Qualitative and Quantitative Dynamics of Vowels*

Thomas C. Bourgeois
Institute for Neurogenic Communication Disorders
University of Arizona
Tucson, Arizona 85721
bourg@cnet.shs.arizona.edu

Richard T. Oehrle
Department of Linguistics
University of Arizona
Tucson, Arizona 85721
oehrle@convx1.ccit.arizona.edu

Introduction

A pervasive and fundamental property of spoken language is the nesting of quasi-periodic structures, ranging from the vibration of the vocal cords to the iteration of syllables, accents, and higher-order prosodic objects. The long-range goal of the research reported here is to bring to bear on the study of this phenomenon the methods and insights of the study of dynamical systems, in the hope that this will increase our understanding of the computation of spoken language. In this paper, we describe this point of view and illustrate the results we have obtained to date in a study of English vowels, within and across individual speakers.
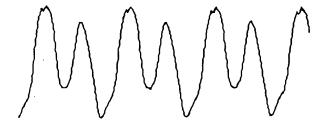
Two perspectives on speech

Speech is a physical event: it is produced by the mechanical actions of the human articulators and propagates itself through physical media. At the same time, speech is the carrier of richly-structured linguistic information. From this latter point of view, speech events constitute tokens of a symbolic system. The fundamental question prompting the research reported here is: what makes this duality possible?

Pretheoretically, an adequate answer to this question should address the interaction in spoken language of variation and stability. Speech is symbolically stable across an impressive range of variation of the physical signal, variation observable within a given speaker, across speakers with a common language, and across the range of dialects and languages. In spite of the scope and pervasiveness of this variation, speech is not entirely arbitrary. For example, the phonological adaptation of speech is not arbitrary: vowel spaces do not cross-cut each other in random fashion. This suggests that the symbolic properties of speech are not the result of a purely conventionalist association between the space of speech sounds and their phonological interpretation. A deeper analysis can be found in the work of Stevens (1972) and Liljencrants & Lindblom (1972). Stevens notes that the vocal tract is constructed in such a way that there are regions of articulatory variation which produce relatively little acoustic variation. Working from the perspective of human action theory, Tuller & Kelso (1991) interpreted Stevens' notion as implying the existence of regions of dynamical stability in the speech production mechanism, and not as implying the existence of invariant acoustic properties within the signal. Liljencrants & Lindblom explore the hypothesis that phonological systems are solutions to the problems presented by ease of articulation, on the one hand, and ease of perception, on the other.
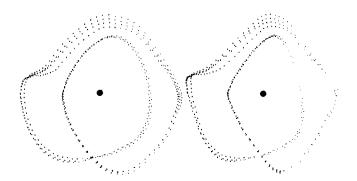
Our approach to the problem focuses on how the intrinsic physical properties of speech are adaptable to the demands of symbolic phonological representation. To explore this question, we base our research on the concepts and methods of dynamical systems (Abraham & Shaw, 1992). Our work investigates the trajectories of vowels in a so-called 'phase space representation' of the associated acoustic wave-

form regarded as a function A mapping time t to amplitude $A(t)$. This reconstruction transforms a point $A(t_0)$ in the wave form to an ordered triple in the phase space of the form $<A(t_0), A'(t_0), A''(t_0)>$ where the y- and z-coordinates correspond to the first and second time derivative; temporally successive values of the function $A(t)$ become successive points in the phase space. Thus time is 'parameterized' in the sense that it is not directly represented in the space, although it can always be recovered by considering only the behavior of the x-axis, which mirrors the original function $A(t)$. The resulting trajectory is a closed (or nearly closed) curve in three dimensions which repeats (or nearly repeats) itself with each glottal cycle. Pitch is also indirectly represented, encoded within the representation as the distance along the trajectory between successive samples of the speech wave (at fixed temporal intervals): higher pitches correspond to more distant successive measurements. In other words, the phase-space reconstruction includes all the information found in spectrographic representations of speech, but normalizes across fundamental frequency variation. An example follows. Consider below a fragment of the waveform produced during one male talker's production of the vowel [u] in the context of the word *who'd*.

The wave form can be characterized qualitatively as having two large peaks, one smaller than the other, which repeat with each period. The phase portrait reconstructed from this wave form appears below in the form of a stereogram (join the two center dots by crossing your eyes to see the three-dimensionality of the resulting image):

In the reconstruction, the largest peak of the wave form constitutes the largest loop, and the smaller peak corresponds to the smaller loop. The trajectory winds twice around the center in the xy plane and twice around the the center in the xz plane. We provide an interpretation of these windings below.
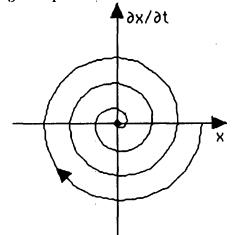
What makes this point of view attractive is, first of all, its physical realism: speech events in fact constitute a dynamical system, and as such, the dynamics of the articulators and the acoustical dynamics they produce in the ambient media around them are directly characterizable as dynamic systems. Equally attractive is that the study of dynamical systems brings together, in a single integrated framework, quantitative and qualitative methods, a feature which has been exploited in the study of physical systems since the pioneering insights of Poincaré. That is, we can study the dynamic aspects of speech to any desired degree of quantitative detail, in the same space that accommodates a non-quantitative, qualitative investigation of behavior. In particular, then, one may identify the phonetic properties of a speech event with the quantitative aspects of its behavior, and ask whether or to what extent the qualitative aspects of this dynamic system support phonological analysis. If these qualitative physical aspects of speech do in fact support phonological analysis, then the simultaneous co-existence of 'phonetic' and 'phonological' properties in the same space provides an interesting alternative to the view that phonological properties are modeled in a discrete space of 'distinctive features' and phonetic realization corresponds to a map from this discrete space to a corresponding space of continuous phonetic parameters. Thinking of phonological properties as the natural qualitative distinctions that exist in the continuous phase spaces of particular speech events makes it

possible to reconcile the apparent abstractness of phonological properties with their intrinsic dependence on such physical parameters as duration, amplitude, and frequency.
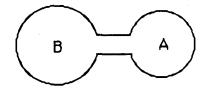
Phase space reconstructions of vowels

As we have said, the phase-space reconstruction makes it possible to study quantitative and qualitative aspects of vowels (in particular, and the full range of speech sounds, in general) in the same space. The qualitative aspects of dynamic behavior correspond to fundamental properties of attractors within the phase-space. The presence of such attractors is revealed by stability in the phase-space trajectory. A continuous phase space can support discretely-structured forms of stable behavior. Thus, one and the same trajectory may be studied from the point of view of the continuous space or from the point of view of the discrete parameters which control the shape of that trajectory. It is this basic duality which we seek to exploit.

An attractor represents a natural limit of a phase-portrait. For example, consider the behavior of a damped pendulum which swings through a series of decreasing arcs until it eventually comes to rest. A phase-space reconstruction of its behavior consists of the set of points (x, y) in the plane, where the x-coordinate represents the displacement---positive or negative---of the pendulum at any given point in time, and the y-coordinate represents its velocity. Since the pendulum swings with decreasing displacement and correspondingly decreasing velocity, its phase-portrait consists of an arc spiraling through the phase space and ending in the origin---a point attractor.

Other kinds of attractors are possible. The attractor for a bowed string, for example, is periodic: a closed curve in the planar phase space. A fundamental question in investigating speech as a dynamic system is the character of the attractors in the phase space.
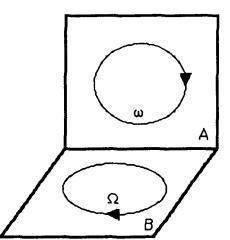
The double-Helmholz resonator model of the vocal tract provides a convenient and straightforward means to introduce the geometry of the vowel trajectory in phase space, the torus. The torus is the product $S_1 \times S_1$ of two circles. Thus, the two dimensions model the oscillatory properties of the two chambers, while the trajectory in the product of the dimensions models their coupling for a given value of their controlling parameters (that is, by hypothesis, for the phonetic value of a particular vowel)[1]. Consider the double resonant cavity schematized below:
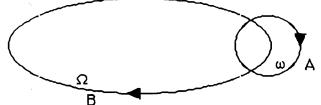

Here, the two chambers A and B are coupled to each other by a connecting tube. For the moment, let us imagine that there is no coupling between the two chambers. Let chamber A have a single resonant frequency ω and chamber B have a different resonant frequency Ω. We can then reconstruct the phase space trajectory as a circle in a plane whose points are determined by the ordered triple

---

[1] Because the trajectory is derived from the actual acoustic waveform, our interpretation of such a trajectory is not in principle restricted by limitations due to the simplicity of the double-resonator model. By changing the parameters of reconstruction we can easily embed the resulting trajectories in a state space of arbitrarily higher dimension. That is, we can relate the trajectory not to a line on the surface of a torus but rather to a rope on that surface, or we can think of this rope as inhabiting the space enclosed by the torus, rather than constraining it to occupy the surface of that torus. Such extensions are straightforward; whether they would be required in a more adequate model remains an open question at this time.
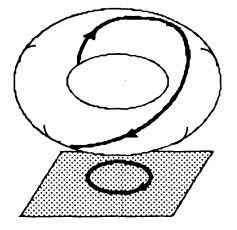
<sin τ, cos τ, −sin τ>; τ = ωt. Similarly, we can reconstruct the phase-space trajectory for B as a circle in a plane whose points are determined by <sin T, cos T, -sin T>; T = Ωt. Because the resonant frequencies ω and Ω are disparate, we can consider the two planar phase spaces to be orthogonal to each other, as below:

If we now translate the two phase space trajectories and adjust the scale appropriately, it is easy to visualize that the space traced out by a point simultaneously constrained to move along the curve described by A and the curve described by B will be the surface of a torus.

An example of such a trajectory is shown below.

The fundamental hypothesis of this paper is that while the vowel space is acoustically continuous, the shape of the trajectories within the phase space representation corresponding to vowels of different quality are topologically distinguishable, and that the trajectories corresponding to vowels of the same quality across talkers are homeomorphic---that is, topologically indistinguishable. We are particularly interested in trajectories which are periodic with respect to both dimensions of the torus and their coupling. Topologically, these trajectories are torus knots of type (m,n), where m and n are relatively prime and m represents the period of the trajectory with regard to one of the circular dimensions of the torus and n represents the period of the trajectory with regard to the other circular dimension (see Crowell & Fox, 1993). These two parameters may be coupled in distinguishable ways as well. Even with m and n quite small, this space of possibilities gives rise to complex varieties of behavior which can be distinguished on simple, discrete grounds. Considering speech within the phase space representation, then, might provide insight into the continuous/symbolic duality which exists in both the production and perception of natural language.

Data

We have collected the vowels [i I E ae ^ u U o a] within the context h_Vd from four adult males, four adult females, and two children (a boy, 11, and a girl, 9) during separate recording sessions, and stored the productions on a digital audio tape sampling at 44 kHz with 16 bits quantization. We then resampled these tokens onto a PC using a separate A/D converter at 22kHz with 8 bits quantization. In order to study the dynamics of these vowels both within the glottal cycle as well as within the syllable, we extracted pitch-periods (one iteration of the closed curve in the phase space) from three regions of the syllable: the first quarter, the middle, and the third quarter. Each of the resulting arrays was transformed into the phase space using the method described in Gibson et al. (1992). At the time of writing, we present results obtained from analyzing one adult male (D.B.) and two adult female talkers.

For reference to other kinds of speech analysis, the table below compiles the fundamental frequency and formant measurements for the three speakers we have

analyzed as a function of vowel quality, averaged across the three positions in the syllable.

Insert table 1 here.

These measurements are in general accord with those presented in Peterson & Barney (1952), and suggest that the vowels within our corpus are phonetically unremarkable.

The phase space trajectories of nine vowels from the male talker (D.B.) are given below:
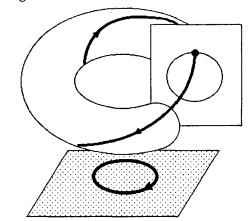
Insert figure 1 here.

As our topology suggests, we discuss two qualitative parameters which serve to distinguish the phase portraits from each other: the smoothness of the trajectory, which we can take to be related to the winding number around the 'tube' of the torus, and the number of trips each trajectory makes about the origin, the 'circle' of the torus. Descending in height across the inventory, the trajectories of those vowels produced with central or posterior articulation [u U o ^ a] orbit the origin in a characteristically smooth fashion relative to their anterior counterparts [i I E ae whose portraits show a great deal more local activity. We can interpret this local activity as an increase in the number of loops around the smaller diameter of the torus. In other words, the winding number m is large for front vowels, and small for back vowels. Comparing now across the trajectories of vowels, the number of rotations around the origin increases as tongue height is lowered. That is, high vowels such as [i I u U] show fewer trips around the origin than low vowels [ae a]. In terms of the state space, we can say that the winding number for the large diameter of the torus n increases as vowel height decreases.

Consider now the following phase space reconstructions of the vowel /u/ spoken by two female talkers (upper two phase portraits) and two male talkers (lower two phase portraits):

Insert figure 2 here.

With regard to the two qualitative parameters we discussed above, certain similarities are preserved within the vowel trajectories across talkers. The winding number of the small ring (i.e. around the 'tube') is small, as is the winding number of the large ring (i.e. around the 'circle'). These qualitative similarities of the winding number can be made quantitative by counting them, using a technique developed by Poincaré (we will consider only the winding number of the 'circle' here).

The Poincaré section provides a means of simplifying the dynamics of a phase portrait by considering not the whole path within the space, as we have done above, but rather a plane which intersects the phase space such that all of the trajectories pass through it. Consistent with common practice, we choose the plane associated with the phase-zero point of the (large) oscillator and, for a given glottal cycle, count how many times the trajectory passes through the plane in a single direction. By this method, we obtain for the continuous trajectory a discrete observation of its winding number. For the example below, the Poincaré section contains a single point, so the winding number would be 1.

The following table summarizes the results we have obtained for Poincaré sections of phase space reconstructions of pitch periods exerpted from 1/4, 1/2, and 3/4 of the way through the vowel portion of the syllable, σ.

		i	I	E	ae	u	U	o	^	a
σ/4	D.B.	1	2	3	4	2	3	3	4	5
	S.O.	1	2	3	4	1	2	3	5	5
	L.W.	1	2	3	3	1	2	3	3	5
σ/2	D.B.	2	2	3	3	2	3	3	4	6
	S.O.	2	3	3	4	1	3	3	3	4
	L.W.	1	2	3	4	2	2	2	4	4
3σ/4	D.B.	2	3	4	5	2	3	3	4	6
	S.O.	2	3	3	4	2	2	2	3	6
	L.W.	1	3	4	4	2	3	2	4	4

As the data illustrate, the winding number increases with a decrease in vowel height, consistent with our qualitative observations for a single speaker. Because the winding number is a measure of trips around the torus and therefore an integer, it provides us with a means of discretizing the vowel space in a way which is not completely arbitrary, but rather reflects the internal structure of the trajectories through the state space itself. Within this space, the trajectories can be grouped together as members of an equivalence class which itself is a function of the controlling parameter of vowel height. Specifically, high vowels [i I u U] can be thought of as being associated with trajectories of winding number $n \leq 2$, mid vowels are associated with trajectories of winding number $3 \leq n \leq 4$, and low vowels are associated with trajectories of winding number $n \geq 5$. Much of our data conform to this generalization with only a few outliers. As for those data which do fall outside of this grouping, it is important to remember that the data given are based on a single Poincaré section for a single pitch period of the relevant vowel. A more thorough analysis would undoubtedly include both P-sections for the phase angles $(0, \pi/2, \pi, 3\pi/2)$ as a means of distinguishing local behavior near the phase plane from the global properties of the trajectory, and for additional pitch-periods in the signal.

We predict then, that a change in the winding number n for a vowel reconstructed as a phase space trajectory will correspond to a change in the perceived phonetic category of the vowel; successively larger values of the winding number n correspond to succesively lower vowel height categories. In the feature system of SPE, a change from $n = 2$ to $n = 3$ corresponds to a change from [+high] to [-high], a change from $n = 4$ to $n = 5$ corresponds to a change from [-low] to [+low].

A reviewer has correctly pointed out to us that the winding number around the large diameter of the torus appears to be correlated with the number of harmonics between the fundamental frequency and the first vowel formant (although it remains to be confirmed, by extension it is most likely the case that the winding number around the 'tube' of the torus is correlated with the number of harmonics between the fundamental frequency and the second vowel formant). Because we have described this number as a means for evaluating the perceived articulatory height of the vowel, it

seems appropriate to determine whether or not there is a precedent in the literature for an interaction between the fundamental frequency and the first formant either in vowel production or vowel perception. The following brief chronology features the highlights of our investigation into this question.

The interaction between F_0 and F_1

Since the early 1950's researchers have observed an interaction between fundamental frequency and vowel perception. Potter and Steinberg (1950), who measured the vowels of male, female, and child speakers, found that an increase in fundamental frequency across talkers was correlated with an increase in the absolute frequency values of the formants within a particular vowel category. While they suggested that fundamental frequency variation might offer a means for normalizing formant frequency values, they decided it was "a dubious possibility" since, among other things, formants are a product of the physical aspects of the vocal tract and have little dependence on fundamental frequency. However, they also found an effect of a change in fundamental frequency on the perception of synthetic vowels whose formant structure remained constant: as fundamental frequency was increased, the perceived frequency of the first formant decreased. That is, a vowel whose formant structure corresponded to a male [a] was perceived as an [a] when synthesized with the fundamental frequency of a male, but as a (child's) [O] when synthesized with the fundamental frequency of a child. Similarly, a vowel whose formant structure corresponded to a male [ae] was perceived as an [ae] when synthesized with the fundamental frequency of a male, but as a (child's) vowel somewhere between [ae] and [E] when synthesized with the fundamental frequency of a child. They report further evidence, albeit anecdotal, that when helium was used as a propagation medium for adult male vowels or an artificial larynx was used to excite the vowel formants of a child (thus raising or lowing, respectively, the fundamental frequency of the subject while leaving the vocal tract constant), that a speaker will "make adjustments in his formant frequencies in order to maintain a given vowel sound." Similar findings are also to be found in Peterson (1961), who reports, again anecdotally, that "if a man raises his fundamental voice frequency to

correspond to that of a child (falsetto), and the higher formants are removed by filtering, the acoustical result corresponds very closely to the |O| of a child with low-pass filtering and may be so interpreted by a listener." More systematic studies of the effects of F_0 on perception of vowels were conducted by Miller (1953), Fujisaki & Kawashima (1968), and Carlson et al. (1975). Each of these studies reported a similar shift in the perceptual boundary between vowel categories as fundamental frequency was changed: an increase in fundamental frequency leads to a decrease in the perceived value of the first formant, and thus an increase in the perceived articulation height of the vowel. In sum, several studies have indicated that a person's fundamental frequency interacts with both vowel production and vowel perception, and that the product of this interaction appears to be under the control of the speaker to some degree.

The observations reported in Scott (1976) provide some insight into predicting the impact on vowel perception of the interaction between F_0 and F_1. Scott explored the perceptual consequences of manipulating the temporal fine-structure of vowel waveforms, and found that the perceptual boundary dividing a synthesized continuum whose endpoints were /i/ and /I/ was correlated with a change in the number of positive-going slopes in the waveform: those stimuli with two positive-going slopes were categorized as /i/ while those with three positive-going slopes were categorized as /I/. In a follow-up experiment, F_0 and F_1 were manipulated in three synthetic continua to produce a series with a waveform change from three cycles of F_1 per fundamental period to four cycles of F_1 per fundamental period at different points along the seven-step continuum. For those continua whose waveform changes occurred near the category boundary (the region where the tokens became ambiguous), the position of the boundary shifted to the stimulus where the waveform change occurred. This research suggests that, at least for ambiguous vowels (i.e., those near the boundary of two distinct phonetic categories), category membership can be decided on the basis of the temporal fine-structure of the wave form.

The Scott study bears close similarity with the dynamic approach discussed here. The temporal fine-structure that Scott manipulated has a direct correlation to the winding number of the phase space trajectory. Specifically, those tokens which contain an extra cycle of F_1 per fundamental period are also those whose winding number n increases by 1. As Scott observed, such a change was detectable perceptually, and its detection corresponds to a change in the mapping of the acoustic stimulus from one phonetic category to another. Relating this to the table of values from the Poincaré analysis, one can see that in natural speech, the change between /i/ and /I/ for a given speaker is consistent with an integral increase in the winding number. Hence, Scott's prediction that temporal fine structure is correlated with the perceptual phonetic category of a synthetic vowel is borne out in our natural speech data as well.

Put in phonological terms, the perceptual distinction between /i/ and /I/ which Scott investigated is typically characterized not as a change in vowel height, but rather as a change in the value of the [tense] or [ATR] feature. Because we have limited our analysis to distinctions of vowel height and not other dimensions which delimit the vowel space, there are clear differences between that study and the predictions we make here; we cannot comment on the extent of those differences at this stage in our research. More work will certainly be required to verify the connection between the parameters distinguishing the vowel space and those aspects of vowel (production) dynamics represented within the wave form. However, the connection between our approach and Scott's results (and the legacy of research which precedes it) is compelling. It suggests the otherwise unanticipated result that the oscillator driving vowel production (the glottal source which produces F_0) and the resonant cavity which determines vowel quality are entrained (coupled) in frequency.

Conclusion

In this discussion we have provided only a very cursory analysis of a small set of talkers, but it nevertheless illustrates the potential power which this theoretical perspective can have as a tool for resolving the continuous/discrete duality we mentioned above. It is important also to note that this technique of phase space reconstruction and subsequent P-section analysis can be obtained without any specialized hardware beyond that needed to discretize the wave form

itself and does not rely on the Fourier transform. As an analysis toolkit, then, this approach offers an augmentation to current spectral analysis techniques by reducing some of the cross-talker variation that such techniques cannot abstract away from via a 'vocal-tract internal' means of normalizing across differences in talkers and situations.

As a final consideration along these lines, we point out two additional curiosities about speech that may also succumb to analysis under the dynamical perspective. First, as early as 1947 French and Steinberg showed that speech could be either low-pass filtered or high-pass filtered at 1.9 kHz while retaining around 68% of its intelligibility. This suggests that the global structure of the vowel's dynamics may in fact be retained in spite of the filtering process at this 'magic' frequency. If this proves to be true, then our approach offers a unique perspective from which a straightforward account of this phenomenon can be obtained. Second, Licklider and Pollack (1948) showed that speech subjected to differentiation followed by infinite peak clipping (which preserves only the zero crossings of the wave form) was also highly intelligible--in fact, about 90% intelligible. Although we have not explored this fully, such a transformation seems intuitively related to the Poincaré section analysis which we have provided above.
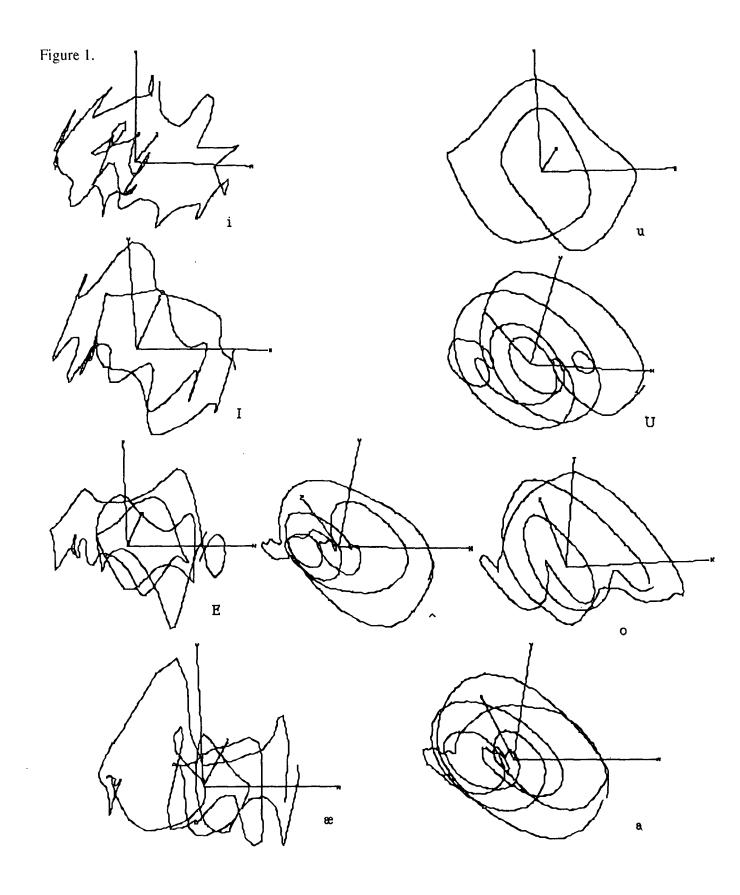
References

Abraham, R., & C. Shaw. (1992). Dynamics: The Geometry of Behavior. Redwood City: Addison-Wesley Publishing Company.

Carlson, R. Fant, G., & B. Granstrom. (1975). Two formant models, pitch and vowel perception. Auditory Analysis and Perception of Speech, G. Fant and M. A. A. Tatham, eds, London: Academic Press.

Crowell, Richard H., & Ralph H. Fox. (1963). Introduction to Knot Theory. New York: Ginn and Company.

French, N. R. & J. C. Steinberg. (1947). Factors governing the intelligibility of speech sounds, JASA, 19, 90-119.

Fujisaki, H. & T. Kawashima. (1968). The roles of pitch and higher formants in the perception of vowels, IEEE Transactions of Audio Electroacoustics, AU-16, 78-80.
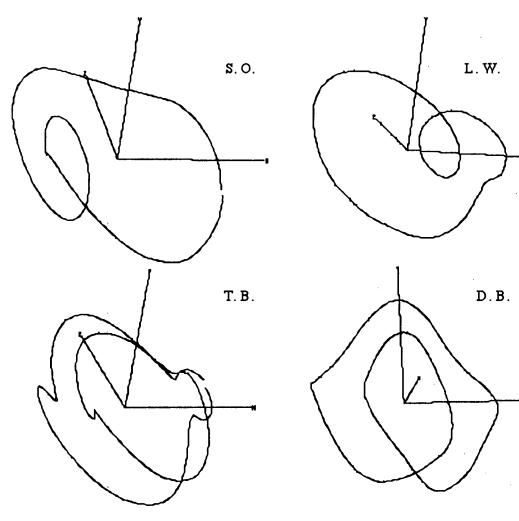
Gibson, John F., J. Doyne Farmer, Martin Casdagli, & Stephen Eubank. (1992). An analytic approach to practical state space reconstruction, Physica D, 57, 1-30.

Licklider, J. C. R, & Irwin Pollack. Effects of differentiation, integration, and infinite peak clipping on the intelligibility of speech. JASA, 20, 42-51.

Liljencrants, J. & B. Lindblom. (1972). Numerical simulation of vowel quality systems: the role of perceptual contrast. Language, 48, 839-62.

Miller, R. L. (1953). Auditory tests with synthetic vowels. JASA, 25, 114-21.

Peterson, Gordon E. (1961). Parameters of vowel quality. Journal of Speech and Hearing Research, 4, 10-29.

Peterson, Gordon E., & Harold L. Barney. (1952). Control methods used in a study of the vowels. JASA, 24, 175-84.

Potter, R. K. & J. C. Steinberg. (1950). Toward the specification of speech. JASA, 22, 807-20.

Scott, Brian L. (1976). Temporal factors in vowel perception. JASA, 60, 1354-65.

Stevens, Kenneth. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. Human Communication: A Unified View, E. E. David Jr. & P. B. Denes, eds., New York: McGraw-Hill.

Tuller, Betty, & J.A.S. Kelso. (1991). The production and perception of syllable structure. Journal of Speech and Hearing Research, 34, 501-8.

Table 1.

	Talker	i	I	E	ae	u	U	o	^	a
F_0	D.B.	131.7	140.9	134.1	125.2	127.1	140.7	121.7	133.5	119.2
	S.O.	165.4	175.3	172.6	166.1	184.7	179.6	168.3	173.5	161.6
	L.W.	216.1	198.5	190	177.7	187.9	192.8	178.7	179.2	175.3
F_1	D.B.	268.9	412.2	537.8	779.1	281.5	442.3	419.7	600.6	801.7
	S.O.	321.7	487.5	826.8	975.1	366.9	583.1	525.2	811.7	1083
	L.W.	444.8	535.3	784.1	904.7	409.6	573	608.2	737.6	841.9
F_2	D.B.	2460	2113	2571	2091	836.9	1252	922.3	1056	1191
	S.O.	3003	2480	2312	2184	990.1	1412	1061	1817	1621
	L.W.	3023	2259	2129	2056	1088	1651	999	1081	1130
F_3	D.B.	3496	2993	4792	2898	2220	2262	2224	2307	2131
	S.O.	4008	3224	3189	2998	1907	2915	2051	3031	2880
	L.W.	4006	3071	3013	2817	2514	2915	*	1686	*

Figure 1.

i

I

E

æ

u

U

^

o

a

80

Figure 2.

S.O.

L.W.

T.B.

D.B.

81