

Formant transitions in normal and disordered speech: An acoustic measure of articulatory dynamics

Björn Lindblom¹, Diana Krull¹, Lena Hartelius² & Ellika Schalling³

¹Department of Linguistics, Stockholm University

²Institute of Neuroscience and Physiology, University of Gothenburg

³Department of Logopedics and Phoniatrics, CLINTEC, Karolinska Institute, Karolinska University Hospital, Huddinge

Abstract.

This paper presents a method for numerically specifying the shape and speed of formant trajectories. Our aim is to apply it to groups of normal and dysarthric speakers and to use it to make comparative inferences about the temporal organization of articulatory processes. To illustrate some of the issues it raises we here present a detailed analysis of speech samples from a single normal talker. The procedure consists in fitting damped exponentials to transitions traced from spectrograms and determining their time constants. Our first results indicate a limited range for F2 and F3 time constants. Numbers for F1 are more variable and indicate rapid changes near the VC and CV boundaries. For the type of speech materials considered, time constants were found to be independent of speaking rate. Two factors are highlighted as possible determinants of the patterning of the data: the non-linear mapping from articulation to acoustics and the biomechanical response characteristics of individual articulators. When applied to V-stop-V citation forms the method gives an accurate description of the acoustic facts and offers a feasible way of supplementing and refining measurements of extent, duration and average rate of formant frequency change.

Background issues

Speaking rate

One of the issues motivating the present study is the problem of how to define the notion of 'speaking rate'. Conventional measures of speaking rate are based on counting the number of segments, syllables or words per unit time. However, attempts to characterize speech rate in terms of 'articulatory movement speed' appear to be few, if any. The question arises: Are variations in the number of phonemes per

second mirrored by parallel changes in 'rate of articulatory movement'? At present it does not seem advisable to take a parallelism between movement speed and number of phonetic units per second for granted.

Temporal organization: Motor control in normal and dysarthric speech

Motor speech disorders (dysarthrias) exhibit a wide range of articulatory difficulties: There are different types of dysarthria depending on the specific nature of the neurological disorder. Many dysarthric speakers share the tendency to produce distorted vowels and consonants, to nasalize excessively, to prolong segments and thereby disrupt stress patterns and to speak in a slow and labored way (Duffy 2005). For instance, in multiple sclerosis and ataxic dysarthria, syllable durations tend to be longer and equal in duration ('scanning speech'). Furthermore inter-stress intervals become longer and more variable (Hartelius et al 2000, Schalling 2007).

Deviant speech timing has been reported to correlate strongly with the low intelligibility in dysarthric speakers. Trying to identify the acoustic bases of reduced intelligibility, investigators have paid special attention to the behavior of F2 examining its extent, duration and rate of change (Kent et al 1989, Weismer et al 1992, Hartelius et al 1995, Rosen et al 2008). Dysarthric speakers show reduced transition extents, prolonged transitions and hence lower average rates of formant frequency change (flatter transition slopes).

In theoretical and clinical phonetic work it would be useful to be able to measure speaking rate defined both as movement speed and in terms of number of units per second. The present project attempts to address this objective building on previous acoustic analyses of dysarthric speech and using formant pattern rate of change as an indirect window on articulatory movement.

Method

The method is developed from observing that formant frequency transitions tend to follow smooth curves roughly exponential in shape (Figure 1). Other approaches have been used in the past (Broad & Fertig 1970). Stevens et al (1966) fitted parabolic curves to vowel formant tracks. Ours is similar to the exponential curve fitting procedure of Talley (1992) and Park (2007).

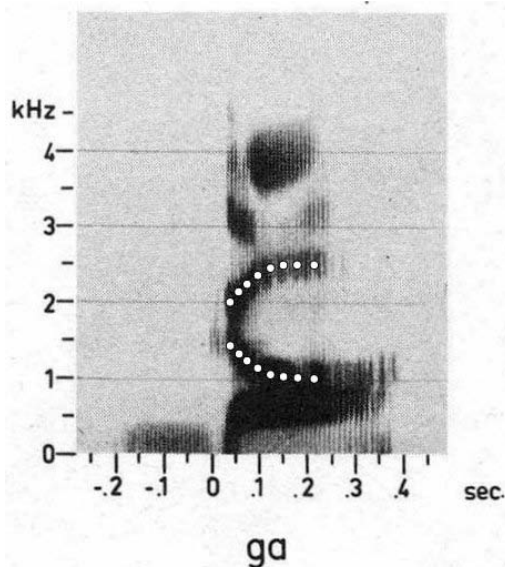


Figure 1. Spectrogram of syllable [ga]. White circles represent measurements of the F2 and F3 transitions. The two contours can be described numerically by means of exponential curves (Eqs (1) and 2).

Mathematically the F2 pattern of Figure 1 can be approximated by:

$$F2(t) = (F2_L - F2_T) * e^{-at} + F2_T \quad (1)$$

where $F2(t)$ is the observed course of the transition, $F2_L$ and $F2_T$ represent the starting point ('F2 locus') and the endpoint ('F2 target') respectively. The term e^{-at} starts out from a value of unity at $t=0$ and approaches zero as t gets larger. The a term is the 'time constant' in that it controls the speed with which e^{-at} approaches zero.

At $t=0$ the value of Eq (1) is $(F2_L - F2_T) + F2_T = F2_L$. When e^{-at} is near zero, $F2(t)$ is taken to be equal to $F2_T$.

To capture patterns like the one for F3 in Figure 1 a minor modification of Eq (1) is required because F3 frequency increases rather than decays. This is done by replacing e^{-at} by its

complement $(1 - e^{-at})$. We then obtain the following expression:

$$F3(t) = (F3_L - F3_T) * (1 - e^{-at}) + F3_T \quad (2)$$

Speech materials

At the time of submitting this report recordings and analyses are ongoing. Our intention is to apply the proposed measure to both normal and dysarthric speakers. Here we present some preliminary normal data on consonant and vowel sequences occurring in V:CV and VC:V frames with $V=[i\ e\ \epsilon\ a\ \alpha\ o\ u]$ and $C=[b\ d\ g]$. As an initial goal we set ourselves the task of describing how the time constants for F1, F2 and F3 vary as a function of vowel features, consonant place (articulator) and formant number.

The first results come from a normal male speaker of Swedish reading lists with randomized VC:V and VC:V words each repeated five times. No carrier phrase was used.

Since one of the issues in the project concerns the relationship between 'movement speed' (as derived from formant frequency rate of change) and 'speech rate' (number of phonemes per second) we also had subjects produce repetitions of a second set of test words: *dag*, *dagen*, *Dagobert* ['da:gøbæt], *dagobertmacka*.

This approach was considered preferable to asking subjects to "vary their speaking rate". Although this instruction has been used frequently in experimental phonetic work it has the disadvantage of leaving the speaker's use of 'over-' and 'underarticulation' - the 'hyper-hypo' dimension - uncontrolled (Lindblom 1990). By contrast the present alternative is attractive in that the selected words all have the same degree of main stress ('huvudtryck') on the first syllable [da:(g)-]. Secondly speaking rate is implicitly varied by means of the 'word length effect' which has been observed in many languages (Lindblom et al 1981). In the present test words it is manifested as a progressive shortening of the segments of [da:(g)-] when more and more syllables are appended.

Determining time constants

To measure transition time constants the following protocol was followed.

The speech samples were digitized and examined with the aid of wide-band spectrographic displays in Swell. [FFT points 55/1024, Bandwidth 400 Hz, Hanning window 4 ms].

For each sample the time courses of F1, F2 and F3 were traced by clicking the mouse along the formant tracks. Swell automatically produced a two-column table with the sample's time and frequency values.

The value of α was determined after rearranging and generalizing Eq (1) as follows:

$$(F_n(t) - F_{nT}) / (F_{nL} - F_{nT}) = e^{-\alpha t} \quad (3)$$

and taking the natural logarithm of both sides which produces:

$$\ln[(F_n(t) - F_{nT}) / (F_{nL} - F_{nT})] = -\alpha t \quad (4)$$

Eq (4) suggests that, by plotting the logarithm of the $F_n(t)$ data – normalized to vary between 1 and zero – against time, a linear cluster of data points would be obtained (provided that the transition is exponential).

A straight line fitted to the points so that it runs through the origin would have a slope of α . This procedure is illustrated in Figure 2.

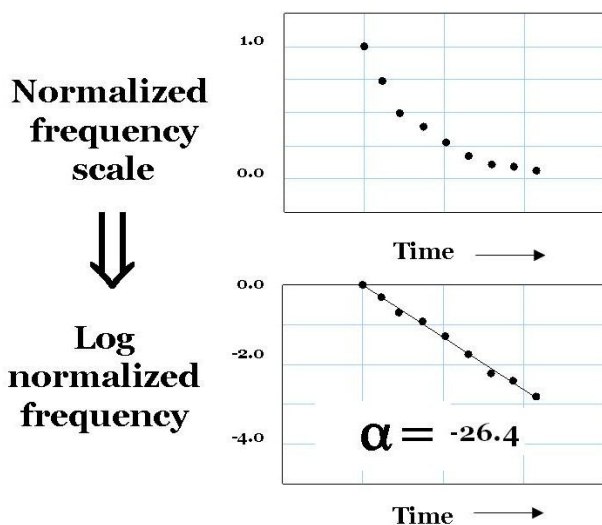


Figure 2. Normalized formant transition: Top: linear scale running between 1.0 and zero; (Bottom): Same data on logarithmic scale. The straight-line pattern of the data points allows us to compute the slope of the line. This slope determines the value of the time constant.

Figure 3 gives a representative example of how well the exponential model fits the data. It shows the formant transitions in [da]. Measurements from 5 repetitions of this syllable were pooled for F1, F2 and F3. Time constants were determined and plugged into the formant equations to generate the predicted formant tracks (shown in red).

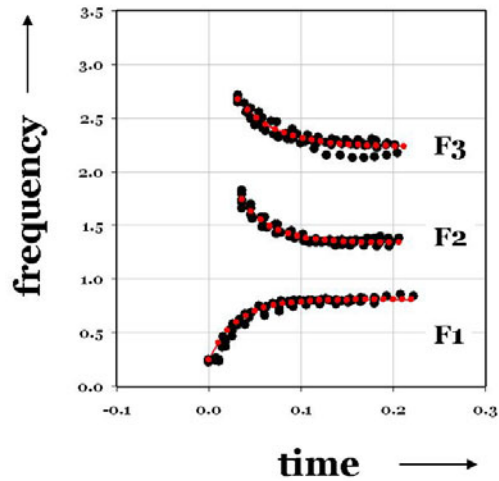


Figure 3. Measured data for 5 repetitions of [da] (black dots) produced by male speaker. In red: Exponential curves derived from the average formant-specific values of locus and target frequencies and time constants.

Results

High r squared scores were observed ($r^2 > 0.90$) indicating that exponential curves were good approximations to the formant transitions.

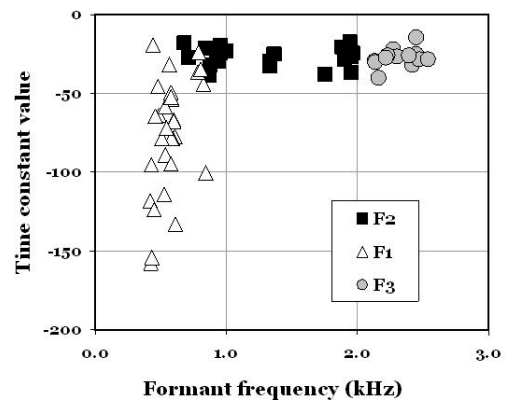


Figure 4. Formant time constants in V:CV and VC:V words plotted as a function of formant frequency (kHz). F1 (open triangles), F2 (squares) and F3 (circles). Each data point is the value derived from five repetitions.

The overall patterning of the time constants is illustrated in Figure 4. The diagram plots time constant values against frequency in all V:CV and VC:V words. Each data point is the value derived from five repetitions by a single male talker. Note that, since decaying exponentials are used, time constants come out as negative numbers and all data points end up below the zero line.

F1 shows the highest negative values and the largest range of variation. F2 and F3 are seen to occupy a limited range forming a horizontal pattern independent of frequency.

A detailed analysis of the F1 transition suggests preliminarily that VC transitions tend to be somewhat faster than CV transitions; VC: data show larger values than VC measurements.

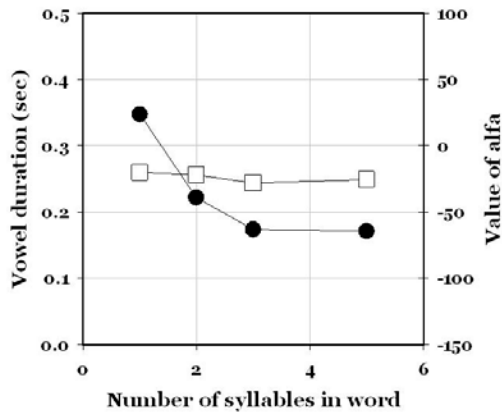


Figure 5. Vowel duration (left y-axis) and F2 time constants (right y-axis) plotted as a function of number of syllables per word.

Figure 5 shows how the duration of the vowel [ɑ:] in [dɑ:(g)-] varies with word length. Using the y-axis on the left we see that the duration of the stressed vowel decreases as a function of the number of syllables that follow. This compression effect implies that the 'speaking rate increases with word length.

The time constant for F2 is plotted along the right ordinate. The quasi-horizontal pattern of the open square symbols indicates that time constant values are *not* influenced by the rate increase.

Discussion

Non-linear acoustic mapping

It is important to point out that the proposed measure can only give us an indirect estimate of articulatory activity. One reason is the non-linear relationship between articulation and acoustics which for identical articulatory movement speeds could give rise to different time constant values.

The non-linear mapping is evident in the high negative numbers observed for F1. Do we conclude that the articulators controlling F1 (primarily jaw opening and closing) move fast-

er than those tuning F2 (the tongue front-back motions)? The answer is no.

Studies of the relation between articulation and acoustics (Fant 1960) tell us that rapid F1 changes are to be expected when the vocal tract geometry changes from a complete stop closure to a more open vowel-like configuration. Such abrupt frequency shifts exemplify the non-linear nature of the relation between articulation and acoustics. Quantal jumps of this kind lie at the heart of the Quantal Theory of Speech (Stevens 1989). Drastic non-linear increases can also occur in other formants but do not necessarily indicate faster movements.

Such observations may at first appear to make the present method less attractive. On the other hand, we should bear in mind that the transformation from articulation to acoustics is a physical process that constrains both normal and disordered speech production. Accordingly, if identical speech samples are compared it should nonetheless be possible to draw valid conclusions about differences in articulation.

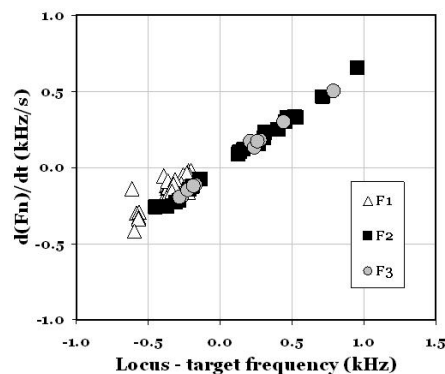


Figure 6: Same data as in Figure 4. Abscissa: Extent of F1, F2 or F3 transition ('locus'-'target' distance). Ordinate: Average formant frequency rate of change during the first 15 msec of the transition.

Formant frequency rates of change are predictable from transition extents.

As evident from the equations the determination of time constants involves a normalization that makes them independent of the extent of the transition. The time constant does not say anything about the raw formant frequency rate of change in kHz/seconds. However, the data on formant onsets and targets and time constants allow us to derive estimates of that dimension by inserting the measured values into Eqs (1) and (2) and calculating $\Delta F_n/\Delta t$ at transition onsets for a time window of $\Delta t=15$ milliseconds.

The result is presented in Figure 6 with $\Delta F_n/\Delta t$ plotted against the extent of the transition (locus-target distance). All the data from three formants have been included. It is clear that formant frequency rates of change form a fairly tight linear cluster of data points indicating that rates for F2 and F3 can be predicted with good accuracy from transition extents. Some of data points for F1 show deviations from this trend.

Those observations help us put the pattern of Figure 3 in perspective. It shows that, when interpreted in terms of formant frequency rate of change (in kHz/seconds), the observed time constant patterning does not disrupt a basically lawful relationship between locus-target distances and rates of frequency change. A major factor behind this result is the stability of F2 and F3 time constants.

Figure 6 is interesting in the context of the 'gestural' hypothesis which has recently been given a great deal of prominence in phonetics. It suggests that information on phonetic categories may be coded in terms of formant transition dynamics (e.g., Strange 1989). From the vantage point of a gestural perspective one might expect the data of the present project to show distinct groupings of formant transition time constants in clear correspondence with phonetic categories (e.g., consonant place, vowel features). As the findings now stand, that expectation is not borne out. Formant time constants appear to provide few if any cues beyond those presented by the formant patterns sampled at transition onsets and endpoints.

Articulatory processes in dysarthria

What would the corresponding measurements look like for disordered speech? Previous acoustic phonetic work has highlighted a slower average rate of F2 change in dysarthric speakers. For instance, Weismer et al (1992) investigated groups of subjects with amyotrophic lateral sclerosis and found that they showed lower average F2 slopes than normal: the more severe the disorder the lower the rate.

The present type of analyses could supplement such reports by determining either how time constants co-vary with changes in transition extent and duration, or by establishing that normal time constants are maintained in dysarthric speech. Whatever the answers provided by such research we would expect them to present significant new insights into both normal and disordered speech motor processes.

Clues from biomechanics

To illustrate the meaning of the numbers in Figure 3 we make the following simplified comparison. Assume that, on the average, syllables last for about a quarter of a second. Further assume that a CV transition, or VC transition, each occupies half of that time. So formant trajectories would take about 0.125 seconds to complete. Mathematically a decaying exponential that covers 95% of its amplitude in 0.125 seconds has a time constant of about -25. This figure falls right in the middle of the range of values observed for F2 and F3 in Figure 3.

The magnitude of that range of numbers should be linked to the biomechanics of the speech production system. Different articulators have different response times and the speech wave reflects the interaction of many articulatory components. So far we know little about the response times of individual articulators.

In normal subjects both speech and non-speech movements exhibit certain constant characteristics.

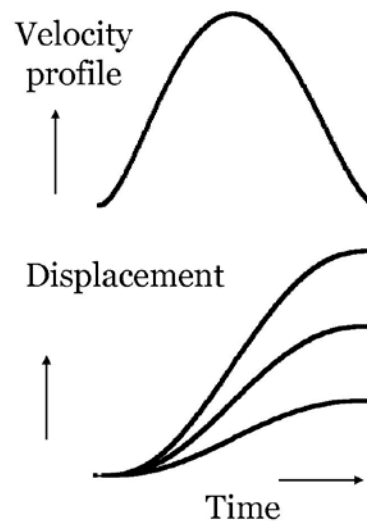


Figure 7: Diagram illustrating the normalized 'velocity profile' associated with three point-to-point movements of different extents.

In the large experimental literature on voluntary movement there is an extensively investigated phenomenon known as "velocity profiles" (Figure 7). For point-to-point movements (including hand motions (Flash & Hogan 1985) and articulatory gestures (Munhall et al 1985)) these profiles tend to be smooth and bell-shaped. Apparently velocity profiles retain their geometric shape under a number of conditions: "...the form of the velocity curve is invariant under transformations of movement amplitude, path, rate, and inertial load" (Ostry et al 1987:37).

Figure 7 illustrates an archetypical velocity profile for three hypothetical but realistic movements. The displacement curves have the same shape but differ in amplitude. Hence, when normalized with respect to displacement, their velocity variations form a single “velocity profile” which serves as a biomechanical “signature” of a given moving limb or articulator.

What the notion of velocity profiles tells us that speech and non-speech systems are strongly damped and therefore tend to produce movements that are s-shaped. Also significant is the fact that the characteristics of velocity profiles stay invariant despite changes in experimental conditions. Such observations indicate that biomechanical constancies are likely to play a major role in constraining the variation of formant transition time constants both in normal and disordered speech.

However, our understanding of the biomechanical constraints on speech is still incomplete. We do not yet fully know the extent to which they remain fixed, or can be tuned and adapted to different speaking conditions, or are modified in speech disorders (cf Forrest et al 1989). It is likely that further work on comparing formant dynamics in normal and dysarthric speech will throw more light on these issues.

References

- Broad D J & Fertig R (1970): "Formant-frequency trajectories in selected CVC syllable nuclei", *J Acoust Soc Am* 47, 1572-1582.
- Duffy J R (1995): *Motor speech disorders: Substrates, differential diagnosis, and management*, Mosby: St. Louis, USA.
- Fant G (1960): *Acoustic theory of speech production*, Mouton: The Hague.
- Forrest K, Weismer G & Turner G S (1989): "Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults", *J Acoust Soc Am* 85(6), 2608-2622.
- Hartelius L, Nord L & Buder E H (1995): "Acoustic analysis of dysarthria associated with multiple sclerosis", *Clinical Linguistics & Phonetics*, Vol 9(2):95-120
- Flash T & Hogan N (1985): "The coordination of arm movements: An experimentally confirmed mathematical model", *J Neuroscience Vol* 5(7). 1688-1703.
- Lindblom B, Lyberg B & Holmgren K (1981): *Durational patterns of Swedish phonology: Do they reflect short-term memory processes?*, Indiana University Linguistics Club, Bloomington, Indiana.
- Lindblom B (1990): "Explaining phonetic variation: A sketch of the H&H theory", in Hardcastle W & Marchal A (eds): *Speech Production and Speech Modeling*, 403-439, Dordrecht: Kluwer.
- Munhall K G, Ostry D J & Parush A (1985): "Characteristics of velocity profiles of speech movements", *J Exp Psychology: Human Perception and Performance* Vol 11(4):457-474
- Ostry D J, Cooke J D & Munhall K G (1987): "Velocity curves of human arm and speech movements", *Exp Brain Res* 68:37-46
- Park S-H (2007): *Quantifying perceptual contrast: The dimension of place of articulation*, Ph D dissertation, University of Texas at Austin
- Rosen K M, Kent R D, Delaney A L & Duffy J R (2006): "Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers", *JSLHR* 49:395-411.
- Schalling E (2007): *Speech, voice, language and cognition in individuals with spinocerebellar ataxia (SCA)*, Studies in Logopedics and Phoniatrics No 12, Karolinska Institutet, Stockholm, Sweden
- Stevens K N, House A S & Paul A P (1966): "Acoustical description of syllabic nuclei: an interpretation in terms of a dynamic model of articulation", *J Acoust Soc Am* 40(1), 123-132.
- Stevens K N (1989): "On the quantal nature of speech," *J Phonetics* 17:3-46.
- Strange W (1989): "Dynamic specification of coarticulated vowels spoken in sentence context", *J Acoust Soc Am* 85(5):2135-2153.
- Talley J (1992): "Quantitative characterization of vowel formant transitions", *J Acoust Soc Am* 92(4), 2413-2413.
- Weismer G, Martin R, Kent R D & Kent J F (1992): "Formant trajectory characteristics of males with amyotrophic lateral sclerosis", *J Acoust Sec Am* 91(2):1085-1098.