FORMANT TRANSITIONS: TEASING APART CONSONANT AND VOWEL CONTRIBUTIONS

S. Y. Manuel and K. N. Stevens
Research Laboratory of Electronics
Massachusetts Institute of Technology, Cambridge, MA USA

ABSTRACT

Acoustic data and models of consonant releases suggest that many CV transitions can be regarded as a sequence of two components: (1) an initial local change due to the release of the primary consonant articulator and (2) slower changes of F2 and F3 as the tongue body and jaw move away from positions as supporting articulators for the consonant constriction to their positions as primary articulators for the vowels.

INTRODUCTION

As is known, the transitions of the formants in consonant-vowel syllables reflect the identity of the articulator that makes the consonantal constriction, and where this articulator is placed. The transitions are also influenced by the articulatory configuration for the following vowel, and how much the vowel configuration is anticipated during the consonant constriction. (e.g., [1], [2], [3], [4], [5], [6]).

The release of a consonant in a symmetric VCV can be viewed as having two components, particularly when the primary articulator is the lips or tongue blade. The first component is the initial local movement of the primary articulator. The second, slower component consists of the movements of the tongue body and/or jaw, and possibly rounding of the lips, toward positions required for the vowel.

Here we explore the contributions of both components to formant transitions, and focus on labials and alveolars.

GENERAL METHOD

We are addressing this problem in two ways: (1) by calculation of formant movements for vocal tract models that are manipulated to change in a stepwise manner from a consonant-like to a vowel-like configuration; (2) by examination of natural speech. The use of modeling techniques augments the analysis of natural speech in several important ways. First, in natural speech, when the area of the constriction is still quite small, the sound may be dominated by a transient or frication burst created at the constriction, making it difficult to determine the natural frequencies of the vocal tract as a whole during this initial part of the release. Second, the initial formant movements can be quite rapid, and measurement is subject to the well-known problem of time-frequency trade-offs in accuracy. This problem is side-stepped with modeling which calculates formant frequencies for each step of the movement. Third, at present it is not possible to completely determine the vocal tract shape from the acoustic signal. Modeling allows one to be quite explicit about the vocal tract shape.

LABIAL STOPS

For labial consonants /b, p, m/, the lips alone can not form a constriction unless there is some jaw raising component, particularly with the jaw in a low position. Nevertheless, producing labials entails relatively little participation of the tongue body and jaw, compared with velar and alveolar consonants. Our preliminary modeling of labial releases suggests that, to a first approximation, the formant trajectories at the release of a labial stop in a symmetric VCV can be modeled by assuming a static vocal-tract shape with a time-varying cross-sectional area of the lip opening.

Using a computational model [7], we have emulated labial constrictions of various cross-sectional areas for several different tube shapes, and calculated the formants as the area at the lips is increased. The total tube length was 16.5 cm, divided into 33 sections that were each 0.5 cm long.

In Fig. 1 we show the calculated formant movements for labial release into /ɛ/. At release, F2 is about 1225 Hz and rises about 200 Hz as the area of the constriction increases to 0.5 cm^2. Assuming a rate of increase of opening

of about 50 cm^2/s, this increase in F2 would take place in the first 10 ms. As the opening increases, F2 continues to rise, though at a decreasing rate, and is still about 70 Hz short of its final value 40 ms after release. These modeled data conform reasonably well to formant values following the release of labials in the natural /ɛbɛ/ and /ɛmɛ/ utterances.

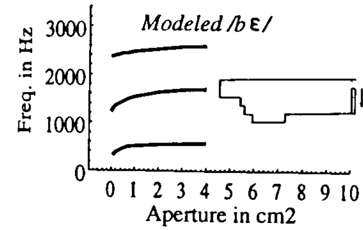

Figure 1. Labial before /ɛ/, showing model with varying aperture on the right, and calculated formants on the left.

For the vowel /i/, the movement of F2 is more extreme. For our modeled /i/ shape, at release F2 is highly sensitive to even very small changes in the labial constriction size, as shown in Fig. 2a. As the constriction area rises from zero to just 0.2 cm^2 (which should take only about 4 ms), F2 jumps from 1040 Hz to about 1800 Hz. F2 rises another 200 Hz as the area increases to about 0.5 cm^2, and then there is very little increase in F2 as the constriction opens further. In the modeled /bi/ the rise in F3 is quite a bit slower than the rise in F2.

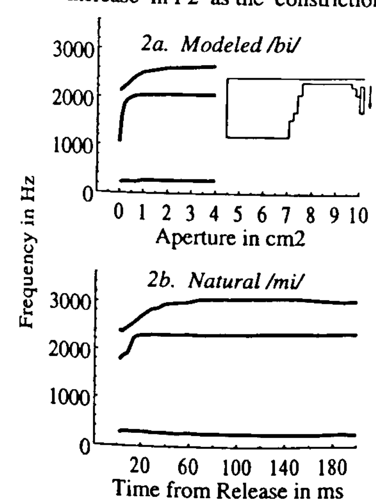

Figure 2. Labial before /i/.

In Fig. 2b we show formants measured at the release of /m/ in natural /imi/. Using nasals avoids the problem of interference of bursts, but still much of the initial rise is not observable as, explained earlier. The first measurable point for F2 was 2 ms after release, and had a value of about 1700 Hz. As in the modeled /bi/, F2 continues to rise another 200 Hz in the next 5 ms or so, and then the rate of rise slows down. F3 continues to rise after F2 completes its movement.

A different picture emerges for a shape similar to a back vowel such as /ɑ/ As indicated in Fig. 3, F2 is relatively flat following release. This is in contrast to the conventional wisdom (but see [4]) that F2 always rises coming out of a labial constriction. The lack of movement in F2 is presumably because the back cavity resonance does not change much during the labial release, and with even a moderate labial constriction, F2 is usually a back cavity resonance. Formant measures made at the release of natural tokens of /ɑbɑ/ and /ɑmɑ/ show a slight rise of F2, and modeling suggests that this is due to tongue body or jaw movement.

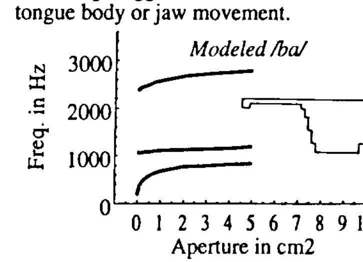

Figure 3. Labial before /ɑ/.

This work is consistent with earlier results (e.g. [4]) for labial releases, and suggests that the general pattern of changes in F2 right after release can probably be attributed to changes in lip aperture, which are superimposed on, and dominate, smaller tongue body and jaw effects.

ALVEOLAR STOPS

We turn now to alveolar stop constrictions (such as /d, t, n/) made with the tongue blade. Due to anatomical constraints, the tongue body must be fronted to allow the tongue blade or tip to make an alveolar constriction. If an alveolar consonant is followed by a back vowel such as /ɑ/, the backing movement of the tongue body should produce a falling F2, and for a following /i/, F2 should rise as the body moves up and forward. However, the initial release of a constriction in the alveolar region should, in theory, result in an initial *rise* of F2. Therefore, one might expect to see a two-part formant trajectory at the release of an alveolar into a back vowel. First one should see an increase in F2, and then a decrease as the tongue moves back.

To examine this expectation, we began our modeling with the vowel /ɛ/, as we expect that in making the alveolar constriction in an utterance like /ɛdɛ/ there is very little tongue body adjustment, with the primary change in vocal tract shape being due to raising the tongue tip. This provides us with an idea of what the tongue-tip constriction itself contributes to the overall formant trajectory. In Fig. 4a the solid lines show what happens to a basic /ɛ/ shape when a constriction of various sizes is made 2 cm back from the front of the tube. The length of the aperture is 0.5 cm. When the aperture is zero, F2 is as low as 1380 Hz, then rises to about 1540 Hz by the time the aperture is 0.3 cm^2. This amount of change in F2 seems large, given the patterns seen in natural speech. We remodeled this constriction with a more anatomically correct tapered tongue tip, and the results are indicated in Fig. 4a as lines with circles. In this case, F2 is only as low as 1590 Hz at release. The raising of F2 when tapering accompanies the tongue-tip constriction is expected from perturbation theory, which predicts an increase in a formant frequency when a narrowing is made near a pressure maximum. The overall pattern is one of a slight rise in F2. We note that all of this movement can be attributed to tongue tip changes, as the tongue body shape was fixed. The natural /nɛ/ in Fig. 4b also shows little F2 movement.

We model /di/ much as we do /dɛ/, with the exception that the emulated tongue-body constriction is tighter and more forward for the /di/. Changing

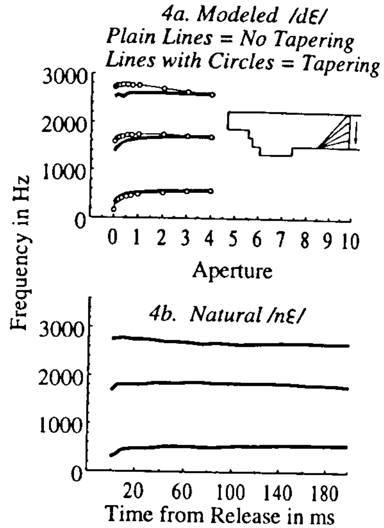

Figure 4 Alveolar before /ɛ/.

tongue-tip aperture alone results in a very rapid increase in both F2 and F3 as the aperture increases from zero to 0.5 cm^2, as shown in Fig. 5a. Not shown is the fact that tapering the tongue tip increases F2 and F3 substantially when the constriction is quite small. This may account for the fact that in natural speech, F2 and F3 do not appear to be particularly low at onset of the formants for /di/ and /ni/. In any case, whether due to actual differences in formant values, or to measurement problems, F2 may be quite similar for /di/ and /bi/. However, F3 is rather different for the two syllables, rising rapidly after the alveolar, but slowly after the labial.

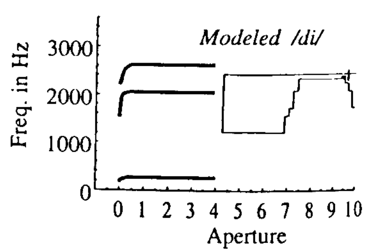

Figure 5. Alveolar before /i/.

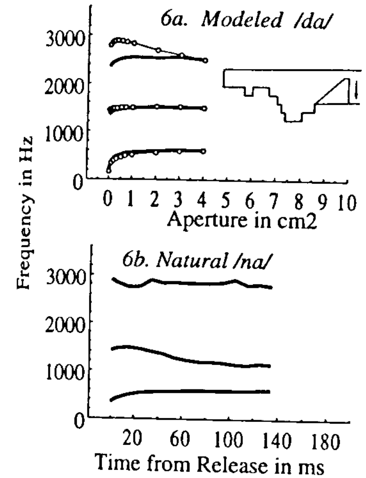

Figure 6. Alveolar before /ɑ/.

For /dɑ/ we begin with a shape similar to that of /ɛ/, but slightly more open in front and more constricted in the back. From this basic shape we adjust the tongue tip constriction, including the tapering, as noted above. As shown in Fig. 6a, there is very little movement in F2 - only a small rise. Of course, one normally expects to see a decrease in F2 as one moves from /d/ to /ɑ/. This model shows that such a drop in F2 must be due to the tongue body moving back, and not to the change in the consonant constriction itself. In the natural utterance /nɑ/ shown in Fig. 6b one can see a two-part F2 transition: a rather short, flat part followed by a longer, falling part. Modeling suggests that the former portion is due to a combination of the release of the alveolar constriction and some tongue backing, and that the latter portion is simply due to tongue backing.

SUMMARY

Formant transitions in the first 20-odd ms following labial releases are primarily a consequence of lip movement, and are probably not greatly influenced greatly by movements of the tongue body. Even after this initial interval, the influence of tongue-body movement is not very great, as the tongue body does not move much.

There are substantial differences in F2 movements following labial releases into front vowels like /ɛ/ compared with back vowels like /ɑ/. There is very little F2 movement for back vowels because the back cavity resonance does not change appreciably during the release. For front vowels there is a significant and rapid upward movement of F2, and a more slowly rising F3.

Formation of an alveolar constriction requires tapering of the vocal-tract area for a few cm posterior to the point of contact. With this tapering, the upward movement of F2 following release of an alveolar constriction is minimized.

Following an alveolar release into a front vowel like /ɛ/, the tongue-body movement is small, and as a consequence of tongue-blade tapering, there is little F2 movement. Following an alveolar release into a back vowel, the combination of the tapered constriction and the backward tongue-body movement leads to an initial flat F2 trajectory followed by a slow downward movement.

ACKNOWLEDGMENT

Work supported in part by NIH grant DC00075.

REFERENCES

[1] Delattre, P. C., Liberman, A. M., & Cooper, F.S. (1955) Acoustic loci and transitional cues for consonants. *J. Acoust. Soc. Am., 27*, 769-773.

[2] Fant, G. *Acoustic theory of speech production.* The Hague: Mouton.

[3] Stevens, K. N., & House, A S. (1956) Studies of formant transitions using a vocal tract analog. *J. Acoust. Soc. Am., 28*, 578-585.

[4] Fujimura, O. (1961). Bilabial stop and nasal consonants: a motion picture study and its acoustical implications. *J. Speech & Hearing Res., 4*, 233-247.

[5] Ohman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *J. Acoust. Soc. Am., 39*, 151-168.

[6] 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.

[7] A program developed by S. Maeda was used to make the computations.