INTERARTICULATOR PROGRAMMING IN OBSTRUENT PRODUCTION*

Anders Løfqvist+ and Hirohide Yoshioka++

Abstract. Most work on speech motor control has been devoted to the spatial and temporal coordination of articulatory movements for successive units, segments or syllables, in the speech chain. An intrasegmental temporal domain has generally been lacking in speech production models, but such a domain is necessary at least for certain classes of speech sounds, e.g., voiceless obstruents, clicks, ejectives. The present paper examines the nature of laryngeal-oral coordination in voiceless obstruent production in different languages using the combined techniques of electromyography, transillumination and fiberoptic filming of the larynx together with aerodynamic and palatographic records for information on supralaryngeal articulations. The results suggest that laryngeal articulatory movements are organized in one or more continuous opening and closing gestures that are precisely coordinated with supralaryngeal events according to the aerodynamic requirements of speech production.

INTRODUCTION

The problem of speech motor control has usually been seen as one of accommodating and coordinating in space and time the articulatory demands for successive segments in the speech chain, and studies of coarticulation have generally been directed towards this problem (Daniloff & Hammarberg, 1973; Kent & Minifie, 1977). Since the articulatory units have usually been taken to be more or less identical with the units of linguistic analysis, the temporal resolution necessary in most speech production models has been of the order of magnitude of the segment. A segmental approach has been further encouraged by the fact that the feature representation of segments at a systematic phonetic level, with few exceptions, contains no intrasegmental temporal domain, and such feature representations have often been taken as the input to the speech production apparatus. For some classes of speech sounds such as voiceless obstruents, clicks, ejectives, and implosives, it is, however, necessary to posit a temporal domain for articulatory movements within one and the same linguistic and/or articulatory unit (cf. Lisker, 1974).

* To appear in Phonetica.
+ Also Lund University, Lund, Sweden.
++ Also University of Tokyo, Tokyo, Japan.

Acknowledgment: This work was supported by NINCDS Grants NS-13617 and NS-13870, and BRS Grant RR-05596.

[HASKINS LABORATORIES: Status Report on Speech Research SR-65 (1981)]
Voiceless obstruent production requires control and coordination of several articulatory systems. The tongue, the lips and the jaw are engaged in the formation of the constriction or occlusion; the soft palate is elevated in order to close the entrance to the nasal cavity and prevent air from escaping that way; the vocal folds are abducted in order to prevent glottal vibrations and, by reducing laryngeal resistance to air flow, contribute to the high air flow and/or buildup of oral air pressure.

Voiceless obstruent production thus involves simultaneous activity at both laryngeal and supralaryngeal levels, and the oral and laryngeal articulations have to be temporally coordinated. The aim of the present paper is to examine the nature of laryngeal-oral coordination in voiceless obstruent production.

METHOD

Laryngeal articulations were monitored simultaneously by fiberoptic filming and transillumination. Filming was made through a flexible fiberscope at a film speed of 60 frames/second. The film was analyzed frame by frame, and the distance between the vocal processes measured as an index of glottal opening. The light passing through the glottis was also sensed by a phototransistor placed on the neck just below the cricoid cartilage. Recordings from 10-15 repetitions of each test utterance were computer averaged. Unless stated otherwise, the average transillumination signals have been integrated over 5 milliseconds. In order to obtain the speed of glottal opening change, the first derivative of glottal displacement was calculated by successive subtractions at 5 millisecond increments in the average transillumination records. Neither transillumination nor fiberoptic films can be calibrated at present. The scales thus differ between experimental runs, and numerical comparisons of glottal opening and velocity should only be made for a given subject within one and the same recording session.

The movement records were supplemented by EMG recordings from the posterior cricoarytenoid and the interarytenoid muscles, in order to determine if observed laryngeal movements were caused by muscular and/or nonmuscular, e.g., aerodynamic forces.

Implosion and release of voiceless stops were determined from records of oral egressive air flow and oral air pressure. Such records are, however, not reliable indicators of beginning and end of oral constriction in voiceless fricatives. Therefore, additional recordings were made using a custom-made artificial palate with implanted electrodes (cf. Kiritani, Kakita, & Shibata, 1977). Six electrodes at the alveolar ridge were connected in parallel; a battery and a resistor were connected in series between the six electrodes and a reference electrode. Onset and offset of tongue-palate contact could then be identified as changes in voltage across the resistor. A more detailed description of the experimental procedure can be found in Yoshioka, Löfqvist, and Hirose (1979), Löfqvist and Yoshioka (1980), and Löfqvist (in press).

The fiberoptic filming was made to assess the validity of the transillumination technique. Temporal patterns of glottal opening variations obtained by fiberoptic filming and by transillumination showed a high correlation and
proved to be practically identical (Yoshioka et al., 1979; Löfqvist & Yoshioka, 1980, in press). We will therefore mainly discuss the information obtained by transillumination. The electromyographic records will not be dealt with apart from the general observation that laryngeal articulatory movements were accompanied by distinct activity patterns in the two muscles investigated, with the posterior cricoarytenoid activated for abduction and the interarytenoid activated for adduction, respectively.

**RESULTS**

In single voiceless obstruents, the laryngeal articulation usually has the form of a single "ballistic" opening and closing gesture, cf. Figure 1. The timing of this gesture in relation to supraglottal articulatory events is tightly controlled, and apparently varies for fricatives and aspirated stops, Figure 1. Peak glottal opening occurs earlier during the fricative than during the stop. The abduction also appears to occur at higher velocity, and peak opening seems larger for the fricative.

In clusters of voiceless obstruents, one or more continuous glottal opening and closing gestures occur, as shown in Figure 2. In general, separate opening gestures are associated with fricatives and with aspirated stops. In a cluster of fricative + unaspirated stop, only one glottal gesture is found with peak glottal opening during the fricative. When several glottal gestures are found in a cluster, their relationship to oral articulations is similar to that found in single obstruents.

Variations in the relative timing of laryngeal and oral articulations are used to produce contrasts of aspiration in stop consonants. This is illustrated in Figure 3 with material from Icelandic, which has a three-way contrast of preaspirated, unaspirated, and postaspirated voiceless stops. The three stops in Figure 3 differ in at least two respects. First, the relative timing of glottal abduction/adduction and oral closure/release is different. For the unaspirated stop, glottal abduction starts at the implosion, and peak glottal opening, i.e., onset of glottal adduction, occurs close to the implosion. The postaspirated category has glottal abduction beginning at implosion and peak glottal opening at oral release. For the preaspirated stop, both glottal abduction and peak glottal opening precede oral closure.

A second difference in Figure 3 is that of glottal opening size. The present material suggests that postaspirated stops have larger glottal opening than their preaspirated and unaspirated cognates. Glottal opening is smaller for the preaspirated type, and very small for the unaspirated one. For the latter, the fiberoptic films showed a small, spindle-shaped opening in the membranous portion of the glottis.

A closer view of interarticulator timing in Swedish voiceless stop production is given in Figure 4. This figure is based on measurements from repetitions of simple CVCCVC nonsense words where the number of segments and the placement of stress were systematically varied. For aspirated stops, peak glottal opening is systematically delayed in relation to stop implosion as the duration of stop closure increases. Unaspirated stops in Swedish have longer closure duration, and peak glottal opening generally occurs closer to stop implosion for unaspirated than for aspirated stops.
Figure 1. Average transillumination signal (GA), interarytenoid (INT) and posterior cricoarytenoid (PCA) EMG records, and audio envelope (AE) of Swedish utterances containing a voiceless fricative (left) and a voiceless postaspirated stop (right).
Figure 2. Glottal area, EMG and audio signals of two Swedish utterances containing different voiceless obstruent clusters.
Figure 3. Glottal area and audio signals of Icelandic utterances containing unaspirated (top), postaspirated (middle), and preaspirated (bottom) voiceless stops.
Figure 4. The interval from stop implosion to peak glottal opening plotted versus closure duration for Swedish voiceless stops in various positions and under different stress conditions. Each data point represents the mean of 20 tokens. Top and bottom graphs refer to two different speakers. Aspirated stops are denoted by X and unaspirated by O.
Figure 5 presents similar measurements for voiceless fricatives and aspirated stops in American English at two different speaking rates. As in Figures 1 and 3, peak glottal opening for aspirated stops occurs at oral release. In fricative production, peak opening occurs close to the middle of the oral constriction. Within the stops, the interval from implosion to peak glottal opening increases with increasing closure duration as in the Swedish material in Figure 4. A similar relationship is found for the fricatives, although the slope of the function is less steep. A comparison between the two speaking rates shows that the two sets of measurements form more or less a continuous function. These results thus indicate that the ratio between the interval from implosion to peak glottal opening and closure duration tends to remain constant.

A more detailed view of the laryngeal opening and closing gesture in voiceless obstruent production is presented in Figures 6, 7, and 8 for three different speakers and languages. The displacement averages were made with an integration of 15 milliseconds, and the velocity calculated by successive subtractions. All curves are aligned with reference to the offset of the preceding vowel. In the velocity plots, positive values indicate abduction and negative values indicate adduction. The linguistic material consisted of single voiceless stops and fricatives as well as clusters of stops and fricatives. The solid lines in the figures represent single fricatives or clusters beginning with a fricative, whereas the broken lines represent single stops or clusters beginning with a stop, irrespective of the nature of the following segments in the cluster. Japanese does not allow consonant clusters, and the Japanese material contains a devoiced vowel following the initial stop or fricative with a single or geminated stop or fricative occurring after the devoiced vowel.

In the displacement plots we observe again a difference in the timing of peak glottal opening with respect to the offset of the preceding vowel, i.e., peak opening occurs closer to the offset of the vowel when a fricative follows immediately after the vowel. From the velocity plots it is evident that peak abduction velocity is higher in the fricative case. The fricative abduction also has a narrow peak in the velocity plots, whereas the abduction gesture in the stop case is broader. For the Swedish subject in Figure 6, the stop abduction has an initial velocity peak followed by a second peak about 50 milliseconds later.

A striking similarity in the velocity plots for the different speakers is that peak velocity of the abduction gesture tends to occur at a fixed distance from the offset of the preceding vowel. This holds true for all the fricative cases, irrespective of variations in speed, size, duration, and timing of the glottal gesture. For the Japanese material in Figure 7, peak velocity of the stop abduction coincides in time with that for the fricatives. In the Icelandic case, Figure 8, peak abduction velocity occurs at two different times for fricatives and stops, respectively, but within the two families of curves, peak velocity tends to occur at the same time.
Figure 5. The interval from onset of tongue-palate contact to peak glottal opening plotted versus duration of oral closure or constriction for American English stops and fricatives in various positions and under different stress conditions at two speaking rates.
Figure 6. Plots of size and speed of the glottal abduction/adduction gesture for Swedish voiceless obstruents. Zero on x-axis represents offset of the vowel preceding the obstruents. Abduction velocity is shown with positive sign, adduction velocity with negative sign. See text for further explanation.
Figure 7. Plots of size and speed of the glottal abduction/adduction gesture for Japanese voiceless obstruents. Symbols as in Figure 6.
Figure 8. Plots of size and speed of the glottal abduction/adduction gesture for Icelandic voiceless obstruents. Symbols as in Figure 6.
DISCUSSION

The present results, as well as those of other studies reviewed in Løfqvist (in press), suggest that the glottis is continuously changing in voiceless obstruent production. Laryngeal articulations are thus organized in one or more opening and closing gestures. Static open glottal configurations rarely seem to occur in speech, and also appear difficult to maintain in some nonspeech conditions (cf. Løfqvist, Baer, & Yoshioka, 1980).

The laryngeal gestures are tightly coordinated with supralaryngeal events to meet the aerodynamic requirements for producing a signal with a specified acoustic structure. Variations in the relative timing of the laryngeal opening and closing gesture and the oral closing and opening gesture are used to produce contrasts of voicing and aspiration (cf. Abramson, 1977).

Initiation of glottal abduction before oral closure in voiceless stops produces preaspiration as shown in Figure 3. If glottal abduction starts after oral closure, prevoicing results, and if the abduction gesture starts at stop release, a voiced (or murmured) aspirated stop is produced. Similarly, a glottal gesture beginning at stop implosion and with peak glottal opening close to the implosion is used for producing voiceless unaspirated stops, whereas a gesture starting at implosion and with peak opening at stop release results in a postaspirated stop. These different obstruent categories are thus basically produced by differences in interarticulator timing.

Differences in the size of the laryngeal gesture seem to co-occur with the timing differences. Variations in size and timing of the laryngeal gesture are best regarded as interacting strategies for achieving a specific acoustic output. An early timing of peak glottal opening together with a small opening can thus be used in producing unaspirated voiceless stops, since they will both contribute to a glottal configuration suitable for voicing at stop release, cf. Figure 3. A comparatively small glottal opening for preaspirated stops could be related to the production of glottal frication noise during the period of preaspiration. Similarly, the size of the glottal gesture for a voiced (or murmured) aspirated stop would be adjusted to produce both glottal vibrations and frication noise. A large glottal opening at the release of voiceless postaspirated stops would not only contribute to the delay in voice onset but also create suitable aerodynamic conditions for noise generation at the oral place of articulation as the articulators are being separated immediately after the release.

The differences in glottal displacement and velocity between stops and fricatives in Figures 6, 7, and 8 are also most likely related to different aerodynamic requirements for stop and fricative production. A rapid increase in glottal area would allow for the high air flow necessary to generate the turbulent noise source during voiceless fricatives (Stevens, 1971). In stops, a slower increase in glottal opening together with the concomitant oral closure could be sufficient to stop glottal vibrations in combination with the buildup of oral air pressure (cf. Yoshioka, 1979). In the Icelandic material in Figure 8, glottal abduction starts considerably later relative to offset of the preceding vowel for stops than for fricatives. Although it is tempting to view this difference as a deliberate action by the speaker to avoid unwanted preaspiration, it is best regarded as a speaker-specific variation, since we
have found similar differences between stops and fricatives for speakers of American English, where preaspiration does not occur. The present results thus indicate that differences exist between stops and fricatives in the initial glottal abduction phase. The magnitude and form of these differences may, however, show some interspeaker variability.

The acoustic consequences of variations in interarticulator timing in obstruent production are complex and spread out over a period of time, involving differences in the sound source and the spectral composition of the signal.

The preaspirated stop in Figure 3 is thus associated with the following sequence of source changes: periodic voicing during the preceding vowel, aperiodic noise, silence, transient noise, periodic voicing during the following vowel. For the postaspirated stop in the same figure, the sequence would be voicing, silence, transient, noise, voicing. At the same time the spectral qualities of the signal would differ according to the nature of the preceding and following vowels and the place of articulation of the obstruent. This complex of acoustic cues, produced by a unified articulatory act, is integrated in speech perception to form a single percept (cf. Liberman & Studdert-Kennedy, 1978; Repp, Liberman, Eccardt, & Pesetsky, 1978).

As interarticulator timing appears to be an essential feature of voiceless obstruent production, one may question the descriptive adequacy and usefulness of feature systems with timeless representations for modeling speech production, whatever their merits may be for abstract phonological analysis. Specifying glottal states along dimensions of spread/constricted glottis and stiff/slack vocal cords (Halle & Stevens, 1971) would thus not only seem to be at variance with the phonetic facts but also to introduce unnecessary complications. The difference between postaspirated and unaspirated voiceless stops is rather one of interarticulator timing than of spread versus constricted glottis. Similarly, the difference between voiceless and voiced postaspirated stops is also one of timing rather than of stiff versus slack vocal cords. Preaspirated stops are naturally accounted for within a timing framework but cannot be readily differentiated from postaspirated ones in a timeless feature representation. It is, of course, possible to translate a timeless representation into differences in interarticulator timing, but if timing is of importance, it seems counterintuitive to derive it rather than represent it directly, especially if feature representations are to have a phonetic basis and describe parameters that the speaker can control independently. The importance of interarticulator timing in obstruent production is not a new idea, e.g., Rothenberg (1968), Lisker and Abramson (1971), Ladefoged (1973), Abramson (1977). It has, moreover, been noted by phonologists who, for reasons not entirely clear, still favor timeless phonological descriptions (e.g., Anderson, 1974).

The tight temporal coordination of laryngeal and oral articulations in voiceless obstruent production exemplified in the present material constitutes an important problem for any theory of speech production.

Models of speech production based on feature spreading (Daniloff & Hammarberg, 1973; Hammarberg, 1976; Bladon, 1979; see also Fowler, 1980) would seem incapable of handling this kind of interarticulator programming, at least
in their current form. One reason for this is that their temporal resolution is limited to quanta of phone or syllable size, whereas laryngeal-oral coordination in obstruents requires a finer grain of analysis. An additional limitation is that such models do not specifically address the general problem of interarticulator coordination in space and time. These limitations of current feature spreading models stem partly from the fact that they take as input the timeless units of abstract phonological theory.

Given the dynamic character of speech production and the need to coordinate different articulators in space and time, it seems rational to view speech production as an instance of control of coordinated movements in general. A powerful theory of motor control has been proposed by Bernstein (1967), and elaborated by Greene (1971, 1972; see also Boylls, 1975; Turvey, 1977; Kugler, Kelso, & Turvey, 1980; Kelso, Holt, Kugler, & Turvey, 1980; Fowler, Rubin, Remez, & Turvey, 1980). Designed to cope with the number of degrees of freedom to be directly controlled, this theory views motor coordination in terms of constraints between muscles, or groups of muscles that have been set up for the execution of specified movements. Areas of motor control where this theory has proved to be productive include locomotion (Grillner, 1975), posture control (Nashner, 1977), and hand coordination (Kelso, Southard, & Goodman, 1979). One merit of this view is that it predicts and rationalizes tight temporal relationships between articulators. In particular, it predicts that some such relationships should remain invariant across changes in stress and speaking rate, and material on oral articulations presented by Tuller and Harris (1980) is in agreement with this prediction. Some aspects of the present results can be rationalized within this theoretical framework.

Peak velocity of the glottal abduction gesture was found to occur almost at the same point in time relative to the offset of the preceding vowel, irrespective of variations in speed, size, duration, and timing of the gesture.

Another aspect is the relationship between laryngeal and oral articulations presented in Figures 4 and 5. Here, peak glottal opening was found to be delayed in relation to the formation of the oral constriction or occlusion as the latter increased. For aspirated stop consonants, this results in a constant temporal relation between peak glottal opening and oral release, ensuring an open glottis at the release to produce aspiration. The ratio between the interval from implosion to peak glottal opening and closure/constriction duration tends to remain constant across changes in overall obstruent duration.

We can regard such constant relationships as structural prescriptions for the articulators, specifying relations that have to be maintained in obstruent production across changes in stress and speaking rate. On the other hand, a metrical prescription specifies the activity levels of articulatory muscles. As suggested by Boylls (1975), the metrical prescription can be regarded as a scalar quantity multiplying the activities of the oral and laryngeal muscles in obstruent production while preserving the structural prescription.
REFERENCES


Boylls, C. C. A theory of cerebellar function with applications to locomotion. II. The relation of anterior lobe climbing fiber function to locomotor behavior in the cat. COINS Technical Report 76-1, (Department of Computer and Information Science, University of Massachusetts), 1975.


Grillner, S. Locomotion in vertebrates: Central mechanisms and reflex interaction. Physiological Reviews, 1975, 55, 247-304.


