STOP CONSONANT PRODUCTION: AN ARTICULATION AND ACOUSTIC STUDY

Kelly L. Poort
Massachusetts Institute of Technology, Cambridge, MA USA

ABSTRACT

Articulation and acoustic data for stop consonant production were examined with the aims of (1) describing the movements and coordination of the articulatory structures and (2) developing procedures for interpreting acoustic data in terms of articulatory movements. Findings were placed in the context of existing acoustic and aerodynamic models.

INTRODUCTION

Articulatory information for the stop consonants, derived from recordings of the physical movements of the articulators, was combined with simultaneous acoustic recordings. The voiceless stop consonants were chosen for detailed study because there is less acoustic information present immediately following the release than for voiced stops. In particular, the labial and alveolar voiceless stops were chosen for in-depth investigation. This paper focuses on the production of the labial voiceless stop consonant /p/. The production of the alveolar stop /t/ will be discussed in the presentation in August. The coupling of articulatory and acoustic information provides an improved understanding of stop-consonant production, leading to knowledge of the sequencing and timing of articulator movements as well as the resultant acoustics. The primary objective of the study is to refine existing acoustic and aerodynamic models to reflect the new level of understanding. In addition, examination of the articulatory information leads to improved interpretation of the acoustic signal, such that in the future the acoustic waveform will be the only information required to determine many of the important aspects of the vocal-tract movements for stops. The results of the investigation are applicable to the areas of speech recognition, speech synthesis, and the study and remediation of disordered speech production.

PROCEDURE

Movements in the midsagittal plane of points on the lower jaw, lips, tongue blade and tongue body were measured using an electromagnetic midsagittal articulometer [1]. Acoustic data were recorded simultaneously. Three normal speaking, normal hearing male subjects spoke single-syllable words /CVt/, composed of one of the voiceless stop consonants /p, t/ followed by one of the vowels /a, i/ and the consonant /t/, imbedded in a carrier phrase. Half the tokens were preceded by the fricative consonant /s/. One example is, "Say spot again." A minimum of three repetitions of each utterance were recorded per speaker.

Events in time, such as the end of the vowel in say, the end of /s/ in the token (if present), the stop release, and the onset of the vowel following the stop, were identified in the acoustic waveform based upon researcher's judgment. The corresponding times were located in the time-aligned articulatory displacement waveforms. In order to preserve the times and magnitudes of each of these event times in the articulation data, as well as the general shape of the displacement waveform between event times, the standard technique of linear time warping was adapted and applied. The repetitions of each utterance for a given speaker were averaged together using the adapted technique. For example, the eight repetitions of "Say spot again." recorded by one of the speakers were averaged together using the modified linear time warping technique to become one, representative displacement waveform for that speaker.

An estimate of the constriction cross-sectional area change with time following the release of the stop consonant was calculated with the aid of the articulation data. The upper and lower lip transducers are located on the vermilion borders of the lips. The movements of these two transducers, averaged as described above, were used to determine rate of vertical lip separation following release. The difference between the upper and lower lip trajectories was used as an estimate of the rate of lip separation, with the difference at the time of release zeroed. The rate of horizontal lip separation was taken from a study by Fujimura [2]. An estimate of constriction cross-sectional area change with time following stop release was obtained by approximating the lip opening cross-sectional area as a rectangle whose height and width change with time according to the vertical and horizontal lip separations, respectively. For example, the cross-sectional area of the lips in the labial /p/, derived in this manner for one subject speaking the token spot, increases by approximately 35 cm^2/sec within the first 5 - 10 msec following the stop release.

ACOUSTIC ANALYSIS

During the first few milliseconds following the stop consonant release, the constriction cross-sectional area change with time can be related to the corresponding acoustics in two important ways: (1) by modeling the vocal tract as a series of tubes of varying cross-sectional area (in the case of a labial stop, a Helmholtz resonator) and calculating the transition of the first formant frequency F1; and (2) by calculating the time course of the burst when the constriction cross-sectional area is used as a parameter in a circuit model of the vocal tract [3].

Fujimura, in a stroboscopic motion picture study [2], observed a three-stage transition in the first formant frequency F1 following labial stop release into a vowel. The first stage occurs during the 5 - 10 msec immediately after the lips begin to open. It consists of an abrupt increase in lip opening cross-sectional area corresponding to a rapid upward shift in F1. Within just the first 5 msec, F1 will shift from 0 Hz (assuming the vocal tract walls are rigid) to a value of 200 - 400 Hz, depending upon the following vowel. The rapid rise of F1 is modeled as a Helmholtz resonator. The short front tube represents the lips during the production of the bilabial stop consonant. The second stage consists of a slower rise in F1, corresponding to a continued increase in lip opening cross-sectional area in conjunction with a downward movement of the jaw. The third stage is the movement of F1 during the initial portion of the following vowel.

The present study has also investigated the F1 transition following the stop release. The constriction cross-sectional area change with time following release, as calculated from the articulation data, is utilized as the rate at which the cross-sectional area of the short front tube of the Helmholtz resonator increases with time. A correction term of 180 Hz has been incorporated into the model to reflect the impedance of the vocal tract walls. From the solution of the Helmholtz resonator, F1 transitions from 180 to 400 Hz during the first 5 - 6 msec following labial stop release in the utterance spot. This study finds that the rate of constriction cross-sectional area change with time immediately after lip opening is less than the 100 cm^2/sec determined by Fujimura. Consequently, F1 does not initially transition upward as rapidly as predicted by Fujimura [2].

An aerodynamic circuit model of the pressures and flows in the vocal tract [3] was employed to calculate the airflow through the lips for the first few milliseconds following the stop release. The derived constriction cross-sectional area was utilized in the determination of one of the parameters in the circuit model, specifically the rate of decrease in constriction resistance following the stop release. From the constriction cross-sectional area of the lip opening and the airflow through the lips immediately following release, the amplitude and duration of the frication noise source can be calculated [4, 5, 6, 7 and others]. The burst contains a peak in amplitude approximately 5 - 6 msec after the release. The duration of the noise burst up to a time where the amplitude is 10 dB down from the burst peak in spot is approximately 8 - 10 msec. Both the shape and location of the peak in the noise burst as well as the duration of the burst derived from the model in this fashion agree well with the shape and duration of the noise burst in the corresponding acoustic waveform for spot.

ARTICULATION ANALYSIS

In addition to the constriction cross-sectional area calculation discussed earlier, the articulation data were examined to determine effects of phonetic context on production. Findings include a constraint on lower jaw position during /s/ production, maximum downward velocity for the lower jaw occurring approximately at the time of the consonant release, and a correlation between increasing distance articulators must travel and faster rates of movement. In addition, the articulators not involved in forming the constriction were found to anticipate the positions required for the upcoming vowel much more so than the constriction-forming articulator(s). For example, the lips are constrained to form the constriction for /p/ in spot; however, the jaw and tongue can and do move to some extent into position for the following vowel during the production of /p/. In a similar fashion, the restriction on jaw movement during the production of /s/ forces the jaw to remain in a high position throughout the duration of /s/. As a result, there is a more rapid downward velocity of the jaw at the time of the /p/ release in spot than in pot. This finding is thought to be a compensating mechanism. In order to reach the low jaw position required for the production of the vowel /ɑ/ in spot in approximately the same amount of time as it takes to reach the /ɑ/ in pot, the jaw increases its downward velocity. Relating the finding to an earlier observation, the jaw must move downward faster for /p/ in spot because it travels farther.

CONCLUSION

One of the primary results of the study is the ability to obtain an estimate from the articulation data of the constriction cross-sectional area change with time for the first few milliseconds following the stop consonant release. For example, this rate for /p/ in spot is 35 cm^2/sec. An acoustic model representing the vocal tract as a Helmholtz resonator yields a rapid transition for F1 following the stop release of 35 - 45 Hz/msec for the same utterance. An aerodynamic model coupled with calculations of the frication noise burst reveal an agreement between calculated (utilizing the articulation data) and acoustic waveform noise burst shapes and durations. The shape is found to contain a peak approximately 5 - 6 msec after release and the duration is approximately 8 - 10 msec for /p/ in spot. This agreement suggests that the original estimate of the constriction cross-sectional area change with time of 35 cm^2/sec for /p/ in spot during the first few milliseconds following the stop consonant release is reasonable.

Inferences, such as those described above for /p/ in spot, made from detailed examination of the articulation and acoustic data will aid in developing a more comprehensive model of stop consonant production. From comparison of model outputs and acoustic data, refinements can be made to the existing aerodynamic and acoustic models to more accurately represent the acoustic signal. The quantitative variations in production resulting from various phonetic contexts can be incorporated into the models in order to broaden their applicability to essentially all stop consonant production. The findings of the study contribute to the goal of a single comprehensive model which incorporates all the acoustic, aerodynamic, and articulatory observations in order to explain the resultant acoustic output.

ACKNOWLEDGEMENT

My deepest appreciation to Professor Ken Stevens for his guidance and support. I also thank Joe Perkell, Melanie Matthies, and Mario Svirsky for the use of the electromagnetic midsagittal articulometer system and their technical assistance with my project. This research has been supported in part by grant DC00075 from the National Institutes of Health.

REFERENCES

[1] Perkell, J., et al. (1992), "Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements," *J. Acoust. Soc. Am.* 92(6), pp. 3078-3096.

[2] Fujimura (1961), "Bilabial Stop and Nasal Consonants: A Motion Picture Study and its Acoustical Implications," *J. Speech and Hear. Res.* 4(3), pp. 233-247.

[3] Stevens, K. (1993) "Models for the production and acoustics of stop consonants," *Speech Comm.* 13, pp. 367-375.

[4] Fant, G. (1960), *Acoustic theory of speech production*, The Hague: Mouton & Co.

[5] Stevens, K. (1971) "Airflow and turbulence noise for fricative and stop consonants," *J. Acoust. Soc. Am.* 50, pp. 1180-1192.

[6] Shadle, C. (1985) *The acoustics of fricative consonants*, RLE Technical Report 506, Massachusetts Institute of Technology, Cambridge, MA.

[7] Pastel, L. (1987) *Turbulent noise sources in vocal tract models*, SM Thesis, Massachusetts Institute of Technology, Cambridge, MA.