GESTURAL PHASING IN /KT/ SEQUENCES CONTRASTING WITHIN AND CROSS WORD CONTEXTS

Mark Tiede,[1,2] *Stefanie Shattuck-Hufnagel,*[1] *Beth Johnson,*[3] *Satrajit Ghosh,*[1]
Melanie Matthies,[4,1] *Majid Zandipour,*[1] *Joseph Perkell* [1,4]

[1]Speech Comm. Group, MIT R.L.E.; [2]Haskins Laboratories; [3]Yale U.; [4]Boston U.
tiede@speech.mit.edu

ABSTRACT

This paper presents results of a kinematic study of the articulatory phasing between successive /k/ and /t/ gestures in English tautosyllabic ("pact op") and heterosyllabic ("pack top") contexts, varied by speaking rate and stress. Although subject responses varied, in general coda clusters are shown to be significantly less variable in timing than heterosyllabic sequences relative to the labial gestures of the carrier context.

Keywords: speech production, articulatory gestures, EMMA

1. INTRODUCTION

Context-dependent timings of articulatory gestures associated with the production of consonant clusters are of interest for their implications for speech organization and planning, yet studies of these phenomena remain sparse. X-ray microbeam data collected at the University of Tokyo were used by Fujimura [1] to argue for the relative invariance of consonantal gestures ("icebergs") afloat in variable vocalic contexts. Browman & Goldstein [2] used the same dataset to support an account of casual speech alternation based on degree of gestural overlap ("perfec[t] memory"), and to examine /spl/ sequences across word boundaries [3]. In the latter study they found that inter-gestural timing variability was less within word onsets (e.g. "pea splots") than across word boundaries ("peace plots"). Several researchers have used EPG to observe tongue-palate contact patterns induced by clusters, including Hardcastle [4], who found rate effects on degree of overlap in /kl/ sequences, and Byrd [5], who confirmed the results of [3] with findings of less overlap and reduced timing variability observed in onset clusters than in codas or heterosyllabic sequences.

But while EPG can be effective for assessing degree of overlap, it is arguably less useful for establishing precise timing relations among cluster elements due to contact saturation during closure. The EMMA point source tracking technique applied here offers in this context the advantage of continuous (rather than discrete contact) sensor tracking, which facilitates the robust identification of velocity extrema associated with maximal consonant constrictions. Although [5] did not find significant differences between /gd#/ and /g#d/ sequences (and [3] did not examine that contrast), it is reasonable to investigate whether more precise identification of gestural timings across a wider range of production conditions would show reduced variability in codas (relative to heterosyllabic sequences) similar to that already observed for onsets. Some supporting evidence is provided by a previous EMMA study (Tiede *et al.* [6]) which established that gestural phasing between /k/ and /t/ in /kt#m/ sequences similar to "perfect memory" was relatively invariant across three production rates, with a wide range of overlap by /m/ apparently licensed by the word boundary.

Accordingly, this work focuses on /kt#/ and /k#t/ clusters produced by several subjects, over different speaking rates and stress conditions. It takes as its hypothesis that relative timing between /k/ and /t/ will be less variable in coda clusters than in heterosyllabic sequences. Expressed in terms of gestural phasing (as outlined below), it is expected that the intergestural phasing within codas will be relatively invariant with respect to production duration, whereas longer instances of heterosyllabic contexts will show a positively correlated increase in /k/ : /t/ phasing as a consequence of the intervening word boundary.

2. METHODS

2.1. Participants

Subjects were nine female and six male young adult native speakers of American English, with normal hearing and no apparent speech deficits.

Each subject participated in a speech production experiment in which their speech and articulatory movements were recorded.

2.2. Materials

Subjects were asked to produce the following two sentences:

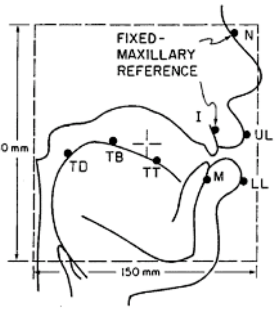
(1) "Say *pack top* for me"
(2) "Say *pact op* for me"

Each of these utterances was repeated ten times within three separate blocks, intermixed with similar stimuli elicited for an experiment wider in scope than that reported here. Different production conditions were used in each block to elicit a range of speaking styles and token durations. In Block 1 trochaic stress was suggested (e.g. ***pack*** *top*). In Block 2 iambic stress was suggested (e.g. *pack* ***top***). Within both of these blocks subjects were instructed to produce the target sentence at their normal speaking rate. Trochaic stress was also suggested in Block 3, but with the instruction to produce each sentence using a 'fast' (twice normal) speaking rate.

2.3. Recordings

An electromagnetic midsagittal articulometer system (EMMA [7]) was used to transduce the location of sensors attached to the subject's speech articulators (see Figure 1). Custom control software was used to sequence the experiment, display stimuli to the subject, and record the movement and acoustic signals for each trial. After appropriate hardware filtering, audio data were recorded at 16 kHz and movement at 500 Hz.

Figure 1: EMMA sensor placement. TD used to characterize /k/ closure, TT /t/ closure, and Euclidean distance UL : LL (Lip Aperture) used for /p/ closure.



2.4. Data Analysis

Raw EMMA voltage signals were first converted to positions over time on the midsagittal plane, and the reference sensors (N, I) were used to rotate and translate each position signal to a consistent maxillary frame of reference. After lowpass filtering for smoothing (Fc = 12.5 Hz) four derived signals were computed: Lip Aperture (LA) and its velocity (from the Euclidean distance between the upper and lower lip sensors), tongue tip (TT) speed, and the vertical velocity of the tongue dorsum (TDy). These signals were used to identify the offsets of four robust articulatory events on each token:

P1 – LA velocity zero-crossing associated with first /p/ closure

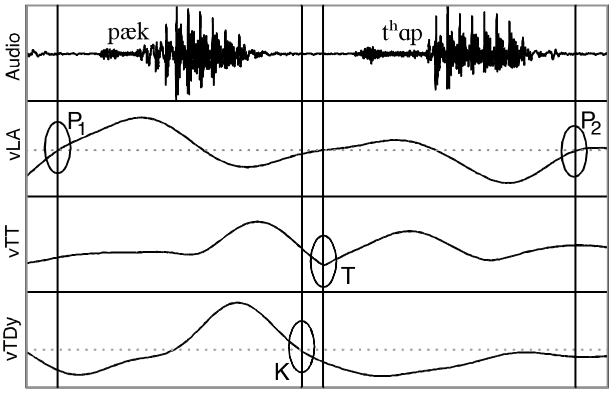
K – TDy velocity zero-crossing associated with /k/ closure

T – TT speed minimum associated with /t/ closure

P2 – LA velocity zero-crossing associated with last /p/ closure

Each offset was determined using an interactive procedure that identified the nearest velocity extremum to a cursor positioned on a display (see Figure 2). In this way any selected location near the event resulted in the same labeled offset.

Figure 2: Example "pack top" token showing labeled extrema.



For each token these four offsets were then used to produce measures of carrier duration ***D*** and relative phasing ***Φ*** between /k/ and /t/ as follows:

(1) $$D = P_2 - P_1$$

(2) $$\Phi = \frac{T - K}{D}$$

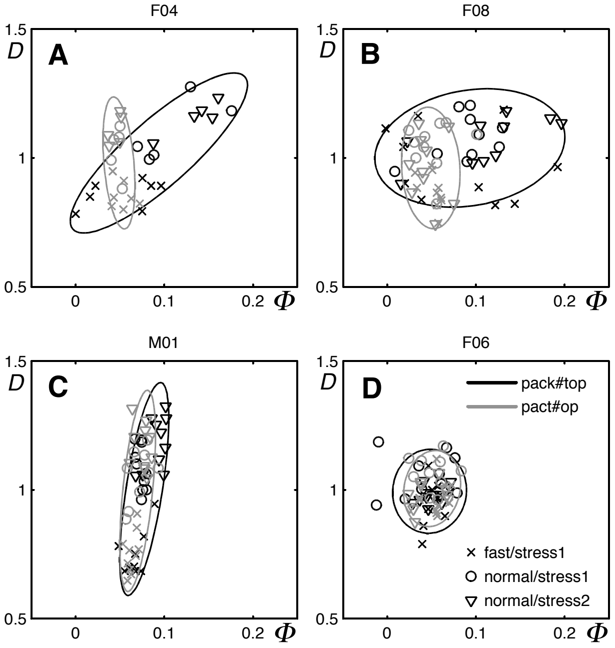
To support cross-subject comparisons each subject's durations were normalized by the mean

P2-P1 duration for that subject across all tokens for that subject.

3. RESULTS

Four patterns of response were observed, exemplified by scatterplots regressing /k/:/t/ phasing $\boldsymbol{\Phi}$ with normalized duration $\boldsymbol{D}$ as shown in Figure 3. In the first group (A) four subjects showed significant correlations for both /kt#/ and /k#t/ contexts, and these slopes were significantly different using confidence intervals computed by Fisher's z-score conversion of Pearson's *r* (p<.05). The six subjects of group B did not reach significance by this criterion, but they did show significant distribution differences between contexts as computed by discriminant analysis. The remaining five subjects showed no significant effect of context on $\boldsymbol{\Phi}$, though four of the five did show differences in duration consistent with the rate instructions (group C).

Figure 3: Scatterplots of data from four representative subjects differentiated by context (/k#t/ in black *vs*. kt#/ in gray) showing 95% confidence ellipses. The vertical axis shows normalized P_1:P_2 duration $\boldsymbol{D}$, the horizontal axis relative /k/:/t/ phasing $\boldsymbol{\Phi}$. Group A (N=4): significantly distinct slopes; B (N=6) significant difference in distribution; C (N=4): expected range in production rate but no difference in phasing between contexts; D (N=1): no variation.



The single subject in group D showed no significant variation in production of any kind. Overall, 12 subjects showed lower variability in the production of coda contexts, as assessed by comparing the area of 95% confidence ellipses.

Effects of context and production condition on relative phasing $\boldsymbol{\Phi}$ were quantified by a within-subjects repeated measures analysis of variance. Results are given in Table 1 and show that the production context factor (/kt#/ *vs*. /k#t/) had the greatest overall influence on phasing. Production condition (rate, stress) was also significant, but not its interaction with context.

Table 1: Repeated measures ANOVA results of /k/:/t/ phasing $\boldsymbol{\Phi}$ as dependent variable, with context (/kt#/, /k#t/) and production condition (rate, stress) selection factors.

Dep: PHASE

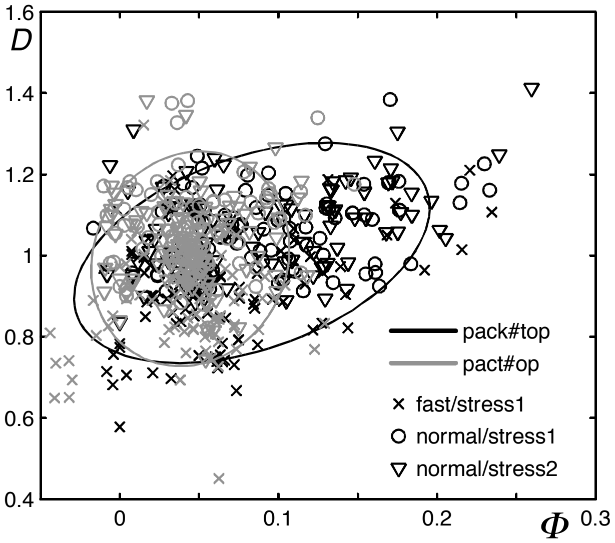
Source	df	SS	MS	F	P
ConTeXt	1	0.013	0.013	36.446	0.000
CONDition	2	0.003	0.002	9.423	0.002
CTX * COND	2	0.001	0.000	0.003	0.997
CTX * SUBJ	9	0.003	0.000		
COND * SUBJ	18	0.003	0.000		
CTX*CND*SBJ	18	4.596	0.255		

4. DISCUSSION

The hypothesis predicted that the phasing between /k/ and /t/ closures (expressed as the difference between their relative offsets within the bracketing bilabials of the carrier) would increase with overall duration in heterosyllabic sequences, but remain relatively invariant tautosyllabically. Although subjects differed in their responses, with some failing to distinguish between context types, no significant inversion of this expected pattern was observed.

Four of the fifteen subjects showed the predicted pattern most clearly, with an additional six showing some form of consistently different pattern for the two context types. Of the remaining five subjects, one showed an essentially uniform pattern across all conditions, while the remaining four produced the expected contrasts in duration but with no significant differences in phasing associated with context. For the ten subjects who did show a difference in phasing between contexts (pooled in Figure 4), the correlation with carrier duration in /k#t/ sequences was positive. This indicates that as production rate increases, overlap between /k/ and /t/ gestures produced across a word boundary also increases. Indeed, for subjects like F04 (group A), at the fastest observed rate /k/ and /t/ constrictions were produced essentially simultaneously. Conversely, at slower rates, separation between gestures across the word boundary is greater than would be expected through scaling by the increased duration. Contrasting with this is the relatively invariant phasing observed between /k/ and /t/ in coda contexts: although subjects (with one exception) did produce a range of durations for /kt#/ stimuli as well, the relative timing between /k/ and /t/ gestures was preserved.

Figure 4: Pooled regression of the 10 subjects that significantly differentiated phasing by context (/k#t/ *vs*. kt#/) showing 95% confidence ellipses. The vertical axis shows normalized P_1:P_2 duration $\boldsymbol{D}$, the horizontal axis relative /k/:/t/ phasing $\boldsymbol{\Phi}$. Significant difference between slopes as assessed by Fisher's *r* to *z* conversion (p<.05).



A possible explanation for the range of observed subject responses might be the deliberately informal approach to speech rate taken in this experiment, in that each subject was free to interpret the instructions for "normal" and "fast" as they saw fit. In general, however, the same subjects who showed phasing differences between contexts were those who produced the largest absolute ranges of carrier durations, which suggests that phase manipulation can be a strategy for enhancing the contrast between sequence context types. This is compatible with the DIVA model of Guenther [8], in which coda phasing invariance follows from word-unit lexical retrieval and execution of learned speech motor programs in fluent speech production. It is also consistent with the phase window approach of Saltzman and Byrd [9] in which these attractor states for intergestural phasing support flexible control of relative timing that can be exploited for contrasting linguistic and para-linguistic purposes.

5. SUMMARY

The relative phasing between /k/ and /t/ closures is in general less variable in coda contexts than when juxtaposed across a word boundary. Although speakers do not react uniformly, a majority of those examined showed a positive correlation of phasing with utterance duration in heterosyllabic sequences, indicating that overlap across word boundaries increases with rate.

6. ACKNOWLEDGMENTS

The authors thank Harlan Lane and Frank Guenther for their help and comments. This work was supported by grant R01-DC01925 from the National Institute on Deafness and Other Disorders, NIH.

7. REFERENCES

[1] Fujimura, O. 1986. Relative invariance of articulatory movements: an iceberg model. In: Perkell, J., Klatt, D. (eds), *Invariance and variability in speech processes*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 226-234.

[2] Browman, C., Goldstein, L. 1990. Tiers in articulatory phonology, with some implications for casual speech. In: Kingston, J., Beckman, M. (eds), *Papers in laboratory phonology I: between the grammar and the physics of speech*. Cambridge: Cambridge University Press, 341-376.

[3] Browman, C., Goldstein, L. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45, 140-155.

[4] Hardcastle, W. 1985. Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication* 4, 247-263.

[5] Byrd, D. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24, 209-244.

[6] Tiede, M., Perkell, J., Zandipour, M., Matthies, M. 2001. Gestural timing effects in the "perfect memory" sequence observed under three rates by electromagnetometry. *J. Acoust. Soc. Am.* 110, 2657.

[7] Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., Jackson, M. 1992. Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *J. Acoust. Soc. Am.* 92, 3078-3096.

[8] Guenther, F. 1995. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psych. Review* 102, 594-621.

[9] Saltzman, E., Byrd, D. 2000. Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science* 19, 499-526.