Formant Transitions of Russian Palatalized and Nonpalatalized Syllables

Stephen M. Howie

Abstract

This paper presents a mathematical model of formant transitions in CV syllables of the form

$f(t) = f_v + [(f_c - f_v) + ct / \tau] \exp(-t / \tau)$

where

$f(t)$ = frequency at time t

f_v = frequency of vowel target

f_c = frequency of formant onset

c = the amount of perturbation from an exponential decay, such that the initial slope is zero for $c = f_c - f_v$.

τ = a time constant.

This model is based on the exponential decay model of Broad & Clermont (1987), but incorporates the additional factor *c*, which describes perturbation from a simple exponential decay. Such perturbation may arise from secondary articulation such as palatalization.

The model was tested on utterances of isolated syllables consisting of the Russian voiced stops /b d g/ and their palatalized counterparts /b' d' g'/, each followed by the vowels /i e a o u/. Formant frequencies were measured at each glottal pulse for the first 300 milliseconds from release, together with the time from release. Curves were then fitted to the measured data and values computed for each of the parameters in the model. Analyses of variance performed on the computed parameter values showed that to a large extent the parameter values of the model reflect the phonetic categories vowel, consonant place of (primary) articulation and palatalization.

1. Introduction

In recent years there has been a strong focusing of interest in articulatory dynamics, as exemplified in the theory of Articulatory Phonology (Browman & Goldstein 1990, 1992, 1994). Articulatory Phonology and the ideas underlying it are themselves an application of the broader theory of Task Dynamics (Saltzman & Munhall 1989). The adoption of the dynamical framework for research and description has been of great benefit to phonetics. It is the aim of the present study to apply some of the principles of dynamics directly to the acoustic domain, that is, to the listener's rather than the speaker's perspective.

The vocal tract cannot change shape instantaneously from the configuration of one phonetic segment to that of the next. During the time the articulators are in motion, the resonance characteristics of the cavity system are necessarily changing. The change in resonance manifests itself as rising or falling of the formants, or resonance peaks, in the acoustic spectrum. Early experiments with speech synthesis discovered that rapid shifts in vowel formant frequencies adjacent to consonants, though occurring too fast to be perceptible as changes in vowel quality, were nevertheless salient cues to the place of articulation of the consonant (Liberman et al. 1954). These shifts occurring at the edge of

a vowel next to a consonant are known as formant transitions. Further experimentation seemed to suggest that there were relatively invariant points on the frequency scale associated with particular places of articulation. Such a point, '. . . a place on the frequency scale at which a transition begins or or to which it may be assumed to "point"' (Delattre et al. 1955, p. 769), was called a locus. The rest of the 1950s was spent mapping the loci for the different formants across various phonetic contexts (Delattre 1961). Unfortunately, the loci are not as invariant as originally thought.

The perceptual experiments with synthetic speech in the 1950's which led to the locus concept must be considered in their historical context. The equipment was crude, and more important, the methodology was designed to find the minimal, most salient acoustic cues for phonetic features. Consequently the stimuli were highly simplified compared to natural speech, the goal being to explore the perceptual effect of one potential acoustic cue at a time. As the stimuli were so impoverished (one might say they were caricatures of natural speech), it is hardly surprising in retrospect that the acoustic cues that were present had to be exaggerated to compensate for the absence of other, supplementary cues. Thus for example the F2 locus at 700 Hz for bilabials is unnaturally low; it simply was the value which could consistently elicit a /p/ or /b/ response and no /t/, /d/, /k/, or /g/ responses from experimental subjects. By the same token the F2 locus at 3000 Hz for velars before front vowels is unnaturally high. In natural speech 3000 Hz is at the high end of the range of the third formant for an adult male; the stimuli in the study by Delattre et al. (1955) had only F1 and F2. Subsequent research investigated the effect of adding F3 (Harris et al. 1958; Delattre 1961).

A classic study of natural speech by Öhman (1966) showed that the exact place of articulation of a consonant in English is strongly affected by the preceding and following vocalic environment, with the formant transition loci reflecting the variability of the place of articulation. Furthermore, the movement of the lips and tongue are fairly independent. During a bilabial closure the tongue can be moving toward a new vowel target, and during an apicoalveolar or dorsovelar closure the degree of lip protrusion or rounding can be changing. Both lip attitude and tongue position affect the formants, and some combinations of lip and tongue movements have opposite effects on a particular formant while other combinations have additive effects. For example, as we shall see, palatalization raises F2 and F3, while labialization lowers F2 and F3. Consequently, F2, and especially F3, rise briefly following the release of a palatalized labial as the lips open, then fall as the tongue moves away from the palate toward the following vowel target. On the other hand, lip rounding and tongue retraction both lower F2 and F3. It is well known that in many English dialects /u/ is pronounced either as a back vowel or as a rounded vowel, but rarely as both back and rounded. Either gesture alone is sufficient to lower F2 below the threshold necessary to distinguish /u/ from /i/. Such interrelation of the acoustic effects of lip and tongue movement is an additional source of variation in formant loci across vowel contexts (Fant 1973).

As the loci are affected by the vocalic environment, one might expect to get interesting results by plotting the onset frequency of a formant in a CV syllable as a function of the vowel's own intrinsic, target formant frequency for various vowels. Such plots typically are nearly linear for consonants of a given place of articulation; the equations describing the best-fit lines through the plotted points are known as locus equations. This was first done by Lindblom (1963a) for a single Swedish speaker. More recently, Sussman and various collaborators have carried out systematic large-scale investigations of locus equations (Sussman et al. 1991; Sussman et al. 1993).

Most studies of formant transitions have dealt with English or other languages in which consonantal secondary articulations are not used contrastively. Sussman et al. (1993) looked at pharyngealization in Arabic and retroflexion in Urdu; however, in these languages the secondary articulation applies to only one place of articulation, viz., coronal. In effect, the distinctive secondary articulation on coronals amounts to having a fourth place of articulation. In Russian, on the other hand, palatalization pervades the entire consonant system. For each manner of articulation (except liquids and /j/) there is a 2 x 3 matrix: a consonant may be labial, coronal or velar, and palatalized or nonpalatalized (Jones & Ward 1969). Although there are severe restrictions on the occurrence of palatalized velars before back vowels and nonpalatalized velars before front vowels, minimal pairs can be found: /satkóm/ 'place for keeping live animals (inst. sg.)' vs /satk'óm/ 'weave (1st. pers. pl. perfective)' and /b'ir'igá/ 'bank; shore (nom. pl.)' vs /b'ir'ig'á/ 'guard (pres. ger.)' (George Fowler, pers. comm.).

During labial closure, the tongue is free to approach the following vowel target or recede from the preceding vowel position; during closure at other places of articulation, the lips are free to round or spread. Also the rate of lip opening may vary depending on whether a rounded or unrounded vowel follows. Because of the relative independence of movement of different parts of the vocal tract (tongue and lips, or even different portions of the tongue), an invariant formant locus for a given place of articulation can be no more than an idealization (Fant 1973, Kewley-Port 1982). It is this very coarticulation which became phonologized in the history of Russian as the palatalized/nonpalatalized opposition. Certain formerly back vowels became front and vice versa, leaving their coarticulatory effects behind as secondary articulation (Carlton 1991). In order to maintain the palatalized/nonpalatalized opposition as part of the phonology of modern Russian, it is necessary to restrict the extent of coarticulation. A consonant of one category (palatalized or nonpalatalized) cannot be allowed to be pushed over the line into the other category by coarticulation. That is, the palatalized or nonpalatalized nature of the consonant must be preserved, the vocalic environment notwithstanding. Earlier historical coarticulatory effects of vowels on consonants, which led to the phonologization of the palatalized/nonpalatalized opposition, have been superseded by coarticulatory effects of consonants on vowels. As we shall see, the dominance of consonants over vowels in Russian, predicted on phonological grounds, appears to have been borne out by Öhman (1966), in contradistinction to the situation observed in English and Swedish.

Öhman (1966) examined coarticulation spectrographically in VCV utterances in Swedish, English, and Russian. While in Swedish and English, the effect of the vowels on the consonants was greater than vice versa, the opposite seemed to be true in Russian. In other words, in Swedish and English the vowels are relatively stable and independent of their consonantal environment, while the effect of coarticulation is to force the consonants to accommodate themselves to the neighboring vowels. In effect, the consonantal gesture is superimposed on a more basic diphthongal gesture. In Russian, on the other hand, it appeared that the consonants were more stable and independent of the vowels, and it was the vowels that accommodated themselves to the consonants through coarticulation. (This trend was mitigated in the case of velars by the phonotactic constraints previously mentioned.) This is what we would expect, given the role of the palatalized/nonpalatalized opposition as a phonologization of coarticulation that existed at an earlier stage of the history of the language, as previously discussed. It was not possible to draw a firm conclusion about the relative stability of vowels and consonants

in Russian, as there was only a single Russian speaker and a smaller corpus. Öhman's result was replicated by Purcell (1979), with a much larger body of Russian VCV data.

Since the significance of formant transitions for speech perception was first discovered by the Haskins experimenters (Liberman et al. 1954, Delattre et al. 1955, Liberman et al. 1956), it has been a common practice to make a hard and fast distinction between 'formant transitions' *per se* as line segments with constant, nonzero slopes, and 'steady-state portions of vowels' with formants as line segments with zero slope. This is primarily the case in speech synthesis, for convenience. For example, the user interface for the Klatt synthesizer (Klatt 1980) is explicitly set up to compute linear transitions between control points, as was the pioneering synthesis-by-rule algorithm of Holmes, Mattingly & Shearme (1964). Since the time of Pierre Delattre, the virtuoso of the Pattern Playback, who painted spectrographic patterns by hand with great skill and verisimilitude, researchers have rarely considered it worth the effort to synthesize formant transitions with smooth curves. Also in spectrographic analysis of natural speech, researchers have typically imposed the transition versus steady-state portion dichotomy, modeling formants with line segments and corners. Of course, no one would claim that this is anything but an approximation, nor has it been a bad one. For most purposes it is simply not practical or necessary to measure more than that. However, if a curvilinear model can be perfected, it will be possible to have both naturalistic, smooth curves and a complete description of the formant trajectory with only a small number of parameters.

Halle (1959) writes 'The effects on the formant transitions of the two types of pharynx configuration [for palatalized versus nonpalatalized consonants] appear to be greater than — and hence capable of obscuring — the effects due to different points of articulation' (p. 151). It has often been noted (e.g., Jakobson et al 1952, Halle 1959, Fant 1970, Pauli & Derkach 1971, Derkach et al. 1983) that the effect of palatalization is to raise the loci of F2 and F3 to frequencies characteristic of the vowel [i]. In fact, Pauli & Derkach (1971) successfully synthesized Russian palatalized consonants of all places of (primary) articulation simply by setting the loci to the [i] targets. That is to be expected in view of the articulatory resemblance between palatalization and [i]. However, although the importance of time-varying information in the acoustic signal over static information has been demonstrated (Kewley-Port 1983, Kewley-Port et al. 1983), relatively little attention has been given to the precise shapes of formant transitions or how they might best be described quantitatively.

The following questions arise:

—In view of Halle's (1959) observation and Pauli & Derkach's (1971) experience, just how much difference is there among formant transitions of palatalized consonants having different points of articulation? If there is no consistent difference among their onset frequencies or loci, might there be some difference in their course toward the vowel target?

—Can the formant transitions at syllable margins be described in a simple and principled way, both in their static aspect (e.g. by locus equations [Sussman et al. 1991]) and their dynamic aspect (formant frequency as a function of time)? Will such descriptions hold for both palatalized and nonpalatalized consonants, and are the two descriptions the same?

—Can the formant loci and transition shapes for palatalized consonants be derived from those of their nonpalatalized counterparts in a simple and principled way, reflecting the concept of palatalization as a secondary articulation superimposed on a

more basic, primary articulation, or must they be separately specified as different, unrelated sounds?

—Does the well-known allophonic variation of Russian vowels, conditioned by stress and by the palatalization of the preceding and/or following consonants, require different articulatory/acoustic targets for the different allophones, or does it fall out naturally from the dynamics of the system through different approaches to the same target? In other words, is this allophonic variation phonological, in the sense of being learned as part of the grammar of the language, or is it the phonetic consequence of mechanical constraints on the speech mechanism?

I will present a model for generating formant patterns for Russian CV syllables. I propose to test this model by comparing such patterns to those obtained from spectrographic analysis of natural speech. The model I will present is essentially a modification of that developed for English CVC utterances in an elegant study by Broad & Clermont (1987). Their model used a pair of exponential decay functions for each formant, one extending forward in time from the release of the initial consonant into the vowel, and the other extending backward in time from the closure of the final consonant into the vowel. The two exponentials were combined, along with a (constant) vowel target, by superposition. The present study incorporates an additional term into the mathematical function to quantify the effect of palatalization, which is a distinctive feature of Russian, in terms of perturbation from a simple exponential decay. It is hoped that, if successful, the model can be generalized to other syllable types, and, ultimately, serve as the basis of a synthesis-by-rule system for Russian, as well as contribute to answering the questions posed above.

A formant begins at or near its locus for a particular place of articulation and rises or falls to the target frequency for the vowel. If the vowel is too short, there may not be time for the formants to reach the vowel target before they start moving toward the targets associated with subsequent segments of the speech signal. Such failure to reach the vowel target is known as undershoot (Lindblom 1963b, Moon & Lindblom 1994); listeners, being implicitly aware of the mechanical constraints of the speech organs, are able to make allowances for it (perceptual overshoot). Research has shown that there is enough information in the shape of the formant movements through time at the beginning and end of the vowel that listeners can correctly identify the vowel even when the central, steady state portion is edited out (Strange 1989).

2. Model

A plausible model of formant transition as a function of time is given by

$$(1) \qquad f(t) = f_v + (f_c - f_v) \exp(-t / \tau)$$

where
$f(t)$ = frequency at time t
f_v = frequency of vowel target
f_c = frequency of formant onset
τ = a time constant.

If the function continued linearly at its initial rate, it would reach the target at $t = \tau$. As it

is not a linear but an exponential decay function, it approaches the target asymptotically. Such functions have been used in formant synthesis by rule systems in English (Rabiner 1968) and Russian (Pauli & Derkach 1971).

The function (1) is the general solution to a first-order differential equation of the form

(2) $dy/dx = -ky$

in which the rate of change of the dependent variable (here *y*; frequency in (1)) is in constant proportion to its value. Such a relation is characteristic of many kinds of natural phenomena, such as decrease of atmospheric pressure with altitude, heating and cooling of objects, radioactive decay, chemical reactions, charging and discharging of inductors and capacitors in electric circuits (see e.g., Lomen & Lovelock 1996). An exponential decay function of the general type in (1) is also the one used by Lindblom (1963b) to describe vowel undershoot quantitatively as a function of syllable duration, and by Broad & Clermont (1987) to model formant frequency as a function of time in naturally produced English CVC utterances.

As illustrated in Figure 3.4-1 in Fant (1970, p. 220), nonpalatalized consonants have transitions that appear to lend themselves to modeling by an exponential function such as (1). Palatalized consonants, however, have transitions with an obviously more complex shape. The effect of palatalization has both a static and a dynamic aspect, judging by the Fant illustration. The static aspect is of course to shift the loci toward frequencies characteristic of the vowel [i]; the dynamic aspect is to keep the formants there for a short time before they start moving toward their respective vowel targets or even, as F3 of /b'a/ in the Fant illustration, move away from the target briefly before beginning to approach it. In mathematical terms, in the (ideal) nonpalatalized case the absolute value of the slope of the formant is greatest at the consonant release, and approaches zero as the formant approaches its target. In the (ideal) palatalized case, however, the absolute value of the slope begins near zero at the release, increases to a maximum, then approaches zero again as the formant approaches its target. Rather than using a function of the type in (1), then, for the formant transitions from palatalized consonants, a more appropriate model is given by

(3) $f(t) = f_v + [(f_c - f_v) + bt] \exp(-t / \tau)$

where the terms are as in (1) with the addition of *b*. It is not possible to relate *b* directly to any phonetic parameter. Its geometric interpretation is as follows. For nonzero values of *b*, the function (3) will intersect the target asymptote at time $t = -(f_c - f_v) / b$, and there will be a peak (for $f_c > f_v$ and $b > 0$) or valley (for $f_c < f_v$ and $b < 0$) at $t = \tau - (f_c - f_v) / b$ with the value $f_v + b\tau \exp\{[(f_c - f_v) / b\tau] - 1\}$. Of course, if $b = 0$ the coordinates of these points are undefined; there is no intercept or peak or valley.

It may be seen that if the parameter *b* in (3) has the value zero, (3) simplifies to (1). Therefore, (1) can be considered simply a special case of (3), and a function of the type in (3) can be used to model the formant transitions from all consonants, palatalized and nonpalatalized alike. In fact, as we shall see, the distribution of the values of *b* appears not so nearly bimodal as the foregoing discussion would suggest.

The function (3) is the general solution to a second-order differential equation of the form

(4) $d^2y/dx^2 + 2dy/dx + (y - y_0) = 0$

A physical system whose motion is described by such an equation in which the coefficients of the terms are in the ratio 1 : 2 : 1, as in (4), is said to be critically damped. If the damping is greater or less than critical, the solution to (4) has a form different from (3), namely

(5) $f(t) = f_v + (f_c - f_v) \exp(-t / \tau_1) + b \exp(-t / \tau_2)$

If the damping is greater than critical, τ_1 and τ_2 will be distinct positive real numbers; if the damping is less than critical they will be a complex conjugate pair. Since (5) contains an additional parameter, τ_2, which does not occur in (3), it will be desirable to assume critical damping until such time as this assumption may be shown to be unjustified. We may note that the Task Dynamic model of Saltzman & Munhall (1989) uses a critically damped second order function. Qualitatively, the significance of damping is that in an underdamped system such as a mass suspended from a spring, the mass having been pulled down and then released will overshoot and oscillate about its rest position (the excursions progressively diminishing). The greater the damping (the smaller the mass or the stiffer the spring in this example), the less will be the maximum velocity attained by the mass in its approach to its rest position. Critical damping represents the upper limit that the maximum velocity can reach without overshoot and oscillation.

Conceptually, the description of formant transitions presented here is not unlike the analysis of electric transients in RLC circuits (i.e., circuits containing resistance R, inductance L, and capacitance C), except for the restriction that the damping be critical (see, e.g., Richmond 1972).

As there is a strong interaction between the variables *b* and τ in (3) in determining the shape of the curve (see Figure 1), let us define $c = b\tau$ and rewrite (3) as

(6) $f(t) = f_v + [(f_c - f_v) + ct / \tau] \exp(-t / \tau)$

Then any variation in τ will merely have the effect of changing the time scale, and *c* will affect the height of any peak (or depth of valley, as the case may be) without any necessity to apportion the relative contributions of *b* and τ. That can always be done later, should it become desirable.

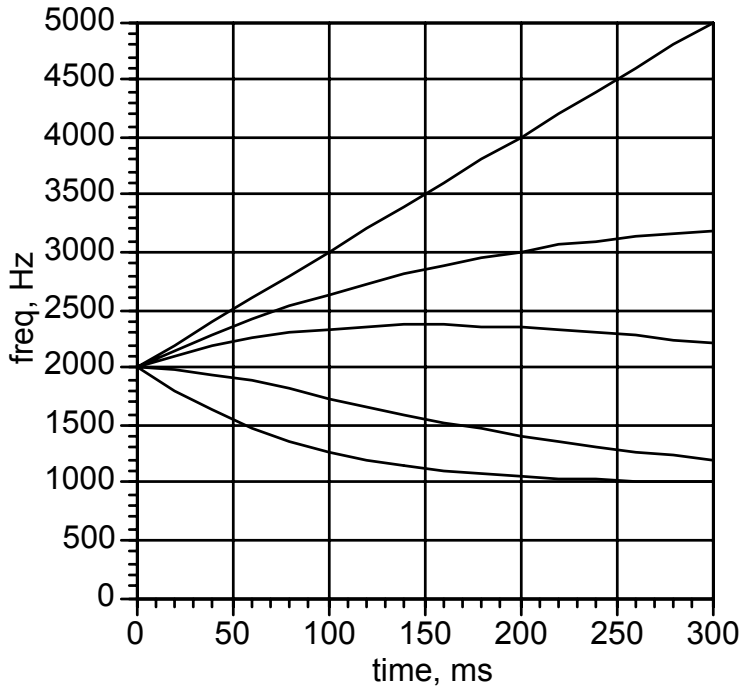

Figure 1. Effect of varying τ in Equation 3, keeping other parameters constant. Values of τ, from bottom curve: 50 ms, 100 ms, 250 ms, 500 ms, 1000 ms, ∞. The graph is linear for τ = ∞. Other parameters: f_c = 2000 Hz, f_v = 1000 Hz, b = 10 Hz / ms.

If the terms *c* and $f_c - f_v$ in (6) are equal, the initial slope will be zero with the peak or valley at t = 0, as in the second curve from the top in Figures 2 and 3. If $|c| > |f_c - f_v|$, the formant will move away from its target briefly, rising to a peak (if *c* and $f_c - f_v$ are positive) or descending to a valley (if *c* and $f_c - f_v$ are negative) before approaching the target. Such is the case with the top curve in Figure 2 and the bottom three curves in Figure 3. If *c* and $f_c - f_v$ are of opposite sign, the formant will approach the target very steeply and overshoot it, only to approach it asymptotically from the opposite direction from which it started. Such is the case with the bottom curve in Figures 2 and 3.

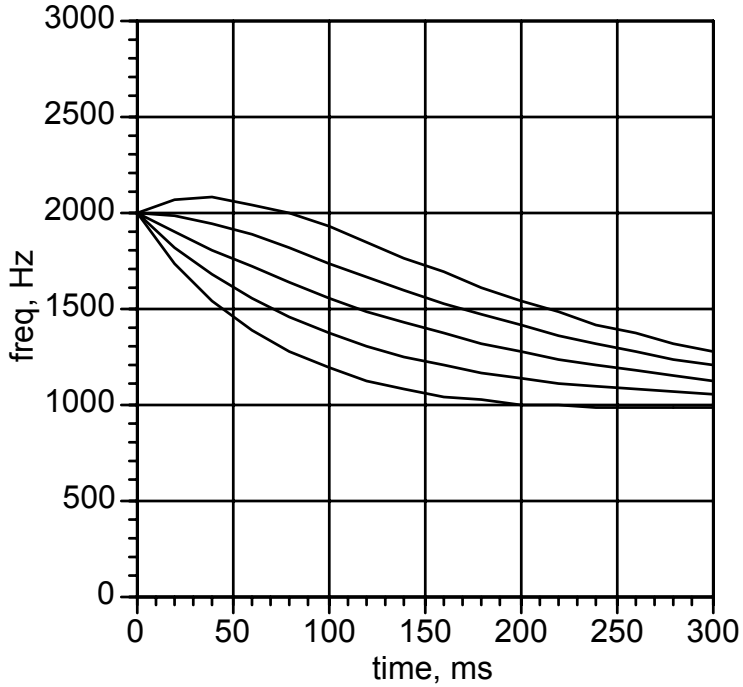

Figure 2. Effect of varying *c* in Equation 6, keeping other parameters constant. Values of *c*, from top curve: 1500 Hz, 1000 Hz, 500 Hz, 0 Hz, –500 Hz. Other parameters: f_c = 2000 Hz, f_v = 1000 Hz, τ = 50 ms.

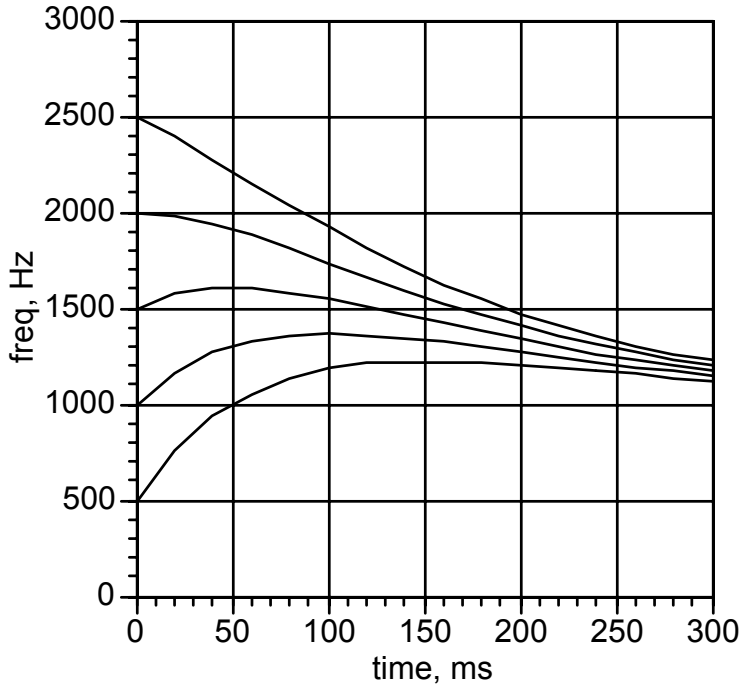

Figure 3. Effect of varying f_c in Equation 6, keeping other parameters constant. Values of f_c, from top curve: 2500 Hz, 2000 Hz, 1500 Hz, 1000 Hz, 500 Hz. Other parameters: *c* = 1000 Hz, f_v = 1000 Hz, τ = 50 ms.

Similar models have been used for predicting vowel targets (Pitermann 2000 & refs. therein), including a revision of the Undershoot theory (Moon & Lindblom 1994); however, I am not aware of any specific application to formant transitions associated

with consonants.

We now wish to define a shape metric, a single number that will quantify the shape of the curve, in terms of its departure from a simple exponential decay, and in terms of its departure from having an initial slope of zero. That is, this shape metric must have one consistent, unique value when the curve is an exponential decay, which we expect for nonpalatalized formant transitions, and a different consistent, unique value when the curve has an initial slope of zero, which we expect for palatalized formant transitions. Two possibilities for defining such a shape metric immediately suggest themselves: one is *c* from Equation 6, the other is the initial slope. We cannot use the initial slope as our shape metric, because for an exponential decay the initial slope is $-(f_c - f_v) / \tau$; it does not satisfy the requirement that the value of the shape metric be consistent and unique for all cases of simple exponential decay, independent of the particular values of f_c, f_v, and τ in any given case. And as a shape metric *c* is not especially informative by itself, but only with reference to $f_c - f_v$. Recall that in the nonpalatalized case we expect *c*, and hence $c / (f_c - f_v)$, to be close to zero, so that the shape of the curve will be nearly a simple exponential decay. In the palatalized case we expect *c* to be approximately equal to $f_c - f_v$, hence that $c / (f_c - f_v) \approx 1$. When $c / (f_c - f_v) = 1$, the initial slope of the curve is zero. It appears, then, that $c / (f_c - f_v)$ would define the shape metric in that it satisfies the requirements presented above for consistent, unique values. However, if we were simply to take the the ratio $c / (f_c - f_v)$, that would not be sufficient. In the first place, there is no restriction on how much greater *c* can be than $f_c - f_v$, although the extent to which *c* might exceed $f_c - f_v$ in any particular case is not necessarily of interest. Second, and more serious, there is nothing to prevent f_c and f_v having the same value (as might be the case for F2 in a syllable with a labial consonant and back vowel; cf. the second curve from the bottom in Figure 3), in which case $c / (f_c - f_v)$ is undefined. Even without actually being zero, a small absolute value of $f_c - f_v$ will make the ratio $c / (f_c - f_v)$ overly sensitive to small variations of *c*. For these reasons, let us use the quantity $(4 / \pi) \tan^{-1} [c / (f_c - f_v)]$ for analysis, instead of merely $c / (f_c - f_v)$. This transformation has the following properties:

When $c = 0$, the result is 0.

When $c / (f_c - f_v) = \pm 1$, the result is ± 1.

When $|c|$ increases without bound, or when $f_c - f_v$ approaches zero, the transform approaches ± 2 as a limit (2 if $c > 0$; –2 if $c < 0$).

The transformation is approximately linear in the range from –1 to 1. Its effects are to reduce extreme values of $c / (f_c - f_v)$ in a simple and principled way, and to provide definitions in the form of limits for cases in which $c / (f_c - f_v)$ is undefined. The latter effect is an advantage which this transformation has over other possible types of transformation that would accomplish the former effect, such as the cube root.

Recapitulating, we have the function (6) giving formant frequency as a function of time:

$$(6) \qquad f(t) = f_v + [(f_c - f_v) + ct / \tau] \exp(-t / \tau)$$

where
f(t) = frequency at time t
f_v = frequency of vowel target
f_c = frequency of formant onset
c = the amount of perturbation from an exponential decay, such that the initial slope is zero for $c = f_c - f_v$.
τ = a time constant.

3. Applying the Model

The proposed model was tested on Russian CV syllables consisting of the voiced stops /b d g/ and their palatalized counterparts /b' d' g'/ preceding each of the vowels /i e a o u/. Subject 1's (age about 45; native of Moscow) production of two tokens of each of the syllable types was recorded, in random order, in the phonetics lab in the Linguistics Department at Indiana University. Subject 2 was from a section of a commercial Russian language instruction record illustrating the contrast between palatalized and nonpalatalized consonants in CV syllables (Pressman 1958). Both subjects were male. Unfortunately there are no tokens of /gi ge g'a g'o g'u/ from Subject 2 as (with the exceptions noted above) these sequences do not occur in native Russian words.

The recordings were digitized at 22 kHz with 16 bit resolution on a Macintosh G4 computer using Signalyze 3.0, and each token was stored as a separate file. In Signalyze, a wide band FFT spectrum was taken at each glottal pulse, and from each spectrum the frequency of the first three formants was extracted along with the time of the sample, for the first 300 ms from the release of the stop. 300 ms was chosen on the grounds that all formants in all tokens had completed their transitions and reached what could be considered a steady state by this time (in fact, usually within 200 ms). Less would have risked losing useful information. The better token of each of Subject 1's types (in the sense that this subject did not have a very 'spectrogenic' voice, and it was often difficult to identify formant peaks) was used for analysis.

No use was made of LPC analysis for data collection, as preliminary experimentation with LPC was unsatisfactory. In this connection it is worth noting the advice in the Signalyze manual: 'CAUTION: USE LPC WITH CARE. LPC formant measures often diverge to a substantial extent from narrow-band or wide-band formant measures. Some laboratories have ceased to use LPCs after years of using it [*sic*]' (Keller 1994, p. 172).

In practice, data collection sometimes fell short of the ideal for various reasons. In some cases the signal quality deteriorated (due to glottal fry or other idiosyncratic irregularities in the glottal source spectrum) before 300 ms to the point where it was not possible to obtain accurate identification of formant peaks, although in all cases at least 200 ms of observations were obtained. In nearly all cases there were a few time points for which accurate measurements were impossible for one or more formants, usually F3. It was not possible to measure F3 for /b'i/ at all from Subject 1, as it formed a single peak with F2 for the entire duration of the syllable. In particular, Subject 1, who was known to be a heavy smoker (as many Russian men are), exhibited spurious peaks and valleys at irregular intervals of time in his speech spectra which interfered with the formant structure. Presumably these spurious peaks and valleys reflected distortions in the glottal source spectrum.

The numerical data from Signalyze were graphed in DeltaGraph 4.0, and the graphs visually compared with the spectrograms in Signalyze. Where there were obvious discrepancies, the data points in question were remeasured. When the graphs were corrected as much as possible, curve fits were applied to Equation 6, using the User-Defined Fits option of DeltaGraph's Curve Fit module. An example appears in Figure 4.

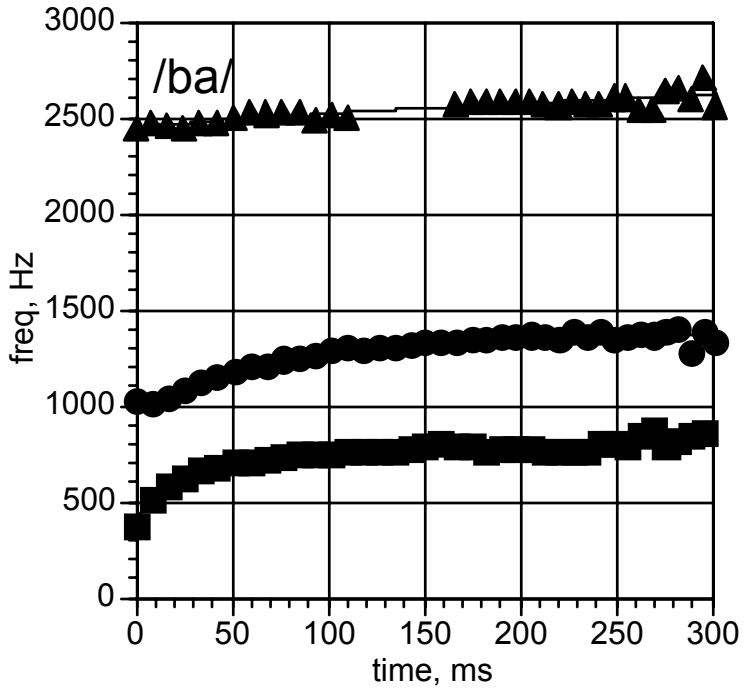

Figure 4a. /ba/ from Subject 2. Parameter values are: (F1) fc = 424 Hz, fv = 797 Hz, c = 0 Hz, τ = 36.3 ms, R^2 = .907; (F2) fc = 1010 Hz, fv = 1369 Hz, c = –249 Hz, τ = 41.5 ms, R^2 = .963; (F3) fc = 2455 Hz, fv = 2756 Hz, c = 0 Hz, τ = 392.4 ms, R^2 = .747.

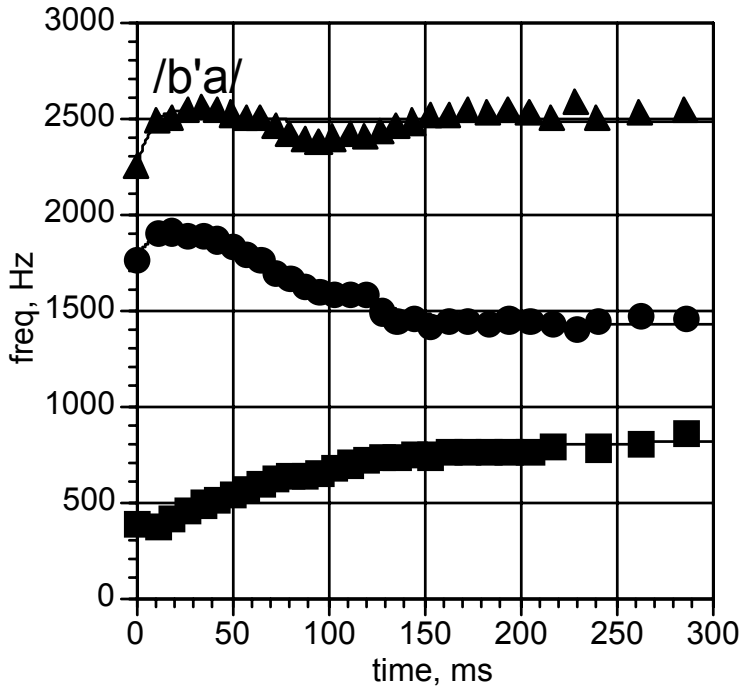

Figure 4b. /b'a/ from Subject 2. Parameter values are: (F1) fc = 372 Hz, fv = 818 Hz, c = –316 Hz, τ = 48.6 ms, R^2 = .987; (F2) fc = 1768 Hz, fv = 1424 Hz, c = 955 Hz, τ = 33.7 ms, R^2 = .981; (F3) fc = 2247 Hz, fv = 2482 Hz, c = 320 Hz, τ = 13.7 ms, R^2 = .439.

A third, limited, data set came from the aforementioned Figure 3.4-1 of Fant (1970, p. 220). Unfortunately there was only the one vowel, /a/, included in the Fant illustration. This figure was scanned on an AppleOne flatbed scanner using Ofoto scanning software on a Macintosh PowerPC computer and saved as a PICT file. The PICT file was then opened in FlexiTrace 2.02 on a Macintosh IIsi computer and data points collected for each formant of each of the syllables /ba b'a da d'a ga k'a/ in the figure at 20 ms intervals, yielding eleven points per formant from zero to 200 ms. The coordinates were then exported to DeltaGraph and curves fitted as for the Signalyze data from Subjects 1 and 2.

4. Results and Discussion

The curve fits were generally quite good. The overall mean R squared value, across all subjects and formants, was .779, with a standard deviation of .261. The mean R squared values, with standard deviations in parentheses, broken down by subject and by formant, appear in the Table 1.

	F1	F2	F3	All Formants
Subject 1	.807 (.201)	.772 (.245)	.526 (.324)	.704 (.286)
Subject 2	.862 (.18)	.87 (.218)	.714 (.261)	.816 (.231)
Subject 3				.978 (.038)
All subjects	.849 (.188)	.835 (.229)	.651 (.309)	.779 (.261)

Table 1. Means and (standard deviations) of R squared values of curve fits, by subject and formant.

Subject 3 is not broken down by formant as the overall mean R squared is so high and standard deviation so low. Recall that there was only one vowel from this subject instead of five. Furthermore, the number of data points for each formant was eleven for Subject 3 versus 20-25 for the other two subjects. We can see right away that R squared was lower overall for F3 than for F1 and F2, and lower overall for Subject 1 than for Subject 2.

For the most part, it was no trouble to obtain close fits with reasonable parameter values. Even where the R squared value was low, the fitted curve looked like an appropriate formant trajectory for the phonetic material in question; with few exceptions the low R squared value was due to the measurement error resulting from shortcomings of the original spectra, as described above. For this reason computed parameter values that were reasonable were not eliminated from the analysis simply because of low R squared values.

There were a few cases, however, where reasonable curve fits could not be made with Equation 6 because one or more parameters either would not converge or would converge toward an impossible value, e.g., a negative vowel target, an initial formant frequency greater than 10 kHz or a tau less than unity. When this happened, a linear fit was applied, which is equivalent to using Equation 6 with a tau of infinity. When tau is infinite, i.e., when $1 / \tau = 0$, (6) reduces to $f(t) = f_c - f_v + bt$, which is the equation of a line with slope *b* and intercept $f_c - f_v$. Such linear fits had slopes with absolute values on the order of 0.1 Hz/ms except that of F2 of /d'e/ from Subject 1, which was close to –1 Hz/ms. Examination of the data shows that the range of values of *b* increases as the inverse of tau; $b \approx 0$ when $1 / \tau = 0$. Such a relationship is what one would expect, since an infinite τ means that the formant has a constant slope for the entire duration of the syllable, where *b* is the slope. The slope must be close to zero if it is constant over the duration of the syllable; otherwise, the formant would reach an impossible frequency by the end of the syllable.

We defined the parameter *c* above such that $c = b\tau$; let us further define *c* such that $c = 0$ when $1 / \tau = 0$. We may justify doing so on the grounds that $b \approx 0$ when $1 / \tau = 0$; therefore, if we assume that $b = 0$ when $1 / \tau = 0$, then $b = c(1 / \tau) = 0 \times 0 = 0$ when $1 / \tau = 0$.

Of course, when a linear fit was necessary, it was impossible to determine the vowel target, f_v, in the same way as with a finite τ, because a line does not approach an asymptote. In the linear cases, then, the target was considered to be the frequency that the fitted line actually reached at 300 ms after release.

A set of three-way analyses of variance was carried out on the data, across all subjects, using StatView 5 on a Macintosh G4 computer. An individual three-way Anova for each subject would have been preferable, but it was not possible because of the limited amount of data. For each formant, the following factors were examined: consonant (b, d, g), palatalization (0 = nonpalatalized, 1 = palatalized), and vowel (i, e, a, o, u). The dependent variables examined were f_c, the formant frequency at the consonant release; f_v, the vowel target frequency (asymptote); $1 / \tau$, the reciprocal of the time constant (the reciprocal was used rather than τ itself, so as not to exclude cases where τ is undefined); and the shape metric, $(4 / \pi) \tan^{-1} [c / (f_c - f_v)]$. As there were four dependent variables being tested, alpha was set at .01 rather than .05 in order to reduce

the probability of a Type I error due to multiple testing.

The results of the Anovas are summarized in Table 2. The columns represent the dependent variables; the rows represent the factors. Each cell lists the formants, if any, for which the factor had a significant effect on the dependent variable. The reciprocal of the time constant, $1 / \tau$, is omitted as no factor had a significant effect on it for any formant. The parenthetical items are effects which are noteworthy in some way, as will be discussed below, but which nevertheless did not quite achieve significance.

There were no significant three-way interactions in any of the tests; consequently, the statistics reported here will be those for which three-way interactions were excluded.

Not surprisingly, the effect of vowel category on the vowel target f_v was significant for all three formants. The results for main effect of vowel were $F(4,42) = 165$, $p < .0001$ for F1; $F(4,42) = 81.5$, $p < .0001$ for F2; and $F(4,41) = 4.53$, $p = .004$ for F3. There were no significant interactions, although for F2, vowel by palatalization came close: $F(4,42) = 3.44$, $p = .016$. This effect appears to result from fronting of /u/ following palatalized consonants. In this environment /u/ may be realized as [ʉ] or even [y], whereas elsewhere it is a fully back [u] (Jones & Ward 1969). Subject 1 exhibited such fronting of /u/ more than did Subject 2; his f_v for F2 was close to 1500 Hz following palatalized consonants but below 1000 Hz following nonpalatalized consonants. Subject 2's f_v for F2 was below 1000 Hz for /u/ in all environments.

There was a significant effect of consonant category on the initial value f_c of F1. For main effect of consonant, the result was $F(2,42) = 23.4$, $p = .0066$. This was rather surprising as the traditional locus for F1 is zero for all consonants. In theory, the lowest resonance of a tube decreases with increasing degree of constriction, independent of the constriction location, reaching zero at complete closure. This, however, applies only to a tube with rigid walls; the vocal tract walls are compliant enough that the minimum F1 even with complete closure is about 200 Hz (Fant 1970). In the present study, the mean f_c for /b/ was 352 Hz; for /d/, 256 Hz; and for /g/, 291 Hz. I have no explanation for this finding, except to suppose that it is an artifact of the limited data set used in this study.

	f_v	f_c	$(4 / \pi) \tan^{-1} [c / (f_c - f_v)]$
V	F1,F2,F3	F2	
C		F1,F2,F3	
pal		F2	F1
V × pal	(F2)		(F2)
C × pal		F2	

Table 2. Formants for which each three-way ANOVA revealed significant (or marginally significant) effects of a factor (rows) on its respective dependent variable (columns). Alpha = .01.

For F2 there was, as expected, a significant effect of both consonant and palatalization on f_c. For main effect of consonant the results were $F(2,42) = 23.4$, $p < .0001$. For main effect of palatalization the results were $F(1,42) = 80.7$, $p < .0001$. Since the starting frequency of F2 is, as previously discussed, a salient perceptual cue to place of articulation (at least for consonants without secondary articulations such as palatalization), we would expect a significant effect of consonant on f_c. And since, as previously discussed, palatalization is known to raise the starting frequency of F2 toward the value characteristic of the vowel [i] (2000 Hz or more), we would expect a significant effect of palatalization on f_c. There was a significant interaction of consonant by palatalization: $F(2,42) = 6.93$, $p = .0025$. Such interaction occurs because palatalization moves the F2 loci associated with different places of articulation closer together. If a consonant has a low locus, such as /b/, palatalization will raise its locus more than would occur for a consonant with a locus that is already relatively high, such as /d/. In other words, palatalization has a greater effect on the F2 loci of some consonants than on the F2 loci of others.

For F3 there was a significant effect of consonant on f_c but not of palatalization. For main effect of consonant, the results were $F(2,41) = 7.28$, $p = .002$. There were no significant interactions. In general, we should expect the results for F3 to be similar to those for F2; that is, that both consonant and palatalization should have significant effects on f_c. However, the typical range of F3 for a given speaker is less than that of F2, and in this study the goodness of the curve fits was somewhat lower for F3 than for F2.

There was no significant effect of vowel on f_c for either F1 or F3. This result suggests the possibility of relatively invariant loci for these two formants in Russian, in keeping with Öhman (1966). Nevertheless, in the present study as in Öhman's, the limited data set precludes drawing a firm conclusion. There was, however, a significant effect of vowel on f_c for F2: $F(4,42) = 6.23$, $p = .0005$. This result would appear to preclude the existence of invariant loci for F2.

It is interesting to compare the results of the present study to those of an elaborate study by Purcell (1979) of Russian VCV utterances, where V = / i e a o u/ and C = /b d g b' d' g'/, as in the present study. F1 and F2 were measured at the steady state of each vowel and at the closure and release of the consonant. A stepwise multiple regression analysis was performed, with coded variables, based on the means of the dependent variables, for categories. Purcell's Table XV (p. 1700) summarizes his results. The CV portion is reproduced here as Table 3:

F1

Transition V2 (R^2 = .518)		Steady-state V2 (R^2 = .703)	
Predictors	**betas**	**Predictors**	**betas**
High/low V2	0.51	High/low V2*	0.84
Pal/nonpal	–0.44		
Place of art.*	0.22		
High/low V1	0.14		

F2

Transition V2 (R^2 = .655)		Steady-state V2 (R^2 = .824)	
Predictors	**betas**	**Predictors**	**betas**
Pal/nonpal*	0.67	Frnt/back V2*	–0.89
Frnt/back V2*	–0.35	Pal/nonpal	0.14
Place of art.*	0.29		

Table 3. Results of stepwise regression analysis by Purcell (1979, Table XV).

The asterisks denote those predictors which were also found significant in the present study. Purcell's 'transition' corresponds to my f_c; Purcell's 'steady-state' to my f_v. Note that Purcell's study did not include F3. Of the dependent variables which Purcell examined, there were none found significant in the present study that were not found significant in Purcell's. On the other hand, the factors which were significant in Purcell's study but not in the present one were those for which Purcell found the lowest absolute values of beta. Beta is the normalized regression slope; i.e., a measure of the importance of the predictor. Those for which Purcell found lower absolute values of beta were not found significant in the present study, which may be simply due to the limited data available. Thus the portion of the present study that overlaps with Purcell's is in fairly good agreement with it.

There was no significant effect on $1 / \tau$ of any of the factors, for any formant. The values of τ showed no recognizable pattern whatever. Of course, because of the nonlinearity of the mapping between articulatory dimensions and formant frequencies (see, e.g., the nomograms in Fant 1970), there is no reason to expect any consistent relation in the value of τ across formants.

As discussed above, we should expect that for nonpalatalized consonants the parameter *c* in equation (6) should have a value close to zero, hence that the formant transition should be approximately an exponential decay. Recall that when $c = 0$, (6) reduces to (1). For a palatalized consonant, we should expect *c* to be approximately equal to $f_c - f_v$, in which case the initial slope of the formant will be zero. A palatalized consonant might even have $c > f_c - f_v$, in which case the formant will initially move away from its target briefly before approaching it. Such is the case for F3 of /b'a/ in the illustration from Fant (1970), where $c / f_c - f_v \approx 5$ (so that $(4 / \pi) \tan^{-1} [c / (f_c - f_v)] \approx 1.76$). Consider that the acoustic effects of lip rounding (complete closure as for a bilabial stop being the limiting case) and palatalization are antagonistic: palatalization tends to raise F2 and F3; labialization tends to lower all formants to some extent, with the effect on F2 and F3 being greater than on F1. Consequently it is not surprising that F2 or F3 of /b'/, while having a higher f_c than /b/, would nevertheless not have as high a f_c as /d'/ or

/g'/; the f_c of /b'/ is being raised by palatalization but this raising is moderated by the lowering effect of labialization. However, since the lips open more rapidly following the consonant release than the tongue moves away from the palate toward its target for the vowel, the formant in question will lose the lowering effect of labialization while still retaining the raising effect of the lingering palatalization. The net result is a brief rise, then fall, which cannot be captured by measuring only the initial and steady-state frequencies of the formants. We have seen that the effect of the consonant on f_c of F3 was significant but the effect of palatalization was not.

While we might expect that /b'/ would differ from /d'/ and /g'/ by having a higher value of $c / (f_c - f_v)$, what is the difference between /d'/ and /g'/ themselves? The formant transitions for Russian /d'/ and /g'/ tend to be similar. This should not be surprising, as both are close articulatorily to a common, palatal point: [ɟ]. The secondary articulation of palatalization interferes somewhat with the primary articulation, apicodental or dorsovelar as the case may be, particularly the latter. The tongue body can't be in two places at once. A palatalized velar is no longer velar *sensu stricto*; when it's palatalized it's palatal. X-ray films show /g'/ much closer to /j/ than to /g/; indeed, the X-ray tracing of the fricative [ç], the palatalized allophone of /x/, appears nearly indistinguishable from that of /j/ (Bolla 1981; Matusevich & Lyubimova 1963).

The shape metric $(4 / \pi) \tan^{-1} [c / (f_c - f_v)]$, as expected, showed a significant effect of palatalization for F1. For main effect of palatalization by vowel, $F(1,42) = 9.2$, $p = .004$. However, it did not, surprisingly, show a similar effect for F2 or F3 at a significant level. For F2 there was a suggestive, though not significant, effect of the interaction vowel by palatalization: $F(4,42) = 3.58$, $p = .013$. A likely explanation for this interaction is allophony of /i/. In a sense, the behavior of /i/ with regard to palatalization of a preceding consonant is the opposite of that of other vowels. As palatalization consists of an [i]-like secondary articulation on the consonant, not surprisingly, a following vowel tends to receive an [i]-like onglide. Foreigners often perceive Russian palatalized consonants as being followed by [j]. But /i/ is, of course, already [i]-like. There is very little movement of F2 of /i/ following a palatalized consonant. Rather, /i/ following a *non*palatalized consonant is repelled from the high-front position in the vowel quadrilateral and acquires a [ɨ]-like (or even [ɯ]-like) onglide. It is usually said that Russian /i/ has the allophone [ɨ] following a nonpalatalized consonant, and the allophone [i] elsewhere (Halle 1959; Jones & Ward 1969), and, indeed, in the present data set the mean F2 target frequency, f_v, of /i/ is significantly lower ($t(8) = -3.5$, p (one tail) = .004) following a nonpalatalized consonant (2200 Hz) than a palatalized consonant (2322 Hz). F2 of /i/ tends to follow a rather sigmoid course following a nonpalatalized consonant, having only a slight slope at the release of the consonant, then rising fairly steeply, then, finally, leveling off again toward its target. Such a sigmoid shape, with the slope initially close to zero, then increasing, then decreasing toward zero again, is precisely what is described by a value of $c / (f_c - f_v)$, and consequently $(4 / \pi) \tan^{-1} [c / (f_c - f_v)]$, close to unity. It is, however, more typical of formant transitions following palatalized than nonpalatalized consonants, for vowels other than /i/. Unfortunately, this is something that cannot be resolved until more data are available. In any event, it is most encouraging that the model produced such a significant effect for F1, both across vowels and across consonants. Even if the shape metric is significant for F1 alone, that result

suggests that it can serve as a quantitative measure of the shape of a formant transition.

The present study has indicated that it is feasible to represent formant transitions with curves derived from a critically damped second-order differential equation model. Such curves are completely described by only four parameters, and there is reason to believe that the parameter values reflect phonetic categories.

5. Suggestions for Further Work

The present work is very much a pilot study, intended to explore the feasibility of the model presented herein. Any future work along this line must use a larger volume of data, both a greater number of speakers, of both sexes, and a greater number of tokens of each type from each speaker. Applying such a data-intensive model as this one is a laborious process. For the three formants of each token, roughly a hundred measurements are distilled down into a dozen parameters—a reduction of nearly an order of magnitude. In addition, it would be desirable to explore further the relationship of the values of a given parameter among the different formants within a single token. Ultimately, the model should be extended to other syllable types, in particular, VC, VCV, or CVC syllables, perhaps along the lines of Broad & Clermont (1987) using superposed functions.

Another worthwhile avenue of exploration would be a perception experiment. It would be a fairly simple matter to generate control parameters for a Klatt synthesizer using equation (6) and varying the values of f_c, f_v, τ and c systematically.

It would be interesting to compare the behavior of the parameters in regard to palatalization in Russian with another language that has phonemic palatalization, such as Gaelic.

Finally, it should not be difficult to apply the model to other types of secondary articulation in other languages, e.g., pharyngealization, retroflexion or labialization. A reviewer has pointed out that Russian consonants tend to be pharyngealized when nonpalatalized, as a strategy speakers can use to magnify the distinction between palatalized and nonpalatalized consonants—a fact not explicitly addressed in the present study. Further work should investigate the modeling of these two opposing types of secondary articulation within Russian.[1]

[1] I am grateful to Steve Jarosz for this suggestion.

References

Bolla, K. (1981). A Conspectus of Russian Speech Sounds. Köln, Wien: Böhlau Verlag.

Broad, David J. & Frantz Clermont (1987). A methodology for modeling vowel formant contours in CVC context. Journal of the Acoustical Society of America 81, 155-165.

Browman, C. & L. Goldstein (1990).Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston and M. Beckman (eds.), Papers in Laboratory Phonology: Between the Grammar and the Physics of Speech. Cambridge: Cambridge University Press. P. 341-376.

Browman, C. & L. Goldstein (1992). Articulatory phonology: an overview. Phonetica, 49, 155-180.

Browman, C. & L. Goldstein (1994). Dynamics and articulatory phonology. In R. Port and T. Van Gelder (eds.), Mind as Motion: Dynamics, Behavior and Cognition. Cambridge, MA: Bradford/MIT Press.

Carlton, Terence R. (1991) Introduction to the Phonological History of the Slavic Languages. Columbus, OH: Slavica.

Delattre, Pierre (1961). Le jeu des transitions de formants et la perception des consonnes. In Proceedings of the Fourth International Congress of Phonetic Sciences, Helsinki, 1961. A. Sovijärvi and P. Aalto (eds.). The Hague: Mouton.

Delattre, Pierre C., Alvin M. Liberman & Franklin S. Cooper (1955). Acoustic loci and transitional cues for consonants. Journal of the Acoustical Society of America 27: 769-773.

Derkach, M. F., R. Ya. Gumetskiy, B. M. Gura & M. E. Chaban (1983). Dinamicheskie Spektry Rechevyx Signalov. L'vov: Izdatel'skoe ob"edinenie Vishcha Shkola.

Fant, Gunnar (1970). Acoustic Theory of Speech Production. 2nd ed. The Hague: Mouton.

Fant, Gunnar (1973). Stops in CV syllables. In Fant, Gunnar (ed.), Speech Sounds and Features. Cambridge, MA: MIT Press.

Halle, Morris (1959). The Sound Pattern of Russian. The Hague: Mouton.

Harris, Katherine Safford, Howard S. Hoffman, Alvin M. Liberman, Pierre C. Delattre, and Franklin S. Cooper (1958). Effect of third-formant transitions on the perception of the voiced stop consonants. Journal of the Acoustical Society of America, 30, 122-126.

Holmes, J. N., I. G. Mattingly & J. N. Shearme (1964). Speech synthesis by rule. Language and Speech, 7, 127-143.

Jakobson, Roman, Gunnar Fant & Morris Halle (1952). Preliminaries to Speech Analysis. Cambridge, MA: MIT Press.

Jones, Daniel & Dennis Ward (1969). The Phonetics of Russian. Cambridge University Press.

Keller, Eric (1994). Signalyze (TM) User's Manual, Version 3.0. Lausanne: InfoSignal (TM) Inc.

Kewley-Port, Diane (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. Journal of the Acoustical Society of America, 72(2), 379-389.

Kewley-Port, Diane (1983). Time-varying features as correlates of place of articulation in stop consonants. Journal of the Acoustical Society of America, 73, 322-335.

Kewley-Port, D., D. Pisoni & M. Studdert-Kennedy (1983). Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. Journal of the Acoustical Society of America, 73, 1779-1793.

Klatt, Dennis (1980). Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America, 67, 971-995.

Liberman, Alvin M., Pierre C. Delattre, Franklin S. Cooper & Louis J. Gerstman (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs 68(8), 1-13.

Liberman, Alvin M., Pierre C. Delattre, Louis J. Gerstman & Franklin S. Cooper (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. Journal of Experimental Psychology 52: 127-137.

Lindblom, B. (1963a). On vowel reduction. Rep. no. 29, The Royal Institute of Technology, Speech Transmission Laboratory, Stockholm, Sweden.

Lindblom, B. (1963b). Spectrographic study of vowel reduction. Journal of the Acoustical Society of America, 35, 1773-1781.

Lomen, David & David Lovelock (1996). Exploring Differential Equations via Graphics and Data. New York: John Wiley & Sons, Inc.

Matusevich, M. I. & N. A. Lyubimova (1963). Al'bom Artikulyacij Russkogo Yazyka. Moskva: Universitet Druzhby Narodov imeni Patrisa Lumumby.

Moon, Seung-Jae & Björn Lindblom (1994). Interaction between duration, context and speaking style in English stressed vowels. Journal of the Acoustical Society of America 96, 40-55.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America 39, 151-168.

Pauli, S. & M. Derkach (1971). Synthesis of some Russian utterances by rules. STL-QPSR 1, 43-49.

Pitermann, Michel (2000). Effect of speaking rate and contrastive stress on formant dynamics and vowel perception. Journal of the Acoustical Society of America 107, 3425-3437.

Pressman, Aron (1958). Living Language Course: Russian. Lesson 4: Palatalized and nonpalatalized syllables. Record 1, side 1, band 4. New York: Crown Publishers.

Purcell, Edward T. (1979). Formant frequency patterns in Russian VCV utterances. Journal of the Acoustical Society of America, 66(6), 1691-1702.

Rabiner, Lawrence (1968). Speech synthesis by rule: an acoustic domain approach. Bell System Technical Journal 47, 17-37.

Richmond, A. E. (1972). Calculus for Electronics. 2nd ed. New York: McGraw-Hill.

Saltzman, E. & K. Munhall (1989). A dynamical approach to gestural patterning in speech production. Ecological Psychology 1, 333-382.

Strange, W. (1989). Evolving theories of vowel perception. Journal of the Acoustical Society of America 85, 2081-2087.

Sussman, Harvey M., Helen A. McCaffrey & Sandra A. Matthews (1991). Investigation of locus equations as a source of relational invariance for stop place categorization. Journal of the Acoustical Society of America 90 (3), 1309-1325.

Sussman, Harvey M., Kathryn A. Hoemeke & Farhan S. Ahmed (1993). Cross-linguistic investigation of locus equations as a phonetic descriptor for place of articulation. Journal of the Acoustical Society of America 94 (3), 1256-1268.