Extensions to the Locus Theory

M. O'Kane
Canberra, Australia

1. Introduction

The Locus Theory originally put forward by Delattre, Liberman and Cooper (1955) proved useful in explaining formant transitions between vowels and consonants. According to this theory, which was developed from experiments with synthetic speech, the formant transitions from the steady state of a vowel to a neighbouring plosive consonant point to a particular frequency or locus depending on the place of articulation of the consonant. Öhman (1966) showed that in real speech the situation was somewhat more complex. He showed that for /b/ and /d/ the position to which the transitions pointed was dependent not only on the place of articulation of the consonant but also on the second formant of the vowel preceding the consonant. Thus, the second formant /b/ focus could be anywhere in the range 500-1400 Hz; being low if the second formant of the preceding vowel was low and high if the second formant vowel was high. A similar situation occurred for /d/ with the locus occurring in the range 1400-1700 Hz. For /g/ the formants pointed to a locus which was determined by the vowel following the /g/ if the vowel preceding the /g/ was a front or central vowel, and to a low locus if the vowel preceding the consonant was a back vowel.

We decided to use formant locus as one of the components in an algorithm to automatically differentiate between plosive consonants occurring in a VCV context in continuous speech. We found, however, that the Locus Theory even with the Öhman modifications was somewhat inadequate for our purposes. Unanswered questions included the following:

How did the Locus Theory apply to Australian English, particularly continuous, conversational Australian English?

Were locus ranges the same for all speakers? In particular were male and female locus range the same?

Did the presence of a word boundary within a VCV combination affect the locus range? VCV combinations in English occur more frequently across word boundaries than within words.

The answers to these questions were sought through an experiment which is described below.

2. Method

The experiment was actually designed to investigate coarticulation, juncture, plosive consonant phenomena and the interaction between all three continuous speech examples of Australian English. Only the results pertaining to the Locus Theory are presented here; results concerning other aspects of the experiment are given in O'Kane (1981).

Lists of two-word sequences were prepared. Each of these two-word sequences was of either of two forms:

1. The first word ended in a VC combination and the second word began with a V, where the vowels could be either /i/ or /ɔ/ (the foremost and backmost vowels in Australian English), and the consonant was one of the six plosives e.g. 'heat ought'.
2. The first word ended in V and the second word began with a CV combination where the vowels could be either /i/ or /ɔ/ and the consonant was one of the six plosives e.g. 'he taught'.

Thus, with the two vowels and six plosive consonants and two juncture positions there were a total of forty-eight two-word combinations. The five male and five female speakers all spoke general Australian English. They were from a geographically diverse area of Eastern Australia and ranged in age from twenty-four to fifty. It should perhaps be pointed out that Australian English is remarkably homogeneous geographically as regards pronunciation, the greatest variation occurring between people from different social and employment groups (Mitchell and Delbridge, 1965).

Each speaker was presented with the list of two-word sequences and instructed not to study the list but to immediately begin saying sentences containing the word sequences. It was impressed on the subjects that the sentences they produced were to be spoken at a conversational speed and that the semantic content of the sentences was not of particular importance. It was even suggested that slightly silly sentences would be perfectly acceptable. If a subject was having particular difficulty in producing a sentence containing any given two-word sequence he was told not to waste time over it but to say something such as 'I can't think of a sentence with ---- ---- in it'. All this was to keep the subject speaking at as conversational a rate as possible. This aim was largely achieved.

Experimental studies on conversational speech are difficult as it is hard to exercise any control over the many interacting parameters and still obtain the characteristic spontaneity of conversational speech. In the experiment described here phonetic and junctural contexts were controlled and almost always the speech of the subjects sounded (in the author's subjective opinion) to be at conversational rate. It is perhaps indicative of the relative newness of continuous speech recognition research that paradigms such as the one used in the experiment outlined above generally have not been systematically developed as a means of controlled study of conversational speech phenomena.

The sentences containing the two-word sequences were recorded on a Nakamichi 550 cassette recorder, using a Bayer microphone. The required VCV tokens were excised from the sentences using a waveform editing routine. These VCV combinations were then analysed using the Interactive Laboratory System waveform analysis package. For the male voices the speech was sampled at 10 kHz and for the female voices the sampling rate was 16 kHz. For each token a linear prediction analysis was done using the autocorrelation technique. Plots of the spectral peaks derived from this analysis as a function of time provided a means of measuring the formants.

3. Results

The data examined in the experiment described above revealed the following results (which can be seen in Figures 1 and 2).

1. That in continuous speech (at least for Australian English) the position of the second formant locus is primarily determined for *all* consonants by the vowel preceding the consonant. *This is more general than Öhman's result.*
2. That the position of the F_2 /b/ locus can range for male voices from 500 Hz for the case when the preceding vowel is /ɔ/ to 1300 Hz for the case where the preceding vowel is /i/.
3. That the position of the F_2 /d/ locus ranges for male voices from 1350 Hz for the case where the preceding vowel is /ɔ/ to 1900 Hz when the preceding vowel is /i/.
4. That there are high and low loci for /g/. If the preceding vowel is /i/ the /g/ locus for male voices is in the range 2000-2300 Hz. If the preceding vowel is /ɔ/ the /g/ locus for male voices is in the range 800-1500 Hz.
5. That, at least for the case where the consonant is /g/, the exact position of the locus is also determined by the vowel following the consonant. Thus if the preceding vowel is a back vowel and the following vowel is a front vowel then the locus position will be at the high end of the low /g/-locus range. But if the preceding vowel is a front vowel and the following vowel is a back vowel then the locus position will be at the low end of the high /g/-locus range.
6. That the higher the locus position is in frequency, the greater is the difference between male and female locus values for a given VCV production. The average locus range for various consonants are given in Table I for both male and female voices.

Table I. Locus ranges of plosive consonants at the three places of articulation for male and female voices. All measurements in Hz

	Labial locus range	Alveolar locus range	Low velar locus range	High velar locus range
Male voices	500-1300	1350-1900	800-1500	2000-2300
Female voices	500-1500	1550-2400	800-1800	2400-2900

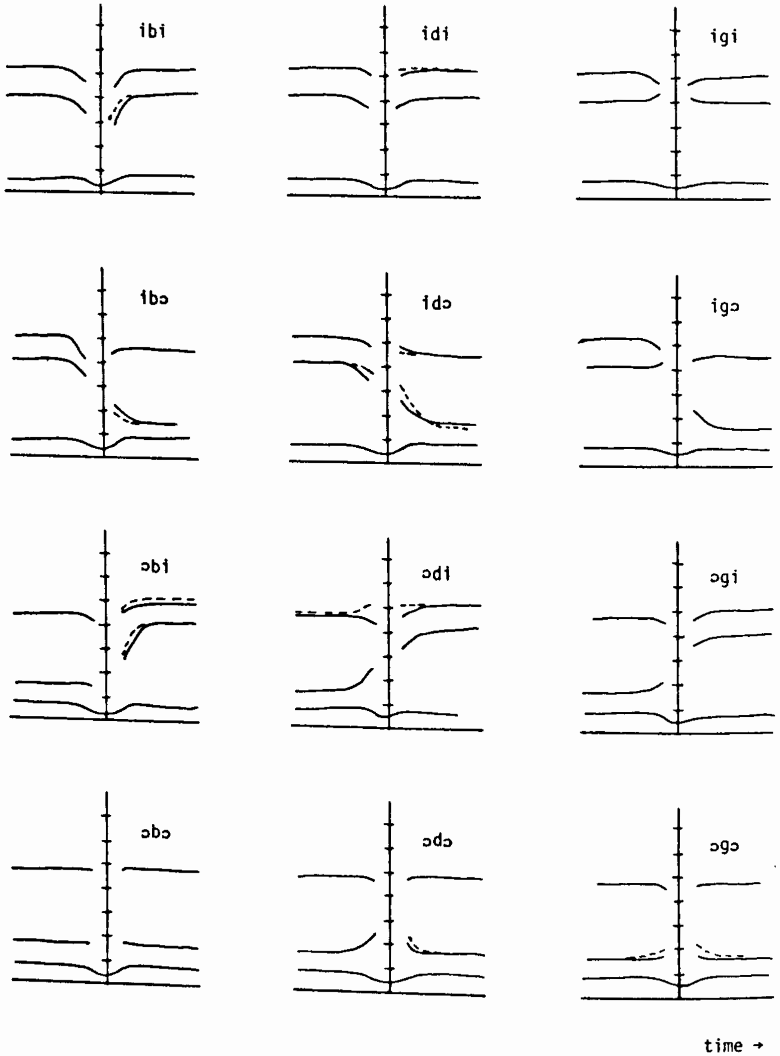

Figure 1. Average, first, second and third formants for male voices VCV sequences in which C is a lax plosive. Where dotted lines are shown they represent the /VCV/ case in contradistinction to the /VCV/ case (full lines). Where only a full line is shown both cases essentially coincide. Frequency marks are 0.5 kHz apart. Time is not strictly to scale but is of the order of 450 msec, in total.

7. That the difference between male and female locus positions are reflections of the differences between formant positions for vowels for male and female voices. This can be seen by reference to Figure 3. Figure 3 shows the average first and second formants of several vowels in Australian English for male and female speakers based on data from Oasa (1980) and O'Kane (1981). The differences in male and female voices is slight for the

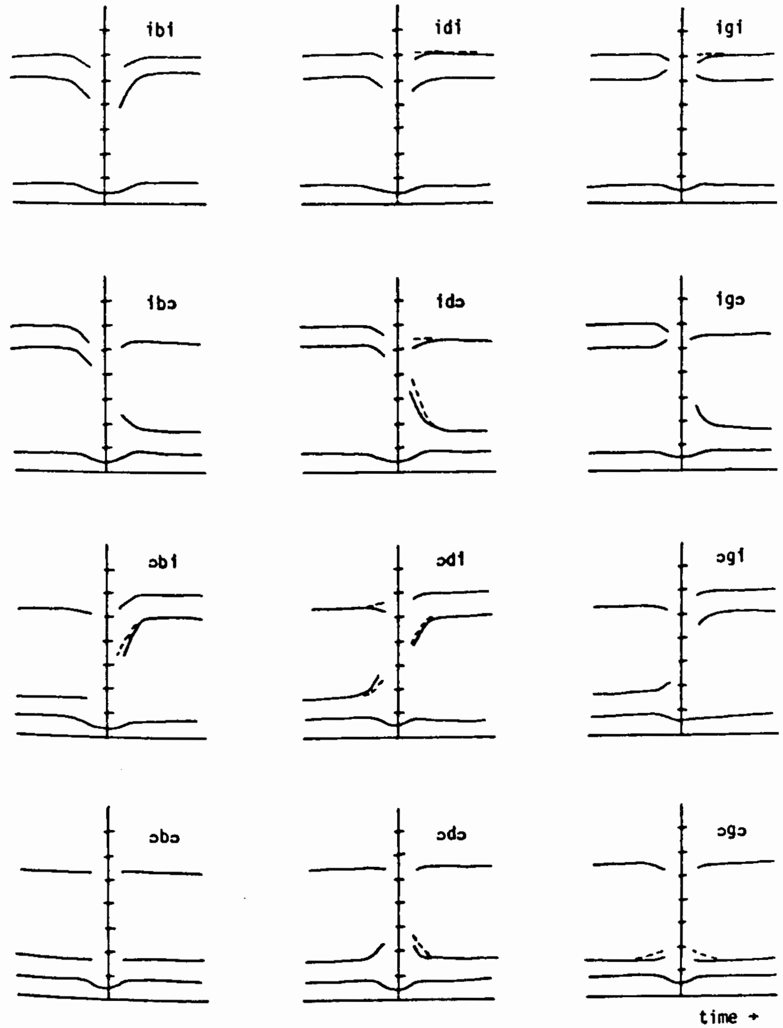

Figure 2. Average, first, second and third formants for female voices for VCV sequences in which C is a lax plosive. Where dotted lines are shown they represent the /VCV/ case in contradistinction to the /VCV/ case (full lines). Where only a full line is shown both cases essentially coincide. Frequency marks are 0.5 kHz apart. Time is not strictly to scale but is of the order of 450 msec, in total.

transitions from /ɔ/ to the /d/ but are considerable for the transition from /d/ to /i/.

8. Inter-speaker differences connected with locus effects within each sex grouping are slight particularly for the labial and velar consonants. In the production of alveolar consonants, however, some speakers produce barely noticeable formant transitions which makes the finding of the locus position difficult.

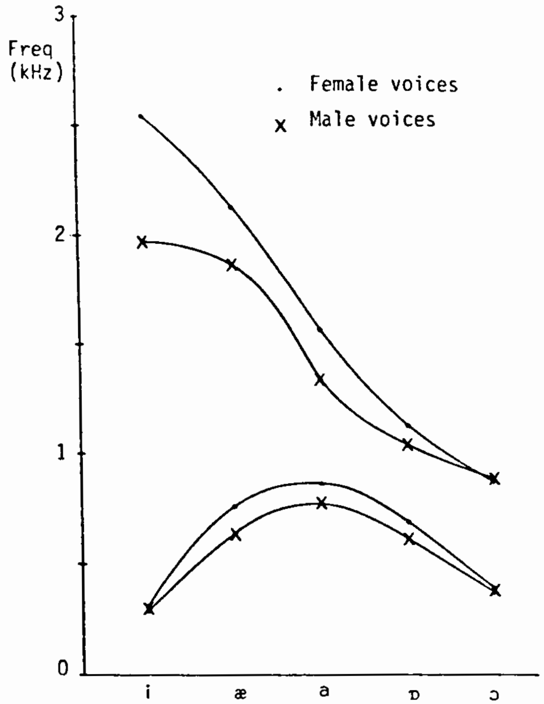

Figure 3. Average first and second formants of the vowels /i/, /æ/, /a/, /ɒ/, /ɔ/ for male and female speakers of Australian English.

9. That differences in the locus positions due to the word-final or word-initial nature of the consonant are only noticeable in a few cases. If an effect is present at all it is often only noticeable in the CV transition. As can be seen in Figures 1 and 2 the most noticeable cases in which juncture effects are present are /ɔbi/, /idɔ/, /ɔdɔ/ and ɔgɔ/.

4. Conclusion

An expanded version of the Locus Theory for plosive consonant transitions has been presented. The results given here can be incorporated as a partial cue for the automatic recognition of plosive consonants. This cue is particularly helpful in cases (5% in the experiment described here) where no burst is produced in the articulation of the consonant. It should also be noted (see Table 1) that the locus cue for plosive recognition is slightly more useful for female voices than for male voices.

References

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

Mitchel, A.G. and Delbridge, A. (1965). *The Speech of Australian Adolescents: A Survey.* Angus and Robertson, Sydney.

Oasa, H. (1980). Regional Variations in Australian English. M.A. Thesis, Australian National University, Canberra.

Öhman, S.E.G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *J. Acoust. Soc Amer.*, **39**, 151-168.

O'Kane, M. (1981). Acoustic-Phonetic Processing for Continuous Speech Recognition. Ph.D. Thesis, Australian National University, Canberra.