Vector Phase Space for Linguistics via Dimensional Analysis

H.M. Hubey, Associate Professor
Department of Mathematics, Computer Science and Physics
Montclair State University
Upper Montclair, NJ 07043

hubey@pegasus.montclair.edu
http://www.csam.montclair.edu/~hubey

Abstract

A vector space using dimensional analysis is produced in which one can show all the phonemes/phones of all languages. Vowels, and consonants can all be shown in this phase space. Furthermore the three-dimensional vector space for vowels, which in simplified form can be shown to be related to the distinctive features can also be compressed to fit in this pahase space for speech. This phase space can be shown to be both based on articulatory/geometric considerations, that is the two-tube model of Fant, and Stevens, and also on the quality/perception arguments based on formant studies, Peterson & Barney, and Clark & Yallop. It can be used to clarify and unify many linguistic phenomena such as child language [Anderson, Jacobson], aphasia, sonority, the cardinal vowel diagram[Jones, Ladefoged], diphthong trajectories [Carre & Mrayati]. It is shown that the sonority scale is directly correlated with this space in that sonority is related to the distance of the phones/phonemes from the origin hence sonority is a function of the magnitudes of the vectors (phonemes/phones) of this space. Dipthong and vowel confusion that crops up when using Artificial Neural Networks [Kohonen] for vowel recognition is easily explicable in this space. The fortition-lenition phenomena and phonological strenths [Foley] are nothing but vector phenomena in this space. The reasons that almost all languages have the phonemes /ptskn/ can be clearly shown in this space as splitting up the available phonological phase volume into nearly equidistant volumes. It is shown that this space provides the ideal space for the discussion of such seemingly disparate phenomena as assimilation, metathesis, haplology, and dissimilation. In short this phase space is the natural phase space for speech.

Introduction: Properties of Consonants and Vowels

We would like to be able to extend the concept of continuous orthogonal vector spaces to consonants or contoids. However, the formants for consonants don't exist, almost by definition. They will certainly not exist in the sense of the formants of vowels. In fact, this can be corroborated easily [see Edwards,1992]. Of course, we can use the distinctive features spaces as derived in earlier sections. However the dimensionality is too high. We'd like to be able to generate a broad transcription space to describe the consonantal sound in a similar way to vowels. This practically limits our dimensions to two or three. One of the most obvious characteristic of consonants (in contrast to vowels) is that the articulatory organs move in time whereas vowels are *steady-state* constructs. This property of consonants is shared by the semivowels, diphthongs and glides. However, at a very broad level we can also imagine a class of consonants that share another property with the vowels. Certain consonant classes, in particular the nasals such as {/m/,/n/}, the liquids such as {/r/, /l/} (referred to as glides by some), the voiced fricatives such as {/v/,/z/} are **ı**-colored -- the Turkish-**ı** is used for schwa-like phones/phonemes; it's written bold to denote a vector-- and maybe called *quasiconsonants*. Since three formants are more than sufficient to approximate the qualities of the vowels, and since this requires at most a three-dimensional vector space, and since we can easily generate eight vowels as the corners of a parallelopiped in this space and which we can approximate using distinctive features (see Hubey[1994,1996]) the eight Turkish vowels which denote an almost perfect match for this space are used as vectors, (see also Hubey, 1996b]. The property that the quasiconsonants share with the vowels is they are also *steady-state* sounds in that the articulatory organs do not move in time-space, although the DOF (degrees of freedom) of contoids such as /l/,/z/, is zero whereas the nasals and /v/ leave the tongue free to move about. That these consonants seem to have vowel like qualities can be seen in their power spectra (see for example, Edwards [1992]). There seem to be high-peaks at a low frequency with an exponentially decaying amplitude which is what we'd expect from an **ı**-colored vowel or consonant; that is, the spectra resemble an **ı** with enough noise (low signal-to-noise ratio) to bury the signal. Of all the vowels the most neutral, and in a way the most *well-behaved* vowel as can be seen on the power spectrum is the **ı** [Edwards, 1992]. Its formants' amplitudes drop off exponentially as one would expect in the absence of a filter. The quasiconsonants all seem to show some evidence of this. In addition, the fricatives are also steady-state sounds, however their sound quality does not show any evidence of vowel like quality. The plosive groups (especially the unvoiced) would best be modeled in the time-domain as Dirac delta functions. Of course, this implies that the power spectra would contain energy at all frequencies and thus would be flat. The voiced plosives are differentiated from the unvoiced essentially by the magnitude of the difference in time between the voicing and the plosion so that they would also be in the consonantal group. Thus from the basic division along the vocoidal-contoidal continuum we can derive a symmetric four-way division; vowel (V), semivowel (S), consonant (C), and quasiconsonant (Q).

Towards a Space

We know that one of the fundamental determiners of the quality of a speech-sound, is the location of the primary constriction. It thus seems that we already have two dimensions in which to represent the consonants. If we denote by S the size of the stricture (i.e. the size of the primary constriction) for the consonant then one of our dimensions would be $Y=|\partial S/\partial t|$ where the vertical bars indicate the absolute value of the derivative. It will be later shown that it's not necessary for the derivative to be partial; indeed it might be more useful otherwise. We can use the location of S as another dimension, say X, essentially a mapping starting from the lips (for the bilabial consonants) and extending back towards the soft palate and pharynx. Although we consider this to be a single dimension extending in curvilinear fashion from the lips toward the pharynx (and maybe even beyond) it will be shown later that X should really have more (physical) dimensions.

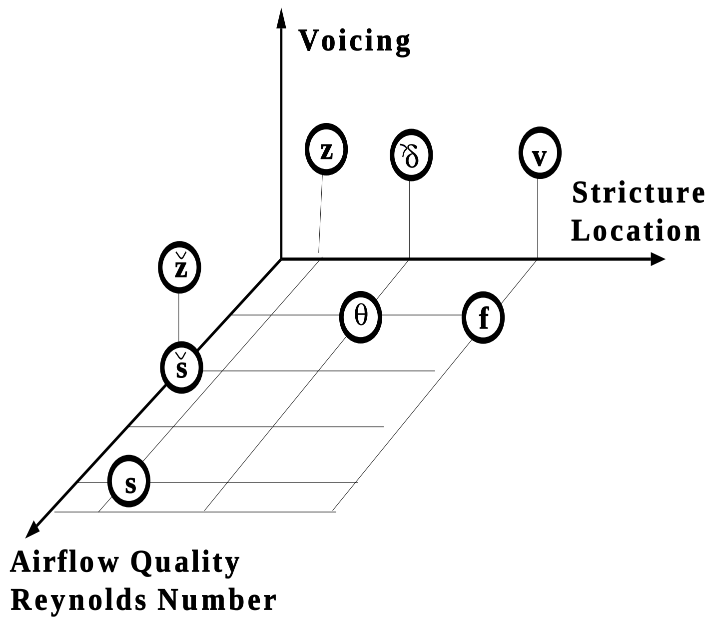


Figure 1: An Intuitive/Suggestive Space

The third dimension for the consonant 3-D vector space (Z dimension) would have something to do with *sound quality* or *airflow quality*. It's essentially the dimension that would distinguish the turbulent-chaotic quality of fricatives and sibilants from the more vowel-like quality of the quasi-consonants. The particular boundary between these sounds is not clearly delineated since some phonemes such as /v/ and especially /ž/ very clearly show evidence of both turbulent or chaotic flow (frication) and laminarity (resonation or periodicity). But it is known that sounds (pressure waves) can scatter from turbulence and there are various methods of extracting signals scattered from turbulence. Various names such as periodic, resonant, fricative, sibilant , liquid, continuant, obstruent dot the linguistic landscape and it is not the intent of this paper to create even more terms; thus the usage here is intended to make a connection between the linguistic terms and the 3-D space being constructed and its dimensions. The example above (Fig 1) shows the relationships of some phonemes but is not drawn to scale. The third dimension (voicing) can be included in the discussion above, if we include the larynx as part of the geometric parameters or the articulatory organs. Therefore we can use some kind of a weighted average of the rates of changes (i.e. time derivatives) of the articulatory apparatus (including the rate of change of the primary stricture) to put the voicing dimension along with Y.

Consonant Vector Space and Dimensional Analysis

Figure 2 shows some common consonants plotted in the X, Y, Z dimensions (i.e. essentially Primary Place of Stricture, Rate of Change of the Articulators, and Quality of Airflow). The drawing is not to scale. It's been distorted to give a general idea of the positions of some of the common consonants. It will be shown in the next section that this arrangement is not fortuitous but obeys very fundamental laws of physics. Real world phenomena take place in *space-time*. These are the *fundamental dimensions;* that is, it takes three space coordinates (dimensions) and one time dimension to describe mechanics events. However in dimensional analysis, the space dimensions are collapsed into one dimension, usually denoted by L. Thus the dimensions of area are L^2 and

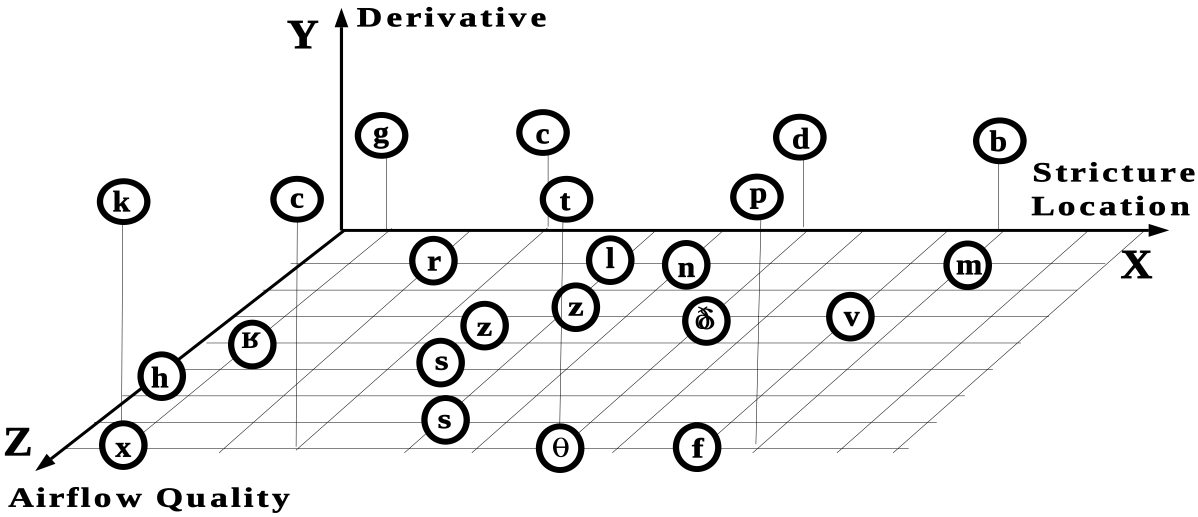


Figure 2 First Attempt at a Three-Dimensional Space

L^3 describes three-dimensional space. The time dimension, of course is denoted by T. However another level of abstraction is needed to describe the fundamental processes of physics. For mechanics, we need one more; usually Force F, or Mass M. They are not independent since they're related by Newton's formula F=ma. For electromagnetic phenomena we need another called Charge and for thermodynamics, Temperature. Speech is a mechanical phenomena and needs three dimensions F, L, and T. *Dimensional analysis* is often used with fluid mechanics [see for example White,1979] where the processes are too difficult to describe simply because they involve many variables and are highly nonlinear. Dimensional analysis has helped physicists to look for groups of variables for which equations or relationships should be sought. Dimensional analysis was first proposed and used by Buckingham and the method used to find the dimensionless groups is called the *Buckingham Pi Theorem.*

If we examine the dimensions of the coordinates X, Y, and Z, we'll see that we're very close to what dimensional analysis would have yielded. The Z coordinate which we called airflow quality corresponds to the dimensionless group in fluid dynamics which discriminates essentially between laminar and turbulent flow. It's called Reynold's number and is given by Kv/υ, where K [L] is the characteristic length; v [LT^{-1}], velocity and υ [L^2T^{-1}], kinematic viscosity. The dimensions of the variables are given inside the square brackets. No specific references were made to the kinematic viscosity because speech only takes place in air and not in other material. The Y coordinate is the time derivative of an area (stricture); thus its dimension is L^2T^{-1}. The third coordinate X is essentially the place of the stricture in one dimension, however it was noted that it would have been better for it to have the dimension L^2. Thus we can take the height dimension into account by modifying the horizontal place coordinate by using the horizontal coordinate multiplicatively, in ways similar to done by Peterson & Barney [1952]. All we have to do now is to multiply by a *characteristic frequency* ω and X will also become L^2T^{-1}. Indeed in the previous section on formants ω $[T^{-1}]$ (radians/second) was used instead of f [Hz]. It is clear now why the correction via a function of ω (having the dimension T^{-1}) is necessary. It is common knowledge that in speech studies it has been found necessary to account for the differences in pitch of various speakers. The method used, the so called *vocal tract length normalization* performs this function, since the smaller vocal tracts result in higher fundamental frequencies for the speaker. Thus there's a functional relationship between the length of the vocal tract and the *characteristic frequency* function that is proposed here. Hence the dimensions of X are L^2T^{-1}.

We can now use this knowledge to redefine the coordinates of the *consonant vector space* to also include vowels and semivowels. It was noted that the Y coordinate need not be only a partial derivative. Indeed the partial derivative will not be able to account fully for even the consonants and especially for certain consonants such as /ç/ since the motion of the articulators (the tongue in this case) is more complicated than what the partial derivative indicates. We should write the stricture function S not as a function of time only but as S=S(X(t),t). The time derivative then is:

(1) $$dS/dt = \partial S/\partial t + (\partial S/\partial X)(dx/dt)$$

Furthermore, we can extend the definition of Y by defining it as a weighted average of the sums of the total derivatives of the strictures involved in the articulation, not only the primary stricture. This means that we are now accounting for the change in the vocal cords also, therefore the distinction in this phase space between the consonants and vowels can be made. Since the average of a sinusoidal wave is zero, the average that is used for the vocal cords should probably be another type, say a root-mean square (rms). All we have to do now is to divide both the X and Y groups of variables by the kinematic viscosity υ (as in the Reynolds number, see, for example, White[1979]) and we'll have a three dimensional vector space consisting of three dimensionless quantities; *Reynolds Number*, and the two dimensionless groups or numbers that we just derived. If we denote the horizontal length as λ and the vertical dimension as η; the strictures as S_1 and S_2, (using the two-tube model) then we have;

(2) $$X = \lambda\eta\omega/\upsilon \qquad \{L^2T^{-1}\}/\{L^2T^{-1}\}$$

which is dimensionless Y coordinate for two strictures could then be

(3) $$Y = \omega \left| (d/dt\,\{S_1(\lambda,\eta) + S_2\}) \right| /\upsilon$$

which is also dimensionless. The stricture S_1 has been written as a function of λ and η, and this is the most general form and correctly describes the procedure, however in practice simpler forms may be used. Since all three coordinates are divided by the kinematic viscosity, in practice it might be eliminated so that the coordinates will be given by the (effectively) dimensionless numbers:

(4) $X = \lambda\eta\omega$; $Y = \omega\,[\,|dS_1/dt| + |dS_2/dt|\,]/$; and $Z = Kv$

It should be noted here that the K (in Z), the characteristic length is also a good candidate for some kind of vocal tract normalization or it may be used as a characteristic of stricture place. These numbers are only suggestive and improvements can be made. For example, for the bilabial plosives the rate of change of the constriction can be positive, negative or both. Thus if we were to use only the absolute value, we'd have a reasonable approximation. On the other hand, if the acceleration is not constant, we'd need to use an average. Since the opening and closing are of different signs the average would be zero unless we used a root-mean-square kind of an average. As for some of the specifics on frication and more easily measurable physical parameters such as pressure; Stevens, along with others, has done research on relationship of frication to pressure drop across a constriction [which can be found in Lieberman[1988]]. Chomsky & Halle [1968] mention that /ptskn/ are rarely absent in any language. It's clear now why even at this level of accuracy as in Fig. 3 They're marked below with arrows. These five contoidal phonemes essentially define the volume of the consonant space, (as will be even clearer in Figure 5) and they roughly divide the phase space into equal intervals/volumes which has implications for distinguishability since the relationship between the articulator positions (and manners of articulation) and the quality of sound (i.e. their perception) must be a deterministic one (although it is a highly nonlinear one). The dotted lines denote the volume in which the vocoids; the vowels, semivowels and polypthongs fall. The little circles are meant to be representative manifestations or instantiations of the sets that compose phonenemes hence neither the phoneme nor the phone symbols are used. In truth they are neither since the results must be generalizable to all languages so that they are representatives of some sounds which can be recognized to belong to natural clusters, and which may be split up differently in different languages, although we are using (American) English as a vehicle for explication of the ideas. It should be understood that the symbols really denote volumes in this space and that their boundaries can be considered to be fuzzy, as in fuzzy sets [Hubey,1994].

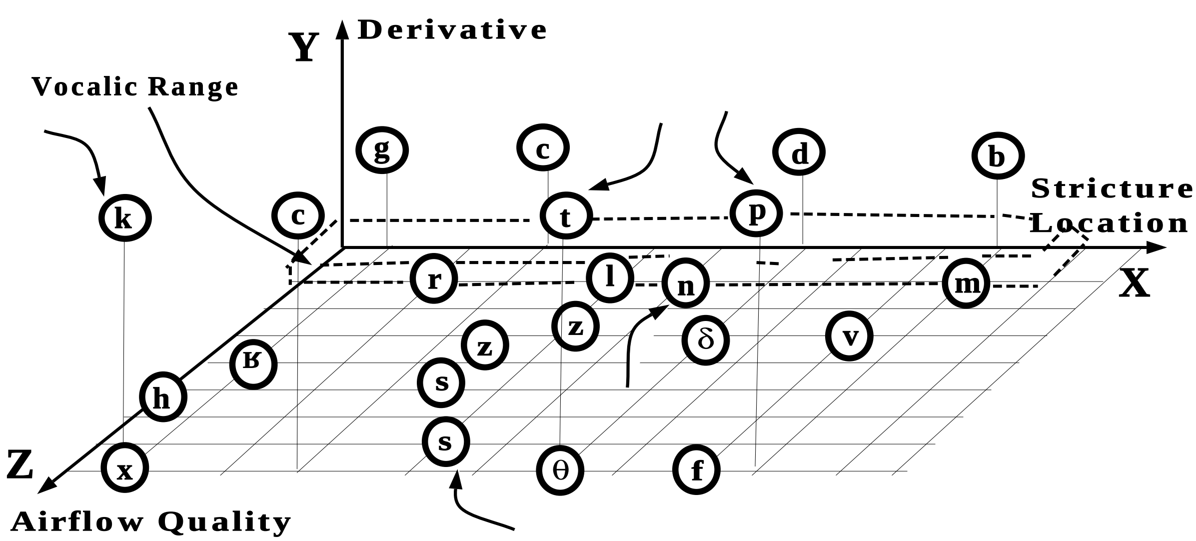


Figure 3 A Slightly Better Placement of the Consonants in Phase Space

Natural Groupings

The groupings in Figure 4 are two dimensional versions (orthogonal projections) of Figure 3, where the phonemes/phones close to one another have been grouped according to their characteristics. The figures give an indication of the way the phonemes/phones cluster in the *speech phase space*. It shows very clearly that these divisions are those that have been described in various ways by phoneticians and linguists for centuries. It’s not clear yet where /ž/, /š/ and /x/ really belong. In fact judging from sound quality the /s/ doesn’t seem to fit into its group either, i.e. with /f/ and /θ/. The final arbiter of the placements of the phones/phonemes has to be the results from acoustic measurements; for example if the major frequency peak is lowered from about 5 KHz to around 2.5 KHz, the listeners’ perception shifts from an /s/ to a /š/. More about the fricatives can be found in Lieberman [1988, p.227]. The grouping in the figure above is intended to show some natural clustering in the *phase space*.

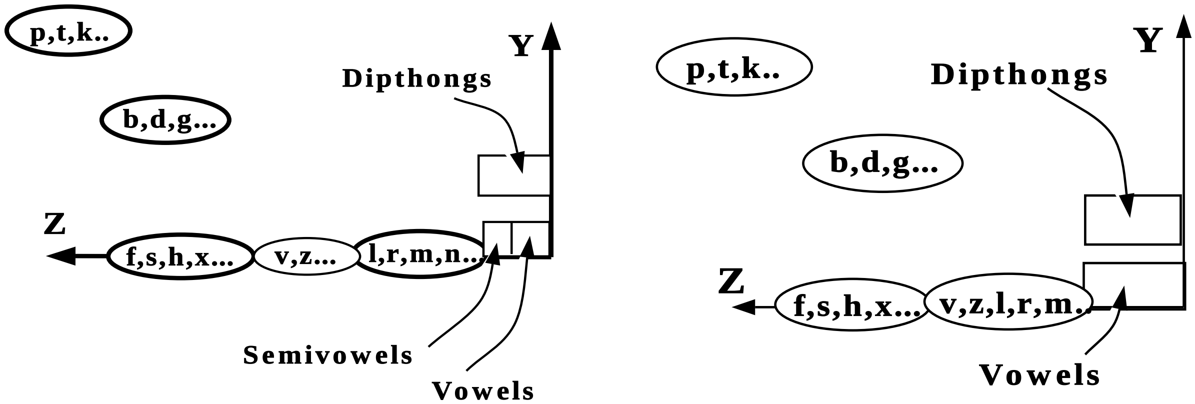


Figure 4 Some Possible Groupings

The last phase space (Fig 3) was derived from the original discussion on consonant spaces. However after the discussion on dimensional analysis, the original dimensions or parameters were altered to take into account the various changes in Y, namely that it is the sum of the derivatives of the strictures and that Z axis has to do with Reynolds number. Now, the semivowels can be considered to be appended to the end (or beginning) of vowels with which to form diphthongs or glides, thus there is a motion of the articulators so that their Y-values are not zero. The vowels are steady-state, therefore they should be solely on the XZ-plane extending very close to and partially mixed with the quasiconsonants since they also display some vowel characteristics (such as being **ɹ**-colored, and being steady-state (continuant)). The diphthongs are defined on the ZX plane in the same area as vowels, except that their Y-values are not zero, thus they will be located above the vowel range. The broken-line boxes indicate the vocalic sounds (vowels, semivowels, glides, diphthongs, and triphthongs). Finally the voiced plosives should now be moved from the Z=0 range since we know that they contain both voicing (vowel-like sounds) and turbulence (the high frequencies that exist in short duration spikes, which can be modeled as Dirac delta functions in

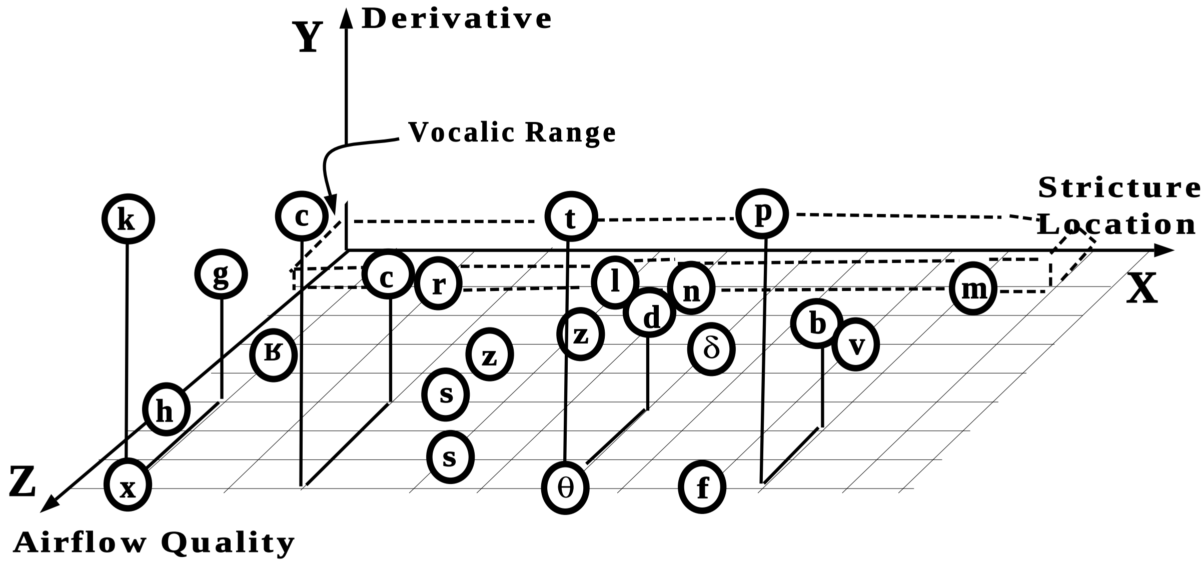


Figure 5 Phase Space for Speech Sounds

time). They should be moved somewhere between the vocalic and the consonantal sounds. All the changes are shown in Figures 5 and 6. It is even clearer now why /ptksn/ are rarely absent in languages. The /p/ is the extreme X (except for /w/); /k/ is the practical extreme for X (minimum) and /n/ defines the minimum Z. Any smaller value in the Z direction than /n/ would fall in the vocalic group. In the same chapter Chomsky & Halle also remark that the **ı** (*full schwa*) should be marked and should get a complexity of 2 along with the compounds like the **æ**. It would seem that the **ı** is the most ubiquitous vowel especially in consonant-cluster laden languages like Slavic and to an extent English and other Indo-European languages and it is spread through

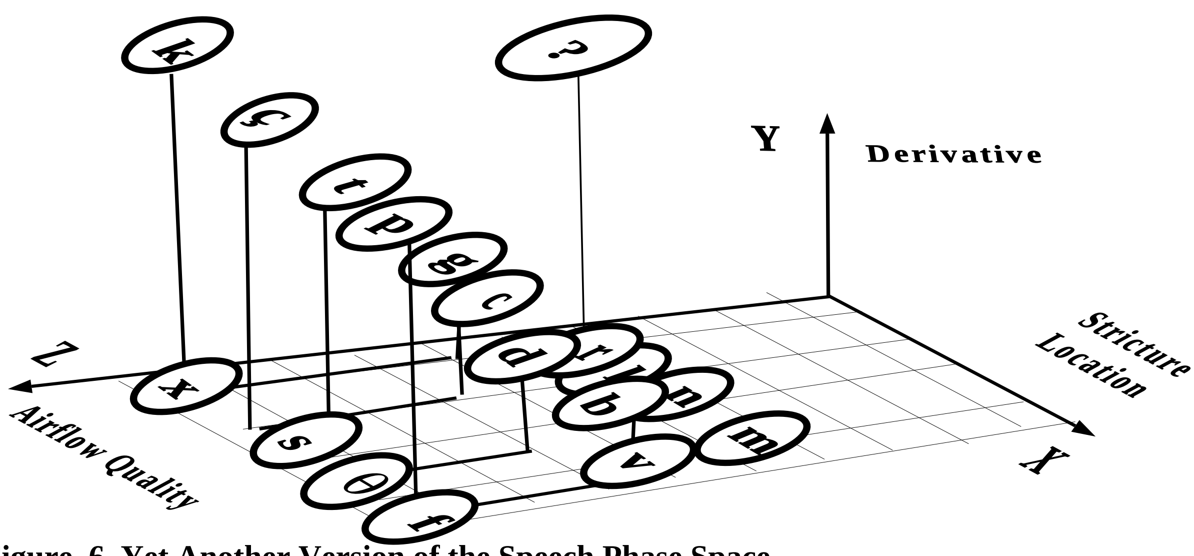


Figure 6 Yet Another Version of the Speech Phase Space

The vowels are probably close to the origin. The Y-axis is probably not to the same scale as the other axes. Thus, the voiceless plosives might be much higher than the voiced plosives.

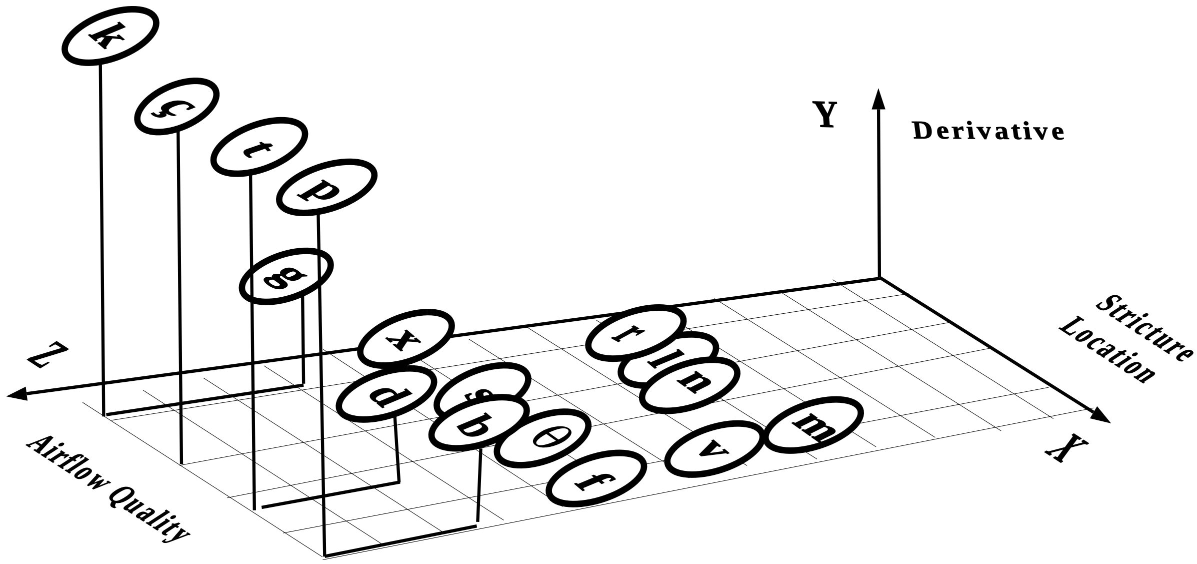
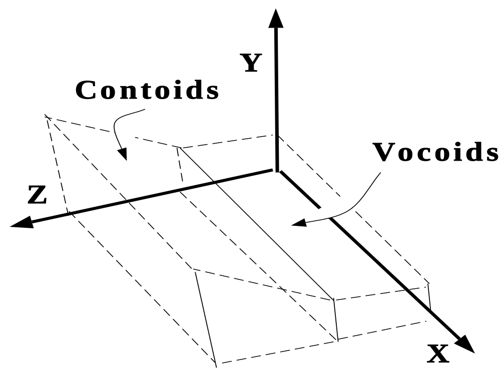


Figure 7 Yet Another Figure Suggestive of the Speech Phase Space

All of the voiced plosives are not shown. The relative locations of the fricatives and plosives have been rearranged slightly.



the region of much of the XZ-plane (except the vocalic parts which are covered by the specific vowels) since the quasiconsonants for all practical purposes are **ı**-colored. It would seem natural to have the **ı** the least marked and the most natural to have in any system. It would seem that, from the place of pride that the basis vowels occupy, their position should be next, right after **ı,** and the others **e, o** and **ü** which can be constructed from the basis vowels could be next. Finally we are left with **ö,** since it requires all three basis vowels **i, a** and **u.** [see Hubey, 1994,1996,1996b] Since these figures are not scaled, and indeed it‘s not possible to know with any degree of precision where some of these phonemes should go, some alternatives are given in these pages. It should be noted that, in general, the relative positions of the phonemes do not change appreciably, however it’s not possible without more evidence to be able to choose among the several competing alternatives. The liquids and nasals should probably be separated by a wider distance because of the more vowel-like quality of the nasals (i.e. nasal murmur).

Path Integrals and Minimization

Many linguistic phenomena can be clearly shown to be the result of some physical optimization effect; that is it can be easily seen to be minimizing the path length in the *phoneme phase space*. From the figure above we can easily explain the phenomena as *path integral minimization.* For example, the phonetic or acoustic realization of the words; *toes, haws, hods, cleans* is with a /z/, but instead we have *huts, tucks, buts, bits*. It's easy to see why from the diagrams. Another example; *magyar* in Turkish becomes *macar*. The figure shows that /c/ (voiced palatal fricative) is between /g/ and /y/ and the trip from a vowel to /gy/ and then back to a vowel is longer than the trip from vowel to just plain /c/ and back to a vowel. The path from a vowel to /sy/ back to vowel (i.e *mission*) is long but the /š/ is only part of the way to /s/. The transitions /tb/→/pb/ (*ratbag* → *rapbag*) , /tm/→/pm/ (*oatmeal*), /vt/→/ft/ (*have to*) also can be explained easily in terms of motion in this space. Since the space here symbolizes the motion of the articulatory organs, the distance in this phase space seems to mimic the actual (real) motion of real objects (i.e. articulators) moving in real space possessing momentum and mass. Thus, the /vt/→/ft/ is actually an *overshoot* which can be explained very easily in terms of physical processes such as momentum, inertia, energy and the force required to execute the motions. The other cases were *undershoot*, since it amounted to cutting the path short. One may make an analogy to making turns with a car; at high speeds, tight corners cannot be taken and will overshoot, and at slow speeds, one can make very sharp turns. Others such as /kt/ →/k/ *(facts→ faks),* /fth/ →/f/ (*fifths →fifs*), /st/ →/s/ (*chest→chess*) involves cutting the zigzag path short by interpolating the zigzag curves and is the same kind of momentum problem in articulation. More examples of tortuous zigzags that have been smoothed; /tr/→/çr/ (*tree→/çriy/*), /dr/→/cr/ (*drive→/crayv/*), *half* but *halves, calm* (no /l/), *psalm* (no /p/). More patterns involving inertia and momentum can be found in *masses, cars, riches, ridges, losses* (all manifesting the ending as /z/). The changes /mb/ → /mbr/, /ml/→/mbl/, and /nb/ → /mb/ can also be seen in terms of the paths in this space. The consonant harmony such as one syllable words having only voiceless plosives such as *pat, pot, cot,* etc. is also explicable in terms of inertia, acceleration and force. The *tenseness* is also easily explained in terms of motion in this space and the duration of the various segments of the path. We can make some general comments about motion in the *phase space*. In so far as it seems to mimic the motion of real articulators in real time-space, we should not expect zigzag paths. If we were to imagine words being constructed as paths in this dimension, we should imagine them as smooth curves since momentum and acceleration effects of physics will inhibit sharp motions because of its cost in energy. Tenseness-laxness can also be explained on this basis. Now, if we were to pass smooth curves (such as fitting *cubic splines*) through these points, we will notice that they'll tend to be distorted helical shapes. If the turns are very sharp, or if the distances too far, they'll tend to

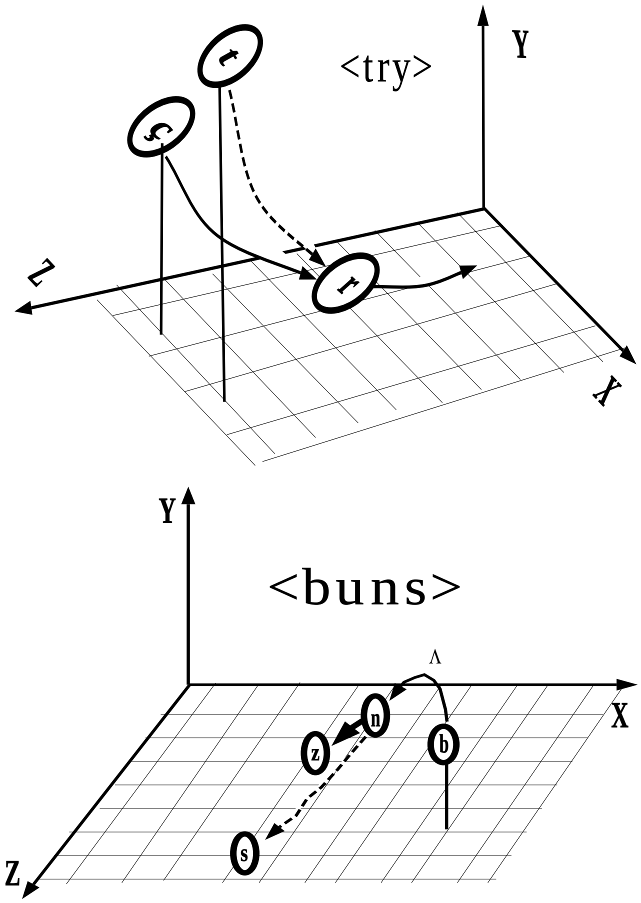


Figure 8 Activity in Phase Space

get smoothed out. On the other hand, tightly wound curves (such as repetitions) will also tend to stretch out. *Assimilation, metathesis, haplology, dissimilation,* and some of the other linguistic effects can be shown in the phase space to be mostly inertia, acceleration and momentum effects. This statement should not be interpreted to prejudice statements regarding the linguistic disambiguation efforts to place separate semantemes in separate phonological spaces. Thus if several words(or phonological manifestations of semantemes; that is, *words* or *lexemes*) collide in the higher-dimensional *phase spaces* then there may be efforts to disambiguate even if it means long paths. Of course, these are due to the phonological constraints of languages. Thus we can think of natural changes occurring in languages due to physiological reasons (ultimately explicable in physics) if not inhibited by the phonological (i.e. phonemic) constraints of languages. Of course, there will be interactions of both physiological and phonological factors. From the previous discussions on the phase space it would be natural to ask if the space is primarily articulatory, acoustic or both. The phase space is both and the dimensions (i.e. the dimensionless groups) can be described in both articulatory and acoustic terms. It shows some evidence that ‘like things’ show up close to one another in this space. The diphthongs are close to vowels; and they also share the property of not being steady-state with the plosives; the voiced plosives are closer to the vowels; and the liquids and nasals are also close to one another. Jakobson thought that the liquids and the nasals functioned as a natural class, and there’s further evidence for this supplied in Anderson & Ewen with respect to the Dutch diminutive suffix selection [1987, p. 153]. The Z direction is essentially inversely proportional to the signal-to-noise ratio, considering the formant peaks as the signal and the frication as the noise.

Phones, Phonemes, Allophones

Until now the words phoneme, consonant, vowel have been used rather loosely in so far as no mathematical definitions have been given. Without getting caught up in the phoneme fights of the early twentieth century phonologists, we can make some observations. Referring to the phase space, it will be easy to describe these concepts in more rigorous fashion. A phoneme is a small volume in the phase space. In other languages in which the same volume is fractionalized, the phonemes of that language will occupy even smaller volumes. More can be seen in Hubey [1994] and Hubey [1996]. Considering words as paths in this space, we can see that we have to make some kind of a decision as to where one phoneme ends and the other starts on the path that describes a word or even a sentence. It’s also clear that this determination will also depend on the particular path; that is, the particular word. Thus, a phoneme or more exactly its particular manifestation (i.e. an allophone) may occupy different small volumes in this space in different words, sometimes extending in one direction more than in other words depending on the direction from which the curve enters and leaves the particular volume representing the particular phoneme. Of course, the concept of the relativity of the phoneme with respect to a given language only means that the boundaries of the phonemes in this space can differ from other languages. If the curve for a word can be constructed from the acoustic signal, then the assignment of phonemes is simply (only conceptually, of course) a process of finding the particular volumes through which the curve passes. These ideas were expressed as early as 1950 as can be found in Joos and Hockett [see for example Saporta & Bastian, 1961]. Hockett writes, in his review of Shannon and Weaver’s book on Information Theory [Saporta & Bastian, p. 51]

> The acoustician examines speech signals and reports that they are continuous. The linguist examines them and reports that they are discrete. Each uses operationally valid methods, so that both reports must be accepted as valid within the limits defined by the operations used, and the apparent contradictions between the reports constitutes a real, not an imaginary problem... The linguist... also, is unable to to examine the speech-signal directly. The ear and the associated tracts of the central nervous system constitute a transducer of largely unknown characteristics...
>
> A continuum can be transformed into a discrete sequence by any of various QUANTIZING operations; ...though the quantizing operations used in electronic communications are all quite arbitrary. Similarly, a discrete sequence can be transformed into a continuum by what might be called a CONTINUIZING operation. Now if the continuum-report of the acoustician and the discrete-report of the linguist are both correct, then there must be, for any given body of raw material, a quantizing operation which will convert the acoustician's description of the raw material into that of the linguist, and a continuizing operation which will do the reverse; the desired quantizing and continuizing operations must be inverses of each other.

In the same paper a very beautiful description of a stochastic process, attributed to Joos by Hockett, is given;

> 'Let us agree to neglect the least important features of speech sound, so that at any moment we can describe it sufficiently well with n measurements, a point in n dimensional continuous space, n being not only finite but also fairly small, say six... Now the quality of sound becomes a point which moves continuously in this 6-space, sometimes faster and sometimes slower, so that it spends more or less time in different regions, or visits a certain region more or less often. In the long run, then, we get a probability density for the presence of the moving point anywhere in the 6-space. This probability density varies continuously all over the space. Now wherever [one]..find a local maximum of probability density, there the linguist finds an allophone; and 'there will be not only a finite but a fairly small number of such points, say less than a hundred.'

These descriptions should be compared to the three-dimensional phase space of this section. It is not yet clear how many of these 'local maximum probability densities' exist in languages. Introspection gives one set of answers and speech recognition researchers give another set. For example, we find in Clark and Yallop [1990, Appendix] that there are 43 English phonemes (21 vocalic and 24 consonantal) which is more than that given in SPE; Kai-Fu Lee uses 48 phonemes in their Hidden Markov Models of speech recognition [Waibel & Lee, 1990, p. 352]; whereas Roucos and Dunham claim that their model uses 270 phonemes [Waibel & Lee, 1990, p. 369]; Churchland categorically states that English has 79 phonemes [Forrest, 1991, p.285]; and Ladefoged gives evidence from Moskowitz, Ohala and Jaeger that 'people can use orthographic knowledge as the basis for forming phonological classes' [Dressler et al, 1988, p.166].

Lenition, Fortition, and Sonority

A concept that we'll need for this subsection is that of a *vector*. There are many different representations of a vector. It is simply an ordered *n-tuple*, and it is called an *array* in computer science. A *feature bundle* is an array or a vector as was shown above. The number of features is the dimensionality of the vector. As long as the operations on the vector are clearly defined they can be applied to many different problems in many different representations. Perhaps the simplest way to think of a vector is to imagine it in pictorially as an object that has a *direction* and a *magnitude*. We can then easily represent it as an arrow. The length of the arrow will be called its *magnitude* and its direction is obvious. Figure A below shows two vectors **A** and **B**. Both the overbar and bold notations are used to denote vectorial quantities (overbar in the figures and bold in the text). Geometrically, one way to add vectors is to put them head to tail and then draw another vector from the left-over tail to the remaining head of one of the vectors. This is shown in Figure A. The sum of the two vectors **A** and **B** is then another vector **C**. Another way to add them geometrically is to put the vectors tail to tail and draw lines parallel to the vectors; the intersection of these lines is the head of the sum of vector **A** and **B** (that is the vector **C**) as can be seen in Figure B. Since these are two-dimensional vectors; (that is, it takes two parameters or variables to represent two-dimensional space) we can represent them in algebraic terms by superposing them on the say XY plane as shown in Figure (10.C.) The endpoints of the vectors are really what we'd call the coordinates of a point.

However points in 2-D have no direction. In order to show this additional property of vectors (i.e. direction) we have to represent them slightly differently. Since, as can be seen in Figures A and B, we can move the vectors around in space without affecting their addition property or their magnitude or direction, we put the vectors with their tails at the origin (Figure C). We can then decompose them into their components along the X and Y axis. Indeed, since they're two dimensional (i.e. array was another representation), these components constitute the vector. These components are denoted subscripts as can be seen in Figure C. The addition of the two vectors **A** and **B** then corresponds to adding up their separate (X and Y) components to yield the X and Y components of the vector **C.** We have to have some notational device to keep the components separate. There are many ways to do this; one way is to represent it as an ordered pair as in $\mathbf{A}=(A_x,A_y)$, $\mathbf{B}=(B_x,B_y)$ and $\mathbf{C}=(C_x,C_y)$. Thus, vector addition rules yield

(5) $\mathbf{C}= (A_x + B_x, A_y+ B_y)$ since $C_x = A_x + B_x$ and $C_y = A_y+ B_y$

The components of vectors are *scalars*, since they possess only the property magnitude. Another notational device that's often used is that of a concept of unit vectors. These are vectors that point in the direction of X and Y but their magnitudes are unity. Thus if we use $\mathbf{u}_x$ and $\mathbf{u}_y$ to denote the unit vectors in the X and Y directions, then the vectors **A** and **B** can be written as

(6) $\mathbf{A}= A_x\mathbf{u}_x + A_y\mathbf{u}_y$; $\mathbf{B}= B_x\mathbf{u}_x + B_y\mathbf{u}_y$ and $\mathbf{C} = C_x \mathbf{u}_x + C_y\mathbf{u}_y$

Thus we have

(7) $\mathbf{C} = C_x\,\mathbf{u}_x + C_y\mathbf{u}_y = \mathbf{A} + \mathbf{B} = (A_x + B_x)\mathbf{u}_x + (A_y + B_y)\mathbf{u}_y$

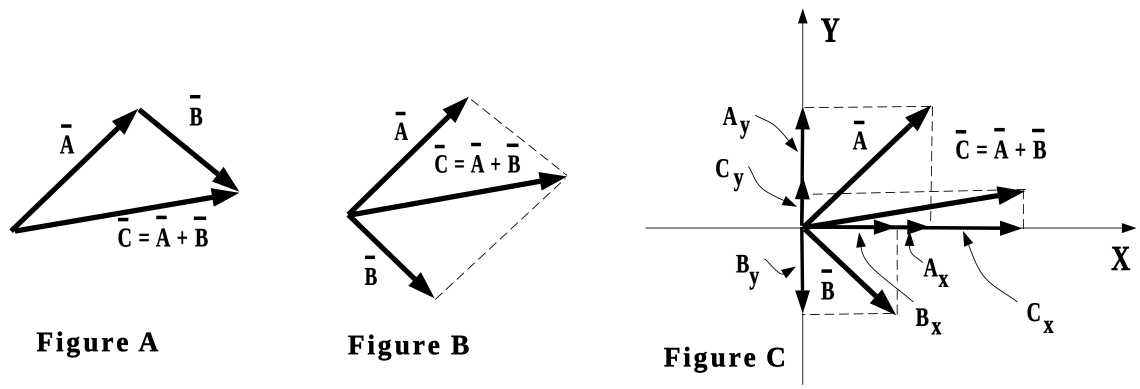


Figure 9

As an application of vectors consider the vectors below in the YZ subspace of the phase space developed. The space is shown below for convenience. Phonemes or even groups of phonemes are indeed vectors in this space; *fuzzy vectors* but still vectors. We might think of the *centers of gravity* (defined in some weighted sense) of the small volumes in this space to be the unfuzzy vectors that we're discussing in this section. Two vectors **P** and **R** are shown in the figure next to the sonority scale. The vector **P** in the figure below points away from the origin and it can be seen that its components both point in the positive Y and Z direction. Vector **R** points towards the origin and both of its components are negative (that is pointing towards the negative Y and negative Z directions). An interesting usage of the concept of vectors will applied to lenition as can be seen in Lass [Lass, 84, p. 177] that gives the phonological rules for lenition and fortition as;

a) Stop > Fricative > Approximant > Zero
b) Voiceless > Voiced

Essentially the same results can be found in Foley [1977]. These results can easily be shown to be derivable quite clearly and unambiguously in the phase space and are related to sonority. We only need two dimensions although three would be better) and the concept of a Cartesian vector to show the essential results. The space shown above is a dimensional subspace of the dimensional phase space of that section. Indeed, the three-dimensional phase space can be considered to be a subspace of the many different feature-bundle spaces discussed in the literature with the caveat that these spaces are not orthogonal and the mapping might not be one-one or linear. We can see immediately from Lass's hierarchy that a) refers to a vector that points in the negative Y direction (Stop > fricative) which is $C_2 > Q_2$ or $C_1 > Q_2$. The second part of a refers to a vector that points in the negative Z direction (i.e. Fricative > Approximant). The third part of a) is also a vector that points in the negative Z direction (i.e. toward the origin of the YZ axes). Part b) refers to a vector that points from C_2 to C_1 (Voiceless > Voiced) and thus is a vector that points in both the negative

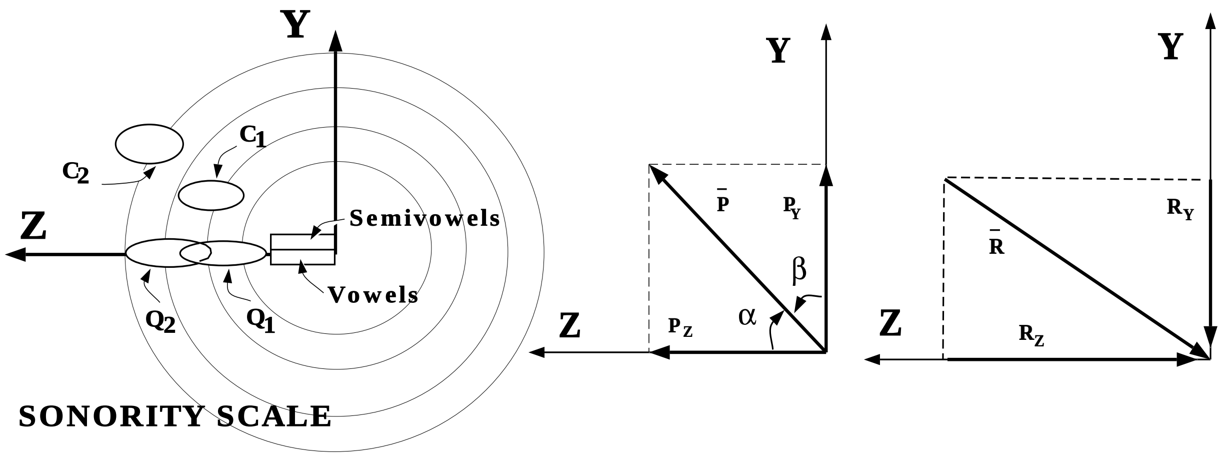


Figure 10 The Vectors in Two-Dimensional Speech Phase Space
Vectors P points away from the origin; and R toward the origin.

Y and Z directions. The vectors that show these concepts are shown above. Since the sonority scale was not originally drawn to scale, the Voiced > Fricative transition shows a slight positive component (on the diagram on the right), however this is only an artifact of the unscaled drawing. Since no measurements have been taken to indicate the scale of the phase space, and no mathematical definitions have been given, at best we can use the data from Lass and Foley as guides to make the phase space reflect reality as closely as possible. The drawing on the right is a slight rearrangement to reflect the data taken into consideration. In the next subsections, data from child language development, aphasia and formant measurements to fill in some of the gaps of the phase space. Meanwhile, it can be seen from the vectors above that all of this phenomena is easily describable in terms of the vectors representing the transitions. Thus lenition is a vector pointing toward the origin. The sizes and shapes in Fig. 11 is not important due to lack of scaling which itself is due to the lack of necessary measurements.

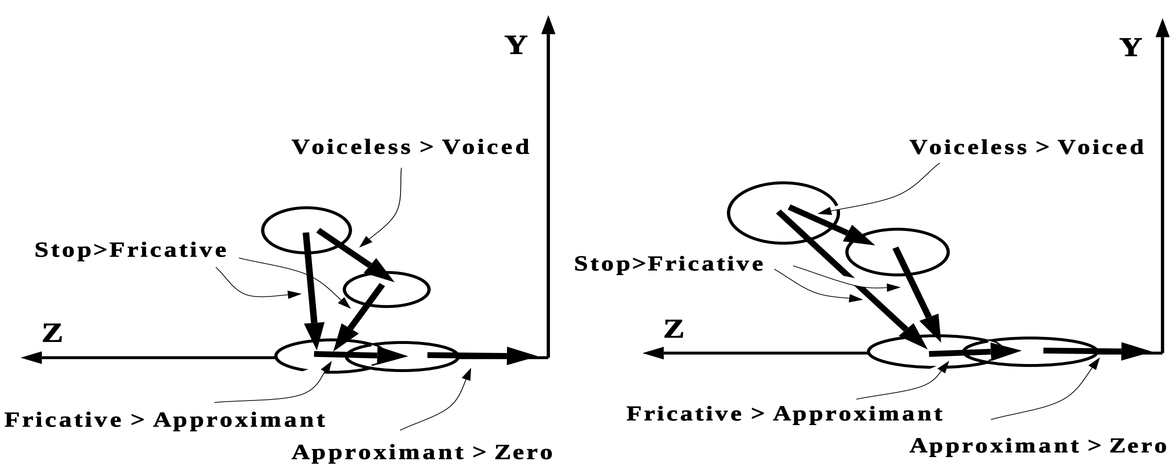


Figure 12 Fortition, Lenition and Sonority

Child Language Development and Aphasia

The study of child language development, although kicked off by Jakobson in his 1941 work, has amassed much date over time. It is summarized in Anderson [1985, p.131] as ;

> ..all children begin with a minimal opposition of a single vowel (roughly [a]) and a single consonant (generally labial [p]). Consonantal distinctions arise with a difference between nasal ([m]) and oral ([p]) segment type; and subsequently with a split in point of articulation between grave (labial) and acute (dental) sounds. Within vowels, the first split is between compact (low) and diffuse (high) segments. With regard to manner of articulation, stops arise before fricatives, and both before affricates. The consonant/vowel distinction precedes the emergence of liquids or glides, and sonorant liquids precede obstruent liquids. Some distinctions, where they are to appear, arise only very late: e.g. nasal vs oral vowels; opposition between liquids; clicks, ejectives, implosives and other nonpulmonary airstream mechanisms, etc. The uniformity of the sequence in which these segmental distinctions are acquired seems quite general.

We can attempt to sketch out where the vowels should fall rather easily from this description based on a very simple algorithm. All we have to do is to assume that children start by distinguishing the most different phones (i.e. those most distant from one another) and then continue to divide this volume into smaller pieces as their power of discrimination increases and as they listen to speech. The sequence is roughly sketched out in the Figure 12. Only the voiceless plosives are shown for

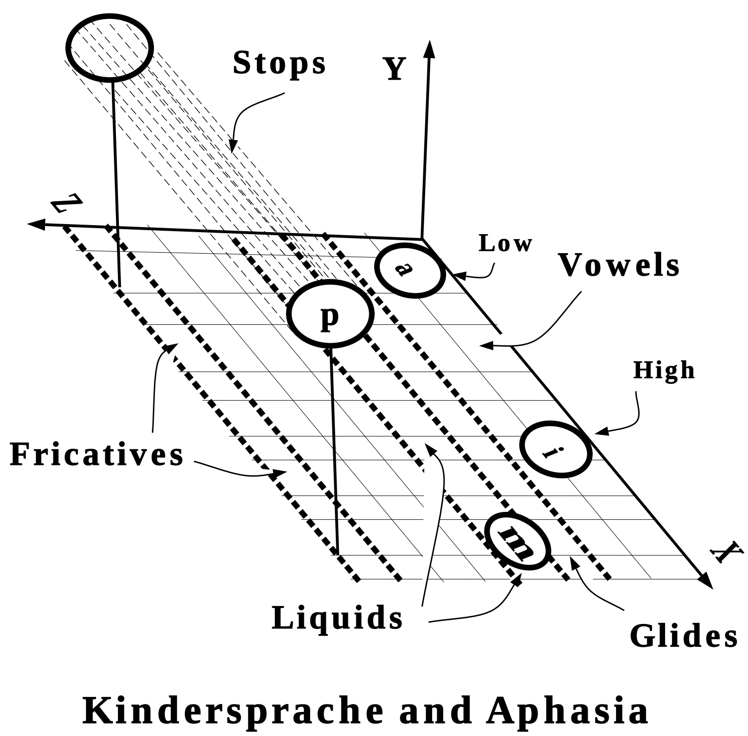


Figure 12 Language Development

the stops. It can be seen that the [a] and [p] start off with the maximal distance at the two extreme ends of the space. Then a nasal [m] is later introduced, an [i] and as the process continues, it seems to further subdivide the phonological volume as if cutting a piece of cake into smaller and smaller pieces. We should note here that there could possibly be other reasons for the order of learning. The [a] is an open vowel i.e. the fact that it is produced with the mouth open means that another channel of communication is available to the child, that is vision. Similar comments can be made about [b], [p], and [m]. The motion of the articulators for the back stops cannot be seen and the child needs more feedback before they are learned. Similar considerations might apply for the learning of the other vowels. Considered in this light, it's not surprising that a child's first utterance seems to be something like [ma], [pa], or [ba]. If the most important factor were intelligibility we would expect the supervowel [i] to be learned first. It should be noted that the liquids and nasals fall in the same general region of distance from the X-axis and the figure does not mean to imply that /m/ is a liquid. However various reasons have been given at different times to justify grouping the liquids and nasals together [for example Jakobson].

Aphasia seems to go in the reverse direction with the last learned being the first lost, thus it seems that there's a stack-like structure (i.e. LIFO= Last In, First Out) in memory. The same phenomena can be observed in boxers during matches. The first thing to go, after a hard punch, seems to be the last things learned (i.e. bobbing, moving from side to side, holding the hands up and finally the inability to stand up). Near death experiences where people see tunnels and bright lights could be due to similar brain processes. Furthermore victims of aphasia never seem to make two featural mistakes (i.e. change two features at once) but rather only one at a time [Lieberman & Blumstein, 1988] which seems to lend credence to the usefulness of the binarity idea of distinctive features or looked at another way, the ability to come *close* to the phoneme. Aphasia is a complex process and its effects (whether it's Wernicke's or Broca's, see for example Lieberman & Blumstein [1988] seem to hinge on the way the brain's memory and neural computation work. Thus not much more can be said about what aphasia implies for problems in speech production. A hint of the whereabouts of the vowels in the phase space has already been given in this subsection. The next subsection will show the full phase space including the vocalic phonemes.

Vowels in Phase Space

In the previous subsections, the vowels were left out of the phase space because of the difficulty of ascertaining their locations. It is difficult, for example, to decide a priori whether an /o/ should be near the front because of the rounding or near the back because of the position of the tongue. However, since the phase space has not only articulatory content but also an acoustic one, it's possible to draw inferences from several results, cull the results and put the vowels in the phase space. Some evidence comes from sonority and yet others from formants. For example, it was shown in the previous subsection that the sonority of the consonantal sounds can be described essentially in two-dimensional space. We can extend the concept of sonority to the full three dimensions of the phase space. For example, Foley [1977] pursues an essentially sonority based course in his descriptions of *phonological strengths.* In his descriptions of the vowels, he derives the phonological strengths [Foley, 1977, p.47] of vowels as {i,1}, {e,2}, {u,3}, {o,4} and {a,5}.

It's also quite interesting that the distances between some of the vowels [Foley, 1977, p. 78] can be derived directly from the binary three-dimensional representation of the vowels; for example, he gives the differences of phonological strength as |**a-o**| = 1, |**e-o**| = 2 and |**i-o**| = 3. In this connection, it should be mentioned that Gilbers [Gilbers, 1992] in his network representation of segments not only uses a binary representation of vowels, but uses binary operations to derive vowels from others via operations of rounding, tensing, laxing etc. Gilbers [p. 130] also reaches the conclusion that " we predict that unarticulated voicing, the articulatory correlate of schwa, is universal. In the area of first language acquisition, we consider schwa to be the first acquired vowel." In a system of *markedness penalties* or *taxes*, the assigns zero penalty to the schwa (**I**), small penalties to **i, u**, **a** and the largest penalty to **ö** [Gilbers, 1992, p.133] which is fully consistent with the results of this paper.

The most important results that are necessary for placing the vowels in the phase space come from acoustic studies. For example, Nearey noted that the *front or acute* vowels ([i,I,e,æ]) have a high F_2 and the *back or grave* vowels ([a,U,u]) have a low F_2. The [i] and [u] with a high tongue position have low F_1 and [a] with a low tongue position has a high F_1 [Lieberman & Blumstein, 1988, p. 222]. This implies that the function that we need to derive the placement of the vowels along the X axis should decrease with increasing F_1 and increase with F_2 yielding some kind of a scaling along a *front-high* dimension. There are many ways of constructing such functions. Only a few simple forms will be given here to produce a rough *complete* phase space. The first step is in scaling the formants. The same scaling will be used as before; i.e.

$$f = (F - F_{min})/ (F_{max} - F_{min}) ;$$

Table 1 shows the results of computations of some candidate functions. Table 2 shows the ordering of the vowels according to the various computed values. In all the cases the various allophones of the acute vowels score near the front and the back allophones have low scores. The ordering essentially seems to show the distance from the origin (which is where the **a-o** apparently belong,

	i	ɪ	ɛ	æ	a	ɔ	U	u
f_1	0	0.26	0.56	0.85	1	0.65	0.37	0.65
f_2	1	0.79	0.68	0.6	0.17	0	0.12	0.02
$f_2(1-f_1)$	1	0.58	0.30	0.09	0	0	0.076	0.001
$f_2 e^{-f_1}$	1	0.61	0.39	0.26	0.06	0	0.08	0.01
$e^{f_2} e^{-f_1}$	2.7	1.7	1.13	0.78	0.44	0.65	0.78	0.53
$f_2(e - e^{f_1})$	1.7	1.12	0.66	0.23	0	0	0.15	0.02

Table 1

radially outwards. The vowels which have approximately equal values are put in circles. Since these vowels don't fall on the corners of the perfect cube but rather are scattered about the corners of a distorted cube, this kind of accuracy (or lack of it) is expected. Thus we can produce a full phase space indicating the whereabouts of the vocalic sounds and phonemes of various languages. It should be noted that since the [ç] seems to be difficult to produce and since it seems to be a combination of two other phonemes' articulations, it might also be possible to indicate this on the phase space.

$f_2(1-f_1)$	i	I	ɛ	æ	U	u	a ɔ	
$f_2 e^{-f_1}$	i	I	ɛ	æ	U	a	u	ɔ
$e^{f_2} e^{-f_1}$	i	I	ɛ	æ U	ɔ	u	a	
$f_2(e - e^{f_1})$	i	I	ɛ	æ	U	u	a ɔ	

Table 2

A rough sketch of the phase space with consonants and vowels is shown below (Figure 13). Anything more accurate than this requires special experiments to determine the values of the phones/phonemes in dimensionless groups. Once again, it should be remembered that the drawing is not to scale since imputing distances using a linear Cartesian distance for sonority will cause problems. However, a three dimensional sonority scale can still be derived from this by weighting the coordinates. Some of the vowels have been placed in the phase space in the next figure. The vowels should probably have been put closer together and nearer the origin however once again it's a matter of scaling and not theory. It is not very clear where the glottal stop should fall, however, it would seem to belong near the origin and is placed there in the figure.

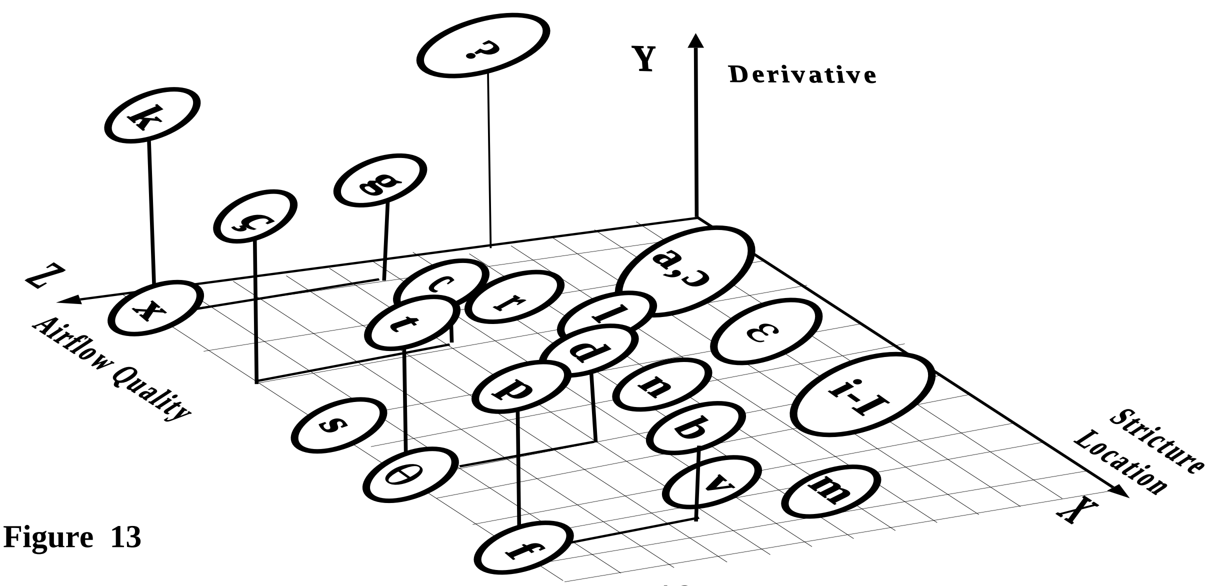


Figure 13

Distance, Birth of New Phonemes and Experimental Evidence from Diphthongs

Ancient Sanskrit was tyramidal (only three vocalic phonemes /iua/) and eventually obtained an /e:/ and and /o:/ at a later time. Since it already apparently had the diphthongs /ai/ (/ay/) and /au/ (/aw/) we might wonder if there is a relationship that can be shown using the concept of distance. The figures below (Fig. V.16) are suggestive of the first three formants of the diphthong /ai/ and

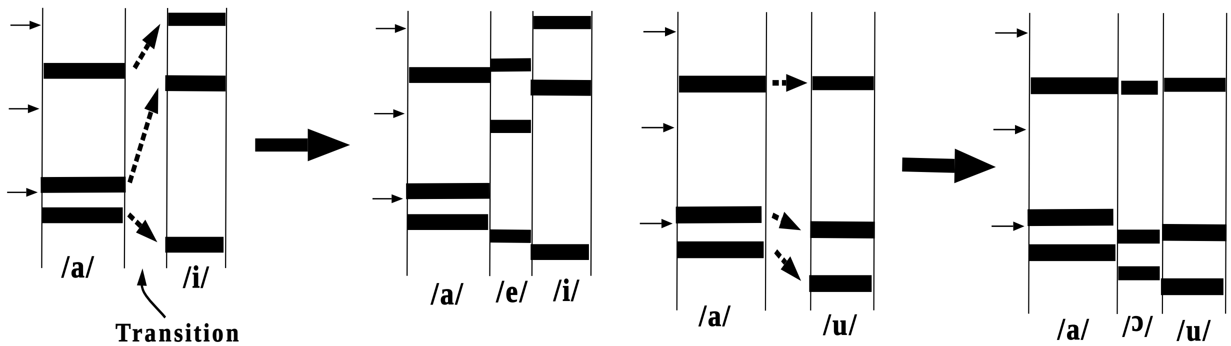


Figure 14 Various Vowel Transitions

the transitions from /a/ to /i/. The horizontal lines indicate the formants. The small arrows indicate a separation of 1 KHz. The formant data are from Peterson and Barney[1952]. Since the beginning of the diphthong begins with a steady state /a/ (which means it's a vector in the n-dimensional formant vector space) and then ends up as another steady-state vector resembling an /i/ (or /y/), the whole phoneme must thus be represented as a vector transition (a dynamic vector or vector velocity) which implies that; a) vector derivatives and vector calculus becomes necessary and b) it must of necessity pass through points some of which might belong to the volume of another phoneme. The second figure indicates what would happen if we substituted the formants for /e/ in the transition zone. The formants of /e/ fall in the zone where the transitions occur. It does not seem to be accidental. If we do the same thing for the /au/ diphthong we get a similar result, as can be seen in the figure above. Of course, the historical interpretation does not seem so clear cut. It's

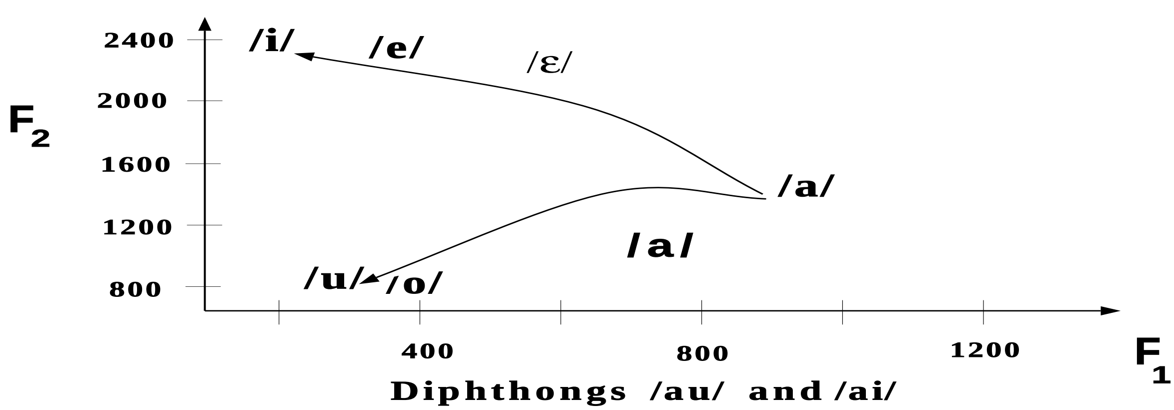


Figure 15 Vowel Transitions; after Carre and Mrayati [1991]

not clear if these changes were innovations or if new language speakers who did not have these diphthongs interpreted the changes as their own vowels /e/ and /o/. These ideas are discussed much more fully in Chapter VIII of Hubey[1994,1996].. And independent confirmation of the perception problems of diphthong transitions comes from an unexpected source. It's from speech recognition research using neural networks. Kohonen, who is a pioneer in research in the use of artificial neural networks for phoneme recognition, reports in his work in Japanese and Finnish that he has found that the network recognizes the diphthong **/au/** in words like /hauki/ (meaning pike) as the **/aou**/ sequence and has found it necessary to introduce a phonological rule to derive /**au**/ from /**aou**/ [Aleksander, 1989, p. 35]. We might surmise from this that the **/ö/** might have developed from the /**üe**/ diphthong the same way that **/e/** developed from **/ai/** and **/o/** from **/au/**. What the perceptual distance implies is that we might draw the relationships slightly differently as can be seen in Figure 14. This would imply that the ordinal vowels **e**, **o** and **ü** do not really fall on the corners of the cube but are rather closer to the diphthongs (i.e. the transitions).

The experimental evidence from Carré and Mrayati [1991] shows that trajectories of the various diphthongs. Of course among these are the diphthongs /ai/ and /au/. Figure 15 indicate the paths that these diphthongs take in the two dimensional formant space. The space is not normalized but it's especially clear that the /au/ passes very near /o/. The figures above were indicative that this result were to be expected.

Implications for Formant-Vowel Space

Figure 16 shows the rough ideal placement of the vowels from Hubey[1994] using the data from Peterson & Barney [1952] and Clark & Yallop [1990] for Australian English vowels, which can also be seen in Hubey [1996b]. It is clear that this three-dimensional view of the vowels is an economic description of many linguistic phenomena. It fits in reasonably well with the traditional introspective vowel descriptions, the newer results from Ladefoged modifying the traditional cardinal vowel diagram, and the latest results of the experimental formant studies. Since it seems possible to represent not only diphthongs but also at least some of the consonants, such as the stops from the formant transitions, the importance of the formant vector space increases. It might be possible to represent speech sounds, with the addition of some other factors such as aspirations noise, burst amplitude, signal-to-noise ratio, within the formant vector space. It might also be possible to use both the formant vector space and the phase space together. The perceptual distance between the diphthongs and the ordinal vowels seems to imply that the ordinal-cube is not really cubic but that it is distorted (and also rotated) in the formant space. This is consistent with the results of the previous sections. We can try to construct this ordinal-cube in the normalized formant [vector] space by using all the information now available. The shape of the ordinal vowels in the formant

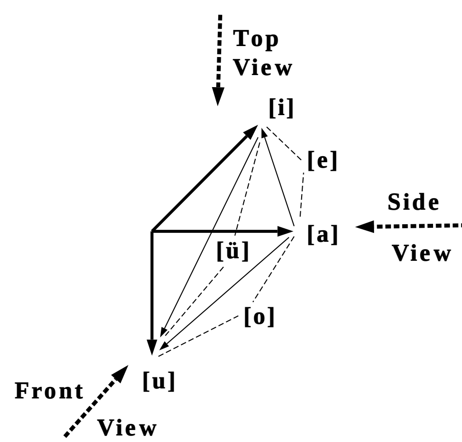


Figure 16

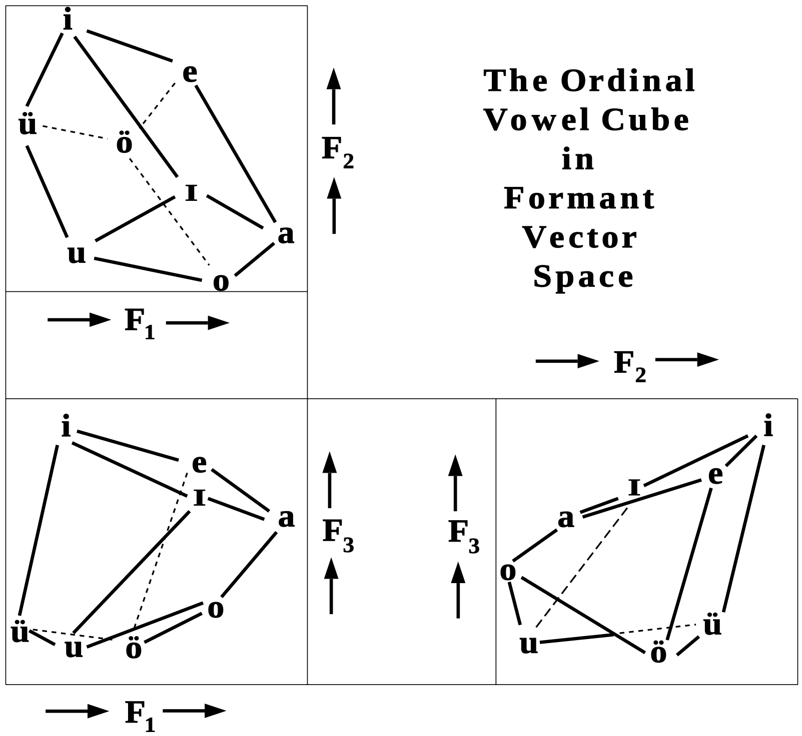


Figure 17 The Ordinal Vowel Cube (Cuboid)

space is shown on Figure 17 (a fuller description of its derivation can be found in Hubey[1994] and Hubey[1996b]. The side view (i.e. F_2 vs F_3) shows some discrepancy with the formant data of Peterson & Barney and Clark & Yallop. The **[e]** has been slightly displaced to show more clearly the shape of the cubic structure. The **[ü]** doesn't show up in the formant studies and its position was estimated from various hints as alluded to in this chapter. It is hoped that a better normalization algorithm or samples of phones more representative of the eight ordinal phones/vowels, say from Turkish, will yield a better fit. As can be seen, the near-cubic shape of the eight vowels resembles the modified vowel diagram given in chapter I, Hubey[1994]. The main problems in constructing this figure is that the [**a**] and the [o] do not separate well and that there's no [**ü**] or [**ı**]. In addition, the [**i**] in English (*heed*) is diphthonged as is the [**u**] (*who'd*). Moreover, the sample for the [**o**] is really from the open-o (as in *hawed*).

Sonority and Scaling

The consonant clusters make it difficult to strictly define syllables in languages such as English versus a language like Turkish, Finnish, or Japanese. The idea of a sonority scale can be explained, first, directly from the graph since the sonority scale seems to go from the vowels toward the plosives so the scale is essentially the distance from the origin of the axis, the voiceless plosives being the least sonorant and the low vowels being the most sonorant; thus being inversely proportional to the distance from the origin (at least in the two dimensions as shown). The three-dimensional version will be developed in later sections. The physical explanation, of course, is that in wave propagation, the high frequencies dissipate faster and thus low frequencies go further. The voiceless plosives are highly spiked and thus contain high frequencies. For example, an ideal spike of zero duration is a Dirac delta function and its Fourier Transform is constant implying equal power at all frequencies. The sonority then can be expressed simply as either one of;

(8a) $\sigma = 1 - \sqrt{(Y^2 + Z^2)}$ $\quad 0 \le Y,Z \le \sqrt{2}/2$

(8b) $\sigma = 2/\{1+\sqrt{(Y^2 + Z^2)}\} - 1$ $\quad 0 \le Y,Z \le \sqrt{2}/2$

(8c) $\sigma = e^{-R}$ where $R = \sqrt{(Y^2 + Z^2)}$; $\quad 0 \le Y,Z \le \infty$

(8d) $\sigma = \log (\sqrt{(Y^2 + Z^2)}$ $\quad 1 \le Y,Z \le \infty$

Obviously, these functions might have various coefficients depending on the units and dimensions used. It should be noted that the definition is only two-dimensional but extension to three dimensions is straight forward. It would thus seem, recalling the positioning of the glottal stop that it might belong closer to the plosives than toward the origin which makes it closer to vowels. Unfortunately, the glottal stop is accepted to be more of a vowel than anything else, thus one solution would be to place the glottal stop near the vowels but with a very large value of Y. It would also be possible to represent the vowel transitions as vectors in a three-dimensional space. Since the pure ordinal vowels can be represented as the corners of three-dimensional cubic structure in the

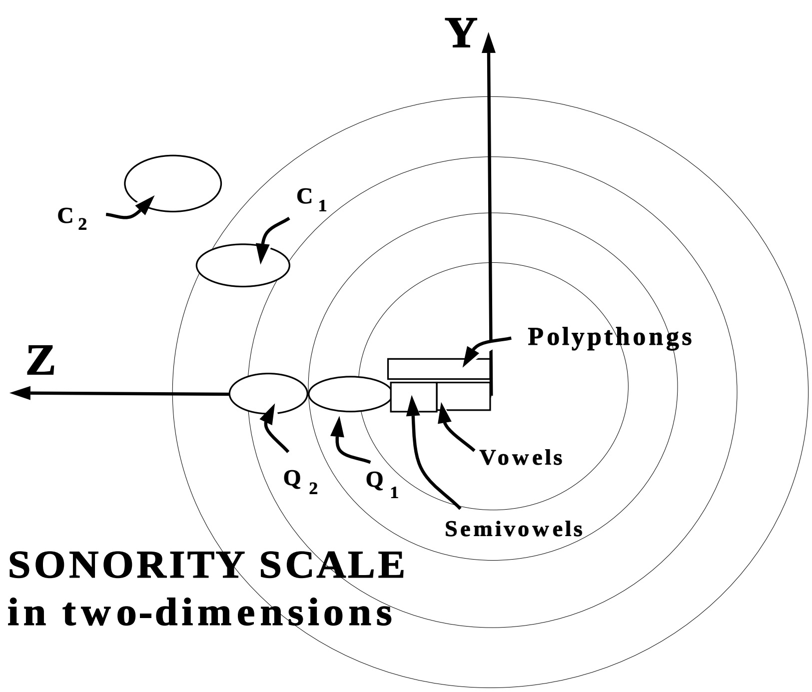


Figure 18 Suggestive Distances from the Origin and Sonority

3-D formant space, the transitions between two vowels whether it's a diphthong or a consonant should still be recognizable as a velocity (i.e. transition). Thus, if only formants are used (extracted from the signal in short intervals, say about 10 ms intervals), then the transitions will be represented as vector derivatives in the formant space. Some consonants show changes in the higher formants although some clearly show transitions in the lower formants. In the phase space of the previous sections, it's clear that vowels are statically representable. It's also possible to view the formant transitions induced by the stop consonants as studied by Stevens & Blumstein[1978], the Haskins Laboratory [Lieberman, 1984], Lieberman & Blumstein [1988] and Blumstein & Stevens [1980]. It has been known since the earliest studies at Haskins Laboratories [Lieberman, 1984] that the stop consonants cannot be isolated in speech from the vowels and that they show up as transitions of the formants in CV syllables such as [ba],[bu],[bi], [da],[du],[di], [ga],[gu],[gi]. Much research has been conducted to look for acoustic invariants representing the consonants in these syllables; and it's still continuing. There are, however, different perspectives that one might use to examine the phenomena of speech perception in terms of mathematical models. A completely integrated and unified mathematical model is lacking. Because of its extreme complexity, the study of speech realization and perception forces the researcher into the divide and conquer mode. The modular research in any scientific field has numerous benefits; among them the simplicity of the phenomena under study. Similarly verbal models of natural phenomena have many advantages, the most important one being the loose usage of words and an even looser usage of analogies. On the other hand, the advantages of mathematical models are too well-known to be listed here in detail. The most important ones are the precision of the explanation and its testability. A verbal model is of course preferable to an incorrect mathematical model or one that because of its simplicity falsifies reality to a great degree. However, it's a well-accepted article of faith in the western world that mathematics is the language of science and all scientific fields, including linguistics, attempt to explain phenomena in terms of mathematical models. A particularly interesting problem in young mathematical sciences is the attempt to formulate connections among the various mathematical models that exist in the field in an effort to weed out inconsistencies in competing theories which might stand up reasonably well in smaller modular domains of a larger integrated field of study.

Conclusion and Discussion:

There's something about the phase space that strikes people as odd; is it articulatory or acoustic? The simple answer, which is the correct one, is that it is both. There's no reason to be surprised about why the articulatory and acoustic parameters should map to one other. If, in fact, it were not so, then we'd be really surprised. The likelihood that this mapping would be highly nonlinear is taken for granted since it's common knowledge that different articulations can give rise to the same sectral pattern or acoustic/phonetic output. And the dimensionless numbers (which are the dimensions of this space) are are really this nonlinear mapping! The nonlinearity is absorbed into the dimensions of the 3-D space and the result is a more tractable space; a simple one, yes. The proposed phase space has both acoustic and articulatory content and why shouldn't it? If there were no correlation between articulation and acoustics how then can we produce the sounds we want? If there was no correlation between acoustics and perception how then can we have any regularity and hence speech as a communication tool? Speech phenomena certainly possesses extreme complexity and it's exactly because it has this complexity of both fluid motion (laminar and turbulent) and wave patterns imposed upon it that it practically begs for dimensional analysis.

Of course, the simple space has limitations; it cannot represent geminates, ejectives, trills, clicks, and , pharyngealization because it is not complex enough. To do that we'd need more dimensions. It might require several tens (or thousands) of nonlinear stochastic differential equations to accomplish reproduce even some of the complexity inherent in speech production. The strength of the model is that it has something that others don't; none of the papers in any phonetics journals as yet shows any rhyme or reason for the data being collected or the scatter plots that are being produced. The most fundamental principle of physics is that the terms of any equation must be dimensionally homogeneous. Dimensional analysis will produce the correct result. This chapter gives an example of the power and usefulness of dimensional analysis by producing the simple and approximate three dimensional space. It doesn't explain everything and it can't but it already "explains" more than anything produced to date in any book or journal and from fundamental physical principles instead of ad hoc curve fitting. Some of the simplifications are that specifically, what is being done not only for the exotic sounds as above but also even for some of the more common ones (such as [š], [ž] and the nasals) is that they are being forced into the 3-D space by tinkering with the extra dimensions that would be necessary for representing them properly. For example, Fant included the information in the third formant by modifying the second formant values to be able to use only the first two formants. Similarly the extra dimensions of the vowels have been squeezed into a single dimension by ignoring the third formant and collapsing the first two formants into a single number. The result can't be anything but an approximate truth but it's still true enough to be novel because of its explanatory power in child language development and sonority scales. It can even be used to understand how the phonemes of languages are distributed over this volume and what we should expect.

The X-axis (place) takes care of both height and length (in a multiplicative way). The fact that the place of constriction has an effect on the acoustic output is undeniable. Looking at it from the point of view of the two-tube model, the place of constriction changes the size of both tubes. Looking at it from the point of view of the source&filter model, it's obviously the mechanism of the changing of the filter which shapes the acoustic output. Looking at it from the point of view of experiments, the data of Stevens & Blumstein [1978] show that the there are cues for the place of articulation in stop consonants. The role of the amplitude of the fricative noise in the perception of place of fricatives can be seen in Behrens & Blumstein [1988]. The role of onset spectra [for example Blumstein and Stevens, 1980] for the stops, and lots of other experimental results indicate that there does indeed exist acoustic correlates of articulation. Of course, for the vowels, the role of the place of constriction in shaping the peaks of the output is clear and the three formants form a left-handed vector space as can be seen chapters III and IV of Hubey [1994,1996]. And if there still exist other acoustic invariants which have not yet been found, it doesn't mean that they don't exist. I think I found some which can be shown on this space in conjunction with the vowel (formant) space, but already implies that we need six dimensions. It will be shown later, that we need yet more dimensions.

The Y-axis definition quite obviously is based on articulation but the acoustic correlates are quite easily found. The stops would probably be best modeled mathematically as the Dirac delta-function. It is well-known that the power spectral density of the delta function is white noise (flat across the spectrum i.e. constant). Of course it is just as well known that no real signal can have power at all frequencies since it would require all energy in the universe, but it works anyway--both in mathematics and physics. In practice in mathematical modeling and physics fat Gaussians (i.e. large second central moment) is used for white noise. The voiceless stops then would have power at all the frequencies (mathematical idealization of course). In practice, we cannot see noise at all frequencies; they'd have to drop off with frequency. This is indeed reasonably well-corroborated (see the power spectra of the voiceless plosives/stops from Edwards). The slight differences among them is no doubt due to the place of the stop as was already mentioned above. Furthermore, the filter (vocal tract) also acts on the frication noise and shapes the output. The high frequencies decay more readily and all of this can be seen in the power spectra (and is known from the study of wave propagation and communication via electrical signals). Now the voiced plosives will have a more complicated spectral density since the output is a result not only of periodicity but also white noise (idealization for frication noise). And this is also easily corroborated [see Edwards]. The spectrum of the voiced plosives is jagged in all cases (it reminds one of von Karman vortex streets). The differences in the rates of closure for the voiceless or voiced stops is also known [Allen & Norwood, 1988, Flege, 1988, etc]. Thus the placings of the voiced and voiceless stops/plosives is motivated by experimental evidence. The differences in the spectral output of the voiced fricatives is also relatively easy to explain using the same ideas. The spectral densities do not show the duration, hence cannot show that the spectrum of a voiceless plosive is of very short duration (like a shock wave) and that the voiceless fricatives are steady-state. Essentially, the duration is taken care of indirectly in this space since the Y-axis is defined as a derivative (impulse) but this mathematical model translated directly into the acoustic domain as described above since its acoustic correlate is indeed white noise (or frication noise). The placement of the plosives/stops is not an accident, and the dual nature of the Y-axis is also clear. It can't be any other way. The dimensionless numbers produce this untangling of the complex phenomena. It is clear that more dimensions have to be added to be able to handle geminates, ejectives, trills voiceless vowels etc. Only three dimensions were chosen to make use of our human intuitive grasp of three dimensional spaces. More dimensions the merrier. The usefulness of this space can easily be expanded by extending it to 7-8 dimensions. Released/unreleased distinctions are already treated since the Y-axis is the absolute value of the some kind of a weighted average of the magnitude. The only reason an average like the rms is suggested is to take care of the case in which the articulators (say the lips) go through one complete cycle where the derivative might increase and then decrease. If the rms value was not used, then we'd compute the value of zero. The released/unreleased opening/closing etc. are lumped into a single dimension only in the positive semi-infinite axis. Separating them will have to be accomplished by probably going to complex numbers. Nasality is also a problem like making the [s] & [š] and [z] & [ž] distinction. Nasality would require another dimension, say for the *nasal murmur*. The differences in the placements of these, as can be seen from Edwards' spectral densities is that they essentially have to do with the shape of the spectrum.

The Z-axis is essentially binary (or at best ternary, considering the transition). It divides the Z-axis into resonant (peaked, or compact) and noisy (flat, or diffuse) spectral shapes. The transition zone has both peakedness and noise. These have been forcefully collapsed into a single dimension in order to avoid having to draw high dimensional spaces on two-dimensional paper. In reality they'd be much better represented as a separate dimensions. The mathematical definitions of the dimensions are inadequate for the task; they'd have to be slightly modified to take these into account. The place of maximum power has been used as a secondary consideration (as a part of the interpretation). Since the Z-axis already divides the phones into resonant (peaked) and noisy (flat spectrum) and since it just so happens that the vowels have their energy towards the lower frequencies, then for the fricatives and stops, the place of the maximum peak of fricative power was used to distinguish between [s] and [š] and [z] and [ž]. So I used an idea similar to Fant, in that I used extra information to modify the definitions of the dimensions. They're not very clearly done but it's good enough as a first approximation. There's a similar problem with forcing the vowels along a single dimension when three are needed. The vowels (which require at least 2-3 dimensions) have been squeezed into a single dimension so that they can be fitted into this 3-D phase space not only to show the properties of the space but also to derive the sonority results and to point to the regularity of language acquisition (really only the phones) in this space. More accurate and more complete space would require about ten dimensions and the results would have to be derived using more sophisticated mathematical tools. Because of these difficulties, it was much easier to use an already existent dimensionless number (Reynolds number) to represent this dimension than to provide the articulatory and mechanical parameters. In all likelihood, a combination of Reynolds number along with Strouhal number will provide a better fit with data. And of course, by extending the dimensions to about ten we'd have a much better description of the system which while mathematically more accurate will lack the intuitive obviousness and attractiveness of the simple idealized three dimensional space presented in this chapter.

In order to make a better case for the placements of the phones/phonemes it is necessary to provide the published experimental data. For example, the placement for /s/ and /š/ is not ad hoc as it might seem at a first glance; none of the placements are ad hoc. They were all carefully placed from reading, re-reading and re-re-reading of phonetics books and journal articles. Short of running exhaustive statistical tests on data collected from a large sample that was the best that could be done, and that's the main reason for being able to obtain only relative positional information about the consonants and not being able to place them according to real data (i.e. numbers.) Without numbers, one cannot experiment with various functional forms to see which types of functions perform best and lead to more consistent results. These numbers (i.e. data) don't exist in the literature. Some of this data have to be collected over the course of the future's experiments with the specific intent of trying to see how [at least some of the more common and less exotic] consonants could be placed in this space. It will take time and there will probably be several types of appropriate functions.The results are due to Stevens [1985] and Delattre et al [1964] and can be seen in Lieberman and Blumstein [1988]. Changing (moving) the peakedness (center of the frication energy) from 5 KHz to 2.5 KHz changes the perception from [s] to [š]. Similar problems occur with the type and place of fricative energy; [s] and [š] have greater amplitude than [f] and [θ] [Stevens,1961]. It also depends on the relative magnitude of the frication compared to the signal (i.e. the spectral peaks of the vowel).

Of course, many simplifications are introduced by the 3-D model. As another example, the "explanations" only say that if a certain path in this space gets cut short it passes through another region. One example says that [sy], say in mission, if cut short goes through [š] and therefore we get /mišïn/ instead of /misyin/; that's all. There are many of these. In fact there are much more say, [mr] > [mbr], [kt] > [ç]. For example, the abrupt rise in amplitude at the consonant release are perceived as stops and a gradual rise as a continuant [Shinn & Blumstein, 1984]. Changing the frication noise can also result in [š] being perceived as a [ç] [Cutting and Rosner, 1974]. Obviously, these must be modeled as "colored noise" not "white noise". That means that the definitions of the X,Y and Z as given in this chapter are idealizations; they can be fixed up with minor twiddling or via extending the 3-D space to five or six, in addition to the 3 for the vowels, so that we are really discussing looking into about ten dimensions. If put all together, then it would be a "grand theory". This space would be very difficult to visualize and only statistical tests could determine the placements of the phones. What has been done in this chapter is that published data, some of which are not totally adequate to produce a convincing argument for the existence of this phase space without interpolation and extrapolation, have been used to create a simple version of this n dimensional phase space for human speech. The alternative would have been to wait (perhaps forever) for the data to appear in journals. But in any case, many scientists have never performed any experiments. They have only given mathematical descriptions of others' experiments. Einstein described Brownian motion and the photoelectric effect, Newton used Kepler's data, so what has been done here is nothing out of the ordinary. This space similarly has been constructed to provide the simplest coherent mathematical space that can be used as an idealized construct to further develop more complicated, more accurate and more sophisticated spaces and to provide a unified perspective for future directions of research and data collection.

The basic notion that is used, the notion of *dynamic* and *steady state* is certainly a simplification. So is the concept of vowel vs. consonant. If a two-way split is good, then a four way split is even better. And the way of distinguishing vowels and consonants is most certainly very easily done using the criteria that is used in this paper. It has to do with making discrete articulations; a vowel can certainly be sustained as long as a human has enough air left in his lungs without making any movements of the articulators. Obviously the initial conditions don't count; it has to start someplace. All the consonants don't have this property, especially the plosives. Of course, the fricatives are steady state but they also don't display the distinct spectral peaks of vowels. Therefore the sounds are divided into four groups; simply extending the binary *consonant-vowel* distinction to semi-vowels (which already forms a part of linguistics) and to quasiconsonants (essentially a steady state consonant). It's all a simplification along the lines of those made by linguists and scientists for centuries and still being made by them.

As another simplification which is used to produce coherence, we note that vowels like /v/ and /z/ are what might be called *schwa-colored*. The statement comes directly from the spectral density of /v/, /z/, the source & filter model, the spectral density of the schwa-like sounds [Edwards,1992], the fact that high frequencies do not travel with the same velocity as low frequency ones (except in simplified linear models in undergraduate texts), and the fact that nonlinear distortion is known to be a chief source of mischief in communication lines. It has to do with the ability to know something about the signal and the process that created it by looking at the power spectral density of the signal. The exponential decay of the spectral density of certain sounds is too obvious to need pointing out. Comparing this to some others where we don't see it tells us a lot about the filter that produced it. The schwa (and its relative /Λ/ in American English) very clearly shows its lack of a filter because of its almost perfect exponential decay [Edwards,1992]. It's not too surprising since all the articulators seem to be in their most neutral position for this phone/phoneme. All we have to do is to take this spectral density and add some noise and we'll see that as the noise level starts increasing the spectrum will start to resemble the nonstop voiced fricatives like [l],[v],[r],[z],[ð]... It's just that some of them have less frication noise. It's not difficult to visualize why this happens from an articulatory perspective. Obviously there are those that might be put in different categories but we don't have much choice except to force them into different pigeonholes when we're generalizing and producing relationships.

As regarding things called degrees-of-freedom [for example /l/ and /z/] the problem is that the simplification is not referring to these sounds for the discussion of normal speech. Obviously the articulators are in constant motion which is very difficult to describe mathematically; otherwise it would have been done by now. Every phonetics theory deals with simplifications. During the production of say, the /l/, we're not discussing the fact that the tongue can hit different spots for a *clear-l* or a *dark-l* or that in some languages the tongue can take on strange slow motion maneuvers like the Russian affricate /ç/ [/tʃ/ for those that prefer it]. The fact is that the target articulation that produces the /l/ can be held in a steady-state, just like a vowel articulation. This cannot be done with a stop/plosive, for example, since if the articulation is held there would be no sound. Even if the tongue moves slightly the essential positions can be held in steady-state and in these positions because of the positioning of the jaw, lips etc., the sound that comes out in the steady state is most closely related to the sound that comes out in the most neutral position. This can be observed in the spectral density of /v/,/z/, etc. in published works. Obviously, the noise level is high and it's hard to see the spectral peaks but the spectral density decays smoothly and exponentially, unlike the fricatives like /s/, or /š/ which are flat or increasing, and unlike the plosives whose spectra are jagged. The only thing that comes close is /ð/ which considering the articulation is not surprising, therefore it could have been included in the list. In fact all the voiced non-stop consonants have similar spectra. The voiced stops display both white noise (only as an idealized property since it's really as colored noise) and also periodicity (resonance); as the spectra is jagged somewhat like an irregular sawtooth curve [Edwards,1992].

As for the boundaries of this phase volume, the sounds described in the chapter do in some way prescribe the boundaries of the volume in this 3-D space. A language like Arabic lacking a [p] doesn't mean the end of everything; other bilabials such as the [b] or [m] will take its place on the boundary. In the later part of the chapter, the usefulness of the space is shown for the particular order of learning of speech sounds by children. One of the first is a bilabial [p], [b], or [m]; all three could have just as easily been used instead of [p].

Similarly the space is useful for demonstrating other phenomena; if the duration of a vowel before a consonant is short, as say in Arabic, it requires a quick traversal in the phase space and since the space does double-duty (in acoustic and articulatory space) it's then required that something should *absorb* the physical momentum of the articulators. And this effect shows up a change in the trajectory in this phase space. If the vowels tend to get lengthened before voiced plosives as in English; that requires the same *motion-in-phase-space* and *conservation-of-momentum-energy* type explanation. It is the very nonlinear constitution of the dimensions of this space that allows the decoupling of twisted nonlinearity of the real phenomena of speech and at the same type allows the space to have the dual (acoustic and articulatory/geometric) properties. Thus the space "explains" too many phenomena to be doing it by accident; that after all is what gives us confidence about science. However it should be repeated again that this space is a simple space, an idealized model of a much more complex reality and behaves something like the *ideal gas* equations.

In the final analysis, there is nothing fanciful about the three dimensional (dimensionless number) space for the consonants and vowels. It is the very complexity of the fluid phenomena that makes dimensional analysis so powerful and useful in that field. The fact is that the space that is produced by using dimensionless numbers is able to represent very complex phenomena in a very simple and intuitive way. It is the very nonlinearity that is introduced by using multiplications and divisions of the parameters that unskewes the complex web of tangled fluid phenomena and allows the experimental physicists and hydrodynamicists to fit "nice" curves to their experiments. It is this amazing power of dimensional analysis that allows this simple space which I constructed to have both *articulatory* and *acoustic* interpretations. There is something to this space and it stands as a simple linear three dimensional space which can be used to unify many "stylized facts" of linguistics.

There could possibly be a brute force approach that could yield better results using dimensionless numbers. It might even produce more dimensionless numbers that could be useful for speech. However it’s likely that the dimensionless numbers of fluid mechanics such as Froude, Euler, Weber, Mach, Prandtl, Eckert, Grashof would somehow show up. The fact that the Strouhal number has to to with oscillating flow could make it useful in conjunction with Reynolds number for the Z-axis. The X and Y axes could certainly be improved upon and only experiments can decide the exact shape of the space. We might try some kind of a combination of the Reynold’s number and Strouhal number for the air quality axis of the space. The alternative is to keep plotting variables against time until doomsday; that won’t accomplish anything. There are very fundamental concepts in physics. One of them is that any equation must be dimensionally homogeneous. That’s the reason for the power of dimensional analysis; it produces combinations of parameters among which experimental relationships must be sought.

References

Abry C., L. Boe and J. Schwartz, *Plateaus, catastrophes and the structuring of vowel sytems,* Journal of Phonetics 17, 1989 pp. 47-54.

Allen, G. and J. Norwood, *Cues for Intervocalic /t/ and /d/ in Children and Adults,* JASA 84, Sep. 1988, pp. 868-875.

Anderson, S., Phonology in the Twentieth Century, Univ. of Chicago Press, 1985.

Anderson, S., The Organization of Phonology, Academic Press, New York, 1974.

Anderson S., and C. Ewen, Principles of Dependency Phonology, Cambridge, Cambridge University Press, 1987.

Assmann, P. and Q. Summerfield, *Modeling the Perception of ConcurrentVowels: Vowels with Different Fundemantal Frequencies*, JASA 88, Aug 1990, pp. 680-697.

Atal, B., M. Schroeder, *Linear Prediction Analysis of Speech Based on a Pole-Zero Representation,* JASA 64, Nov 1978, pp. 1310-1318.

Bailly, G, R. Laboissiere and J. Schwartz, *Formant trajectories as audible gestures: an alternative for speech synthesis*, Journal of Phonetics 19, 1991 pp. 9-23.

Banks, S., *Signal Processing, Image Processing and Pattern Recognition,* Prentice-Hall, 1990.

Behrens, S. and S. Blumstein, *On the Role of the Amplitude of the Fricative Noise in the Perception of Place of Articulation in Voiceless Fricative Consonants,*JASA 84, Sep 1988, pp. 861-866.

Brainerd, B., Introduction to the Mathematics of Language Study, Elsevier, New York, 1971.

Broad, D. and H. Wakita, *Piecewise Planar Representation of Vowel Formant Frequencies,* JASA 62, Dec 1977, pp.1467-1473.

Browman, C. and L. Goldstein, *Gestural specification using dynamically-defined articulatory structures,* Journal of Phonetics 18, 1990 pp. 299-320.

Carre, R and M. Mrayati, *Vowel-vowel trajectories and region modeling*, Journalof Phonetics 19, 1991 pp. 433-443.

Catford, J.C., Phonetics, Clarendon Press, London, 1988.

Chiba, T, and J. Kajiyama, The Vowel: Its nature and its structure, Tokyo,Tokyo-Kaiseikan Publishing Company, 1941.

Clark, J. and C. Yallop, Phonetics and Phonology, Blackwell, Oxford, 1990.

Damasio, A. and H. Damassio, Brain and Language, Scientific American, Sep. 92, p. 88.

Dayhoff, J., Neural Network Architectures, Van Nostrand Reinhold, 1990.

Edwards, H., Applied Phonetics: The Sounds of American English, Singular Pub., 1992.

Fant, G., Acoustic Theory of Speech Production, Blackwell, Oxford, 1990.

Fischer, R. and R. Ohde, *Spectral and Duration Properties of Front Vowels as Cues to Final Stop-Consonant Voicing,* JASA 88, Sep 1990, pp. 1250-1259.

Flanagan, J., Speech analysis, synthesis and perception, New York, Springer-Verlag, 1972.

Flege, E., The *Development of Skill in Producing Word-Final English Stops: Kinematic Parameters*, JASA 84, Nov 1988, pp. 1639-1652.

Foley, J., Foundations of Theoretical Phonology, Cambridge Univ. Press, Cambridge, 1977

Fromkin, V. (ed), (1985), Phonetic Linguistics: Essays in Honor of Peter Ladefoged, Academic Press, New York.

Ganong, W., R. Zatorre, *Measuring Phoneme Boundaries Four Ways*, JASA 68, Aug 1980, pp. 431-439.

Goldsmith, J., Autosegmental and Metrical Phonology, Blackwell, Oxford, 1990.

Gopal, H., *Effects of speaking rate on the behavior of tense and lax vowel durations,* Journal of Phonetics 18, 1990 pp. 497-518.

Greenberg, J. (ed), Universals of Human Language: Phonology, Stanford University Press, Standford, 1978.

Hankamer, J. A. Lahiri and J. Koreman, *Perception of consonant length: voiceless stops in Turkish and Bengali*, Journal of Phonetics 17, 1989 pp. 283-298.

Hawkins, J.(ed), Explaining Language Universals, Basil Blackwell, New York, 1988

Henton, C., *One vowel's life (and death ?) across languages: the moribundity and prestige of /L/,* Journal of Phonetics 18, 1990 pp. 203-227.

Hillenbrand, J., G. Canter and B. Smith, *Perception of Intraphonemic Differences by Phoneticians, Musicians and Inexperienced Listeners*, JASA 88, Aug 1990, pp.655-662.

Hubey, H.M., (1996a) *Catastrophe Theory and Speech*, submitted to Journal of Nonlinear Dynamics, Psychology, and Life Sciences.

Hubey, H.M., (1996b) *Speech Realization, Fuzzy Sets, Differential Equations and Categorical Perception*, submitted to Journal of Nonlinear Dynamics, Psychology, and Life Sciences

Hubey, H.M.,(1993) *Psycho-socio-economic Evolution of Systems* , in Mathematical Modelling and Scientific Computing, (X. Avula, ed.), Principia Scientia, St. Louis.

Hubey, H.M., (1996) Mathematical and Computational Linguistics, to be published. First edition published in 1994, Mir Domu Tvoemu, Moscow, Russia, ISBN

Hyman, L., Phonology: *Theory and Analysis*, Holt, Rinehart and Winston, New York, 1975.

Jackson, E., *Perspectives on nonlinear dynamics 1*, Cambridge University Press, Cambridge 1991.

Kelso, J., E. Saltzman and B. Tuller, *The Dynamical Perspective on Speech Production: Data and Theory*, Journal of Phonetics, 1986, Vol 14, pp. 29-59.

Kelso, J., E. Saltzman and B. Tuller, *Intentional Contents, Communicative Context, and Task Dynamics:* a Reply to Commentators, Journal of Phonetics, 1986, Vol 14, p. 171-196.

Kewley-Port, D. and B. Atal, *Perceptual Differences Between Vowels Located in Limited Phonetic Space*, JASA 85, Apr 1989, pp. 1726-1740.

Klatt, D., *Software for a Cascade/Parallel Formant Synthesizer,* JASA 67, Mar 1980, pp. 971-995.

Ladefoged, Peter, A Course in Phonetics, Harcourt, Brace & Jovanovic, 1975.

Ladefoged, Peter, Elements of Acoustic Phonetics, Univ. of Chicago Press, Chicago, 1962.

Ladefoged, Peter, Preliminaries to Linguistic Phonetics, Univ. of Chicago Press, Chicago,1971.

Ladefoged, P. and I. Maddieson, *Vowels of the world's languages*, Journal of Phonetics 18, 1990 pp. 93-122.

Ladefoged, P., *Some reflections on the IPA*, Journal of Phonetics 18, 1990 pp. 335-346.

Lass, Roger, Phonology, Cambridge University Press, Cambridge, 1984.

Lieberman, P., The Biology and Evolution of Language, Harvard University Press, Cambridge,1984

Lieberman, P, and S. Blumstein, Speech Physiology, Speech Perception and Acoustic Phonetics, Cambridge University Press, Cambridge , 1988

Lindblom, B., *On the notion of "possible speech sound"*, Journal of Phonetics 18, 1990 pp. 135-152.

Lindblom, B. and O. Engstrand, *In what sense is speech quantal?*, Journal of Phonetics 17, 1989 pp. 107-121.

Lindblom, B. and P. MacNeilage, *Action Theory: Problems and Alternative Approaches*, Journal of Phonetics, 1986, Vol 14, pp. 117-132..

Lubker, J., *Articulatory Timing and the Concept of Phase*, Journal of Phonetics, 1986, Vol 14, pp. 133-137.

Lyons, J., Theoretical Linguistics, Cambridge University Press, Cambridge, 1968.

Makkai, V,(ed) Phonological Theory, Holt, Rinehart and Winston, New York, 1972.

Manuel, S.,*The Role of Contrast in Limiting Vowel-to-Vowel Coarticulation in Different Languages*, JASA 88, Sep 1990, 1286-1298.

Miller, J., *Auditory-Perceptual Interpretation of the Vowel*, JASA 85, May 1989, pp. 2114-2133.

Moore, B., R. Peters and B. Glasberg, *Auditory Filter Shapes at Low Center Frequencies*, JASA 88, July 1980, pp. 132-140.

Mowrey, R. and I. MacKay, *Phonological Primitives: Electromyographic Speech Error Evidence*, JASA 88, Sep 1990, pp. 1299-1312.

Nearey, T., Phonetic features for vowels, Indiana Univeristy Linguistics Club, Bloomington, Indiana, 1978.

Nearey, T., *The segment as a unit of speech perception*, Journal of Phonetics 18, 1990 pp. 347-373.

Ohala, J. and J. Jaeger, Experimental Phonology, Academic Press, New York, 1986.

Ohala, J., *There is no interface between phonology and phonetics: a personal view*, Journal of Phonetics 18, 1990 pp. 153-171.

Ohala, J., *Against the Direct Realist View of Speech Perception*, Journal of Phonetics, 1986, Vol 14, pp. 75-82.

Peterson, G., and H. Barney, *Control methods used in the study of vowels*, Journal of the Acoustical Society of America, 24, 1952, pp. 175-184.

Picone, J., *Continuous Speech Recognition Using Hidden Markov Models*, IEEE ASSP Magazine, July 1990.

Repp. B, *Perception of the [m]-[n] Distinction in CV Syllables*, JASA 79, June 1986, pp. 1987-1999.

Schroeder, M., *An Integrable Model for the Basilar Membrane*, JASA 53, pp. 429-434.

Secker-Walker, H. and C. Searle, *Time-Domain Analysis of Auditory-Nerve-Fiber Firing Rates*, JASA 88, Sept 1990, pp. 1427-1436.

Shailer, M., B. Moore, G. Glasberg and N. Watson, *Auditory Filter Shapes at 8 KHz and 10 KHz*, JASA 88, July 1990, pp. 141-148.

Shirai, K, and T. Kobayashi, *Estimation of articulatory motion using neural networks*, Journal of Phonetics 19, 1991 pp. 379-385.

Silverman, H. and D. Morgan, *The Application of Dynamic Programming to Connected Speech Recognition*, IEEE ASSP Magazine, July 1990.

Sinha, N., B. Kuszta, Modeling and Identification of Dynamic Systems, Van Nostrand Reinhold, New York, 1983.

Siski,R., *Stochastic Differential Equations in* Modern Nonlinear Equations (T. Saaty, ed.), McGraw-Hill, New York, 1967.

Sondhi, M. and B. Gopinath, *Determination of Vocal-Tract Shape from Impulse Response at the Lips*, JASA 49, pp. 1867-1873

Stevens, K., and S. *Blumstein, Invariant Cues for Place of Articulation in Stop Consonants*, JASA 64, Nov 1978, pp. 1358-1368.

Stevens, K., *On the quantal nature of speech*, Journal of Phonetics 17, 1989 pp. 3-45.

Strange, W., *Evolving Theories of Vowel Perception*, JASA 85, May 1989, pp. 2081-2087.

Sundberg, J. and B. Lindblom, *Acoustic Estimation of the Front Cavity in Apical Stops*, JASA 88, Sep 1990, pp. 1313-1317.

Sussman, H., *Acoustic Correlates of the Front/Back Vowel Distinction: A Comparison of Transition Onset Versus "Steady State"*, JASA 88, July 1990, pp.87-96.

Syrdal, A. and H. Gopal, *A Perceptual Model of Vowel Recognition Based on the Auditory Representation of American English Vowels*, JASA 79, April 1986, pp. 1086-1100.

ten Bosch, L. and L. Pols, *On the necessity of quantal assumptions. Questions to the quantal theory*, Journal of Phonetics 17, 1989 pp. 63-70.

Traunmüller, H., *Analytical Expressions for the Tonotopic Sensory Scale*, JASA 88, July 1990, pp. 97-100.

Treiman, R., J. Gross and A.Cwikiel-Glavin, *The syllabification of /s/ clusters in Engli*sh, Journal of Phonetics 20, 1992 pp. 383-402.

Visch, E., A *Metrical Theory of Rhythmic Stress Phenomena*, Doris Pub., Providence, 1990.

Van Son, R., and L. Pols, *Formant Frequencies of Dutch Vowels in a Text, Read at Normal and Fast Rate*, JASA 88, Oct 1990, pp. 1683-867.

Waibel, A., and K. Lee (Ed), Readings in Speech Recognition, Morgan Kaufman Publishers, San Mateo, CA, 1990

White, F., Fluid Mechanics, McGraw-Hill, 1979

Witten, I., Principles of Computer Speech, Academic Press, New York, 1982.

Winter, I. and A. Palmer, *Temporal Responses of Primary-like Anteroventral Cochlear Nuclear Units to the Steady-State Vowel /i/*, JASA 88, Sept 1990 pp. 1437-1441.

Zwicker, E. and E. Terhardt, *Analytical Expressions for Critical-Band Rate and Critical Bandwidth as a Function of Frequency*, JASA 68, Nov 1980, pp. 1523-1524.