from: Proceedings of the 1992 International Conference on Spoken Language Processing (ICSLP '92) Addendum, Banff, Alberta, Canada. Edmonton, Canada, Priority Printing, 1992.

831

A PERTURBATION STUDY OF LIP-LARYNX COORDINATION

Elliot Saltzman, Anders Löfqvist, Jeff Kinsella-Shaw, Philip E. Rubin

Haskins Laboratories
270 Crown Street, New Haven, CT 06511-6695, USA

Bruce Kay

Department of Psychology, Brown University,
Providence, RI, USA

ABSTRACT

In this paper we examine the sequential dynamics governing bilabial and laryngeal gestures, using a mechanical perturbation paradigm that entails the application of randomly timed downward forces to the lower lip during productions of discrete speech sequences (/pəsæpæpl/). In addition to labial and jaw kinematics and speech acoustics, laryngeal movements were monitored using transillumination techniques. Results indicated that the temporal intervals between successive bilabial closing gestures were affected systematically by the perturbations. Most of the timing changes occurred during the lip opening phase of these intervals, while the closing phases were relatively resistant to temporal distortion. Additionally, perturbations induced systematic changes in the timing of peak laryngeal abduction for the second /p/, supporting the hypothesis that the laryngeal adduction gesture is coordinated temporally within the oral closing and opening gestures for the /p/ closure. While the effects of perturbations on the bilabial gestures appear to be relatively direct, the effects on the laryngeal gesture appear to be indirect, i.e., the perturbations advance and/or delay the timing of bilabial events which, in turn, alter the timing of the laryngeal gesture.

1. INTRODUCTION

This preliminary report addresses the general issue of how to characterize the dynamics that underlie the temporal patterning of speech gestures. The experimental methodology entailed the application of unpredictable mechanical perturbations to the articulatory periphery during speech sequences, and the analysis of changes induced in the temporal or phasing structure of the sequences [1, 2, 3, 4, 5, 6]. Previously, we have reported data regarding the dynamics of intergestural phasing between gestures involved in forming and releasing successive bilabial constrictions, i.e., between successive gestures defined in the same portion of the vocal tract. In particular, we examined the temporal changes induced by mechanical perturbations delivered to the lower lip during the production of repetitive (/pæpæpæ.../) and discrete (/pəsæpæpl/) speech sequences [5, 6]. Comparison of the repetitive and discrete data indicated consistent patterns of temporal change across the different sequence types. This commonality allowed us to conclude that the same sequential dynamic processes underlie both the productions of repetitive sequences and discrete words. In particular, these data implied that a hypothesized central timing network for speech does not simply drive the articulatory periphery in a unidirectionally coupled manner; rather, central and peripheral dynamics must be coupled bidirectionally, so that feedback information from the biomechanical periphery also can influence the state of the central "clock".

In the present paper, we extend our previous work to examine as well the sequential dynamics governing the formation and release of constrictions in different regions of the vocal tract, i.e., between bilabilal release and laryngeal devoicing gestures [3, 4]. The data described below represent work in progress, and were collected as part of a larger perturbation study of laryngeal-oral coordination involving both discrete and repetitive speech sequences. To date, only the discrete utterances of a single speaker of American English have been fully analyzed, and these data are the focus of this paper.

2. METHOD

2.1 Equipment and data processing.

The experimental setup is shown in Figure 1. The subject sat in a dental chair with the head fixed. Perturbations were delivered to the lower lip through a paddle that rested on the lip. The paddle was connected to a DC brushless torque motor that applied a small tracking load of approximately 3 g throughout the experiment. At a predetermined point in time, the motor generated a 50 g load that was used to perturb the lip downward.

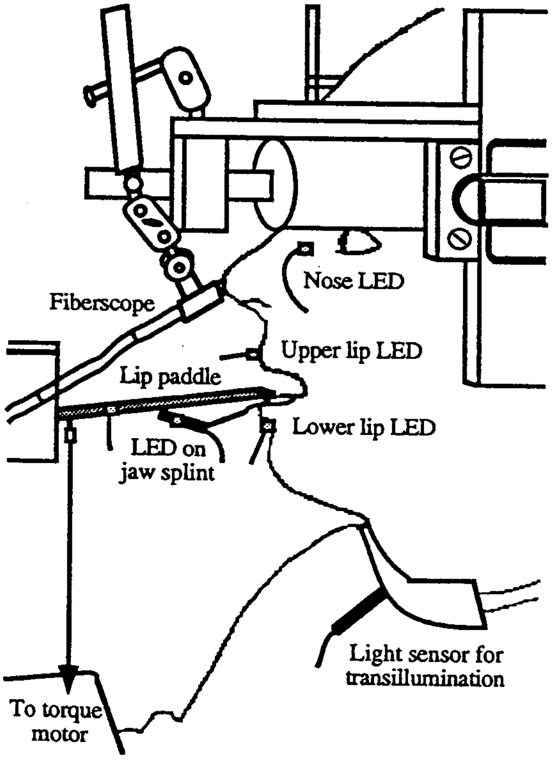

Figure 1. The experimental setup. Light-emitting diodes (LED's) are attached to articulators to be tracked. A fiberscope provides light for transillumination used to record laryngeal behavior. Perturbations are delivered to the lower lip via a lip paddle connected to a torque motor.

Oral articulatory movements were recorded using an optoelectronic system. Light-weight infrared light-emitting diodes were placed on the upper and lower lip. To track jaw movements, a custom-made splint was securely fastened to the lower premolars on both sides of the mouth. A diode was placed on a wire attached to the splint and protruding from the mouth. In addition, a diode was placed on the bridge of the nose; this signal was used to correct for head movements. Laryngeal articulatory movements, abduction and adduction, were recorded using transillumination of the larynx. A fiberscope introduced through the nose and placed in the pharynx provided illumination of the larynx. The light passing through the glottis was sensed by a photo-transistor placed on the neck just below the cricoid cartilage. During the experimental session, the fiberoptic image of the larynx was displayed on a video monitor to allow assessment of an unobstructed pathway for the ligth and no fogging of the fiberscope lens.

The physiological signals, together with the control signal to the torque motor, were recorded on FM tape for subsequent processing. A conventional acoustic recording was made simultaneously on the tape recorder.

For processing, the audio signal was sampled at 10 kHz, while all other signals were sampled at 500 Hz. The movement signals were smoothed with a 40 ms triangular window. After smoothing, the velocity of upper lip movement was obtained using a three-point central difference algorithm; the obtained velocity signal was smoothed once more using the same window.

2.2 Experimental protocol.

Two experimental sessions were conducted, each lasting approximately 3 hours, and 12 blocks of 25 trials were performed per session. Blocks alternated between *discrete* and *repetitive* experimental conditions. In the discrete condition, each trial consisted of the single "word" /pəsæpæpl/ ; in the repetitive condition, each trial consisted of a sequence of approximately 15-20 repetitions of the syllable /pæ/, spoken at a syllable rate comparable to that used in the discrete trials. Perturbations were delivered during a random sampling of 80% of the trials.

To date, only the discrete condition has been fully analyzed (211 perturbed trials and 66 control trials). On each perturbation trial of this condition, the duration of the perturbation was preset to equal the subject's average interval between maximum lower lip lowerings for the first and second /æ/, measured during a set of pretest productions. Pretest measures were also used to parameterize a random timing circuit for controlling perturbation onset. This circuit was triggered by the release burst of the initial /p/, and allowed perturbation offsets to occur at z% of the pretest interval between maximum jaw lowerings for the first and second /æ/ (z varied randomly from 1 to 100). The subject was instructed to not actively resist the perturbation, and to continue speaking as normally as possible.

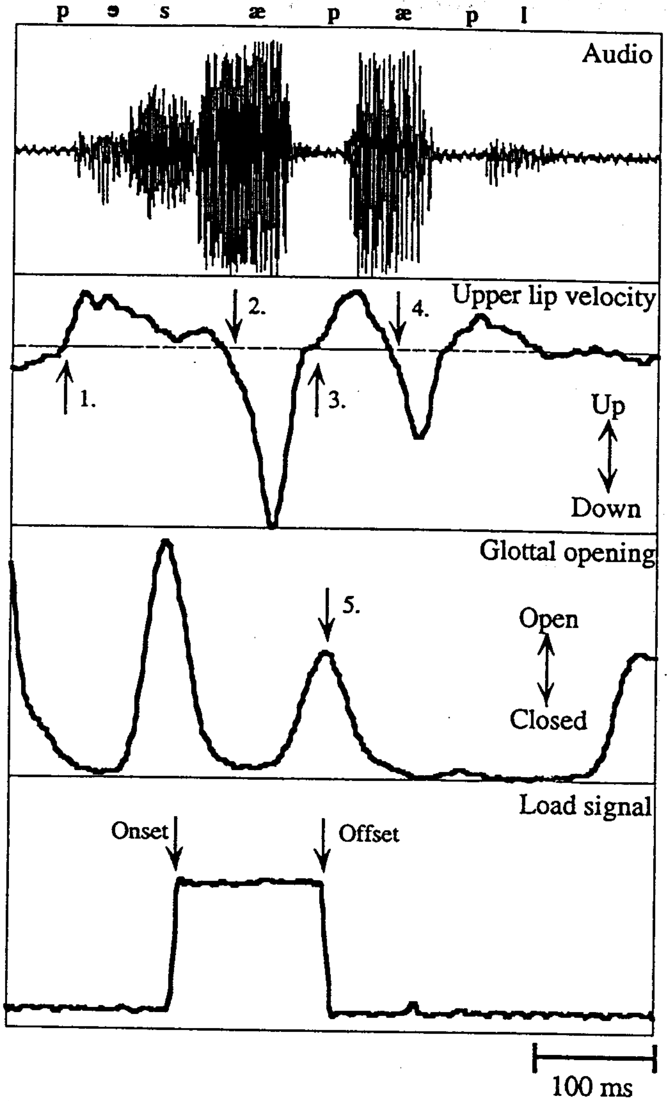

Figure 2. Plot of audio signal, upper lip velocity, glottal opening, and load signal during a representative perturbed trial. The events used for delimiting articulatory intervals are: (1) onset of upper lip raising for the initial /p/; (2) onset of upper lip lowering for the closure of the second /p/; (3) onset of upper lip raising at the release of the second /p/; (4) onset of upper lip lowering for the third /p/; (5) peak glottal opening during the second /p/; and onset and offset of the load signal. These events were used to define the set of intervals that served as the focus of the present analyses: precycle (event 2 minus event 1); vcycle (event 4 minus event 2); vcycle1 (event 3 minus event 2); vcycle2 (event 4 minus event 3); totdur (event 4 minus event 1); glo (event 5 minus event 3); and pertoff (load offset minus event 4).

3. RESULTS

In this report, we describe how different articulatory intervals change as a function of the temporal offset of the lip-opening mechanical load. Oral articulatory intervals were identified in the upper lip velocity signal. The upper lip signal was used since it is mechanically unaffected by the load applied to the lower lip. To obtain a criterion for movement onsets, the maximum peak velocities of upper lip raising and lowering recorded during the experimental session were identified. Onsets were then identified algorithmically as the point at which the upper lip velocity reached 10% of maximum peak velocity. Peak glottal openings were identified using a simple peak-picking procedure. The audio signal, and the articulatory and torque events for a single representative trial are shown in Figure 2.

Data for the single subject across two different experimental sessions are presented in. Figure 3. For the perturbed trials, the intervals were partitioned into four bins and averaged according to the interval *pertoff*, the time of perturbation offset minus the time of onset of upper lip lowering for the third /p/ (see also Figure 2).

To normalize for differences in speaking rate between the two experimental sessions, *pertoff* was divided by the interval *totdur*, the average duration of the interval between events 1 and 4 for the control, nonperturbation trials of the respective session. Changes in the duration of the articulatory intervals were similarly normalized by taking the difference between each loaded trial and the mean control duration of that interval for the respective session, and dividing this difference by *totdur* for the respective session. T-tests were computed for each articulatory interval comparing whether the mean normalized duration change in each of the four perturbation bins differed from zero. To adjust for an elevated Type I error rate due to multiple comparisons, α-levels were selected by dividing .01 and .05 by the number of comparisons made. To facilitate interpretation of these data, Figure 4 displays in schematic form the manner in which the different articulatory events and intervals (Figure 2) are related temporally to the normalized perturbation time bins (Figure 3).

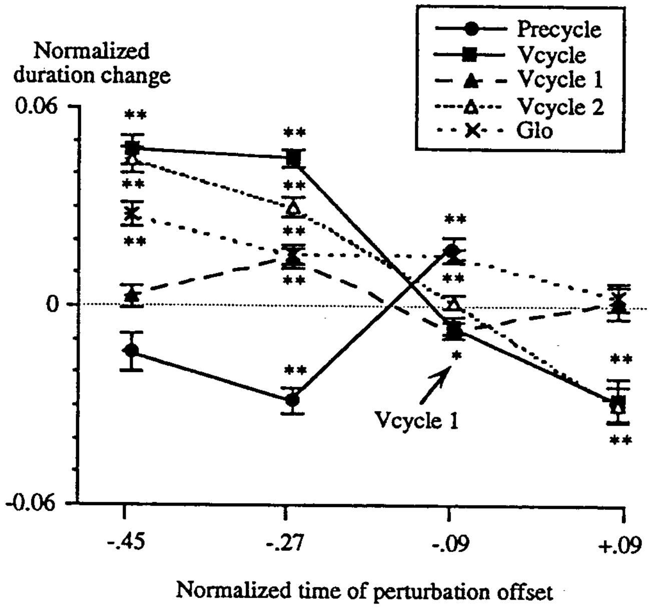

*Figure 3. Normalized duration changes of oral and oral-laryngeal articulatory intervals plotted as functions of binned, normalized perturbation offset time. Perturbation time axis labels indicate the centers of these normalized time bins. There are no precycle data reported for bin ".09", since only three tokens fell into this bin meeting the criterion that perturbation onset occur before the end of precycle. Normalization and binning procedures are described in the text. Starred data points denote significant perturbation-induced changes in interval duration (** indicates $p < .01$; * indicates $p < .05$).*

4. DISCUSSION

How do lip-opening perturbations induce the changes in articulatory intervals displayed in Figure 3? To address this question, we have assumed that these changes are the results of advances and/or delays in the five measured articulatory events (Figures 2 & 4), and that these event shifts are systematically related to the timing of the perturbations in relation to the events. We interpret the interval changes in each time bin in terms of event shifts as follows:

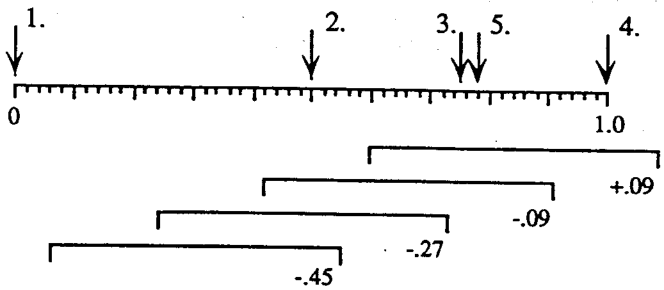

Figure 4. Control values averaged across both sessions for the five articulatory events (see Figure 2) are identified by arrows and plotted on a time scale normalized from zero to one. Time normalization was performed by dividing the average articulatory event time from event 1 by the average value of totdur, the average value of the time from event 1 to event 4. Plotted below and vertically aligned with this time scale are four lines whose lengths represent the average perturbation duration across both sessions, and whose rightmost borders are located at the center of the normalized perturbation time bins shown in Figure 3.

a) bin "-.45". *Vcycle2* and *vcycle* lengthened due to the delay of event 4, and *glo* lengthened due to the delay of event 5;

b) bin "-.27". *Precycle* shortened due to the advance of event 2, *vcycle1* lengthened due to the advance of event 2 and a smaller advance of event 3, *vcycle2* lengthened due to the advance of event 3 and the delay of event 4, *vcycle* lengthened due to both the advance of event 2 and the delay of event 4, and *glo* lengthened due to the advance of event 3 and the delay of event 5;

c) bin "-.09". *Precycle* lengthened due to the delay of event 2, *vcycle1* shortened due to the delay of event 2 and a smaller delay of event 3, *vcycle2* did not change due to delays of both events 3 and 4, *vcycle* did not change due to delays of both events 2 and 4, and *glo* lengthened due to the delay of event 5 and the smaller delay of event 3;

d) bin ".09". *Vcycle2* and *vcycle* shortened due to the advance of event 4.

The pattern of event shifts just described are consistent with the following account of the effect of introducing lip-opening perturbations at various times during the analyzed discrete utterance. We begin by considering the effects of perturbation on the phasing of labial activity. The first point we wish to make is that "actively controlled" gestures resist temporal disruption from perturbations applied to the articulators that participate in those gestures. For example, the bilabial closing gesture defined over the *vcycle1* interval will actively resist temporal disruption from a perturbation applied to the lower lip, since the lower lip is part of the bilabial closure synergy of upper lip, lower lip, and jaw. During such gestures, the phasic onset or offset of a perturbation, or the tonic presence of a perturbation will tend to have relatively small effects on the durations of the gestures. This tendency toward temporal invariance can be seen in the fact that changes of *vcycle1* tend to be smaller and of opposite sign than changes of *precycle* in the earliest three bins ("-.45", "-.27", "-.09"), and is consistent with the hypothesis that event 3 adjusts to "compensatorily" track changes in event 2 but may not be totally successful. If event 3 did not shift temporally, *vcycle1* changes would simply have been equal and opposite to *precycle* changes. We view the tendency toward temporal invariance demonstrated by actively controlled gestures in the present study to be analogous to the tendency toward the spatial invariance shown by such gestures in so-called "remote compensation" studies [7, 8, 9, 10, 11].

Second, perturbations applied during controlled gestures do, however, appear to alter the durations of subsequent "uncontrolled" gestures, that is, gestures whose articulatory synergies do not include the perturbed articulator. In particular, if the onset of perturbation occurs during the controlled gesture, the following uncontrolled gesture will be shortened; if the offset of perturbation occurs during the controlled gesture, the following uncontrolled gesture will be lengthened; and if the perturbation is tonically present throughout the controlled gesture, the following uncontrolled gesture will be shortened. A particularly salient example of this effect can be found in the shortening of *vcycle2* in the last bin ("+.09"), following the onset of the perturbation during *vcycle1*.

Third, perturbations applied during uncontrolled gestures affect the durations of these gestures. Lengthening is induced when the onsets or offsets of perturbations occur during the gestures; no duration changes are induced, however, if the perturbations are tonically present throughout the gestures. Thus, in bin "-.09", where the perturbation onset occurs during the opening gesture into the first /æ/, event 2 is delayed (see Figures 3 & 4).

Finally, we will consider the effect of perturbations on the relative phasing of oral and laryngeal events. From Figure 2, it is evident that the onset of glottal adduction (event 5) occurs during the oral closure for the /p/, as is commonly found in syllable-final voiceless stops [12, 13]. Perturbation data from the present study suggest that the onset of glottal adduction is coordinated temporally with the oral closing (*vcycle1*) and opening (*vcycle2*) gestures for the /p/ closure. Additionally, while the effects of perturbations on the bilabial gestures appear to be relatively direct, the effects on the laryngeal gesture appear to be indirect, i.e., the perturbations advance and/or delay the timing of bilabial events which, in turn, alter the timing of the laryngeal gesture. For example, when the oral opening gesture (*vcycle2*) is lengthened, the interval between peak glottal opening and the onset of the opening gesture (*glo*) is lengthened proportionately. This pattern is evident in the first bin ("-.45"), where the delay of event 4 contributes directly to the lengthening of *vcycle2*; the delay of event 4 also appears to induce a proportional delay of event 5 which, in turn, serves to lengthen *glo*. The changes in *glo* are asymmetric, however. Lengthening is allowed, as just discussed for the first bin, but shortening is not. This absence of shortening may be interpreted as a type of saturation nonlinearity, and is reflected in the last bin where *glo* does not change even though event 4 advances and *vcycle2* shortens considerably.

ACKNOWLEDGMENTS

This research was supported by NIH Grants #DC-00121 and #DC-00865, and NSF Grant #BNS-88-20099.

REFERENCES

[1] J. Folkins, & G. Zimmerman. "Lip and jaw interaction during speech: Responses to perturbation of lower-lip movement prior to bilabial closure," *Journal of the Acoustical Society of America*, 71, 1225-1233, 1982.

[2] V. L. Gracco, & J. H. Abbs. "Sensorimotor characteristics of speech motor sequences," *Experimental Brain Research*, 75, 586-598, 1989.

[3] A. Löfqvist, & V. L. Gracco. "Discrete and continuous modes in speech motor control," *Perilus XIV, Institute of Linguistics, University of Stockholm, Stockholm, Sweden*, 27-34, 1991.

[4] K. G. Munhall, A. Löfqvist, & J. A. S. Kelso. "Laryngeal compensation following sudden oral perturbation," *Journal of the Acoustical Society of America*, 80, Suppl. 1, S109, 1986.

[5] E. Saltzman. "Biomechanical and haptic factors in the temporal patterning of limb and speech activity," *Human Movement Science*, 11, 239-251, 1992.

[6] E. Saltzman, B. Kay, P. Rubin, & J. Kinsella-Shaw, J. "Dynamics of intergestural timing," *Perilus XIV, Institute of Linguistics, University of Stockholm, Stockholm, Sweden*, 47-55, 1991.

[7] J. H. Abbs, & V. L. Gracco. "Sensorimotor actions in the control of multimovement speech gestures," *Trends in Neuroscience*, 6, 393-395, 1983.

[8] J. W. Folkins, and J. H. Abbs. "Lip and jaw motor control during speech: Responses to resistive loading of the jaw," *Journal of Speech and Hearing Research*, 18, 207-220, 1975.

[9] J. A. S. Kelso, B.Tuller, E. Vatikiotis-Bateson, & C. A. Fowler, C. A. "Functionally-specific articulatory cooperation following jaw perturbation during speech: Evidence for coordinative structures," *Journal of Experimental Psychology: Human Perception and Performance*, 10, 812-832, 1984.

[10] K. G. Munhall, & J. A. S. Kelso. "Phase-dependent sensitivity to perturbation reveals the nature of speech coordinative structures," *Journal of the Acoustical Society of America*, 78, Suppl. 1, S38, 1985.

[11] S. Shaiman. "Kinematic and electromyographic respones to perturbation of the jaw," *Journal of the Acoustical Society of America*, 86, 78-88, 1989.

[12] A. Löfqvist. "Interarticulator programming in stop production," *Journal of Phonetics*, 8, 475-490, 1980.

[13] A. Löfqvist, & H. Yoshioka. "Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production," *Speech Communication*, 3, 279-289, 1984.