure. There is a significant pressure difference across the closure. The sudden release results in a 'burst' of air coming out of the mouth. This burst is like an explosion similar to pricking a blown balloon with a pin. Actually, it is this burst that is generally associated with a stop sound. See the burst marked in Figs.6 and 7. Energy in the case of a burst is restricted to a short interval. There may be repetitive bursts as seen in Fig.6. In SS, the action of release has not been mentioned. Perhaps it is implied.

Paradoxically, there is sound output when there is no longer a contact. Although the external effort is said to be at the glottis, the noisy burst is generated at the *sthana* as in the case of the fricatives.

(ii.c.) After the sudden release, the articulators continue to move gradually towards the next sound to be produced. See discussion on transition interval given below.

(ii.d) After the burst, a finite interval of time has to elapse before voicing can begin for the immediately following voiced sound. The interval from the release to the initiation of voicing is called the voice onset time (VOT). Such a delay is required since the trans-glottal pressure has to exceed 3 cm of water in order to begin voicing.

Note-1: The closure duration, though consists of silence, is not perceived as a pause or a break in the sequence of sounds. VOT is also not perceived as a pause or a break.

Note-2: The speaker may release the articulator before the air pressure above the glottis reaches the lung pressure. Such a premature release results in a weak burst. In the case of a voiced stop, the burst gets superposed on the continued low amplitude voicing (Fig.7).

Note-3: In the case of a word initial voiced stop, air pressure may be built-up behind the closure even before beginning to speak. Thus, pre-voicing may be absent for a word initial

`

voiced stop. This has been noted in the case of English voiced stops. Experiments have to be made to confirm if this is also true in the case of spoken Sanskrit.

IV.E.2. Sub-process 2: Follow-up after release: *maha-prana* and *alpa-prana*

The production of a burst doesn't mark the end of the production of a stop sound. One of the two possibilities occurs immediately after the release: production of either *maha-prana* or *alpa-prana*. As noted above, immediately after release, the pressure difference across the glottis is still small. In order to produce a stop with *maha-prana*, an additional intentional or extra effort is exerted and air is expelled through a narrow opening in the glottis with some force (Fig.6). This breathy noise that is generated at the glottis is referred to as 'aspiration'. *Alpa-prana* is the absence of an extra effort to produce aspiration. After the occurrence of burst the immediate next sound begins after a silence interval and hence the term '*alpa-prana*'.

IV.E.3. Transition Interval

The positions of the articulators to be attained for any given sound are referred to as articulatory targets. After the release of a stop sound, the articulators take a finite time to transit from their current positions to reach the articulatory targets of the immediate next sound. This is referred to as the 'transition interval'. The transition interval exceeds the sum total of the interval of burst, interval of aspiration, if any, and voice onset time, if any. The influence of a stop sound persists even during a part of the following sound. See Sec.IV.E.5 and description of Fig.8 given below.

IV.E.4. Co-articulation and Contextual Influence

Co-articulation: Assume that you have to pronounce the syllables, 'ka' (as in 'caught') or 'ki' (as in 'key') or 'ku' (as in 'cool'). It is understood that the syllables are of Sanskrit language though the words cited as examples are for English language. Observe the shape of the lips for the initial sound 'k' just prior to uttering the syllable,. The shape of the lips are those of the following vowel even while producing the initial 'k'. This is called 'co-articulation' or 'anticipatory co-articulation'. In general, such a co-articulation may happen with respect to any one of the articulators, not necessarily the lips. In a word like 'school', both 's' and 'k' are influenced by the following vowel. The author is not aware of any reference to the concept of co-articulation phenomenon in SS.

Contextual effect: Articulatory positions of the present sound are influenced by (transiting from) the preceding and (co-articulation effect of) the following sound. For example, in the utterance of word *'mantra'*, since both 'm' and 'n' are nasals, velum continues to be open during the mid-vowel 'a', which is a contextual effect. As a consequence, the vowel 'a' is nasalized. In the utterance of *'pachati'* spoken rapidly, the sound *'ch'* may become voiced, i.e., realized as *'pajati'* because of the influence of voicing in the preceding and following sounds [Ref.1., Sec.120]. The voicing of 'ch' may be realised as a mixture of *'samvara-vivara'*, i.e., vocal folds vibrating only partially along their lengths and the rest remaining apart. [See Quote from *Rik-Pratishakya* in Ref.1 Sec. 1.120].

IV.E.5. Visualization of transition interval and the influence of co-articulation

Fig.8 shows the waveform in the upper window and the so called 'wideband spectrogram' in the lower window of the uttered syllables 'kaki' ('kaakey'). Wideband spectrogram is a standard technical term used in the literature. A spectrogram is a graphic representation of the relative energy distribution with respect to time and frequency. Here, the x-axis represents the duration. The duration of the utterance is about 560 msec. The y-axis represents the frequency. The frequency range is shown from 0 to 4000 Hz. Horizontal lines are drawn every 1000 Hz. The energy present in the signal around a chosen time and frequency, (t, f), is shown by a patch whose darkness (gray scale level) is proportional to the energy. For example, at the instant (i) corresponding to the burst of 'k', preceding vowel 'a', energy is distributed mainly between 1000 to 2000 Hz as shown by a dark patch enclosed by an oval and energy elsewhere (below 1000 Hz and above 2000 Hz) is very low as shown by light gray or white patches that are hardly visible. Around the instant marked (v), corresponding to the burst of 'k', preceding vowel 'i', the energy concentration is between 2500 to 3500 Hz as shown by a dark patch enclosed by an oval. Thus, the energy of the burst for the same sound 'k' differs depending on the context (following sound). This is due to co-articulation effect. The differences arise because in the production of 'k' in 'ka', the tongue moves backwards and downwards from the velar position whereas in the production of 'k' in 'ki', the tongue moves forwards from the velar position. Such differences in energy distribution of the ‘k’ burst don't affect the auditory perception of 'k'.

For an isolated steady vowel, with steady articulatory positions, one expects a steady pattern in the energy distribution across frequency as seen in a spectrogram. In the initial part of vowel 'a' between instants marked (ii) and (iii) there is a significant change in the pattern of energy distribution due to the movement of articulators from 'k' to 'a'. Also, there is a change in the pattern throughout the vowel 'i' between instants marked (vi) and (vii).

`

When only a part of the signal between (ii) and (iii) of vowel 'a' is played, a listener perceives the sound 'k' though this part has no burst within it. The transitions act as auditory cues about the place of articulation of the preceding stop sound. It is as if a listener introspects and infers what could have been the preceding sound based on the transitions in the perceived segment (i.e., between ii and iii).

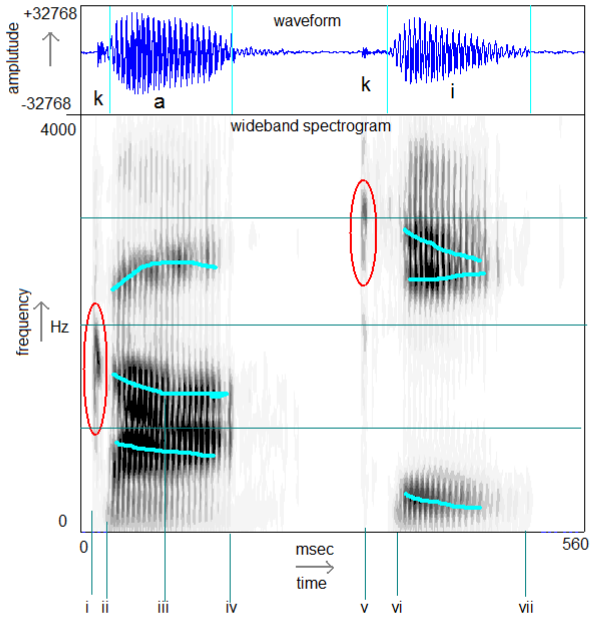

Fig.8. Waveform and wideband spectrogram of the utterance 'kaki' ('kaakey'). Note the differing distribution of energy of the two 'k' bursts (enclosed by an oval at instants i and v). Note the transitions in the energy concentrations with respect to time that begins with vowel onset and continues almost till the end of the vowel indicative of continuous movement of the articulators.

A wideband spectrogram emphasizes the effect of vocal tract (consequence of internal effort) relatively more than the effect of periodicity in the signal (consequence of external effort *'samvara'*). The use of narrow band spectrogram that emphasizes the effect of periodicity is illustrated in Sec.IV.H.

IV.F. Sound Quality of Nasals

Nasals are grouped along with stops since a nasal shares the same internal effort of *sprustam* and *sthana* as that of a homogenous stop (i.e., of same *varga*). Since for a nasal there is no sound output from the mouth why should the internal (within the mouth) effort has to be specified for a nasal?

`

For nasals, since the passage to nasal tract is kept open there is no pressure build-up and release (burst) as in the case of stops. Unlike stops, nasals are always voiced and can be uttered ('m', 'n') for a prolonged duration. Despite such a difference, nasals are grouped with the stops of the respective *'vargas'* based on the common internal effort.

Since there is no sound output from the mouth and since the geometry of nasal tract is fixed, one would expect the sound quality to be the same for all the five nasals. For nasals, although air comes out only via the nostrils, yet the mere branching of air into the closed oral cavity alters the sound quality. Geometry of the entire space occupied by the vibrating air particles (i.e., both the nasal tract and the branching closed oral cavity) determines the sound quality. For this reason, the internal effort for nasals is specified though there is no sound output from the mouth.

In this context another related topic may be mentioned here. Sound waves are movement of air particles and represent a weak form of mechanical energy. Mechanical coupling of sounds to chambers of chest and/or head also influences the overall sound quality. [5]. The indirectly coupled chambers of chest and head act as passive resonators. In fact, if a sensor is placed on the chest, the recorded signal is heard as speech, intelligible though distorted. These passive resonators need not be interpreted as registers of voice [Ref.1, Sec.1.10].

IV.G. Why Only Five Places of Articulation and Five Internal Efforts?

One may wonder why only five places (*kanta* or velar etc) have been chosen along the axis of vocal tract. Recall that the axis is the curved mid-line from the glottis to the lips along the vocal tract. The term 'place' doesn't strictly refer to a precise location along the axis. Rather, it is understood that the ‘place’ refers to a small region. Strictly, instead of 'place of articulation' it could have been referred to as 'region of articulation'. There are several reasons why only five 'places' of articulation are specified. Firstly, if place refers to a specific location then it demands an extreme motor skill on the part of the speaker to precisely position the articulators. That would slow down the rate of speech. Secondly, different speakers have different anatomical structures of the facial region, especially the length of the pharynx, lateral dimensions, arch of the mouth etc. For a specified place, the geometry of the Vocal Tract differs from speaker to speaker. If 'place' refers to a precise location along the axis, then due to varying geometry, the auditory quality for one and the same sound will differ for different speakers. Thirdly, during fluent speech, articulators don't abruptly jump from one set of positions to another set of positions. Rather, the articulators are continually moving from one set of targets to another set of targets and hence even within one sound the 'place of articulation' along the axis of Vocal Tract is changing. See discussion on

`

'Transition Interval' in Sec.IV.C.5 above. Fourthly, from the auditory point of view, as long as the narrowest gap is within a given region around the place of articulation, then the quality of sound is perceived to be the same. On the other hand when the narrowest gap shifts from one region to another then there is an abrupt shift in the perceived sound category. This has been established via experimental perceptual studies using synthesized syllables. This phenomenon is referred to as 'categorical perception'. This is especially true for the perception of stop consonants. Similar abrupt changes in the quality of a vowel sound when the place of articulation shifts across regions is referred to as 'quantal effect'. Hence, 'place' refers to a 'region' and not a precise location along the axis of vocal tract.

One may wonder why in the case of internal effort only five types of gap formations (*sprustam*, *isad-sprustam* etc) have been chosen for speech production? Why not more than five with finer gaps? This would demand a skilful control on the movement of an articulator on the part of a speaker. Small changes in the size and shape of a gap may produce only small changes in sound quality thus demanding extreme auditory skills on the part of a listener. Speech sounds should be as distinct as possible in order to facilitate unambiguous communication. The distinction amongst speech sounds is further enhanced because of the different types of sound energy (external effort) used in speech production such as voiced, voiceless, pressure build-up and release, *alpa-prana* and *maha-prana*.

IV.H. Measurement of Duration and Pitch of *Svaras*: An Example

There are divergent views, especially, as related to the tone pattern of *svarita svara* [Ref.1, Sec.3.24] though all views concur that *svarita* is not a steady tone. See also the review article on Vedic accent [13]. One view states that the first half of *svarita* is *udatta* and the remaining half there is a falling tone. In another view it is said that first half of *svarita* is at the level of *udatta* and the second half is at the level of *anudatta*. Yet another view states that the entire *svarita* consists of a continuously falling pattern. In another view it is said that the **first half *matra* of *svarita* is higher than *udatta*** (see Results below) and the rest is *anudatta*. Other descriptions for *svarita* are '*un-nicha*' (High-Low'), '*dvi-yama*' (two pitch values), 'udattaanudatta madhye' (pitch is between that of *udatta* and *anudatta*). Such divergent opinions on *svarita* arise since it is difficult to subjectively judge the relative durations and pitch levels of a dynamically changing pattern of a tone. By using modern day tools of experimental phonetics such ambiguities may be resolved. In this Section we illustrate one such example.

In a previous study [14] we have reported measurement of pitch of the three *svaras*. In that study, three chanters participated as subjects of which one of the chanters was a trained

`

professional who learnt chanting in a traditional school for about twelve years. We considered samples of vowels, 'aa, 'ee' and 'oo' uttered in isolation as well as in Vedic passages. In this Chapter, we take one sample from that study, viz., the recording of vowel 'aa' in isolation with accents in *anudatta*, *udatta* and *svarita* chanted by the professional.

The upper window of Fig.9 shows the waveform of the chosen sample. Any desired part of the displayed signal may be 'highlighted' by placing the cursor at the starting location, clicking the left button and dragging it horizontally till the ending location. The highlighted part of the signal appears in the reverse colour and its duration is displayed on the Screen in text format. Alternately, by noting the time locations of the cursor at the beginning and ending instants of an event, the duration of the event may be computed.

The lower window of Fig.9 shows the so called 'narrow band spectrogram' for the chosen sample. Narrowband spectrogram is a standard technical term used in the literature. A narrowband spectrogram emphasizes mainly the periodicity in the signal, a consequence of the external effort *samvara*. The y-axis represents the frequency. In this display, the spectrogram has been shown for the frequency range of 0 to 1000 Hz. At any given instant of time (i.e., along the y-axis), alternate dark and white bands are seen and the mid-part of the successive dark bands correspond to the fundamental frequency (F0) and its harmonics (integer multiples of F0). For illustration, short horizontal lines are shown on the right hand side at F0, 2F0, ...,6F0 where F0 corresponds to the value at 1630 msec.

Fig.10 shows a graph of the measured F0 Vs time. An advanced signal processing algorithm called 'autocorrelation method' has been used for measuring F0. The measured F0, scaled up five times is shown superposed on the narrow band spectrogram in Fig.9.

Result - *Svaras*: F0 around the mid-part of steady *anudatta*, *udatta* and *svarita* are 130, 145 and 160 Hz, respectively, as shown in Fig.10. There is a difference of 15 Hz between the two tones. Note that F0 of the steady part of *svarita* is higher than that of *udatta*. At the beginning of *anudatta* and *udatta* there is an abrupt rise and dip in F0 that arises due to the physiological adjustment. The transition from *udatta* to *svarita* shows a raising curve, reaches a peak (maximum) around 195 Hz and there after it falls and reaches a plateau. The maximum value of 195 Hz is greater than the level of steady parts of *udatta* (145 Hz) and *svarita* (160 Hz).

Svarita is known as a mixed-tone or High-Low tone and not three level tone of Low-High-Low as seen in the Figures. Thus *svarita* corresponds to the part between the instants (iv) to (vi). The rising part transition between the end of *udatta* and the beginning of *svarita* might have arisen due to the requirements of physiological adjustment.

`

The value of F0 of steady part of *svarita* is higher than that of *udatta* and *anudatta* and the peak value of F0 in *svarita* is the highest of all tones.

Result - Duration: The measured duration for *anudatta* (between i and ii) and *udatta* (between ii and iii) is 870 msec. In this experimental data, the three isolated 'a's are all prolonged (*'pluta'*). One-third of these values may be taken as representative of the duration of a *hrasva* vowel. Duration of a *hrasva* vowel corresponds to one *matra kala*. Thus a representative value of one *matra kala* is about 290 msec as per this experiment.

There are two parts of *svarita*: Up-down movement and a steady part. The total duration of up-down movement is 290 msec, i.e., the above estimated value of one *matra kala*. If we consider *svarita* to begin with the maximum (peak) value then the duration from the maximum (peak) to reach plateau is one-half (*ardha*) *matra kala* as mentioned in one of the ancient texts. The duration of the steady part of *svarita*, i.e., from (v) to (vi) is about 730 ms. If we measure the duration of *svarita* between the maximum (peak) to the end, i.e., between instants (iv) to (vi), the duration comes out to be 875 msec, which matches within about 5 msec of the duration of *audatta* and *udatta*. This further confirms that *svarita* begins from the maximum (peak) value of F0 and *svarita* is the mixed tone or High-Low tone.

`

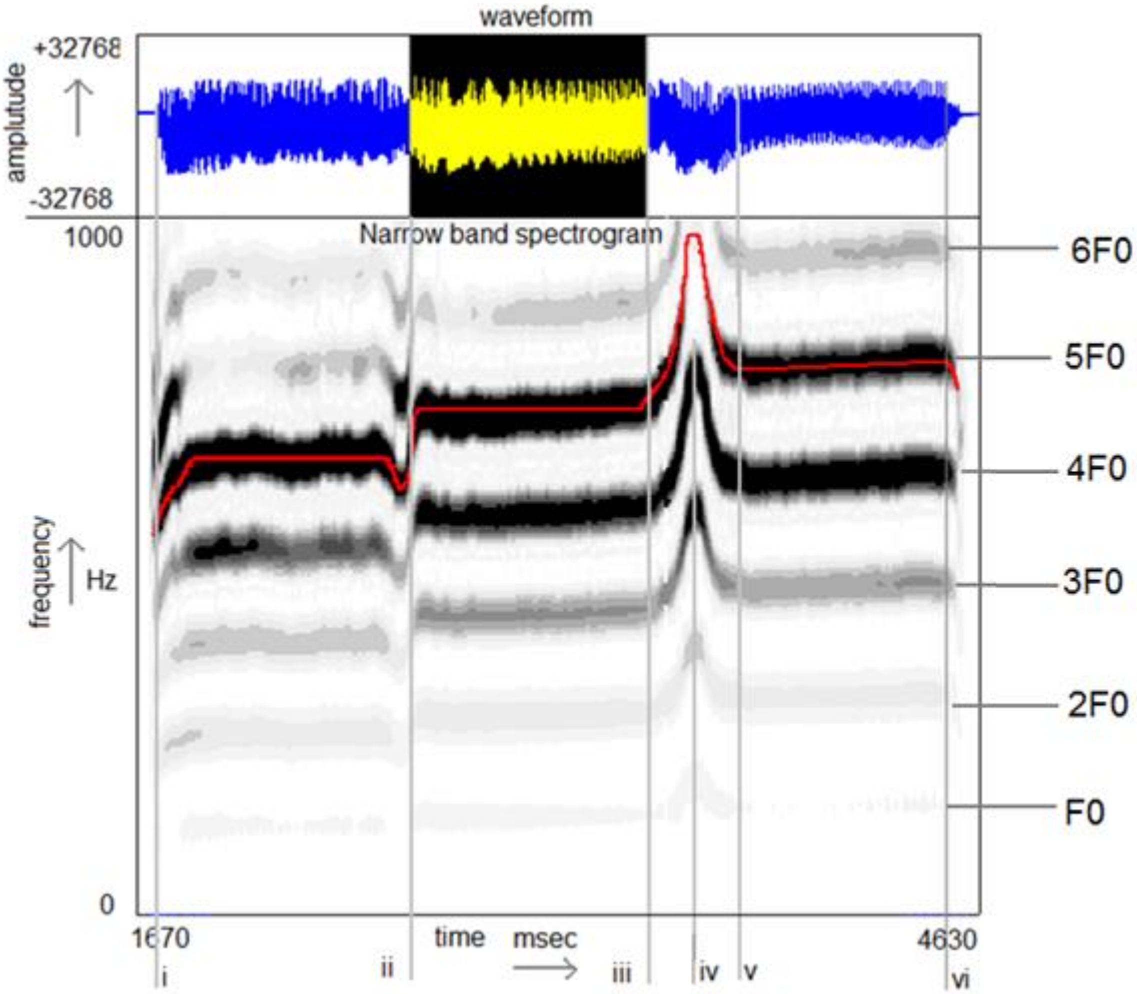

Fig.9. Waveform (upper window) and narrowband spectrogram of prolonged vowel 'a' chanted by a professional in anudatta (i to ii), udatta (ii to iii) and svarita (iv to vi) svaras.

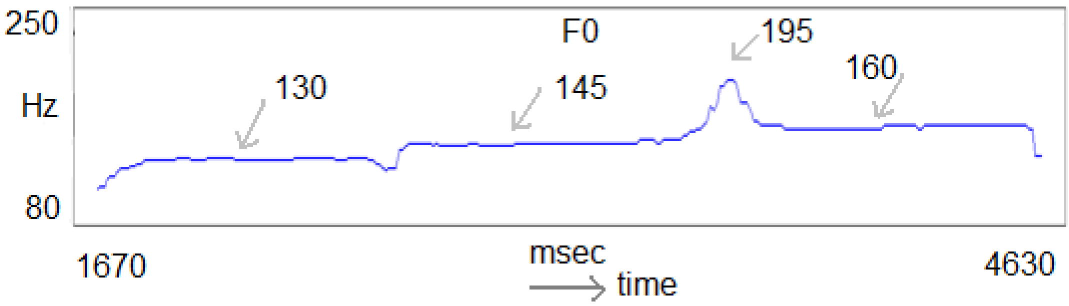

Fig.10. Measured fundamental frequency, F0 Vs time for the same chant as used for preparing Fig.9.

Comments: The results presented in this Section on F0 for the three *svaras* and the estimated duration of *matra kala* are applicable only to this specific sample and should not be generalized. In actual practice, such measurements have to be made on a large number of passages and as chanted by a large number of subjects. Statistics on the means and

`

standard deviations have to be computed including intra and inter subject variability. Nevertheless, the results presented here may be considered as representative values. This example illustrates the power of the tools available in experimental phonetics in resolving differences of opinion on certain concepts in SS.

IV Conclusion

A broad review on the pronunciation of speech sounds as given in SS has been presented along with the contemporary knowledge of the acoustic theory of speech production. We could appreciate the knowledge given in SS on the structure and function of lungs, larynx and vocal organs in view of the latest findings. Although a detailed description of the internal and external efforts have been mentioned in SS, details on the mechanism of production of voicing, fricative noise, stop bursts and dynamics of speech events, transitions, coarticulation etc seem to be missing.

The external effort of *samvara* (closed) for voiced sounds is said to be at the glottis. This should NOT be interpreted literally. A speaker would choke if the glottis gets closed while speaking. Closure of the glottis is only a pre-condition for voicing. Based on the latest findings, the mechanism of generation of periodic vibrations of vocal folds, viz., phonation for voiced sounds, has been explained and illustrated.

In SS, the **external** effort *vivara* (open) for voiceless fricatives and voiceless stops is said to be at the glottis (external to mouth cavity). However, based on recent knowledge it has been argued that the location of the source for voiceless sounds is at the *sthana* (**internal to mouth cavity**) except for the voiceless sound 'h'.

In SS, the internal effort for stop sounds is said to be '*sprustam'* (contact). In actuality, there is no sound output during the contact. Based on recent findings, sound is produced only when the contact is released. A detailed description on the mechanism of the production of stops has been presented. Various sub-intervals, effect of coarticulation and articulatory dynamics have been illustrated using visualization tools of experimental phonetics.

An illustrative example on the measurement of tone heights (*svaras*) and estimation of *matra kala* has been presented. It is shown that diverse opinions expressed in the ancient texts may be resolved using the latest tools.

Perhaps *vivara-samvara* distinction in the external effort and the various internal efforts such as *sprustam* etc given in SS might have been proposed only for the sake of categorizing of the sounds and were not meant to explain the mechanism of production of speech sounds. By the light of latest knowledge and visualization tools we now understand that these

`

concepts in SS should not be interpreted literally. Similarly, the statements in all our *Shastras* shouldn't be interpreted literally but these have to be examined critically and interpreted properly using the latest knowledge and tools, where applicable.

Almost all the statements given in SS may be verified using the modern technological tools of experimental phonetics. Verifying the pronunciation of Sanskrit sounds as in Chanting and spoken Sanskrit may demand cine-radiography, MRI, Palatography etc. This is relatively more involved. However, measurements on duration of segments and pitch of *svaras* may be done relatively easily. At present, Vedas are being chanted at various tempos. Standards may be established on the duration of *matra kala* and thereby the proper tempo to be used for chanting can be suggested. The reference pitch level to be used for *svaras* of the Rik, Yajur and Sama Vedas has to be established. Other aspects in Vedic Chanting, such as *deerga svarita*, have to be understood. Analysis of Sama Veda chanting and its relation to the evolution of modern musical scale is yet another interesting topic. Such studies help in the preservation of the heritage of Sanskrit in a scientific manner.

Sanskrit *varnas* and their groupings as in Shiva Sutras are related to Yoga and mantra *shastra*. These topics, though very interesting, are beyond the scope of the present work.

An inter-disciplinary workshop has to be called to expose researchers in the area of speech science on the nuances of SS and at the same time expose Sanskrit scholars to the latest tools available in experimental phonetics so that these tools may be used to teach/learn Sanskrit in a scientific manner.

The author has developed several software tools like Speech Science Lab (SSL) for speech analysis, synthesis and visualization, Vagmi Voice and Speech Therapy, Vagmi Nada for learning musical notes etc. Readers may visit www.vagmionline.com for more details about these software tools. Using Speech Science Lab, chanting by well trained professionals may be analyzed to measure important parameters. Vagmi therapy software may be used to give a visual feedback in order to learn the correct pronunciation of Sanskrit speech sounds. For example, the word *'shanti'* is often is pronounced as *'santi'*. Pronunciation module of Vagmi therapy may be used to learn to make the proper distinction between 's' and 'sh'. In one of the software tools called Vagmi-Nada, a student can learn to sing the notes at the correct pitch (*svara shuddhi*) by using a visual feedback. This tool meant for learning Carnatic music may be extended to develop a tool for learning Vedic chanting.

References

1. Allen W. S., *Phonetics in Ancient India*, Oxford University Press, London 1953. (Available as an open source)

`

2. Subrahmanya Sastri P. S. , *Lectures on Patanjali's Mahabhasya*, Vol.1, Kuppuswami Sastri Institute.

3. Ballantyne J. R., *Laghu Kaumudi, A Sanskrit Grammar by Varadaraja, English Version* . (Available as an open source)

4. Sowmya Krishnapur, *Laghu Kaumudi*, YouTube, Lectures 2.1 to 2.3

5. Fant G., *Acoustic Theory of Speech Production*, 1960, Mouton Hague

6. Stevens K. N. Acoustic Phonetics, MIT Press, Cambridge, USA, 2000

7. Ladefoged P and Johnson K, *A-Course-in-Phonetics*, Wadsworth, 2010 (Available as an open source)

8. Jakobson R., Fant C. G. M. and Halle M., *Preliminaries to speech analysis: The distinctive features and their correlates*, MIT Press, Cambridge, USA, 1967

9. Chomsky N. and Halle M, *The sound pattern of English*, Harper and Row, NY, 1968. (Available as an open source)

10. Perkins W. H. and Kent R. D., Functional Anatomy of Speech, Language and Hearing, Allyn and Bacon, 1986, Reprint 1991.

11. Elena Rizzo Riera. R1 ORL HUSE, Larynx Anatomy, URL: https://sborl.es/wp-content/uploads/2018/01/Larynx-anatomy.pdf. Last visited 2nd Nov 2022.

12. Diehl C. F., Introduction to the Anatomy and Physiology of the Speech Mechanisms, Springer Ill., C. C. Thomas, 1968.

13. Varma Siddheshwar "*The Vedic accent and the interpreters of Panini*', J. of Bombay Branch of Royal Asiatic Society, vol. 26. 1950. . (Available as an open source)

14. Ananthapadmanabha T. V., Silverman K., and Prasad M. G., “Vedic chanting and vowel intrinsic pitch: Evidence from an ancient source,” Ac. Soc. Am., 117th Meeting, Syracuse. May 1989.

`

APPENDIX-A

Concept of Sound at the Physical Level

At the physical level, sound is a form of energy, called acoustic energy. Sound propagates in the atmosphere as waves. Concept of a 'wave' is central to an understanding about quality of a sound. When a pebble is thrown into a pond, waves appear on the surface of the pond. When a vessel containing water is tapped on the outer surface, ripples can be observed on the surface of water. Similarly, a sound produces invisible spreading waves or ripples in the air. For example, when a string of a Tanpura (Tambura) is plucked, the up and down movement of the string moves the surrounding air thus setting up waves of sound. When the surface of a drum (Mrudangam) is struck, the movement of the surface of the drum moves the surrounding air. Just as mass of water on the surface of a pond carries a wave, air in the atmosphere carry sound waves.

Amplitude, Wavelength, Period and Frequency of a wave: When a pebble is thrown into the middle a pond, waves are seen on the surface of water. A **snap shot** (**still image**) shows waves as made up of crests and troughs that are spread out on the surface of water. Usually, the height to which water rises at a crest or the depth to which water sinks at a trough decreases far away from the centre of disturbance. The height of the crest or the depth of a trough is referred to as the 'peak amplitude of the wave'. The distance between two consecutive crests is called the 'wave-length of the wave'. The unit of the wavelength is the same as that of a distance measurement; example: centi-meter.

If a **video is taken at a fixed location** on the water surface, one sees that the water rises and falls cyclically in the course of time. In order to see this clearly one can put a piece of floating object like puffed rice or sponge. Assuming highly regular cyclic up and down movement, the time interval between two instants when the wave reaches the top-most point is called the 'time period of the waves'. The unit of time period can be expressed in second or milli-second (1/1000 th of a second). Alternately, the number of times the surface of the water reaches the top-most point (crest) in one second interval can be counted. The number of cycles per second at a given location on the surface of the water is referred to as the 'frequency of the wave'. Counting need not be made exactly for one second though the unit is stated as number of cycles per second. For example, to get the frequency, the number of counted cycles has to be multiplied by 10 (or divided by 3) if counting is done for 1/10th of a second (or 3 seconds). The unit of frequency, 'cycles per second' is presently referred to as Hertz (abbreviated as Hz). The frequency of a wave is the reciprocal of the time period. The frequency of a wave and the wavelength are also related to each other via the velocity of the propagation of sound waves.

`

Spectrum: When the disturbance is highly regular in time then the crests and troughs are regularly located in space with the same heights and depths and the wave is said to consist of only one frequency component. An irregular disturbance (say, a splash) consists of several waves of differing frequencies and differing peak amplitudes. Given an irregular disturbance, it is possible to derive mathematically the number of distinct waves present and their respective frequencies and amplitudes (or energy). Such a mathematical procedure (Fourier analysis) is called 'Spectral analysis'. Different speech sounds differ in terms of their spectra (plural form of spectrum). This is the basis for perception of different speech sounds. Today's instrumentation provides a means of visualizing the spectrum of any given sound. In a graphic representation of a spectrum, the x-axis represents the frequency and the y-axis represents the energy (E). Since the relative energy associated with a very feeble sound to a very loud sound differ by a very large proportion, energy is expressed in a compressed scale, called deci-Bell (dB), which is 10log(E). A spectrogram is a 3-Dimensional graphic representation depicting changes in the spectrum with respect to frequency. Here, the x-axis represents the time, the y-axis represents the frequency and the gray scale (relative darkness) represents the energy in dB at any given time-frequency.

Audible sounds: Heat, light, radio waves, x-rays etc are all of the same form of energy, called electro-magnetic (EM) energy. Only a certain band of frequencies of EM waves are visible to the human eyes. Similarly, acoustic waves of a certain band of frequencies (20 to 20000 Hz) are audible as sound. Acoustic waves of very high frequencies are used in ultra-sound scanning. Acoustic waves of very low frequencies are not audible.

Instrumentation: In principle, sound can be transformed into an electrical signal by a device called a microphone (see below). An electrical signal can be recorded and saved in various ways. In one of the methods, called digital recording, an electrical signal can be transferred to computer memory. Using software tools a digital signal stored in computer memory can be displayed or processed. Thus it is possible to visualize a sound or its spectrum. Some illustrative figures are presented within the main text and are interpreted.

Principle of a microphone: A vibrating string sets up waves in the air. Conversely, when a singer sings a note of correct frequency close to a Tanpura (Tambura), one of the strings vibrates (sympathetic vibrations). This shows that vibrating air particles associated with a sound carry mechanical energy that can move a string or a thin membrane. This is the basic principle of a microphone. In a microphone, the movement of a thin membrane in a magnetic field generates electrical current. Instant to instant changes in the value of the current of the electrical signal are proportional to changes in the air pressure caused by a sound falling on the diaphragm. There are various types of microphones.

`

About the author:

Education: **T. V. Ananthapadmanabha** obtained his B.E. degree in Electrical Engineering from the Bangalore University in 1970 and B.E., M.E. and Ph.D. degrees in Electrical Communication Engineering from the Indian Institute of Science, Bangalore in the years, 1972, 1974 and 1978 respectively.

Career: **Research & Teaching**: He has worked at world renowned labs such as **CMU, Pittsburgh; Royal Institute of Technology, Stockholm; AT&T Bell Labs, Murray Hill; MIT, Cambridge**. He worked as a **professor and dean of research at MSRSAS**, Bangalore during 2003-08. He has research collaborations with the electrical engineering department of the Indian Institute of Science, Bangalore and the All India Institute of Speech and Hearing, Mysore.

As an entrepreneur he has developed pedagogical and clinical software products in the area of voice and speech which are widely used. These products have been recognized by the Rehabilitation Council of India and finds mention in the 108^{th} Rajya Sabha proceedings of the Government of India. Vagmi software was nominated for the President’s award.

Publications: He has published a large number of international standard papers noted for their originality. He convened a three day workshop on '*Shiksha Shastra* and Experimental Phonetics' on behalf of the Academy of Sanskrit Research, Melkote. As a part of TIFAC activity at MSRSAS, he convened a National Conference on Image Processing with medical applications (NCIP-2005).

Notable Awards: He is a recipient of **Alumni award of IISc** for the year 1974 and the **Sir CV Raman award** of the Acoustical Society of India for the year 1992 for best research publication.

Affiliations: He was the **founder chair of IEEE EMBS** Bangalore chapter during **2004-08**. He is a member of several professional bodies; Acoustical Society of America (1988-1996), International Phonetics Society, Acoustical Society of India, Indian Speech & Hearing Association, Dravidian Linguistics Association, Association of the Phono-surgeons of India.

Allied interests: He is a disciple of Sri Sriranga Mahaguru (www.ayvm.in). He has a keen interest in understanding the contribution of the ancient Indian (Sanatna Bharateeya) Rishis. His allied interests are in the areas of yoga and philosophy.

He has authored two general books ‘Inner Workings During Yoga Practice’ and ‘Geometry of Srichakra’.

Free download: Software for *Sound-to-Form Transformation (Vagmi Tonoscope)*, research note on the *Acoustic Analysis of Pranava*, *Second Edition of Geometry of Srichakra*, Articles: *Reality of Self and Self as Reality*, Article: *A critical review of instrumentation to record nadi* etc are available for free download from the website www.vagionline.com (visit About Us - > TVA ->Allied interests).