Automatic Selection of Syllable Nuclei

Now that a solution to the problem of gross errors has been found, we can turn our attention to the selection of a syllable nucleus. In order to test different nucleus selection criteria, six speakers were selected from the 1977 corpus, three males and three females. Together, approximately 1000 syllables were measured for these six speakers.

The first criterion that was considered for the selection of a syllable nucleus was the spectrum at which the first two formants were changing the least quickly. This seems reasonable for French since it has monophthongal vowels and any rapid formant transitions are likely to be consonant transitions. A coefficient of change is defined for each spectrum in the syllable, except the first and the last. The coefficient of change is the sum of the absolute values of the percentage changes in the first two formants from the present spectrum to the previous spectrum plus the sum of the absolute values of the percentage changes in those two formants from the present spectrum to the following spectrum. Thus, where C_i is the coefficient of change of the ith spectrum and $F1_j$, $F2_j$ are F1 and F2 of the jth spectrum, respectively, we have

$$C_i = \frac{|F1_i - F1_{i-1}| + |F1_i - F1_{i+1}|}{F1_i} + \frac{|F2_i - F2_{i-1}| + |F2_i - F2_{i+1}|}{F2_i}$$

Figure 3-2 is a graph of formant frequency versus time for the word tout. Interpolated formants appear as triangles. Reassigned formants are indicated with squares. The arrow points to the spectrum that was chosen as the syllable nucleus because it had the lowest coefficient of change in the syllable.

This steady state criterion was applied to the one thousand syllable test corpus and standard deviations of the formant frequencies were calculated for each speaker-allophone combination. Since sixteen word classes[5] were used and since there were six speakers in the experiment, this gave 96 within-speaker, within-allophone standard deviations for each formant. These standard deviations ranged from approximately from 25 to 200 Hertz for F1 and from 40 to 280 Hertz for F2.

To evaluate the result, this can be compared to another criterion. Another criterion that has been suggested for choosing vowel nuclei is the point of maximum first formant frequency in the syllable, corresponding to the point of maximum opening of the articulators. This criterion was applied to the same one thousand syllables and the 96 within-speaker, within-allophone standard deviations were recalcu-

[5]These sixteen allophonic classes are /i/, /e/, /ɛ/ after /r/, /ɛ/ after segments other than /r/, /a/, /ɑ/ in the word pas, /ɑ/ in words other than pas, /ɔ/ before /r/, /ɔ/ segments other than /r/, /o/ in checked syllables, /o/ in free syllables, /u/ after coronal segments, /u/ after non-coronal segments, /y/, /ø/, and /œ/.

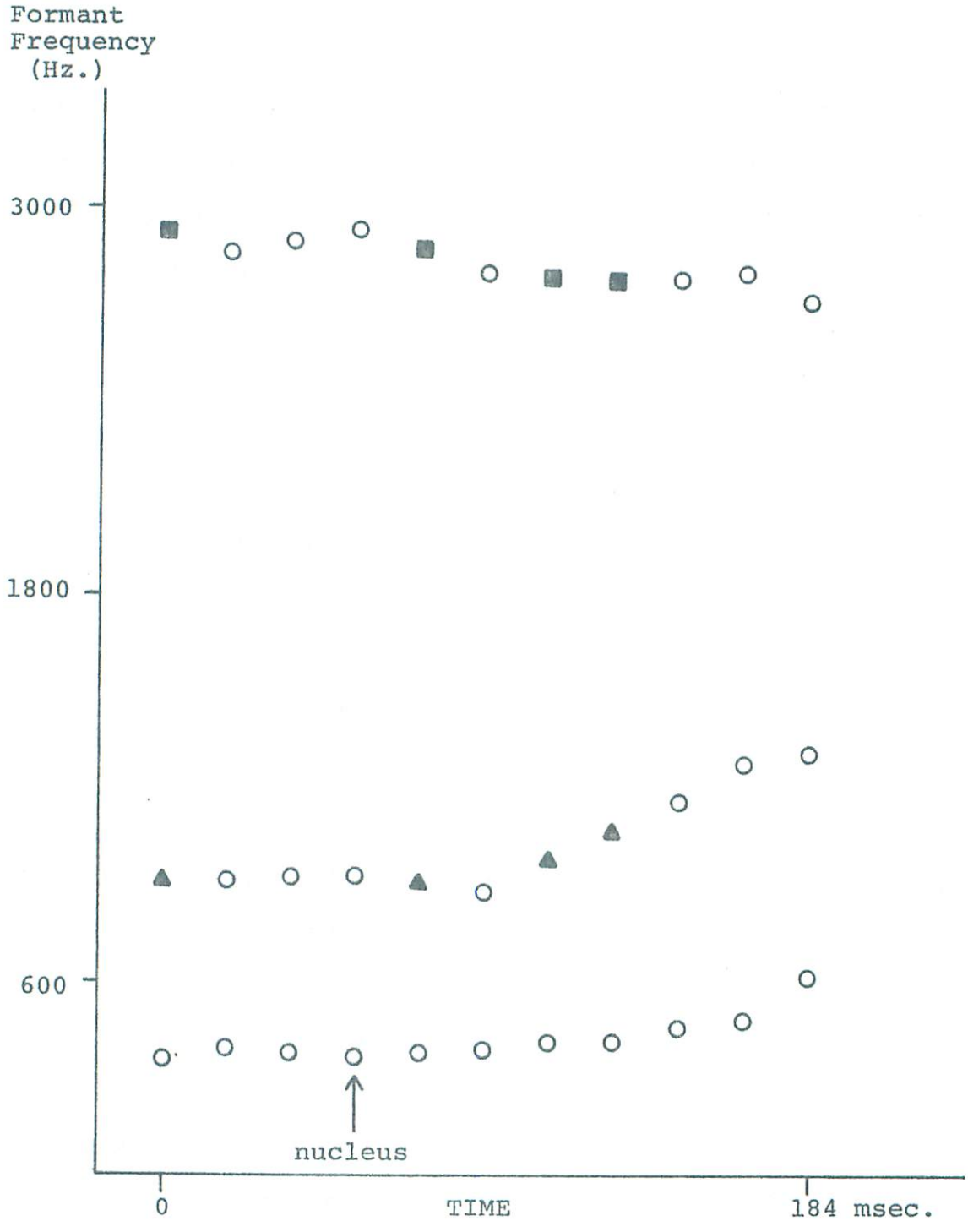

Figure 3-2. Final estimates of formant frequencies plotted against time for one token of French tout produced in running speech. ■ = reassigned formant, ▲ = interpolation. Arrow shows nucleus.

lated for each formant under this maximum F1 selection criterion. An F-test was used to compare the within-speaker, within-allophone variances produced under the two criteria. For most of the speaker-allophone combinations there was no significant difference in variance between the two methods of nucleus selection. Both methods seemed to perform equally well in choosing a nucleus whose formants were minimally affected by neighboring phonetic segments. However, in the cases where a significant difference in variance between the methods occurred, the steady state criterion seemed to perform better most of the time. For those speaker-allophone combinations in which a significant difference in variance occurred, the steady state criterion achieved the lower variance in 57% of the cases for F1, 87% of the cases for F2, and 75% of the cases for F3.

Returning to Figure 3-2, a human observer would never mistake the final spectrum, which represents a central offglide, for the nucleus. Whereas an automatic algorithm working under the maximum F1 criterion would mistakenly choose the last spectrum as the vowel nucleus, the steady state criterion would not make such an error. While both nucleus selection criteria seem to work well, the steady state criterion seems better adapted to a language like French, which has nearly monophthongal vowels. This is the method that was used to select the nuclei of all the tokens in the 1977 sample.