

# SYLLABLE AFFILIATION OF FINAL CONSONANT CLUSTERS UNDERGOES A PHASE TRANSITION OVER SPEAKING RATES

*Philip Gleason, Betty Tuller, & J. A. Scott Kelso*

Program in Complex Systems and Brain Sciences  
Center for Complex Systems, Florida Atlantic University

## ABSTRACT

Previous work (Tuller & Kelso, 1990) reported a change in syllable affiliation, which was reflected by a change in the relative phase, of glottal and lip movements, as the rate of speaking a VC syllable increased. That is /ip#ip#.../ becomes /pi#pi#.../ as speaking rate is systematically increased. Here we report a change in the syllable affiliation of part of a final consonant cluster; /opt#opt#.../ becomes /top#top#.../ at higher speaking rates. The relative phase between the time series generated by two supralaryngeal structures, the lip and tongue tip, was measured in three ways, by discrete Fourier transform, by the cospectrum, and by dynamic time warping. Speaking rate was controlled using a metronome that began at 1 Hz and increased to 3.25 Hz in 0.083 Hz steps every 12 repetitions. Subjects repeated the target word "opt," "hopped," or "top" in time with the metronome while movements of their tongue, tongue blade, lower lip, and jaw were tracked by an alternating magnetic field device, the Articulograph AG100. Phonetic analysis confirmed the presence of a transition from VCC to CVC syllable at higher speaking rates. The perceptual change corresponds to a statistically significant change in the measurements of relative phase, indicating that perceived syllable affiliation is determined, to some extent, by relative timing of articulatory events. This effect is consistent across the five subjects whose results are reported here. Three methods of measuring relative timing are compared, and the theoretical issues behind these methods are discussed.

## 1. METHOD

### 1.1. Articulatory Recording

An Articulograph (AG100, manufactured by Carstens Medizintechnik GmbH) was used to track movements of the tongue, lips, and jaw. This device uses three radio transmitters mounted on a helmet, with the transmitters distributed equally around the helmet on the midline, to track the movements of up to five small (3 mm X 3 mm X 1 mm) receivers (referred to as pellets) which are glued to a subject's mouth using a surgical adhesive. A pair of thin wires from each pellet led out of the subject's

mouth to a preamplifier clipped to a towel around the subject's neck.

The first pellet was placed under the upper lip on the maxilla just below the frenulum as an unmoving reference point. The second receiver was placed on the corresponding point on the mandible to track jaw movements, the third on the vermilion border of the lower lip on the mid-sagittal plane, the fourth on the centerline of the tongue blade 3 cm. from the tongue tip, and the fifth on the lower surface of the tongue tip so as not to interfere with the production of /t/.

The Articulograph signal was recorded by a micro-computer, while the subject's speech was recorded by a VAX mini-computer. The two recordings were synchronized by a signal at the start of recording, and by start and stop time stamps written by the recording software.

### 1.2. Subjects

Subjects were undergraduates at Florida Atlantic University who participated in the experiment for course credit and graduate student volunteers. Ten subjects were recorded and five of those subjects were analyzed. Subjects were rejected because of instrumental failure, or because the subject failed to follow the metronome, even at slow speeds or failed to produce syllables which could be evaluated as "opt" or "top".

### 1.3. Procedure

During the experimental session, the subject was asked to repeat a word in time with a metronome click presented over plastic airline headphones. The metronome's rate was increased by 1/12 Hz every 12 beats, from 1.0 Hz to 3.25 Hz. giving a total of 28 plateaus of steady speaking rate. The subject was seated in a sound insulated booth, wearing the Articulograph helmet and the airline headphones. A boom microphone was placed approximately 3 inches in front of the subject.

The subject inevitably had to take several breaths during a trial and was instructed to stop speaking while inhaling, to resume speaking at the

metronome frequency, and to keep pace with the metronome. The subject was also told that her pronunciation might not be as good at higher speeds. If she felt a change of pronunciation occurring she should allow it to happen, and concentrate on keeping pace with the metronome.

Five trials of the target word "opt" were recorded, followed by five trials of the target word "hopped", followed by five trials of "top." This order of blocking trials was intended to keep the subject from switching to "top" deliberately. In what follows, we omit discussion of "hopped" trials. These three target words were chosen because lip closure for /p/ and tongue contact with the alveolar ridge for /t/ are clear points of demarcation in both the Articulograph trace and the acoustic recording.

## 2. ANALYSIS

### 2.1. Three Measures of Relative Phase

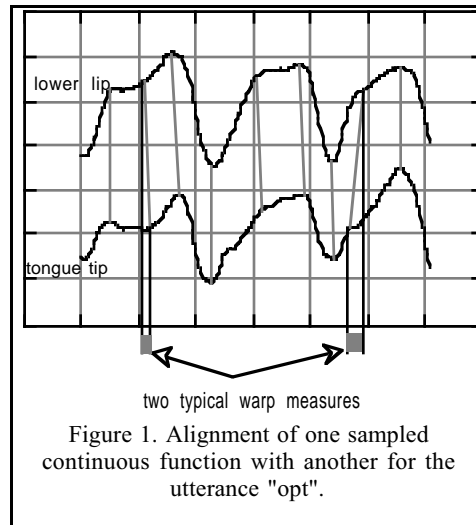
The **Discrete Fourier Transform** is applied to each syllable, as defined by the peak-to-peak distance of the y direction of the tongue tip trace. A DFT is calculated for both the tongue tip y motion, and the lower lip y motion over the same time window. The difference between the phase of the first components of these two DFTs yields the relative phase.

The DFT method suffers from the use of arctan, but for this data it is a good method if the movements of the articulators are roughly sinusoidal. To the extent that the shape of a cycle deviates from a sinusoid the phase value is significantly altered by redistributing weight between the real and imaginary components. An advantage of the method is its insensitivity to high frequency noise.

**Cospectrum** is used to calculate relative phase by taking a peak to peak interval in one time series which determines a window. A cospectrum is calculated between two time series over that window, and the relative phase between the two series is the phase of the first component of the DFT.

This method is analytically the same as the first method, but the resulting values are different for numerical reasons. The statistical results based on this method were the same as those based on the first method.

**Dynamic Time Warping (DTW)** is an algorithm which finds the optimal (lowest cost) order-preserving alignment between two sequences for a given cost function. Fig 1 shows the DTW alignment of two time series of Articulograph data. Note that peaks are aligned to peaks, valleys to valleys, and that inflection points are also aligned.



The DTW algorithm produces a time series of roughly the same length as one of the input time series, while the DFT and cospectrum produce one value per cycle. In the analysis the mean value per cycle (defined as a peak to peak window) of the DTW was used as the "relative phase".

### 2.2. Plateau Alignment

Subjects varied from trial to trial in the speaking rate at which they made the transition from producing a syllable perceived as "opt" to one perceived as "top". Plateaus were aligned across trials in order to align the first transition point, defined as the first instance in which the subject produced an "opt" followed by three "top"s. The plateau in which the transition occurred was labeled plateau 0, and previous plateaus were numbered -1, -2, etc. to -6. Following plateaus were numbered 1, 2, etc. to 6. Thus a window of 13 plateaus centered the point of first transition was used in the analysis. All 5 of a subject's "opt" trials were aligned and pooled. "top" trials were paired with "opt" trials and aligned on the basis of speaking rate, then pooled across trials to form a set of 13 plateaus which could be compared to the 13 plateaus from the "opt" trials.

### 2.3. Pre-transition & Post-transition Analysis

Separate statistical analyses were run on pre-transition plateaus, and on post-transition plateaus. Subjects' productions were separated on the basis of whether a listener perceived the syllable as "opt" or "top". In the pre-transition case productions were either "opt" or "top". In the post-transition plateaus there were three types of productions: "opt", "top" and "opt>top" which is a "top" production which occurred in an "opt" trial as the result of rapid speech. The pre-transition analysis included 5 subjects, and 6 plateaus (-6..-1); the post-transition analysis included 5 subjects and 7 plateaus (0..6).

### 3. RESULTS

The DFT and DTW measures and the cospectrum measure gave nearly identical results as far as the main effects of interest here are concerned. For the sake of brevity we will report only the cospectrum measure.

In the pre-transition plateaus there were main effects ( $p < 0.05$  or better) of subject,  $F(4,17)=37.87$ , and percept  $F(1,11)=8.89$ . There was no significant main effect of plateau,  $F(5,83)=0.08$ , and there were no significant interactions.

In the post-transition plateaus there were significant main effects of subject  $F(4,17)=28.83$  and of production  $F(2,28)=31.48$ . There was no significant effect of plateau nor were there any interactions. In addition Tukey's Studentized Range (HSD) Test revealed that "top" productions from "opt" trials were no different from "top" productions from "top" trials, but both were significantly different from "opt" productions.

Figure 2 shows the distribution of the mean relative phase between tongue and lip motion, averaged across trials of "opt" and "top", pre- and post-transition. Note there are clear differences in the timing relations for the two utterances. Also, when "opt" switches to "top", the relative phase adopts the appropriate value of the "new" syllabic form. There was no consistent increase in the variance of the relative phase in the plateaus immediately preceding the transition plateau. However, the mode of the relative phase distributions for each pre-transition plateau showed a slow and consistent shift toward the relative phase values associated with productions of "top".

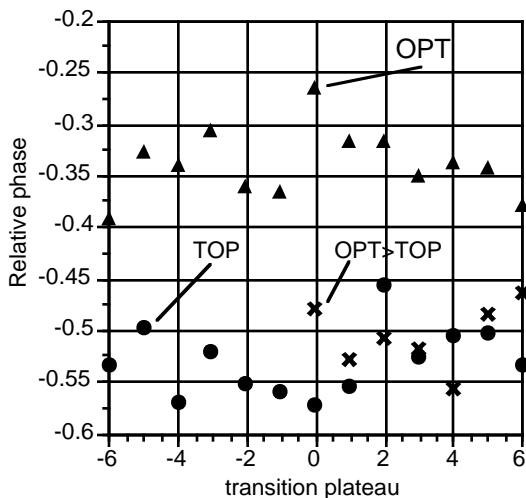


Figure 2. Mean relative phase (radians) of tongue and lip motion around the transition plateau. Triangle = "opt", circle = "top", X = "opt>top".

### 4. DISCUSSION

Relative phase is clearly a correlate of syllable affiliation. Measuring relative phase is a difficult issue because jaw, lip and tongue movements are not sinusoidal (see Fig 1), and so using Fourier analysis is an approximation to the shape of speech gestures. Nevertheless, with a metronome and iterated speech, a speaker's movements are sufficiently regular that the significance of relative phase to perception has been demonstrated. A more fine grained method of analysis, the DTW, proved to be too sensitive although it is possible that some other measure derived from the DTW would be better than the mean value per cycle.

The data did not show an increase in variance near the point of transition. Thus the notion that the change of syllable affiliation reflects a second order phase transition is not supported (or the experimentally determined timescales are not optimal for detecting such transitions). Instead, the trend of the mode of the per plateau distributions indicates a smooth transition between syllable affiliations. This view is supported by the phonetic observation that there are intermediate states between "opt-opt" and "top-top", i.e. the /t/ becomes ambisyllabic. This smooth transition is reminiscent of other motor behaviors, such as the relative coordination between the arms and legs which exhibit slow shifts in the phasing due to eigen frequency differences among the components. (Kelso & Jeka, 1995).

**Acknowledgements** This work was supported by National Institutes of Mental Health Training Grant in Complex Systems and Brain Sciences MH 19116.

### 5. REFERENCES

1. Jeka, J. J., & Kelso, J. A. S. (1995). Manipulating Symmetry in the Coordination Dynamics of Human Movement. *Journal of Experimental Psychology: Human Perception and Performance*, 21(2), 360-374
2. Tuller, B., & Kelso, J. A. S. (1990). Phase Transitions in Speech Production and Their Perceptual Consequences. In Jeannerod (Ed.), *Attention & Performance XIII*, (pp. 429-451).