

Mechanisms of stop consonant release for different places  
of articulation

Patricia Keating, John Westbury, and Kenneth N. Stevens  
Research Laboratory of Electronics, M.I.T.  
Cambridge, MA 02139

This paper is a slightly expanded version of one presented at the 99th meeting of the Acoustical Society of America in Atlanta, GA, April 1980.

The following sequence of events occurs at the release of a stop consonant. A complete closure is made at some point along the length of the vocal tract; pressure builds up behind the constriction; the constriction is released, and this release is accompanied by a rapid airflow and by the generation of a brief burst of noise due to turbulence in this flow. At the same time, the vocal folds can either continue vibration through the consonantal closure interval and the release, or cease to vibrate during the closure interval and resume vibration at some time following the release. It has often been observed that vocal-fold vibration begins with greatest delay for velars, less delay for alveolars, and least delay for labials. Also, the time envelope of the burst of sound at stop release differs from one place to another. In particular, double bursts often have been observed for velars, and occasionally for alveolars.

The first aim of this paper is to provide detail about these differences across place of articulation. The second is to <sup>begin to</sup> explain these facts for voiceless unaspirated stops with a model of stop consonant production devised solely in terms of mechanical, aerodynamic, and pressure variations at consonant release, and consistent with existing data on these variations. Results to date will be presented.

Figure 1 shows data from English and Polish on the effect of place of articulation on VOT (top half) and the incidence of double bursts (bottom half), for the voiceless unaspirated stops in each language.<sup>1</sup> Both languages show the overall effect that velars have higher average VOT's and many more double bursts. Apicals have slightly longer VOT's than labials and a few double bursts.<sup>2</sup>

Further acoustic analysis of English voiceless unaspirated stop bursts<sup>3</sup> indicates that the location in time of the amplitude peak also varies with place of articulation. Figure 2 shows the first-differenced waveform and aligned amplitude contour for a sample labial. The amplitude contour represents the rms value of the waveform averaged over 0.6 msec. The amplitude peak occurs at 2.5 msec, as shown by the arrow. For a sample alveolar (Figure 3), the VOT is longer but the amplitude peak is earlier, at 1 msec. For velars, however, the amplitude rise is more gradual and the peak is later, as shown at the top of Figure 4. The middle and bottom waveforms are examples of double bursts, the middle one with a late amplitude peak and the bottom one with an atypical early peak.

How can one account for these various acoustic characteristics in terms of the mechanics of consonantal release for various places of articulation? We begin with a model based on Rothenberg's work (1968), and illustrated for a labial consonant in Figure 5. The relevant components of the model are the mechanical characteristics of the vocal-tract walls (their mass  $M_w$ , their compliance  $C_w$ , and their resistance  $R_w$ ), the acoustic compliance of the air volume,  $C_a$ , and the time-varying acoustic resistances at the glottis,  $R_g$ , and at the constriction,  $R_a$ . The more detailed circuit in the upper right can be reduced to the one in the lower right, since, except for very rapid changes in the intraoral pressure, it is possible to neglect the mass and the resistance of the vocal tract walls.  $C_o$  in the lower circuit represents the sum of the compliance of the vocal tract walls and the acoustic compliance of the air volume. If the time variations of the resistances at the glottis and the constriction are known, and the subglottal pressure  $P_s$  is given, it is a relatively straightforward matter to use this lower circuit model to calculate the glottal airflow  $U_g$ , the airflow through the constriction  $U_a$ , and the intraoral pressure  $P_m$ .

If the glottis is in a configuration appropriate for voicing, then vocal-fold vibration may commence spontaneously when the intraoral pressure  $P_m$  drops enough to create a pressure drop of about 3 cm H<sub>2</sub>O across the glottis (Baer, 1975). If the time variations of volume flow and cross-sectional area at the constriction are known, it is possible, by applying theories of noise generation due to turbulence (Stevens, 1971), to calculate the envelope of the noise burst.

Values for the elements in the model have been fixed according to data from the literature.<sup>4</sup> The glottal resistance  $R_g$  is fixed at a constant 80 acoustic ohms for computational convenience. However, as is the case for many of the values used in our implementation of this model, calculations indicate that the glottal resistance can be constant or time-varying over a wide range of values without substantially affecting output. A fundamental assumption in our modeling to date is that  $R_g$ , whether constant or time-varying, is the same for all three places of articulation (Lindqvist, 1972).

Prior to release, subglottal pressure and intraoral pressure are equalized at 10 cm H<sub>2</sub>O. Following release, intraoral pressure decreases, eventually reaching atmospheric. Vocal fold vibration begins when  $P_m$  drops to 7 cm H<sub>2</sub>O after release.

The principal determinant of the various acoustic and aerodynamic events immediately following release is the resistance at the constriction,  $R_a$ .  $R_a$  is itself determined by the constriction geometry and by the volume velocity of the airflow through the constriction. The way in which these parameters vary with time depends upon the place of articulation of the consonant. Midsagittal tracings for velar, alveolar, and labial stops are shown in Figure 6. Note that velars have a much greater amount of contact than labials and alveolars. Cross-sectional area functions, that is, functions that show cross-sectional area vs. time for each constriction, have been estimated from these midsagittal tracings, from dental cast sections for a single

male speaker, and from assumptions about the acceleration and velocity of the relevant articulator immediately following the consonantal release.<sup>5</sup>

It turns out that good data on the motion of the different articulators at consonantal release are not available, with the possible exception of some high-speed cine data on lip movement reported by Fujimura (1961). We are reduced, then, to making guesses about articulator movements.<sup>6</sup>

In sum, from this circuit analog, we are attempting to model the place effect on VOT and on the envelope of the burst for voiceless unaspirated stop consonants. The input to our implementation is a time function for the area of the constriction. We will report our results for labials, where we must account for voicing onset at 11-12 msec and an amplitude peak at 2-3 msec after the release.

The question is, then, what area functions consistent with what few facts we know about articulator movement will generate the desired outputs in intraoral pressure (to give VOT) and in amplitude envelope (to give the burst)? Suppose, as on the left side of Figure 7a, we smoothly accelerate the lips apart via muscle contraction at a constant rate of  $4000 \text{ mm/sec}^2$  from  $t=0$ . The upper right panel shows how intraoral pressure will decrease, with voicing onset occurring 13-14 msec after release, a fair approximation to our mean measured VOT of 12 msec. Volume velocity through the constriction will increase as shown mid-right. However, the lower right panel shows that the calculated radiated sound pressure peaks about 11 msec after release, well after the 3 msec observed for real speech.

How can we get an earlier amplitude peak? Suppose we move the lips apart at a constant velocity of 125 mm/sec, as shown on the left side of Figure 7b, dotted lines. The bottom right panel shows that indeed the amplitude peak is early enough now, but the top right panel shows that the pressure drop, and voicing onset, is also earlier now, in fact, too early, at 5 msec, given our assumption about the

pressure drop required to initiate vocal fold vibration. This is a general result of the model: if the <sup>rate of change of the</sup> /cross-sectional area function is monotonic, VOT and the burst peak will occur at roughly the same time. Thus a principal conclusion from the model is that the time change in the cross-sectional area of the constriction cannot be characterized either by constant acceleration or constant velocity, since such movements cannot account both for the quick rise and fall in sound pressure and the (assumed) slow fall in intraoral pressure.

Suppose, now, that we move the lips apart at a constant 100 mm/sec for a 2.5 msec interval following release, again as shown in the left half of Figure 7c, ~~the~~ dashed lines, and only subsequently accelerate them apart at a constant rate. Then we find that  $P_m$  decreases more rapidly to a level permitting voice onset at 7-8 msec post-release, while radiated sound pressure reaches a maximum at the point where movement velocity decreases markedly. Thus we see that the area function for labials, and for alveolars even more so, must include both a relatively high velocity early after release to provide a quick peak in acoustic amplitude, and a low velocity later to allow a slow drop in  $P_m$  -- something like the last function shown. Is this reasonable? The lips, and tongue tip, have low masses that respond quickly to the aerodynamic force available to move them at the moment of release. The high velocity would come from an initial stage when the articulators are essentially blown apart, and the later constant acceleration would come from a second stage where they are pulled apart by muscle forces. Movement velocities for the lips have been found to be exceedingly high immediately after the release (Fujimura, 1961).

On the other hand, velars involve the dorsum, which has a large mass that does not respond rapidly to the force due to intraoral pressure. We might expect the rate of change of constriction area for velars to be more nearly monotonic -- in particular, smoothly accelerated like the first function shown -- and this type of area function does give a somewhat

longer VOT and the required later amplitude peak. Still, the VOT calculated from the model for this function is not near the observed 24 msec, and no double bursts are predicted. However, the gradual area function plus the longer constriction creates conditions favorable for a Bernoulli force to operate, much as it does for the vocal folds. This could delay the separation, with an added delay in VOT, and at the same time result in the possibility of a second closure, leading to a double burst.

Further work on the model will include exploring the consequences of various area functions for each place of articulation, and incorporating the proposed Bernoulli effect for velars. More precise calculations on the possible contributions of pressure reduction at the surface due to the Bernoulli force at each place of articulation are required, although preliminary calculations are encouraging.

In conclusion, we are attempting to model VOT and burst amplitude in voiceless unaspirated stops at labial, alveolar, and velar places of articulation. Our intention is to account for the effects of place solely in terms of peripheral constraints, not timing control mechanisms. Our attempt to do so suggests that for labials and alveolars, the articulators initially have a high movement velocity, leading to a relatively short VOT and an early peak in amplitude. For velars, this is not the case; rather, a Bernoulli force between the tongue surface and the palate acts to slow down the release and sometimes produce a second closure and a second release burst.

## Footnotes

1) VOT was measured for English and Polish pre-stress word-initial stops in real-word disyllables. Stops were followed by all of each language's vowels where phonologically possible (12 English and 8 Polish vowels). The graph shows data from 150 English tokens of one speaker and 151 tokens total of 2 Polish speakers recorded in Łódź, Poland. VOT was measured from a computer oscillographic display of waveforms which were not first-differenced. The number of bursts in each token was determined from a first-differenced waveform; in some cases information from LPC spectra was also used.

2) The number of speakers is too small to allow the cross-language differences apparent in the graph to be interpreted, since both VOT and the frequency of double bursts appear to vary with speaker. For example, the incidence of double bursts shown here for Polish velars comes from the tokens of one speaker with relatively few double bursts (32%) and one speaker with relatively many (88%). The English speaker shown has 36% double bursts for /g/, but another English speaker, <sup>not</sup> shown here, has 80%. Thus the question of cross-language differences must await further study. However, to the extent that similar stops in two languages could have slightly different places of articulation, for instance dentals vs. alveolars, we might expect these differences to have an effect on VOT and bursts as well.

The higher VOT values for velars are not attributable to the mere addition of a second burst. The mean VOT value of velars with only one burst is nearly identical to the mean value of velars with two or more bursts.

The place effects on both VOT and bursts are also found for English long lag /p,t,k/ and Polish prevoiced /b,d,g/.

3) These data are derived from 20 tokens each of /bʌ/, /dʌ/, /gʌ/ produced by one English speaker.

4) The compliance of the vocal tract walls was estimated from measurements made by Ishizaka et al. (1975) for the outer surface of the neck. It turns out that the wall compliance is considerably greater than the acoustic compliance of the air volume, and consequently  $C_o$  in the model is dominated by the effects of  $C_w$ .

5) We model these constrictions as rectangular slits, each with two dimensions (length and width) fixed and one dimension (height) variable.

6) Characteristically, the tongue tip exhibits the greatest velocity, and the tongue dorsum the least. The lips, when considered together, can be assumed to have a net opening velocity equal to that for the tongue tip, though the lower lip, when considered alone, generally moves more slowly than the tongue tip (though not as slowly as the dorsum).



## References

- Baer, Thomas. 1975. Investigation of Phonation Using Excised Larynxes. Unpublished Doctoral Thesis, M.I.T.
- Fujimura, Osamu. 1961. Bilabial stop and nasal consonants: A Motion picture study and its acoustic implications. Journal of Speech and Hearing Research 4, 233-47.
- Ishizaka, K., J.C. French, and J.L. Flanagan. 1975. Direct determination of vocal tract wall impedance. IEEE Transactions of Acoustics, Speech, and Signal Processing 4, 370-3.
- Lindqvist, J. 1972. Laryngeal articulation studied in Swedish subjects. STL-QPSR 2-3, 10-27.
- Rothenberg, Martin. 1968. The Breath-Stream Dynamics of Simple-Released Plosive Production. Bibliotheca Phonetica 6. Basel: S. Karger.
- Stevens, K.N. 1971. Airflow and turbulence noise for fricative and stop consonants: Static consideration. J. Acoust. Soc. Amer. 50, 1180-92.

Figure 1

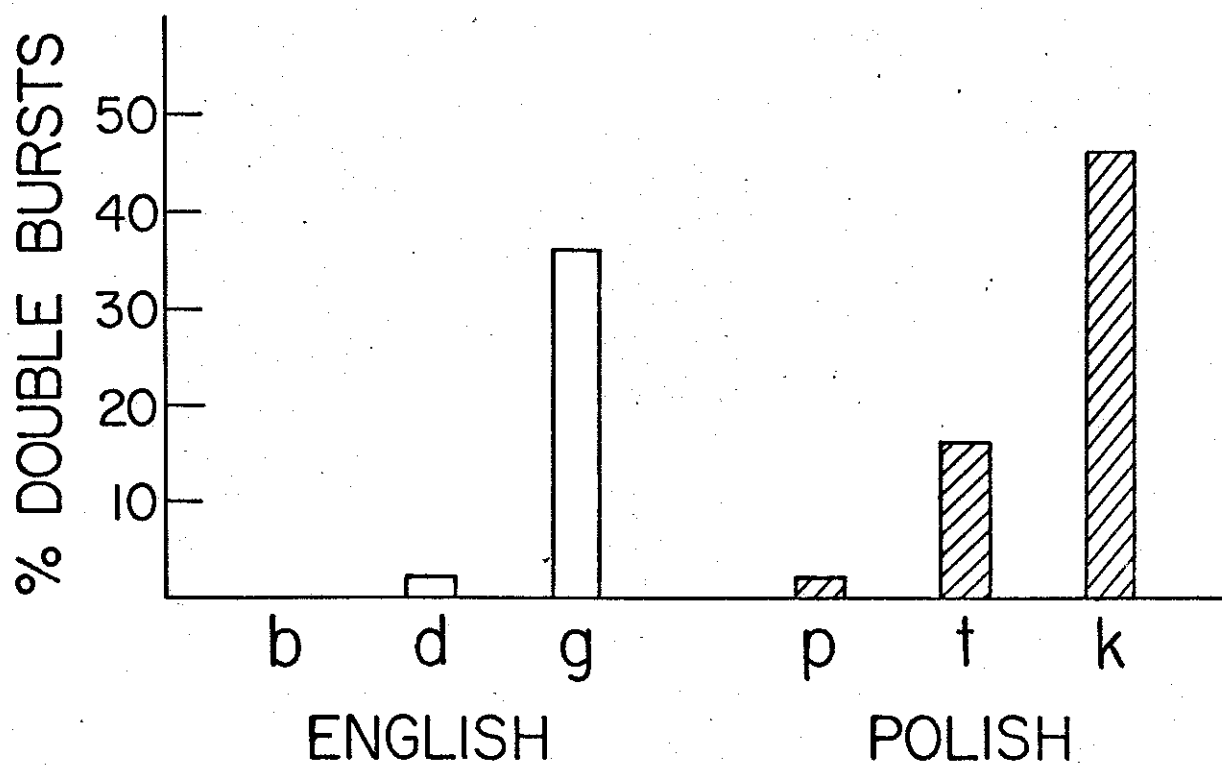
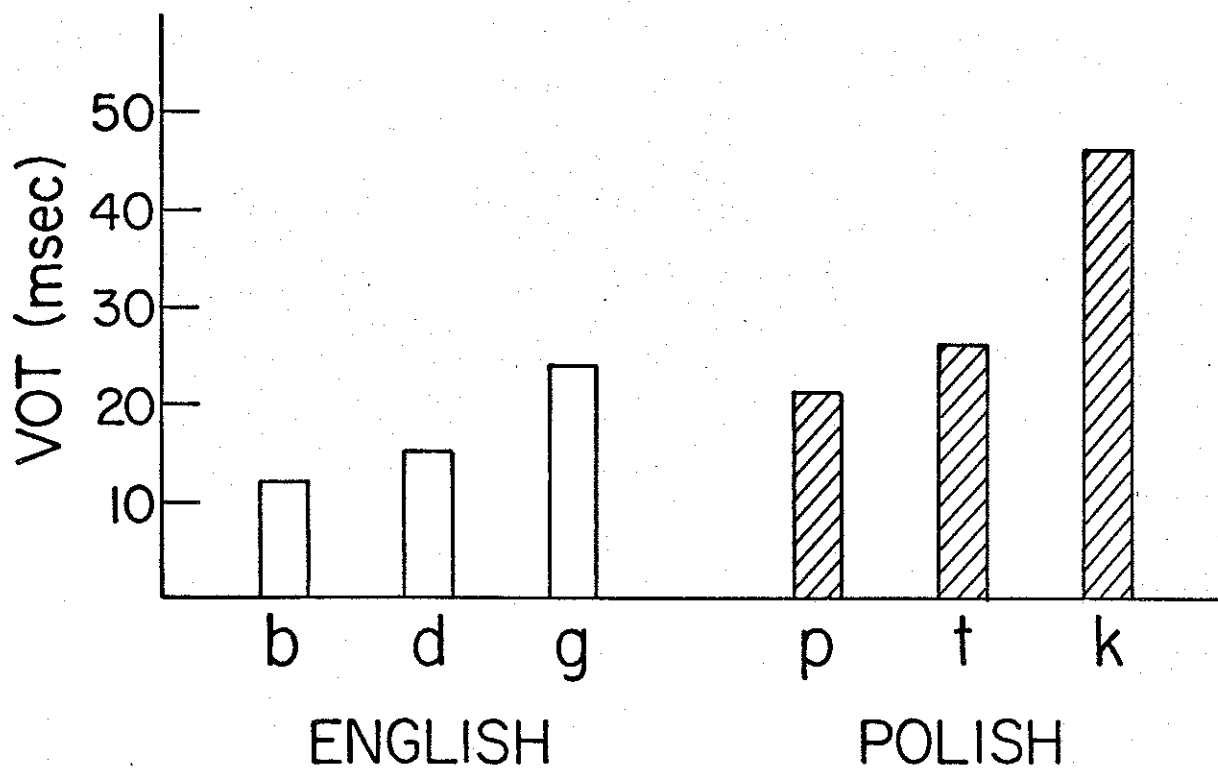


Figure 2

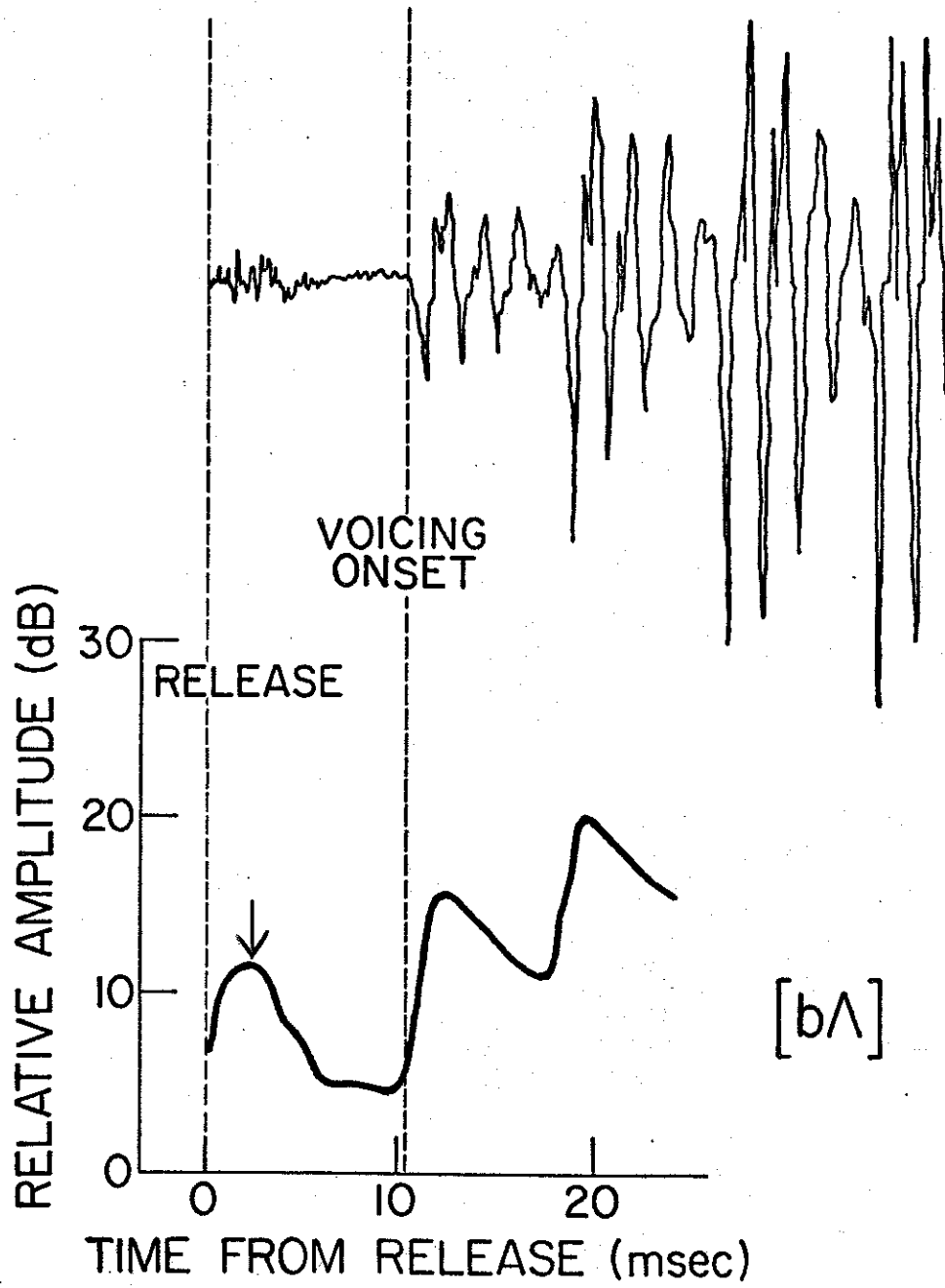


Figure 3

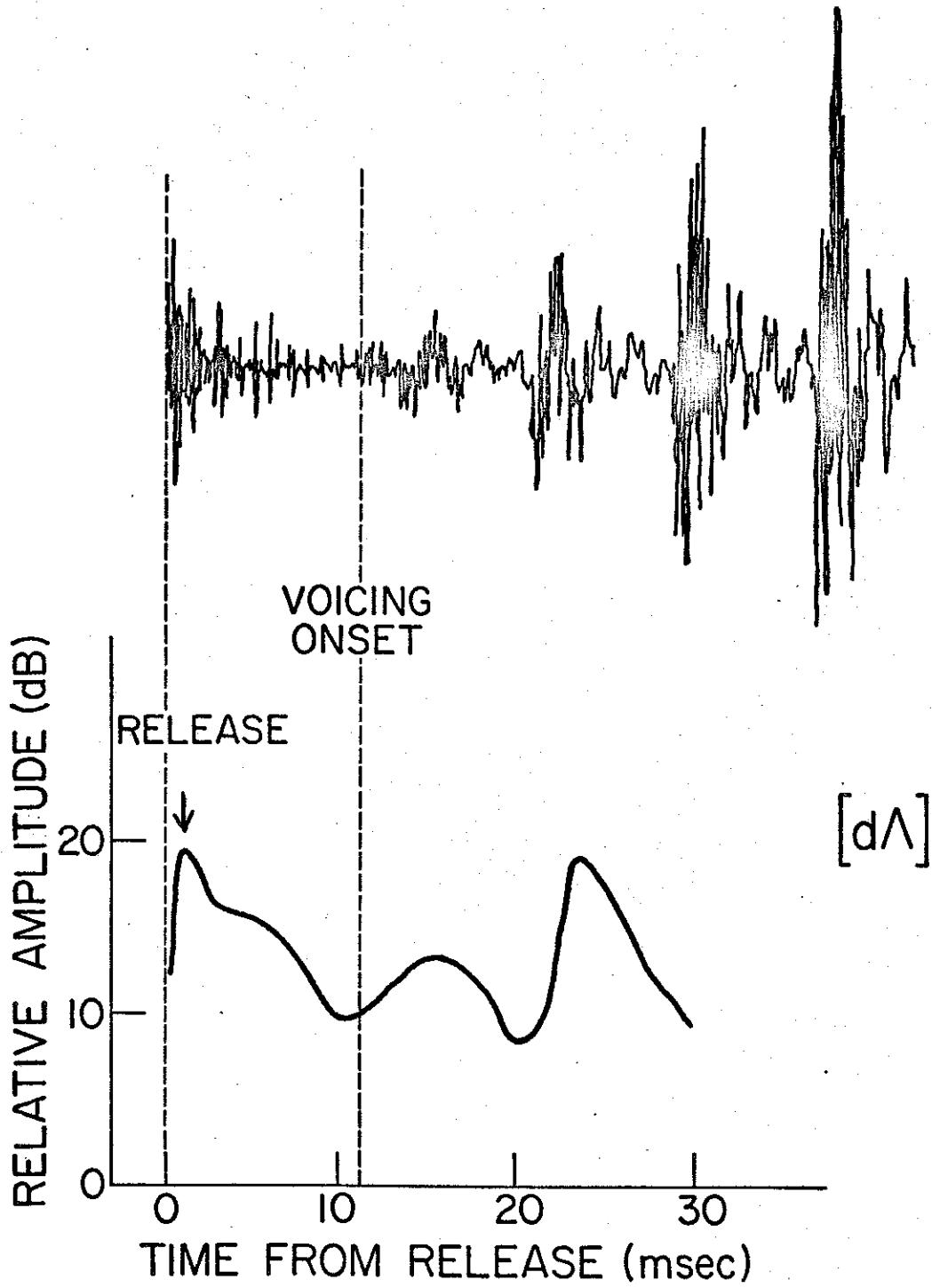


Figure 4

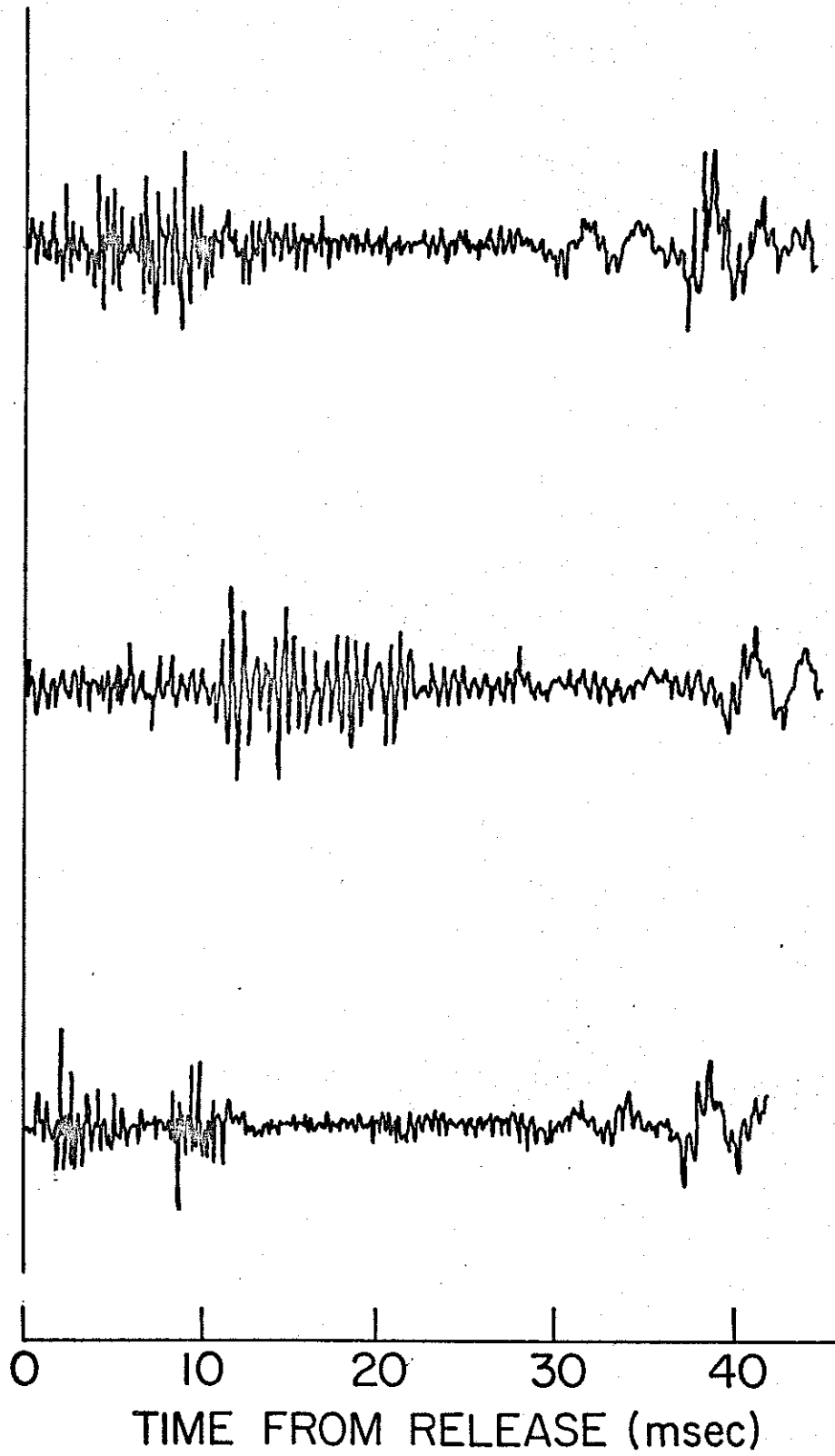


Figure 5

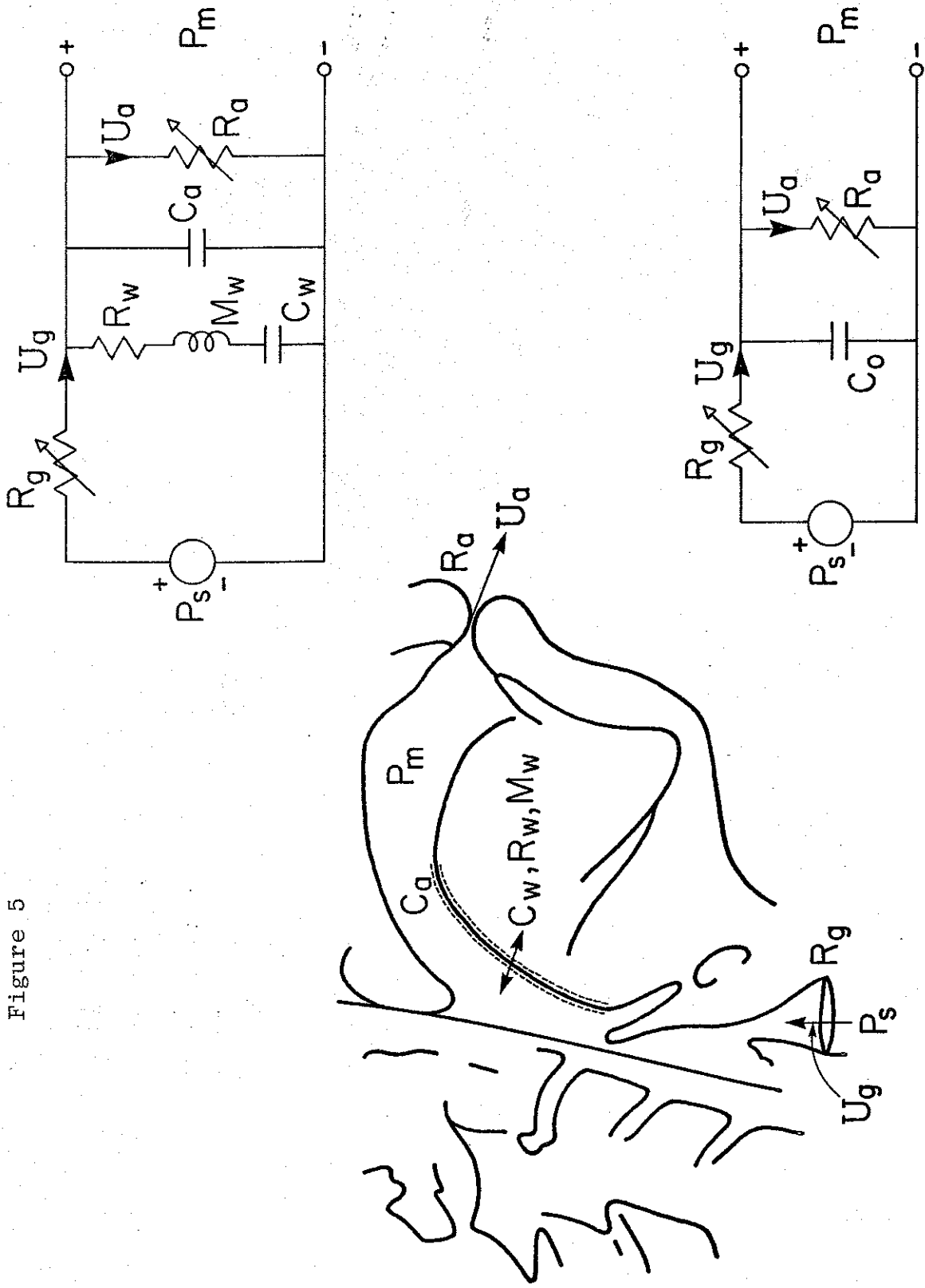
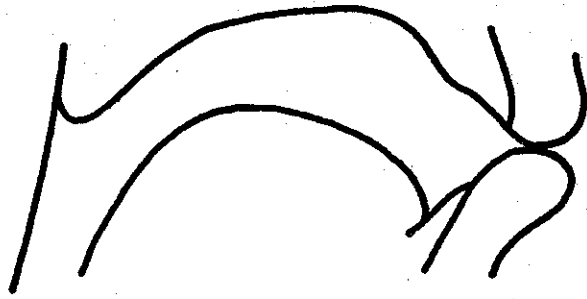
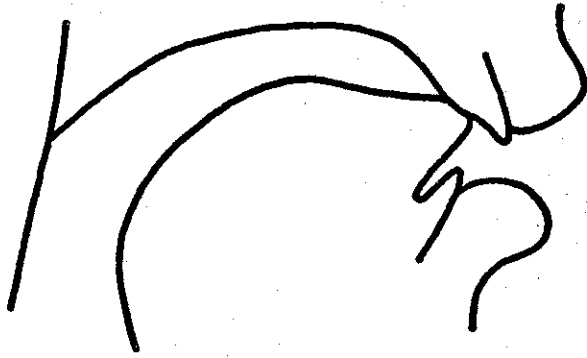


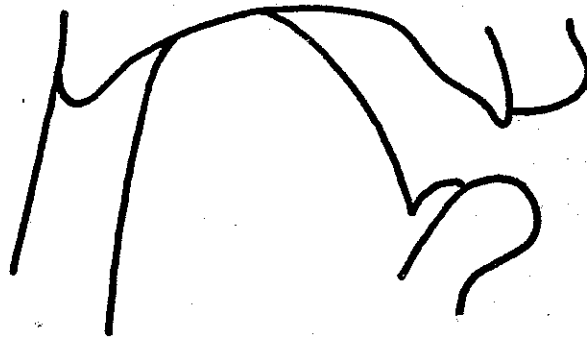
Figure 6



LABIAL



ALVEOLAR



VELAR

Figure 7a

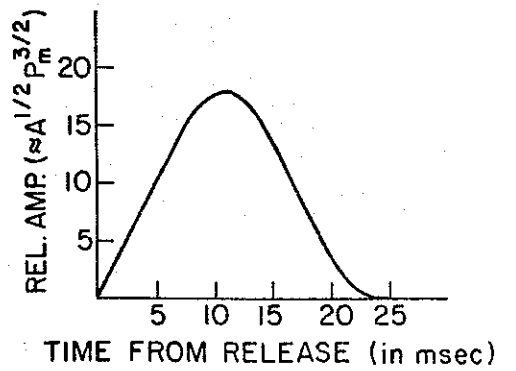
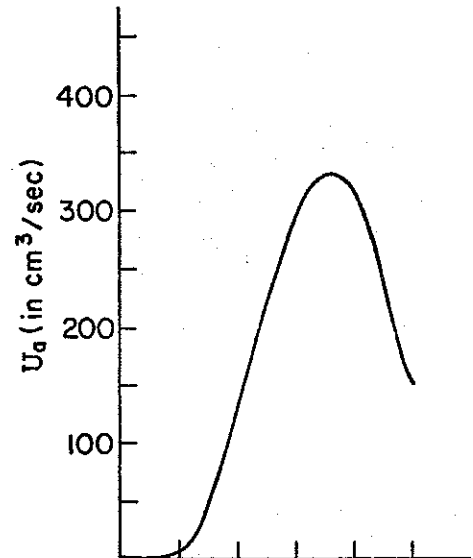
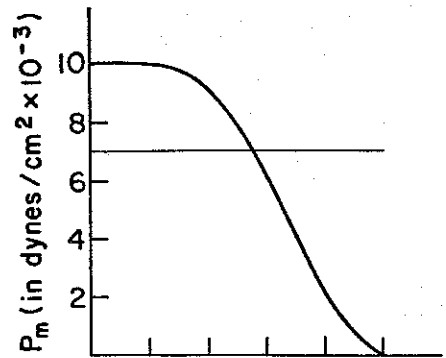
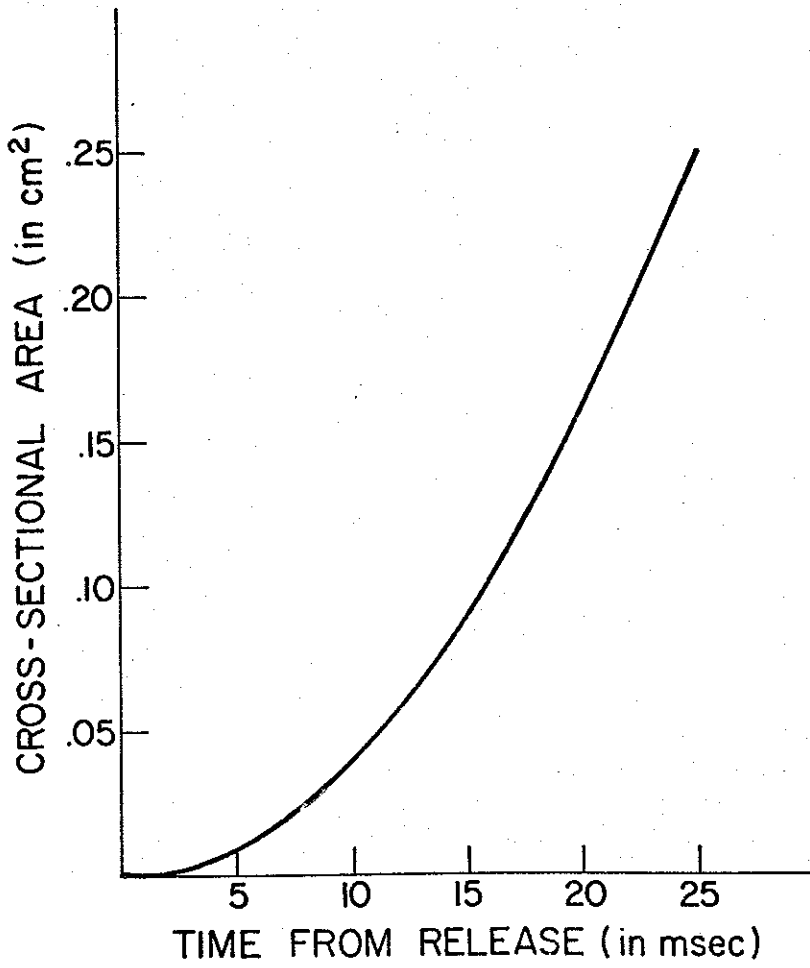




Figure 7c

(= Figure 7a plus second overlay)

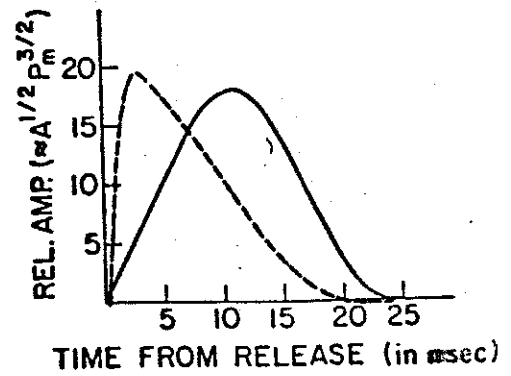
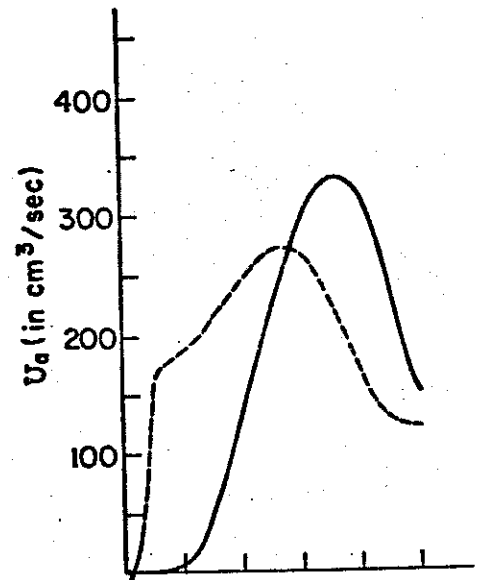
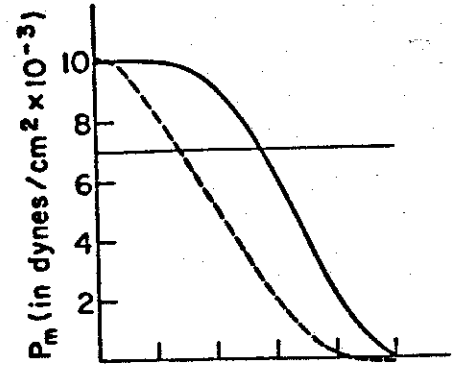
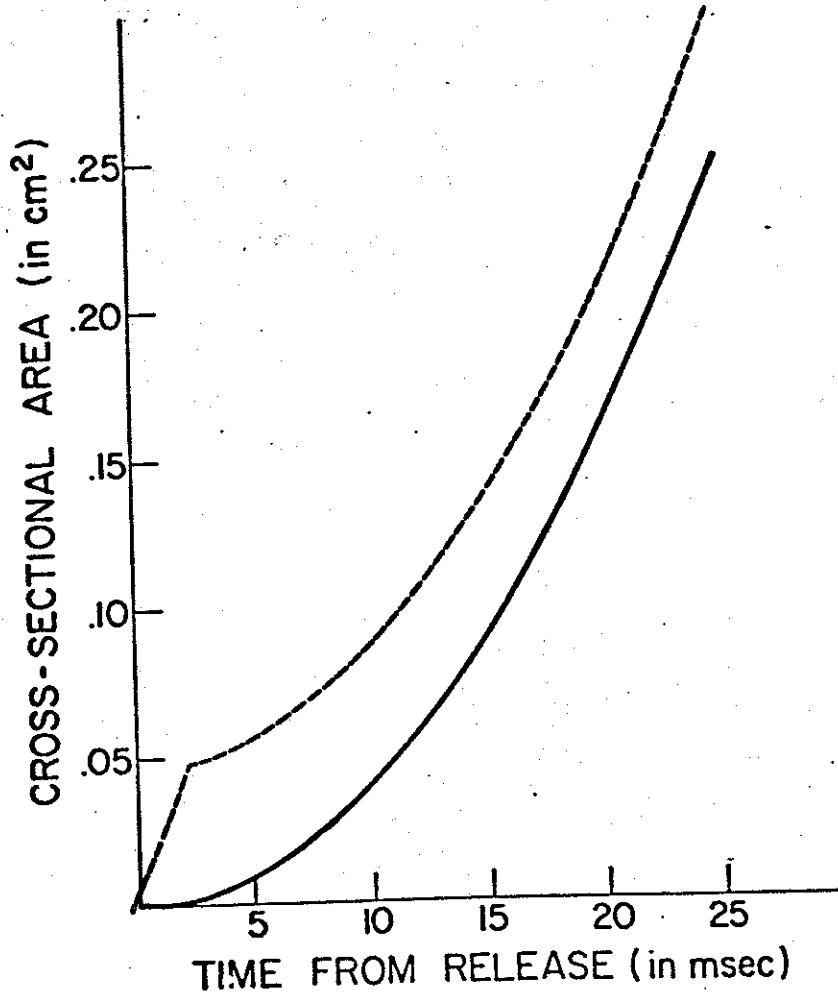


Figure 7b  
(= Figure 7a plus first overlay)

