MODELS FOR PRODUCTION AND ACOUSTICS OF STOP CONSONANTS

K. N. Stevens

Research Laboratory of Electronics and
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology, Cambridge, MA 02139

ABSTRACT - Stop consonants are produced by forming a closure in the vocal tract, building up pressure in the mouth behind this closure, and releasing the closure. Models of the mechanical, aerodynamic, and acoustic events in the vicinity of the stop consonant are described, and examples of calculations of the airflow and of various components of the radiated sound are given.

INTRODUCTION

A stop consonant that occurs between two vowels is produced by creating a complete closure in the vocal tract and then releasing that closure into the following vowel. The closure is formed by a particular articulator — the lips, the tongue blade, or the tongue body. Examples of the midsagittal configuration of the vocal tract just prior to the consonant release for stop consonants produced with each of these articulators are displayed in Fig. 1.

Figure 1. Midsagittal sections through the vocal tract during the production of stop consonants produced by closing the lips (left), raising the tongue tip to the alveolar ridge (middle), and raising the tongue body (right). (From Perkell, 1969).

During the interval of closure, we assume that no air escapes through the constriction, and pressure builds up in the vocal tract behind the closure. When the closure is released, there is a rapid flow of air through the vocal tract, and the intraoral pressure decreases rather abruptly. Within the closure interval, as well as immediately preceding and following this interval, the configuration of the glottis and the stiffness of the vocal folds can be adjusted in various ways either to permit continued vocal-fold vibration or to inhibit vibration.

The time from the initiation of movement toward closure of the vocal tract in the first vowel to completion of the movement from closure into the second vowel can be as long as 200 ms. Consequently acoustic information about the various features that underlie the production of a stop consonant in running speech can be distributed over this time interval. Acoustic analysis procedures that aim to identify these features must therefore examine attributes of the speech signal that extend over 200 ms or more.

The purpose of this paper is to review theoretical models that have been developed to account for the various acoustic, mechanical, and aerodynamic events that occur during the process of producing stop consonants with different articulators. One motivation for this exercise is to provide a more solid basis for machine synthesis of these consonants. Another is to gain insight into the acoustic properties resulting from the articulatory movements that generate stop consonants so as to develop a rational basis for the design of signal processing procedures that can lead to identification of the features of the consonants. Still another motivation for developing a model of consonant production is to provide a starting point for understanding impaired production of consonants.

LOW-FREQUENCY MODEL

We begin by developing a model that examines vocal-tract movements, airflows, and pressures that occur during stop-consonant production. That is, we consider the low-frequency behavior of the vocal tract. In order to estimate the flows and pressures, we model the vocal-tract during consonant production as a tube with two constrictions — one at the glottis and one formed by an articulator within the vocal tract as shown at the top of Fig. 2. The subglottal pressure P_s is assumed to be constant. The glottal constriction has an average cross-sectional area A_g and the area of the supraglottal constriction is A_c.

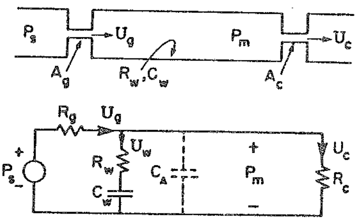

Figure 2. Top: Model for estimating airflows and pressures for consonants. Bottom: Equivalent circuit for model. The acoustic compliance C_A of the air volume is neglected in calculating low-frequency airflow.

The walls of the vocal tract are not rigid, and they move in response to a change in pressure within the vocal tract. At low frequencies, the impedance of the walls can be approximated by a resistance R_w in series with a compliance C_w. A complete circuit diagram of the system is given at the bottom of Fig. 2. The principal driving source for the flow is the subglottal pressure P_s. Elements are included to represent the resistances of the constrictions and the wall impedance. The acoustic compliance C_A of the vocal-tract volume is small compared with C_w, and can be neglected here. The pressure in the mouth is represented by P_m. The resistances of the glottal constriction R_g and the supraglottal constriction R_c are nonlinear, and, over much of the range of constriction sizes, can be approximated by a dynamic resistance that is proportional to volume velocity.

Both the average area of the glottal constriction and the area of the supraglottal constriction change with time in the vicinity of the time when the supraglottal constriction is formed or is released. We will show here the behavior of the model when there is no active adjustment of the glottal configuration during the production of the consonant. When pressure builds up above the glottis during closure for a consonant, outward forces are exerted on the vocal-fold surfaces, causing a passive increase in the glottal area. The rate of change of cross-sectional area of the supraglottal constriction near the times of consonantal closure and release can be estimated from cineradiographic, photographic, and airflow data, although further data of this kind are needed. At the consonantal release, initial rates of change of area are taken to be 100 cm^3/s for labial and alveolar consonants, and 25 cm^3/s for velar consonants.

Calculations of the airflow immediately following the consonantal release for labials and alveolars (left panel) and for velars (right panel) are shown in Fig. 3. The airflow U_c at the constriction, which rises rapidly at the release, has two components — a component U_w due to

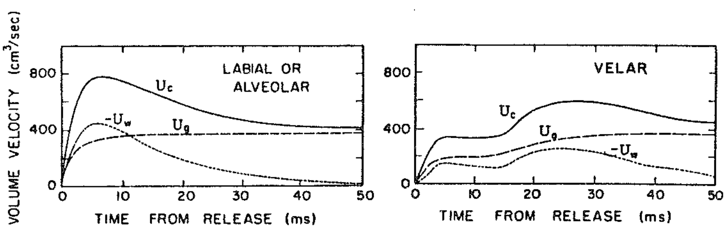

Figure 3. Calculated airflows as a function of time following the release of voiceless unaspirated stop consonants. The constriction area follows a trajectory of the form $A_c = A_{max}(1-e^{-t/\tau})$ for labial and alveolar releases (solid lines), where A_{max} is taken to be 1.0 cm^2, and $1/\tau$ = 100 sec^{-1} for labials and alveolar releases and 25 sec^{-1} for velars. The glottal area is assumed to be constant at 0.1 cm^2. Initial intraoral pressure is 8 cm H_2O. Negative U_w indicates that the walls of the vocal tract displace inwards following the release. Note that $U_g = U_c + U_w$.

the rapid inward movement of the walls after the pressure is released, and the glottal flow U_g. It is noted that the component U_w is a significant part of the total flow during the important time interval a few milliseconds following the release when, as will be discussed below, a burst of turbulence noise is being generated. The rise in flow and the drop in pressure (not shown in the figure) is slower for the velar release than for the labial and alveolar. If we assume that the vocal folds are vibrating in a modal fashion prior to the consonant closure, then the increased intraoral pressure immediately following closure will cause a decreased transglottal pressure and hence a decrease in the amplitude of the glottal pulses. Estimates of the shapes of the glottal flow pulses near the consonant closure indicate that two or three glottal pulses of reduced amplitude can occur before vibration ceases, for this condition in which there is no active adjustment of glottal configuration or vocal-fold stiffness. With different laryngeal and pharyngeal adjustments, glottal vibration may be either extended for a larger time or inhibited more abruptly. At consonantal release, glottal vibration resumes and the pulses achieve full amplitude after 20-odd milliseconds, based on estimates from the model. Of course, if the glottis is adjusted to be in an abducted configuration near the time of release, there will be an interval of aspiration before glottal vibration begins.

ACOUSTIC EVENTS AT RELEASE

We turn now to a model of the sound generation following the release of a stop consonant. The airflow through the constriction and through the glottis in the few tens of milliseconds following the release gives rise to a sequence of four types of sources, as schematized in Fig. 4. Immediately following the release, there is a brief transient as the air that has been compressed in the vocal tract discharges through the opening constriction. In terms of the equivalent circuit of Fig. 2, this initial transient is the flow from the acoustic compliance C_A. The amplitude and spectrum of this volume-velocity transient is determined by the intraoral pressure prior to release and by the rate of increase of the area of the constriction.

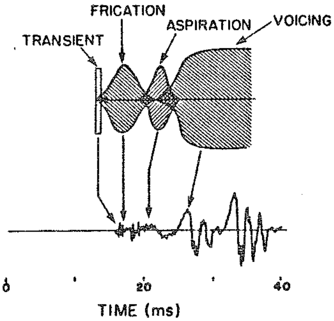

Figure 4. Schematic representation of sequence of events at the release of a voiceless unaspirated stop consonant.

Following this initial transient (with a duration less than 1 ms) the rapid airflow through the constriction (shown in Fig. 3) gives rise to turbulence and hence to a sound source immediately downstream from the constriction. This source is identified in Fig. 4 as frication noise. The amplitude, spectrum, and space distribution of this sound source can be estimated approximately, based on the work of Fant (1960), Stevens (1971), Shadle (1985), Pastel (1987), and others. The turbulence noise source is usually represented as a sound-pressure source near an obstacle downstream from the constriction. In some situations there may also be fluctuations in the flow through the constriction, giving rise to a volume-velocity source. The amplitude of the source is roughly proportional to $U_c{}^3 A_c{}^{-2.5}$, based on empirical data, but with some theoretical support. Based on the flow calculations in Fig. 3, the amplitude of the frication noise source can be calculated. The results of this calculation for the labial or alveolar release are shown as the solid line in Fig. 5. This noise source rises to a peak within 1-2 ms following the release, and then decreases as the area of the constriction becomes larger. The spectrum of the source is based on the work of Shadle (1985).

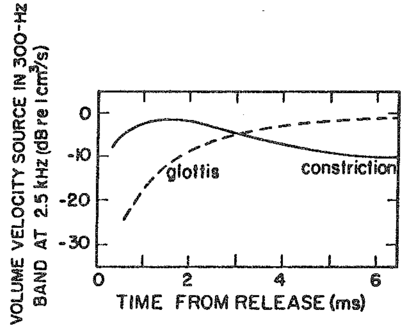

Figure 5. Calculation of relative levels of turbulence noise sources at the constriction (solid line) and at the glottal opening (dashed line) during release of a voiceless unaspirated labial stop consonant. Levels are calculated from trajectories of airflow and area given in Fig. 3. An obstacle is assumed to be positioned directly in the airstream, downstream from the constriction. Corrections must be applied to these curves to account for the less efficient generation of noise at the labial and glottal constrictions.

Rapid airflow through the glottis also gives rise to turbulence noise, labeled as aspiration noise in Fig. 4, with source characteristics similar to those for frication noise. The calculated amplitude of this source for a labial or alveolar release is given by the dashed line in Fig. 5. Vocal-fold vibration begins simultaneously with or immediately following the aspiration noise, as indicated in Fig. 4.

Each of the sources in Fig. 4 is filtered by the vocal tract, and the spectrum of the radiated sound is the product of the source spectrum, the vocal-tract transfer function, and the radiation characteristic. For the transient and frication sources, the transfer function is dominated by the cavity in front of the constriction. There is no cavity in the case of a labial release, and the cavity length can be in the range 1.5-7 cm for alveolar and velar consonants. The output for the transient and frication sources has no dominant spectral peak for labials, whereas there is a spectral prominence in the range 1000-6000 Hz in the case of alveolars and velars. Examples of calculated spectra of the transient and frication components of the radiated sound at the release of a voiceless unaspirated alveolar stop consonant are given by the spectra labelled *T* and *F* in Fig. 6. Both spectra have a peak at about 4500 Hz, representing the resonance of a cavity of length about 2 cm, anterior to the consonant constriction. The spectra at low frequencies are different because of the differences in source types and spectra.

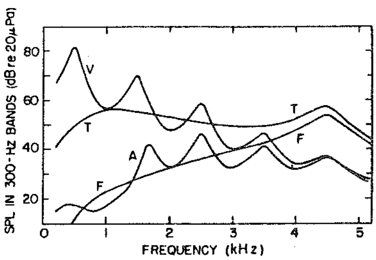

Figure 6. Calculated spectra of several components in the sequence of acoustic events following the release of a voiceless unaspirated alveolar stop consonant. The ordinate is the sound-pressure level in 300-Hz bands at a distance of 20 cm from the mouth opening. The curves are labeled with letters indicating the different components as schematized in Fig. 4. *T*: initial transient at consonant release; *F*: sound due to frication noise generated at or near the labial constriction; *A*: sound due to aspiration noise generated near the glottis; *V*: vowel spectrum after onset of vocal-fold vibration.

When the source reverts to the laryngeal region, all of the vocal-tract resonances are excited to various degrees depending on the source spectrum and the resonator losses. During the time interval when the source is near the glottis, there are transitions in the formants reflecting the changing vocal-tract shape as the consonant articulator moves away from the constricted position and the vocal tract assumes the configuration for the following vowel. Immediately following the release, the formant frequencies are the natural frequencies of the constricted vocal tract (as in Fig. 1). The first-formant frequency for these configurations is always low (about 200 Hz when there is a complete closure). The second- and third-formants are at frequencies (*F2* and *F3*) that depend on the position of the constriction in the vocal tract and also on the configuration of the tongue body and lips that is assumed in anticipation of the following vowel.

Measurements of *F2* and *F3* immediately following the release of labial, alveolar, and velar stop consonants before different stressed vowels have been reported by several investigators (Kewley-Port, 1982; Sussman, 1991; Stevens, House, and Paul, 1966). These results quantify the well-known dependence of the formant transitions on the place of articulation of the consonant and on the formant frequencies for the following vowel.

In Fig. 6, spectrum *A* is the calculated spectrum of aspiration noise just following the release of an alveolar consonant. The glottis is assumed to be slightly abducted, and the turbulence noise source is assumed to be distributed over a region 1-3 cm downstream from the glottis. The prominences corresponding to the second and higher formants are evident in the spectrum, and are at frequencies appropriate for the alveolar place of articulation. The weak low-frequency amplitude for this spectrum is in part a consequence of weak coupling between this source and the lowest two or three modes of the vocal tract. Spectrum *V* in Fig. 6 is the calculated spectrum for a hypothetical vowel with the lowest three formant frequencies at 500, 1500, and 2500 Hz. This spectrum has the downward sloping shape that is expected for vowels.

The series of spectra for voiceless unaspirated alveolar consonants in Fig. 6 show that the spectra of the transient and the frication components at high frequencies exceed the spectrum of the following vowel in this frequency range by 10-20 dB. Similar series of calculated spectra for labial and velar consonants show different patterns. The transient and frication spectra for labials have no spectral prominences, and are somewhat weaker than the spectrum at the onset of the following vowel at all frequencies. For velars, there is a prominence in the transient and frication spectra in the *F2* or *F3* range, depending on whether a front or a back vowel follows the consonant. The calculated spectrum amplitude of this prominence is within a few dB of the spectrum amplitude of the corresponding spectral peak in the following vowel. The rate of increase of the constriction area for a velar consonant is considerably less than that for a labial or alveolar consonants. Coupled with the longer constriction for the velar, this fact leads to slower transitions of the formants (particularly $F1$) into the following vowel.

DISCUSSION

Acoustic spectra sampled during the time interval following the release of different stop consonants are in general agreement with spectra like those in Fig. 6 that are based on calculations from models of stop-consonant production. However, the transient and frication spectra (such as *T* and *F* in Fig. 6) are often sufficiently close together that it is difficult to measure them separately. The calculations do show, however, that, at least for alveolar and labial consonants, a true indication of the ``burst'' spectrum, uncontaminated by resonances of the vocal tract posterior to the constriction, can be obtained only if a time window of a few ms, centered on the release, is used to estimate the spectrum.

It should be noted that some variability is to be expected in the various spectra of the type shown in Fig. 6. Depending on the rate of release of the closure, the subglottal pressure, and vocal-tract configuration for the following vowel, the amplitudes and spectra of the different components can be different. In all cases, however, acoustic events in the 50-odd ms following the release provide evidence for (1) the lowest natural frequency of the acoustic cavity anterior to the constriction (or the nonexistence of such a cavity), (2) the configuration of the vocal tract posterior to the constriction, as signalled by the formant transitions; and (3) the rate of change of the constriction area. These three types of acoustic data provide evidence indicating whether the consonant was produced by the lips, the tongue blade, or the tongue body.

ACKNOWLEDGEMENT

This research was supported in part by grant DC00075 from the National Institutes of Health.

REFERENCES

Fant, G. (1960) *Acoustic theory of speech production*, (Mouton & Co., the Hague).

Kewley-Port, D. (1982) *Measurement of Formant Transitions in Naturally Produced Stop Consonant-Vowel Syllables* , J. Acoust. Soc. Am. 72, 379-389.

Pastel, L. (1987) *Turbulent Noise Sources in Vocal Tract Models* , SM Thesis, Massachusetts Institute of Technology, Cambridge, MA.

Perkell, J.S. (1969) *Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study* , (MIT Press, Cambridge, MA).

Shadle, C. (1985) *The Acoustics of Fricative Consonants* , RLE Technical Report 506, Massachusetts Institute of Technology, Cambridge, MA.

Stevens, K.N. (1971) *Airflow and Turbulence Noise for Fricative and Stop Consonants* , J. Acoust. Soc. Am. 50, 1180-1192.

Stevens, K.N., House, A.S. & Paul, A.P. (1966) *Acoustical Description of Syllable Nuclei: An Interpretation in Terms of a Dynamic Model of Articulation* , J. Acoust. Soc. Am. 40, 123-132.

Sussman, H.M., McCaffrey H.A. & Matthews, S.A. (1991) *An Investigation of Locus Equations as a Source of Relational Invariance for Stop Place Categorization* , J. Acoust. Soc. Am. 90, 1309-1325.