LONG-DISTANCE COARTICULATORY EFFECTS OF ENGLISH /l/ AND /r/

By
PAULA WEST
Christ Church, University of Oxford

A dissertation submitted to the Committee for
COMPARATIVE PHILOLOGY AND GENERAL LINGUISTICS
in partial fulfilment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

TRINITY TERM

2000

LONG-DISTANCE COARTICULATORY EFFECTS OF ENGLISH /l/ AND /r/

Paula West, Christ Church
D.Phil. Trinity 2000

This thesis explores the coarticulatory effects of English /l/ and /r/, examining their articulatory basis, acoustic manifestation and perceptual relevance. It demonstrates that there are perceptually relevant coarticulatory differences associated with the distinction between /l/ and /r/. Two perceptual experiments, an articulatory experiment and a modelling study were conducted. Both perceptual experiments used a modified gating technique. The first experiment demonstrates that the coarticulatory effects of /l/ and /r/ on surrounding vowels and consonants can sometimes be used by listeners to identify an /l/ or /r/ which has been deleted and replaced by noise. The second perceptual experiment shows that the cues for an /r/ are more perceptually salient than those for an /l/. The articulatory experiment used simultaneous electromagnetic articulography, electropalatography and acoustic recordings to investigate the coarticulatory effects of /l/ and /r/. In /r/ contexts, relative to /l/ contexts, raising and retraction of the tongue, lip rounding and lowering of F_3 were found, up to two syllables preceding and following the /r/. The extent of this coarticulatory effect is far greater than commonly acknowledged in the coarticulation literature. Phonetic and phonological theories fail to predict or account for effects of this extent. The theory that coarticulation can be modelled as overlap of articulatory gestures was tested in a modelling study. A subset of the articulatory data was modelled numerically using dynamical descriptions of articulatory gestures from an approach developed at Haskins Laboratories. The modelling showed that long-distance coarticulatory effects could not be adequately accounted for by gestural overlap alone. Feature-spreading models, such as Keating's window model of coarticulation, are also unable to account for these effects adequately. The results of this thesis pose a challenge to current phonetic and phonological theory, as they show that coarticulatory effects have greater extent than commonly recognised.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, John Coleman, for his support and guidance throughout this research. Many thanks to Andrew Slater, for patient and comprehensive technical assistance, and to Mario Cortina and Burt Rosner for help with statistical analysis.

Thanks are also due to Alan Wrench and colleagues at Queen Margaret University College in Edinburgh for making the EMA experiments possible. Martin Barry provided help with a perceptual experiment run in Manchester. I am especially grateful to my three EMA subjects, who gave a great deal of their time.

For financial support I am grateful to Christ Church, Oxford, and the Association of Commonwealth Universities, who made this research possible.

I would also like to thank my family and friends for their support, especially those in the Phonetics lab for company over numerous coffee and lunch breaks. Many thanks to my father for last minute proof-reading, and to Magnus, for unfailing support, encouragement, and meals.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1. Introduction

This thesis explores the extent and nature of coarticulatory patterns associated with English /l/ and /r/, examining their articulatory basis, acoustic manifestation and perceptual relevance. It demonstrates that coarticulatory differences associated with the distinction between /l/ and /r/ exist up to two syllables remote from the /l/ or /r/. Coarticulatory effects are found in articulation (differences in tongue and lip placement) and acoustics (mainly F_3 differences). These coarticulatory patterns are perceptually relevant and can sometimes be used by listeners to identify an /l/ or /r/ which has been deleted and replaced by noise. The extent of this coarticulatory effect is greater than most reported, and is highly problematic for current phonetic and phonological theory, which fails to predict or account for effects of this extent. Although coarticulation is an area which has been extensively studied, research has focussed on the coarticulatory effects of vowels on adjacent segments or across consonants. The effects of consonants on adjacent vowels have been reported in detail (*e.g.* Stevens and House 1963), but consonantal coarticulatory effects (C to V) are generally acknowledged to be smaller, with a more restricted temporal range than V to C effects (Farnetani 1990). This is despite the claim that the consonants /l/ and /r/ have long-distance coarticulatory effects, or "resonances" (Kelly and Local 1989). This thesis is an experimental verification of some of Kelly and Local's claims; a demonstration of the existence of long-domain resonance patterns. This chapter begins by giving an overview of research on English liquids, followed by a discussion of existing work on, and hypotheses about, long-domain resonances. These are then discussed in terms of the extensive literature on coarticulation. An outline of the structure of the thesis concludes the chapter.

1.2. English /l/ and /r/

Pronunciation of the liquid approximants /l/ and /r/ in English varies from dialect to dialect, and possibly speaker to speaker. Liquids have been the subject of considerable cross-linguistic research, and have evoked interest because of their complex articulations and clear acoustic correlates. The articulation and acoustics of these sounds have been the subject of research ranging from early impressionistic observations to modern instrumental techniques, such as electropalatography and magnetic resonance imaging. This research has shown that /l/ and /r/ have a wide range of articulations in English, with implications for surrounding speech.

1.2.1. The allophones, traditional accounts

Traditional phonetic descriptions of English /l/ identify two allophones which differ in resonance: clear or light /l/ and dark /l/ (Jones 1972). The application of the term 'resonance' to English liquids derives from its use in physics to refer to a vibrating system which has reached a condition in which the frequency of the driving and the driven system coincide (Fry 1979). This applied to acoustics gives the formant frequencies of speech. Liquids, which have formant structure, may thus be described as having resonances.

Clear and dark /l/ are usually defined as differing in secondary articulation. Clear /l/ is palatalised, or has a secondary articulation made by raising the front of the tongue towards the hard palate. Dark /l/, whether velarized (tongue raised towards the soft palate) or pharyngealized (tongue root retracted towards the back wall of the pharynx), has a backer secondary articulation (Abercrombie 1967, 63). This front/back distinction is also described in terms of vocalic categories:

> The difference between 'clear' varieties of l and 'dark' varieties of l is thus simply a difference of vowel resonance. In clear varieties of l there is a raising of the front of the tongue in the direction of the hard palate (in addition to the tongue-tip articulation), while in dark varieties of l there is a raising of the back of the tongue in the direction of the soft palate. In other words, clear l-sounds have the resonance of front vowels, whereas dark l-sounds have the resonance of back vowels. (Jones 1972, 176)

Jones describes English dark /l/, which is produced finally and before consonants other than /j/, as having the resonance of a back vowel approaching /u/. Southern English clear /l/, which occurs pre-vocalically, has resonances approaching /i/. There are a few other cases of different vowel resonances in /l/, depending on the following vowel, and different varieties of English have a different distribution of /l/ sounds. For example, Irish English has clear /l/ in all positions, and Scottish English usually has dark /l/ (Jones 1972, 176).

Descriptions of British English /r/ note that the commonest variant [ɹ] is often labialised, or produced with some lip protrusion, particularly in stressed position (Abercrombie 1967; Brown 1981; Jones 1972). Other variants described for British English include the trill [r] and flap or tap [ɾ]. Jones reports that the flap tends to occur in unstressed intervocalic position for RP speakers, and that trills are generally found in the north of England and Scotland (Jones 1972). Traditional accounts do not identify clear and dark variants of /r/.

1.2.2. Dialectal variation

Existing descriptions of English dialects do not pay much attention to characteristic secondary articulations or resonances of /r/ and /l/, although Hughes and Trudgill (1979, 66) note that /l/ is clear in all environments in Tyneside speech. Accounts of dialect are largely concerned with grammar, and the modern focus is a sociolinguistic one, concerned with education and dialect prestige. A list of resources for dialect study shows the range of literature available (Edwards 1993). Standard texts agree that pronunciation of /r/ and /l/ are amongst the defining characteristics of English dialects. Some features frequently commented upon are amount of /l/-vocalisation (largely vocalised in the south east, especially London) and the rhoticity of dialects, i.e. presence or absence of post-vocalic /r/ (present in Lancashire and South West England) (Trudgill 1990, 51-62). Wells (1982, 313-317) writes that /l/-vocalisation leads to various neutralisations of contrast (such as *full*, *fall* and *fool*) in the London area, although in other cases vowel contrasts are preserved. /l/-vocalisation is also called loss of post-vocalic /l/ and compared to the similar /r/-loss in non-rhotic dialects (Hughes and Trudgill 1979).

An unusual dialectal feature involving /l/ is reported for the Bristol area, which is a rhotic area: words ending in a vowel in other dialects are pronounced with a final intrusive /l/, so *idea* and *ideal* would be expected to be homophonous. This phenomenon is not identical to linking /r/, as the intrusive /l/ may appear sentence-finally or in words in isolation, but linking /r/ occurs only inter-vocalically across a word boundary (Trudgill 1990, 73; Wells 1982, 344).

1.2.3. Instrumental studies

Magnetic resonance imaging (MRI), electropalatography (EPG), electromagnetic articulography (EMA or EMMA), cinefluorography and X-ray microbeam are instrumental techniques which have been used to study liquid approximants in American and British English. Giles and Moll (1975) report a cinefluorographic study of allophones of American English /l/. Based on x-ray tracings of data from three speakers, they conclude that pre- and post-vocalic /l/ are best treated as separate categories. Post-vocalic /l/ has a less anterior lingual position than pre-vocalic /l/, which exhibits faster movement and more frequent anterior lingual contact.

Gartenberg (1984) conducted an EPG study of allophonic variation in /l/ articulations using one speaker of British English (from the West Midlands). He found the two-way clear/dark distinction countered by his data: the range of pre-vocalic (clear) and post-vocalic (dark) /l/ articulations overlap. In some cases intervocalic /l/s more clearly resembled post-vocalic /l/s. However, his study did not include acoustic data, and whilst it is suggestive, no conclusions about the relative clearness and darkness of the liquids can be drawn, as changes in articulation may not result in acoustic differences. He also notes that prevocalic, postvocalic and intervocalic /l/s formed separate groups in articulatory terms (largely based on duration patterns).

Sproat and Fujimura (1993) present acoustic and X-ray microbeam data for English /l/ (as produced by three speakers of American English and one of British English). They show that the clear-dark distinction for these speakers rests primarily in the degree of retraction and lowering of the tongue dorsum and the timing of the retraction and lowering in relation to the apical gesture, with darker /l/ having greater and earlier retraction and lowering. Huffman (1997) extends Sproat and Fujimura's model to include onset /l/s, with acoustic data recorded from eight American English speakers. She shows that onset /l/s may vary in relative darkness (as shown in F_2 values). From her study of postconsonantal /l/ and intervocalic /l/ following a schwa (in *e.g. blow* and *below*), she concludes that variation in the quality of onset /l/ is linked to duration (which differs with subjects), and can be explained by phonetic and timing factors.

Narayanan *et al.* (1997) report from MRI and EPG data that tongue body shapes are similar for clear and dark /l/, but that overall tongue body position is different. They confirm Sproat and Fujimura's (1993) findings that velarization is not a consistent characteristic of the production of dark /l/, whilst decreased uvular and upper-pharyngeal area due to retracted tongue root and/or raising of the posterior tongue body is consistent. In a further articulatory study using MRI and EPG, Alwan *et al.* (1997) report that American English /r/ tends to be produced with secondary constriction in the pharynx caused by retraction of the tongue root. Despite a wide range of articulatory positions for /r/ they found no systematic differences in 3-D vocal tract and tongue shape for word-initial and syllabic /r/. However an acoustic study of American English by Olive *et al.* (1993) shows that both /l/ and /r/ have light and dark variants, with different distributions. Light /l/ and dark /r/ appear syllable initially, dark /l/ and light /r/ appear after a vowel in the same syllable.

Westbury *et al.* (1995) report great variation in articulation found in an X-ray microbeam study of the syllable initial /r/ in *row* produced by 55 American English speakers. However, the very large differences in tongue shape at phonation onset of *row* were not accompanied by statistically reliable differences in formant frequencies. Neither was a strong relationship between tongue shape and oral cavity size and shape found. A later paper, by a different set of authors (Guenther *et al.* 1999), did find evidence for articulatory trade-offs which reduce acoustic variability in American English /r/ production. Based on acoustic theory, they hypothesise that trading relations might exist between constriction length and front cavity length, front cavity length and constriction area, constriction length and constriction area. These trading relations were translated into correlations between the position of electromagnetic midsagittal articulometer (EMMA) coils placed on the tongue. For all seven speakers recorded, they find that at least one trading relation is used, and furthermore show, using statistical analysis of variance in F_3, that excluding the articulatory covariances in an estimate of variance significantly increases the variance of F_3. They conclude that the trading relations exist to reduce acoustic variance and maintain the primary acoustic cue for /r/ (low F_3), and that the different vocal tract shapes used by subjects suggest that speakers aim for an acoustic target and not a vocal tract shape in /r/ production.

Another instrumental study of American English liquids, using electromagnetic articulography, examines /l/ and /r/ from the perspective of articulatory phonology, in which gestures, and not acoustic goals, are paramount (Gick 1999a; Gick 1999b). Gick compares the liquids and glides, and replicates

Sproat and Fujimura's finding of a delayed tongue tip gesture in syllable-final /l/ for one subject, although the difference does not reach statistical significance. Gick distinguishes between initial and final /l/, and /l/ which he terms "ambiguously syllabified" - word-final /l/ followed by a vowel-initial word. The stimuli he compared were *hall hotter* with final /l/, *ha lotter* with initial /l/, and *hall otter* with "ambiguously syllabified" /l/. He recorded two subjects producing these phrases in the frame sentence 'Say — again'. Although the results did not reach statistical significance, and he could not measure the tongue dorsum gesture for one subject, he found that "ambiguously syllabified" /l/ patterned consistently with final /l/ in terms of tongue tip displacement. A similar study of /r/ (using the stimuli *par hotter*, *pa rotter* and *par otter*) shows gestural reduction of the tongue tip/blade gestures (depending on speaker strategy in producing the /r/) in final position, with less fronting and more raising of the tongue tip. Intermediate levels of reduction were found in the "ambiguously syllabified" *par otter* case.

1.2.4. Acoustic properties

Most acoustic studies of /r/ and /l/ agree on the following fundamental properties of these sounds, whether produced in General American or Southern British English: they have acoustic structure similar to vowels, with formant-like resonances, and the transitions of these formant structures are important for perception (Ladefoged 1993; Lieberman and Blumstein 1988). Liquids cannot be modelled with a single tube model, as the tongue in the oral cavity is shaped in such a way as to split or bifurcate the airway (Stevens 1998). This bifurcation has acoustic consequences in the frequency range above 1500 Hz (additional poles and zeros), which distinguish liquids from vowels and glides.

Stevens and Blumstein (1994) note that laterals are distinguished from other consonants by a second formant (F_2) with decreased prominence and a wider bandwidth. For /l/ the first formant (F_1) is lower than 500 Hz, somewhere around 250 Hz, with clear /l/ having lower values than dark /l/. F_2 is in the range 900-1600 Hz, with lower values indicating dark /l/s, and the third formant (F_3) is around 2400 Hz. F_3 is usually steady and is similar for dark and clear /l/, whilst the second formant falls. Higher formants are considerably reduced in intensity, and spectral minima or antiformants are often found between the first and second formant and in the region of the third formant (Bladon 1979; Cruttenden 1994; Ladefoged 1993; Ladefoged and Maddieson 1986; Sproat and Fujimura 1993; Stevens 1998).

Narayanan *et al.* (1997) relate the acoustics of /l/ to their articulatory findings, and suggest that the low F_1 is a result of resonance between the large back-cavity volume and the oral-constriction space. They associate F_2 with the half-wavelength resonance of the back cavity which accounts for the reduction in F_2 when the cavity is lengthened by retraction or raising of the tongue, as in the production of dark /l/.

/r/ has a first formant in the range 120-600 Hz, second formant 700-1200 Hz and third formant fairly close to the second formant, so fairly low. The formant structure is similar to vowels, but the steady-state formants are relatively weak and the formants are usually moving. Steeply rising transitions of the first four formants to a following vowel are characteristic (Cruttenden 1994; Ladefoged 1993; Lehiste 1964). Alwan *et al.* (1997) suggest that low F_1 in this case is a result of a large cavity behind the oral constriction, as with /l/. Low F_3 values are associated with the large front cavity volume between the oral constriction and the lips, influenced by slight lip rounding (Alwan *et al.* 1997; Stevens 1998).

Some early studies (Lisker 1957; O'Connor *et al.* 1957) focused on a larger group of (American) English consonants, which are usually called glides: /r, l, w, j/. The studies attempted to synthesise these consonants and to discover cues for perception. O'Connor *et al.* searched for shared acoustic features of these sounds to distinguish them as a natural class, on the grounds that they form a group in terms of their articulation (degree of oral stricture) and phonotactic distribution (*e.g.* appearing as the third element in a consonant cluster). Their work is an example of the search for simple acoustic invariants corresponding to articulatory descriptions, by which sounds could be identified and reproduced. As a class this group of sounds could be synthesised using similar transition durations, starting from steady-state onsets. /w/ and /j/ could be distinguished from /r/ and /l/ by using transitions of the first two formants, a third formant needed to be added to distinguish between /r/ which required a rising third formant, and /l/ which required a steady-state third formant (O'Connor *et al.* 1957).

Both studies found that /l/ was particularly difficult to synthesise. There was no common area of agreement (in terms of a range of fixed formant values) for the various vowels it was synthesised between (Lisker 1957). Different vowels required different starting frequencies for both /r/ and /l/ formant transitions, which had interesting implications for locus theory: if the consonants began their transitions at a consonantal locus (as locus theory required), the locus had to be variable (O'Connor *et al.* 1957). This work shows that acoustically the sounds labelled /r/ and /l/ (and the other consonants considered) are not simple combinations of invariants, but are far more complicated and depend on surrounding context.

In a more recent study of /r/ and /l/ in a single context, Polka and Strange (1985) synthesised the words *rock* and *lock*, varying F_1, F_2 and F_3. They found that trading relations existed between the cues to /l/s and /r/s: with /r/ cued by a lower F_3 and a shorter steady state F_1 with a more gradual rise than /l/. They conclude from their study that listeners discriminate between /l/ and /r/ on the basis of an integrated phonetic percept.

1.2.5. Phonological accounts

The terminology "liquid" which covers English /l /and /r/, implies that these sounds pattern together in a group. Walsh Dickey (1997) argues that liquids, defined as rhotics and sonorant laterals (or non-nasal sonorants), are in fact a well-defined phonological group, and proposes the feature [liquid] to cover the group. Evidence for this group comes from phonotactic restrictions, morpheme structure constraints and dissimilations. Using a version of Feature Geometry, Walsh Dickey shows that liquids may be differentiated from each other by place of articulation features and that manner features such as lateral are unnecessary or "phonologically invalid". Liquids are all sonorant consonants with complex or branching place nodes: the laterals have a complex corono-dorsal articulation and rhotics a non-primary laminal node. This work is a recognition within phonological theory of the importance of secondary articulations of liquids.

Carter (1999) is also concerned with the phonological system into which liquids enter. Following Kelly and Local's (1986; 1989) description of different resonance patterns (clearness and darkness) associated with liquids in different varieties of English, Carter investigates the resonance patterns associated with liquids in four British varieties of English. His findings, based on measurements of F_2 frequencies and duration of the liquids from two rhotic dialects (Tyrone and Fife) and two non-rhotic dialects (Wearside and SE Lancashire), show that clearness and darkness of liquids is both structure-dependent and dialect-specific. Initial laterals are clearer than final laterals in all dialects. In the non-rhotic varieties there is no contrast in liquid in final position. In initial position, Wearside has a clear lateral and dark rhotic, SE Lancashire has a darker lateral and clearer rhotic. These patterns are as suggested by Kelly and Local (1986). The rhotic dialects show fixed patterns of resonance, with rhotics darker than laterals in initial position and clearer than laterals in final position. Carter concludes that this is evidence for a single liquid system with dialect and structure-dependent interpretation at different points in syllable structure.

Sproat and Fujimura (1993), followed by Huffman (1997), suggest that the clear and dark allophones of English /l/ are not distinct phonological or phonetic units. They suggest that /l/ is a single phonological entity with clearer or darker realisations depending on factors such as position in the syllable, duration of the surrounding prosodic context and strength of a contiguous boundary (intonational or morphological). In a similar vein, Hardcastle and Barry (1995) have suggested that vocalisation of /l/, which typically occurs before back consonants and more frequently if also preceded by a front vowel (*milk* would be a good candidate), is not a categorical process either but partly caused by coarticulatory undershoot. These authors all agree that the allophones of English /l/ are not discrete units, but form part of a continuum of graded variation, contextually determined.

Carter (1999) argues that Sproat and Fujimura's gestural alignment model, in which the dorsal and apical gestures of /l/ are aligned to syllable structure, and clearness or darkness is a result of this gestural alignment, cannot explain his data. If clearness and darkness is a result of gestural alignment, then alignment of gestures must be dependent on dialect-specific phonetic interpretation of the phonological system of liquids. However, the data only weakly support this view: the SE Lancashire dialect has a relatively dark initial lateral, but it is not significantly different (in F_2) from the final lateral. The Fife dialect also has no significant difference in F_2 between initial and final laterals and the other two dialects have clear initial and dark final laterals, in keeping with Sproat and Fujimura's observations about clearness and darkness and syllable structure. Nevertheless, Carter's measurements do show that there are dialect-specific differences in the implementation of clearness and darkness, which need to be accounted for. As Nolan writes,

> ... the variation in degree of darkness of laterals and rhotics is more than could be accounted for by general principles ... The problem is how to account for phonetic detail which is neither contrastive (and hence has no place in a phonological representation as traditionally conceived) nor

> predictable from the physiology, dynamics, and so on of the speech mechanism (Nolan 1999, 5).

Carter argues for a single phonological system of liquids with systematic constraints on phonetic interpretation, dependent on structure and dialect, and his data show that there are indeed dialect specific differences in resonance, or clearness and darkness of English liquids.

1.3. Long-domain resonance

1.3.1. Kelly and Local's observations

Kelly and Local (1986) note that clearness and darkness is not only associated with /l/. They quote a description of English /n/ as 'duller' than German /n/, and refer to work which mentions palatalised /s/ and /p/, or dark /t/ and /f/. In fact, Kelly and Local suggest that the clear/dark terminology which is used for various segments can be viewed as the reflex of one underlying phonetic or phonological phenomenon: resonance.

> The term resonance is used to refer primarily to those features of consonants which are the auditory concomitant of configurations of the oral cavity and/or movements of the tongue other than those which relate to principal articulation. The kinds of resonance that we are concerned to reflect in our records are treated under the heading of secondary articulation in the phonetics literature. The possible number of such oral cavity configurations and corresponding auditory categories is large (Kelly and Local 1989, 72-73).

They propose notations for seven degrees of resonance for consonants (although they regard resonance as a continuum): palatalised, clear, half-clear, central, half-dark, dark and velarised. Furthermore, resonance categories for consonants need not be static as they may change while a primary articulation is maintained (Kelly and Local 1989, 73-4).

Kelly and Local (1986) report on a study conducted on five speakers of non-standard English dialects. These speakers were asked to produce around 200 utterances containing /l/ or /r/ sounds in various phonological contexts. Kelly and Local used spectrographic data, in particular F_2 frequency, to verify impressionistic observations about clear and dark resonance (the lower F_2, the darker the sound). They found that each of the speakers (who came from Epsom, Stockport, Salford, Haltwhistle and Cullercoats) had two liquids differing in resonance category: one dark and one clear.

An interesting result of this study was the discovery that for all speakers the domain of resonance was usually considerably larger than the syllable, so that neighbouring vowels (and perhaps consonants) have different qualities depending on the resonance domain into which they fall. They call this a long-domain effect, and report that the domain of resonance is all subsequent non-accented syllables, i.e. the phonological foot. Preceding unaccented syllables also showed resonance effects, but a preceding accented syllable is not part of the domain, for all dialects studied. Although they considered only front vowels in the study, they report an informal finding that back vowels show the same resonance effects. They report that accents appear to differ in the resonances produced when two liquids occur in the same resonance domain.

Kelly and Local suggest that changes in resonance value mean different things for different segments phonetically: fronter or backer articulation for dorsals and for vowels, different secondary articulations and fronter or backer place of approximation for apicals (Kelly and Local 1986). They argue that these changes should be viewed as the interaction of values of a resonance parameter on other non-resonance parameters in speech. The point is elaborated by Kelly:

> Almost invariably the phonetic phenomena associated with the presence of one or other resonance category have extent within the speech complex of a kind that overlaps changes of state on other, non-resonance, strands of this complex. They are best viewed, then, as parametrically associated with these (Kelly 1989, 57).

In this 1989 paper Kelly reports on an electropalatographic investigation, designed to complement the earlier spectrographic one (Kelly and Local 1986). Again instrumental techniques are used to supplement impressionistic judgements. The results given are preliminary and based on EPG data from one subject: the author. Kelly proposes a theory of "acme articulation" in which the phonetic item acting as acme articulation has its typical resonance value, which it shares with neighbouring sounds (i.e. domination of one category over another). EPG data shows that velar contact is fronter in the utterance 'Barry came to my mind' than in 'Ballet came to my mind' which is in keeping with the fact that the author's dialect has clear /r/ and dark /l/. However data for geminate consonants appears quite complex, and Kelly notes that an investigation into the rules governing acme functions is necessary.

Kelly suggests that long-domain resonance plays an important role in speech perception and is useful in lexical identification, although he does not provide experimental verification.

> If resonance effects of the kind described here work in the way we suspect they do they are clearly of great importance as carriers of information about such things as syllable groupings within and across structural boundaries at other levels, accentual prominence, and rhythm. They may accordingly play a role for listeners in establishing syntactic and lexical identifications. (Kelly 1989, 59).

Resonance patterns associated with inter-vocalic consonants differ with (English) dialect: dialects from the North West Midlands have clear /r/ and dark /l/, Tyneside and the home counties accents have dark /r/ and clear /l/ (Kelly and Local 1989, 212-214). Each of the dialects that Kelly and Local studied has one dark and one clear liquid in this position. Kelly and Local suggest that resonance difference might well be one of the features that is involved in accent identification and that it plays an important role in the phonology of English dialects (Kelly and Local 1989, 215).

Kelly (1995) discusses /l/ and /r/ resonances in three varieties of English (Stockport, Haltwhistle and Epsom), based on data from three speakers. He replicates the finding that /l/ and /r/ resonance patterns depend on dialect, and finds further that the extent of resonance may differ from dialect to dialect. The findings are based on "impressionistic observations", and Kelly suggests that they should be complemented by instrumental studies. However, he concludes that resonance is "a variable component of sound-complexes that is both essentially independent of its co-variables and wide-ranging in its values" (Kelly 1995, 347).

In a paper on Tyneside speech, Local (1990) demonstrates the phonological importance of resonance. He shows that apparently free vowel variation is explicable if resonance patterns, rhythmic-quantity characteristics and voicing properties are taken into account. The vowel qualities under consideration are those in the final syllable of words such as *city* and *geordie*. When preceding gestures have clear resonance, these vowels have closer, more peripheral qualities. More retracted or central qualities occur after gestures with dark resonance. The quality of this vowel, when prepausal, can be fairly accurately predicted if resonance is amongst the environmental features considered.

Although Kelly and Local use the 'resonance' concept to make fairly radical claims about phonology, the concept is not a new one and its development can be traced in the literature. Jones' use of the term in relation to clearness and darkness of English /l/ has been described above. Delattre writes (1965, 108) that tongue shape and point of articulation for apical consonants play a large role in creating the auditory impression of a language, as they contribute to the degree of "frontal resonance" of the language.

Work on other consonants also demonstrates resonance effects. Scott (1984) reports that members of American English word pairs such as *coating* and *coding* when produced with alveolar flaps (and thus no voicing contrast) are correctly distinguished by listeners. She demonstrates that there is a significant difference in vowel quality between the /t/ and /d/ words, not a significant vowel length difference (although total word length difference is significant), and suggests that this is the basis upon which listeners discriminate. Kwong and Stevens (1999) discuss the pair *writer, rider*, and demonstrate that cues to the voicing contrast are found in vowel quality.

The implication of Kelly and Local's results is that the traditional segmental description of speech is not accurate and may be misleading, and they suggest that a different model of speech be used: one in which slices are taken horizontally across stretches of speech rather than vertically along the time axis. They argue for a model of speech without segments as primitives. Kelly and Local's work on resonance has important implications for theories of the organisation of speech sounds. Their suggestions are largely the result of extrapolation from impressionistic observations, which can and should be tested experimentally, both on natural and synthetic speech.

1.3.2. Perception of long-domain coarticulatory effects

The acoustic nature and perceptual relevance of resonance effects have been studied in some detail by Hawkins and colleagues (Hawkins and Slater 1994; Nguyen and Hawkins 1999; Tunley 1999). This work forms part of a larger research program aimed at creating natural and robust synthetic speech (Hawkins *et al.* 1998; Heid and Hawkins 1999), based on the premise that synthetic speech will be more intelligible if it includes the systematic variation that is found in natural speech. Hawkins has argued that the fragility of synthetic speech is due to its unnatural quality (Hawkins 1995). Modelling systematic variation makes synthetic speech acoustically more coherent and thus perceptually more coherent, improving word recognition in noisy conditions (Hawkins and Slater 1994; Heid and Hawkins 1999; Tunley 1999). The argument has empirical support: Hawkins and Slater show that modelling consonantal resonance effects on vowels in synthetic speech improves intelligibility: including resonance effects on vowels caused by /z/ (palaticity) and /r/ (rhoticity) improved phonemic intelligibility in noise by about 15% (Hawkins and Slater 1994). They attribute this improvement to the increased naturalness of the synthetic speech, due to inclusion of variation that is typically unknown and therefore not usually included in synthetic speech.

In a detailed study of the coarticulatory influences of liquids on vowels in English, Tunley found that incorporating coarticulatory detail in synthetic speech over /ərVCə/ sequences improved segmental intelligibility in noise by 7-28% (Tunley 1999). However, results with a similar sequence containing /l/ instead of /r/ were quite different: incorporating the coarticulatory detail did not improve intelligibility. Tunley hypothesises that this is largely the result of the Infovox synthesis system, which has a highly intelligible /l/ based on American English formant frequencies. Her attempts to model a British English /l/ were not satisfactory, and thus the changes made did not in fact make the synthetic speech more natural. Her acoustic study showed that the coarticulatory effects of /l/ on adjacent vowels are less systematic than those of /r/, which led her to focus on the temporal course of rhotic, and not lateral, effects. She found that non-adjacent unstressed vowels showed /r/-induced lowering of F_2 and F_3 when compared with vowels in an /h/ context, but stressed vowels did not. These findings were not used by Tunley to construct perceptual tests, and this work remains to be done.

Hawkins and Nguyen examined /l/s in syllable onset position and showed that they are slightly longer and darker (lower F_2) in syllables with voiced as opposed to voiceless codas (Hawkins and Nguyen forthcoming). They show that cues to the phonological voicing contrast are distributed across a whole syllable. Interestingly, some of these cues are confined to the /l/, and were not extensions of coarticulatory cues in the following vowel. Spectral centre of gravity was significantly lower in frequency before a voiced than a voiceless coda at the midpoint of the /l/, but not at its offset or in the initial part of the following vowel. These coarticulatory cues contribute to word recognition, as they demonstrate with a speeded lexical-decision task using cross-spliced /l/s in /(C)lVC/ syllables which differed in coda voicing.

Listeners' reaction times were longer to cross-spliced /l/s with greater acoustic changes from a voiced context placed in a syllable with a voiceless coda (Hawkins and Nguyen forthcoming; Nguyen and Hawkins 1999). This demonstrates that tampering with fine acoustic detail in non-adjacent phonetic segments can disrupt lexical access.

The work of Hawkins, Slater, Nguyen and Tunley suggests that so-called redundancy or coarticulatory detail in speech fulfils a useful function, particularly in difficult perceptual conditions. Acoustic cues for perception might well differ in importance depending on perceptual conditions; redundancy might be extremely useful, if not essential, for accurate speech perception and not merely a result of articulatory organisation. Hawkins and Nguyen take this further and support a model in which words are recognised in a non-segmental manner by mapping a fine-grained auditory representation of the acoustic input directly onto the lexicon (Hawkins 1995; Nguyen and Hawkins 1999), following Klatt's LAFS model (Klatt 1979). In this model long-domain coarticulatory information is central to word recognition, as it makes speech more resistant to background noise, allows faster word recognition and may help listeners to backtrack and correct identification errors. Whether the details of the model are correct or not, the research it has stimulated shows that long-distance coarticulatory cues do play a role in speech perception.

1.4. Coarticulation

1.4.1. Studies of coarticulation

Resonance effects might be described as a process of coarticulation, in which the resonance quality of the liquid influences other segments. Coarticulation has been extensively studied (*e.g.* Bell-Berti and Harris 1979; Delattre *et al.* 1955; Fowler 1980; Fowler 1981; Hawkins and Slater 1994; Lubker and Gay 1982; Magen 1984; Magen 1997; Manuel 1990; Moll and Daniloff 1971; Öhman 1966; Recasens 1989). The effects of coarticulation have been shown to be extensive. Coarticulatory effects of vowels on other vowels (shifted F_1 or F_2) may be seen across consonants (Öhman 1966) across syllables (Recasens 1989) and across foot boundaries (Magen 1997). A consonant may affect vowel quality in non-adjacent syllables as in the case of "r-lowering" of F_2 (Hawkins and Slater 1994). A summary of experimental studies of coarticulation can be found in Hardcastle and Hewlett (1999).

A striking result which emerges from some of the above studies (*e.g.* Bell-Berti and Harris 1979; Lubker and Gay 1982; Magen 1997; Recasens 1989) is the existence of considerable inter-speaker variability in coarticulatory patterns. Ellis and Hardcastle (2000) report considerable differences in coarticulatory behaviour for speakers producing assimilated, partially assimilated and non-assimilated /nk/ sequences. Nolan (1985) points out that the experimental study of coarticulation has produced no straightforward generalisations over different dimensions of coarticulation (lip rounding, nasalisation, *etc.*). He suggests that part of the reason for this untidiness may be the idiosyncrasy of coarticulatory strategies: speakers may be free to choose from a range of timing options, resulting in different patterns of coarticulation. Nolan found that extent of coarticulation of an /l/ with a following vowel varied with individuals. The implication for /l/ and /r/ resonances is clear, one should expect individual variation in the extent of resonance.

Rosner and Pickering (1994) discuss coarticulation effects on vowel production and perception. They note that both adjacent and more distant segments can influence the articulation of a vowel, but that distant coarticulatory effects are much weaker than immediate ones. Immediate context can change F_2 formant frequencies, but there is little effect on F_1. The effect of immediate context on production is strongest along the front-back dimension of place of articulation. Distant coarticulatory effects are weaker, and relatively few studies have been concerned with effects on production. In the context of resonance in English, it is important to note that it appears to be changes in F_2 which reflect clear and dark resonance, in keeping with Rosner and Pickering's findings.

Tunley's study of the coarticulatory influence of liquids on vowels in English (Tunley 1999) considers the influence of liquids on following vowels and the role of metrical influence (stress and foot size) on liquid coarticulation. Using /h/ as a standard of comparison, she found that /l/ exerted very little influence on following vowels (although there was some F_2 lowering in a following /ɪ/ and /ɛ/). /r/ induced strong F_2 and F_3 lowering in following vowels, but the amount of lowering varied quite considerably with vowel quality: /ɪ/ shows more lowering/colouring than /i/ and /ɛ/ which show more colouring than /æ/. Mean differences between the /h/ and /r/ context range from 19 Hz to 390 Hz for F_2 and 135 Hz to 425 Hz for F_3. Tunley goes on to compare /rV/ and /lV/ sequences in different metrical structures, dropping the comparison with /hV/ as differences between /l/ and /r/ contexts had been established. She finds that temporal and spectral properties are clearly related. However stressed syllables show no spectral variation related to foot length even though they undergo temporal compression in longer feet, whereas unstressed syllables show lower F_3 and F_4 values in longer feet (3 to 5 syllables, compared to 2-syllable feet). Tunley attributes these findings to active control of articulatory undershoot, with speakers exhibiting more undershoot in unstressed syllables than stressed syllables.

Tunley (1999) also investigated the extent of rhotic resonance effects over /V_1CV_2r/ and /rV_3CV_4/ sequences, compared with /V_1CV_2h/ and /hV_3CV_4/ sequences respectively. Stimuli sentences containing the sequences of interest were created, and three speakers were recorded producing five repetitions of each sentence. Factors in her analysis were consonant type (/r/ or /h/), stress of the syllable containing the consonant, direction of influence (anticipatory or perseverative) and vowel quality. Tunley measured F_2, F_3 and F_4, but found that difficulty in measuring F_4 led to large variation and no statistically significant results involving F_4. Rhotic resonances found in this experiment were limited: there were no significant differences in formant frequency in V_1 and V_4 when they were stressed. Small, but significant, F_2 and F_3 lowering in V_1 and V_4 (by 13 and 44 Hz respectively) were found when these vowels were unstressed. However it appears that the lowering was limited to examples containing the vowel /i/ and, surprisingly, not found when the vowel of interest was /ə/. There was no difference between perseverative and anticipatory effects, direction of influence was not a significant factor. Tunley also found no difference in coarticulatory influence dependent on the stress of the syllable containing the /r/ or /h/. Tunley's results suggest that long-domain resonances do exist, but point to the need for further detailed experimentation.

In a recent study following on from Tunley's work, Heid and Hawkins (2000) explore the anticipatory effects of /l/ and /r/. Examining 994 utterances produced by a male speaker of southern British English, they find resonance effects (lowered F_3 and F_4 in the /r/ context relative to the /l/ context) up to five syllables before the liquid. The effects are quite subtle, and interact with stress and segment type. Four carrier phrases with two stress patterns were designed for the l/r pairs of interest: 'We heard it might be —', 'We heard it could be —', 'We heard it might be a(n) —' and 'We heard it could be a(n) —'. In the first stress pattern *heard*, *might* (or *could*) and the l/r word were stressed, in the second pattern only *heard* and the l/r word were stressed. The l/r words differed in number of syllables and position of stress. The main findings of the study were resonance effects in *be* and *it*, independent of the metrical stress and phonological weight of the conditioning syllable. No resonance effects were found in *could*, and the effects for *might* are not reported in detail, but appear to consist mainly of F_3 lowering. The size of effects in *be* differed with segmental context and formant under consideration, although effects were all in the same direction: formant lowering in the /r/ context. Resonance effects in *it* were found only when the preceding *might* or *could* was stressed, thus the effects may spread across stressed vowels. The effects decreased as the number of syllables between the conditioning (l/r) syllable and the target vowel increased, with resonance effects in *be* no longer significant when more than one syllable intervened. Effects in *it* did not interact with the number of intervening syllables and were significant when as many as four syllables intervened between *it* and the l/r syllable. Heid and Hawkins have demonstrated that l/r resonances exist in the anticipatory domain, and that the effects are complex and interact with metrical structure and segment type.

1.4.2. Theoretical proposals

There are several different theoretical accounts of coarticulation; some of the more recent theories will be discussed here. A more comprehensive survey can be found in Farnetani and Recasens (1999).

Bladon and Al-Bamerni (1976) developed a theory of coarticulation to account for data from an experimental investigation of RP /l/. They introduce the notion of coarticulation resistance, which refers to the amount of coarticulation which a segment may exhibit, in other words how greatly it is influenced by sounds adjacent to it (measured in terms of second formant frequencies). They study RP /l/, concluding that clear /l/s show the least coarticulatory resistance (to surrounding vowels), and syllabic dark /l/s the most. Pairs of adjacent laterals (separated by word boundaries) show stronger coarticulatory effects from the leftmost to the rightmost lateral. This data is interesting in the face of Kelly and Local's claims about resonance domains: it suggests that the darker lateral extends its influence to a following lateral, although the effects of stress are not considered. The authors suggest that degree of coarticulatory resistance is language specific. One might argue that what underlies this, in the case of English /l/ at least, is the domain of resonance. Further support for this idea is found in Recasens *et al.* (1996) where acoustic data shows that German clear /l/ does not undergo more coarticulation than velarised Catalan /l/, despite the greater articulatory constraints on the velarised /l/. They interpret this as meaning that coarticulatory resistance does not depend on articulation, but do not produce an alternative hypothesis.

In later work, Recasens and colleagues (Recasens and Pallarés 1999; Recasens *et al.* 1997) propose a model of lingual coarticulation called the degree of articulatory constraint (DAC) model. This model aims to predict how much coarticulation is allowed by a given segment and the extent to which it is likely to affect surrounding segments. The DAC scale ranks phonetic sound categories on the basis of tongue dorsum constraint from maximally to minimally constrained on a scale from 3 to 1, where dark /l/ has a DAC value of 3 and /ə/ a DAC value of 1. The DAC value of segments should predict the extent and degree of coarticulation found. Recasens *et al.* (1997) consider V_1CV_2 sequences, and predict that the salience of V_2-dependent anticipatory effects varies inversely with C-to-V_2 perseverative effects. Similarly, V_1-dependent perseverative effects should vary inversely with the C-to-V_1 anticipatory effects. Dark /l/, which has strong anticipatory tongue dorsum effects, blocks vocalic perseverative effects more than anticipatory ones. The model thus defines the relationship between the direction of vocalic and consonantal coarticulation over a VCV sequence.

Another approach to coarticulation, which has a long history (*e.g.* Henke 1966), refers to distinctive features. An overview of these theories can be found in Kühnert and Nolan (1999) and Farnetani and Recasens (1999). A recent model of coarticulation as feature spreading is Keating's "window" model (Keating 1990a), which is discussed further in section 6.4.3. Keating is concerned with modelling phonetic coarticulation, that is coarticulation that involves quantitative interactions in continuous time and space. In this model, for each articulatory dimension, each feature of a particular segment has a range of possible spatial values (a maximum and minimum) associated with it. The value for the segment (in any context) must fall within this window, and the window itself is independent of context. Coarticulation arises as the most efficient pathway (interpolation) through successive windows is found. Narrow windows reflect very little contextual variation. Wide windows reflect a large amount of contextual variation, and some degree of phonetic underspecification. Keating argues that phonetic underspecification is gradient, i.e. that segments can have varying degrees of specification, and that this gradience is captured well by the window model which allows windows of different sizes.

A very different view of coarticulation is proposed by Fowler and colleagues (*e.g.* Fowler and Saltzman 1993). This work is based on a task-dynamic model of speech production developed by Kelso and colleagues (Kelso *et al.* 1986; Kelso *et al.* 1983), in which each segment has associated articulatory control structures. These structures, or gestures, are thought to be consistently present and relatively stable with regard to segmental context (Boyce *et al.* 1990). Coarticulation reflects the temporally staggered activation of coordinative constraints for different phonetic gestures. In this theory, coarticulation is the result of low-level (below the level of the speech plan, i.e. not part of the speech plan) interaction of gestures due to the time course of an utterance. Fowler (1981) argues that coordinative structures are phonological segments and that coarticulation is the result of the coproduction of coordinative structures which have a temporal dimension; the interaction of the timing of these structures. Thus it is not the gestures of speech which are variable, but the relative timing of gestures.

Opposition to this view stems from the belief that coarticulatory strategies are driven by competing constraints which are both kinematic and acoustic in nature (Perkell 1991; Perkell and Matthies 1992; Perkell *et al.* 1993; Perkell *et al.* 1997). Using articulatory and acoustic data, Perkell and colleagues find evidence of "motor-equivalent" goals, i.e. acoustic goals that are reached using more than one articulatory strategy, or trading relations between articulators (Guenther *et al.* 1999; Perkell *et al.* 1993). There is considerable variation between subjects, with different subjects using different strategies. Perkell *et al.* (1997) also show that auditory feedback plays a role in adult speech production, suggesting that it is used to maintain the parameters of an internal model of the relation between motor commands and sound output, and to monitor acoustic transmission conditions. In this framework, coarticulation cannot be caused solely by variation in timing of intrinsically stable gestures; acoustic-perceptual goals must also play a role.

Whalen (1990) also argues that coarticulation performs a useful role in speech perception. He provides evidence that talkers structure their utterances to include coarticulation, hypothesising that this is because coarticulation serves a communicative function. He argues that anticipatory coarticulation requires planning and is not the result of automatic processes. He reports an experiment in which speakers were required to begin reading a nonsense utterance before they had seen the whole string to be produced. Segments known before articulation began exerted normal anticipatory coarticulatory influence, those only seen after the onset of speaking did not.

None of the theoretical proposals discussed here seriously anticipates the existence of long-distance coarticulation, and thus they do not account for long-distance coarticulatory effects. The main focus of debate is the nature of the underlying cognitive units of speech (features or gestures) (Farnetani and Recasens 1999). The focus of research on the extent of coarticulatory effects has been on anticipatory coarticulation. The theories of coarticulation discussed above can be divided into three types: time-locked models in which coarticulation is based on gestures and occurs over fixed time stretches; look-ahead models in which coarticulatory effects spread to preceding segments which are unspecified for the feature in question; and hybrid models in which coarticulatory strategies include a time-locked and a look-ahead component (Perkell 1990, Farnetani and Recasens 1999). Heid and Hawkins (2000), who report long-distance coarticulatory effects, suggest that these effects consist of two components: a short-range component with a big acoustic effect that interacts with segmental context and a long-range component that is smaller and interacts less with segmental context. This proposal is similar to a suggestion by Hewlett and Shockey (1992) that coarticulation be divided into two areas of study: short-term effects and longer-term settings. These proposals resemble the hybrid model of coarticulation, but are not worked out in any detail and are descriptive rather than explanatory.

In their survey of models of coarticulation, Farnetani and Recasens (1999) conclude that existing models succeed only partially in accounting for speakers' behaviour, and that the topic is open for future research. The results of this thesis add to our knowledge of coarticulatory effects. The implications of these results for theories of coarticulation will be discussed in chapter 6, where it will be argued that existing theories of coarticulation require extensive modification in order to account for these effects.

1.5. Scope of this study

This thesis demonstrates the extent of the coarticulatory effects of English liquids and their perceptual relevance. It begins with two experiments investigating the perceptual relevance and relative salience of coarticulatory cues associated with /l/ and /r/. Although two dialects (Manchester English, denoted by ME, and standard southern British English, labelled RP) were initially investigated, the remainder of the thesis is restricted to standard southern British English, as the perceptual results of chapter two were strongest for this group. The perceptual results show that listeners can identify an /l/ or an /r/ even when that segment and parts of adjacent vowels are missing. This raised questions about the extent and direction of coarticulatory cues, which were examined in an articulatory experiment using simultaneous electromagnetic articulography (EMA) and electropalatography (EPG). Chapter 3 describes the experimental method and the strong local coarticulatory effects, both in articulation and acoustics, which were found. Chapter 4 describes the long-distance coarticulatory effects, which are shown to extend up to two syllables remote from the liquid and to involve both the lips and the tongue body. In chapter 5 the articulatory data is modelled, using a model of gestures as damped second order differential equations. The modelling shows that the long-distance coarticulatory effects cannot be adequately accounted for by gestural overlap or coproduction. In chapter 6, the findings of the thesis are summarised and directions for further work suggested. The coarticulatory effects demonstrated in this thesis pose a challenge to current phonetic and phonological theories, which cannot account for their extent. Chapter 6 contains a discussion of these challenges and the directions in which current theories might have to change to account for the data presented in this thesis.

CHAPTER TWO

THE PERCEPTUAL RELEVANCE OF LIQUID RESONANCES

2.1. Introduction

The relevance of long-distance coarticulatory cues in speech perception has been demonstrated by Hawkins and colleagues (Hawkins and Nguyen forthcoming; Hawkins and Slater 1994; Nguyen and Hawkins 1999; Tunley 1999). Their findings, which were discussed in chapter one, focus on the role which long-distance coarticulatory cues play in the intelligibility of synthetic speech and in word recognition. Their work on liquid resonances is restricted to synthetic speech: they do not test whether resonance cues can be used to identify liquids in natural speech. This chapter describes two perceptual experiments which investigate the salience of resonance cues in natural speech, demonstrating that they are perceptually available to listeners, and that in some cases listeners can use these cues to identify a liquid which has been deleted and replaced by noise.

2.1.1. Phoneme restoration

Kelly and Local (1986) suggest that the long domain of resonance could account for a speech perception effect known as phonemic restoration, an "illusory effect ... in which listeners restore masked segments of speech on the basis of context" (Warren 1996).

> When a speech sound in a sentence is replaced completely by an extraneous sound (such as a cough or tone), the listener restores the missing sound on the bases of both prior and subsequent context. This illusory effect, called phonemic restoration (PhR), causes the physically abstract phoneme to seem as real as the speech sounds which are present. (Warren and Obusek 1971, 358)

A survey of techniques used to explore the phoneme restoration effect and the results obtained from the experimental paradigm can be found in Samuel (1996b).

One of the principal assumptions of cognitive psychology is that humans routinely perform complicated processing without any awareness of the tasks they perform

(Goldinger *et al.* 1996). Thus although listeners perceive a missing phoneme as present, this may be an illusory effect based on their processing of the audible signal.

Warren and Obusek (1971) report experiments designed to explore the phonemic restoration effect. They are aware of the importance of consonant-vowel transitional cues for speech perception and account for these in their experiments by removing transitional zones around sections of speech (*i.e.* consonants) that they delete. In Warren and Sherman (1974) the excised phoneme was deliberately mispronounced in order to remove acoustic cues, where 'mispronunciation' consisted of the replacement of /b/ with /s/ or /v/, /v/ with /b/, /n/ with /t/, /m/ with /d/ and /s/ with /f/. The authors designed their studies specifically to eliminate obvious acoustic-phonetic cues that they were aware of. Phoneme restoration is a puzzling phenomenon if a purely segmental theory of speech is assumed: one needs to explain why speech perception occurs despite absence of segments caused by noise in and deletion of the speech signal.

Warren and his collaborators' work argues that phonemic restoration is a process of auditory induction, in which the restored sound is extracted or induced from noise in the acoustic signal. An alternative view is that restoration is in fact an auditory illusion accompanying the completion of some abstract phonological representation. Repp (1992) conducted a series of experiments to test these competing views. He required subjects to compare the timbre of a target noise that replaced a fricative in a word with the same noise preceding or following the word. He reasoned that if phonemic restoration involved reallocation of sections of the noise, there would be a noticeable change in timbre between the noise in the word and that outside it. Out of five experiments, the last three reported showed no significant change in timbre of the noise. Repp interpreted these results as evidence that phonemic restoration is illusory and does not interact with auditory processing.

Repp comments that "accepting phonemes as abstract linguistic units without specific auditory properties may go a long way toward clarifying the intriguing phenomenon of phoneme restoration" (Repp 1992). He argues that restoration of phones (as opposed to phonemes) may not actually occur, because perception of auditory segments is not a prerequisite for phoneme perception. His view of phoneme restoration is that it is not a purely auditory process but is a result of higher-level linguistic processing.

Samuel (1981 and 1996a) approaches phoneme restoration from a slightly different perspective. He considers the role of the acoustic signal in phoneme restoration and argues that this is significant. His experiments show that the phonemic restoration effect is considerably stronger when the phone and the sound replacing it are similar: when white noise replaces a sound fricatives are restored better than vowels; the opposite obtains when a pure tone is used. Samuel points out that phoneme restoration is the result of the interplay of a listener's knowledge and the acoustic signal; if there is confirmation in the acoustic signal that the expected sound is present then the restoration effect is strong, or speech is generated.

The problem of phoneme restoration is considerably diminished, however, if Kelly and Local's view regarding long-domain cues is correct. According to them, cues to perception occur throughout an utterance and small stretches of deleted or distorted signal do not have the same significance that they have in a purely segmental model. Listeners may actually find the necessary information present in the spectrum (despite removal of immediate transitions), as excision of immediate context does not necessarily remove or even minimise ancillary information (Kelly and Local 1986). Thus it might not be necessary to invoke linguistic processes (alone) as explanation of the phoneme restoration effect.

2.2. Experiment One

The first experiment was designed to investigate whether long-domain resonance distinctions remote from an /l/ or /r/ could be used to identify the /l/ or /r/. Two English dialects (Received Pronunciation (RP) and the accent used in Manchester (ME)) were chosen as they have contrasting resonance patterns for /l/ and /r/. In syllable-initial (and inter-vocalic) position RP has clear /l/ and dark /r/, whereas ME has dark /l/ and clear /r/ (Kelly and Local 1986). The choice of two dialects was intended to determine whether the domain of resonance differs between dialects with contrasting resonance patterns, as Kelly and Local suggest. An experiment using progressive replacement of speech with noise was designed and carried out with the aim of establishing how much linguistic material needs to be present for listeners to discriminate correctly between minimal pairs differing (phonologically) only in /l/ *vs.* /r/.

2.2.1. Experimental design

The experiment was divided into two parts, one conducted in Oxford using RP stimuli and subjects, the other in Manchester using the local Mancunian dialect and subjects from the area. Both parts of the experiment had identical design and each was run on six subjects. Subjects listened to tape recordings made from a speaker of the appropriate dialect, and responded in a booklet of 30 answer sheets, identifying one of a choice of two words (*e.g. belly vs. berry*) as the stimulus they had heard. Percentages of correct identification were calculated from the resulting answers, to determine how accurately subjects were able to identify words in which the liquids had been replaced by noise.

The experiment is an extension of Warren's technique for investigating phoneme restoration (Warren and Obusek 1971). Stimuli with a certain extent of noise replacing relevant sections of the utterance were presented to subjects, who were required to perform a two-category decision task. The technique provides valuable information about speech perception by demonstrating whether listeners are able to make use of cues present in the audible material for speech perception, and thus establishes the presence of information in the signal even when the acoustic properties are not known.

2.2.1.1. Stimuli

The stimuli were prepared from a list of thirty six minimal pairs containing either /l/ or /r/ in inter-vocalic position. The words were all di- or tri-syllabic and were obtained from a machine readable dictionary (Mitton 1992), supplemented by words from Kelly and Local (1986; 1989). Each word was recorded in the frame sentence 'No, I utter(ed) — today'. The form *utter* was used before words beginning with a consonant. The past tense form *uttered* was used before words beginning with a vowel, to prevent occurrence of linking /r/. In order to ensure a fixed pattern of sentence intonation, a script was produced for the recordings and the speaker was given a numbered list of the sentences to be read. Each sentence was prompted by a question read by the experimenter: 'Did you utter Y today?', where Y was a word not containing an /r/ or an /l/ and not in the word list. The list of experimental stimuli is given in Table 2.1. Stimuli with consonant clusters in the first consonantal position after the liquid (*i.e.* 3/4, 7/8, 15/16, 21/22, 25/26, 29/30, 35/36, 49/50, 71/72) were not included in the data analysis, as they formed too small a subset for separate statistical analysis.

Table 2.1: Stimuli used in the first perception experiment.

1. No, I uttered alight today.	2. No, I uttered aright today.
3. No, I uttered alive today.	4. No, I uttered arrive today.
5. No, I uttered allayed today.	6. No, I uttered arrayed today.
7. No, I uttered allows today.	8. No, I uttered arouse today.
9. No, I uttered aloe today.	10. No, I uttered arrow today.
11. No, I utter Bally today.	12. No, I utter Barry today.
13. No, I utter belated today.	14. No, I utter berated today.
15. No, I utter believed today.	16. No, I utter bereaved today.
17. No, I utter belly today.	18. No, I utter berry today.
19. No, I utter bestially today.	20. No, I utter bestiary today.
21. No, I utter collected today.	22. No, I utter corrected today.
23. No, I utter dearly today.	24. No, I utter deary today.
25. No, I uttered elected today.	26. No, I uttered erected today.
27. No, I uttered Ella today.	28. No, I uttered error today.
29. No, I uttered Ellen today.	30. No, I uttered Erin today.
31. No, I utter fairly today.	32. No, I utter fairy today.
33. No, I utter fallow today.	34. No, I utter farrow today.
35. No, I utter filing today.	36. No, I utter firing today.
37. No, I utter finally today.	38. No, I utter finery today.
39. No, I utter Halley today.	40. No, I utter Harry today.
41. No, I utter hallowed today.	42. No, I utter harrowed today.
43. No, I utter jelly today.	44. No, I utter Jerry today.
45. No, I utter malicious today.	46. No, I utter Mauritius today.
47. No, I utter mallow today.	48. No, I utter marrow today.
49. No, I utter mandarin today.	50. No, I utter mandolin today.
51. No, I utter masterly today.	52. No, I utter mastery today.
53. No, I utter molasses today.	54. No, I utter morasses today.
55. No, I uttered odourless today.	56. No, I uttered odorous today.
57. No, I uttered ovary today.	58. No, I uttered overly today.
59. No, I utter pilot today.	60. No, I utter pirate today.
61. No, I utter queerly today.	62. No, I utter query today.
63. No, I utter Selina today.	64. No, I utter Serena today.
65. No, I utter Shelly today.	66. No, I utter sherry today.
67. No, I utter teller today.	68. No, I utter terror today.
69. No, I utter telly today.	70. No, I utter Terry today.
71. No, I utter tiling today.	72. No, I utter tiring today.

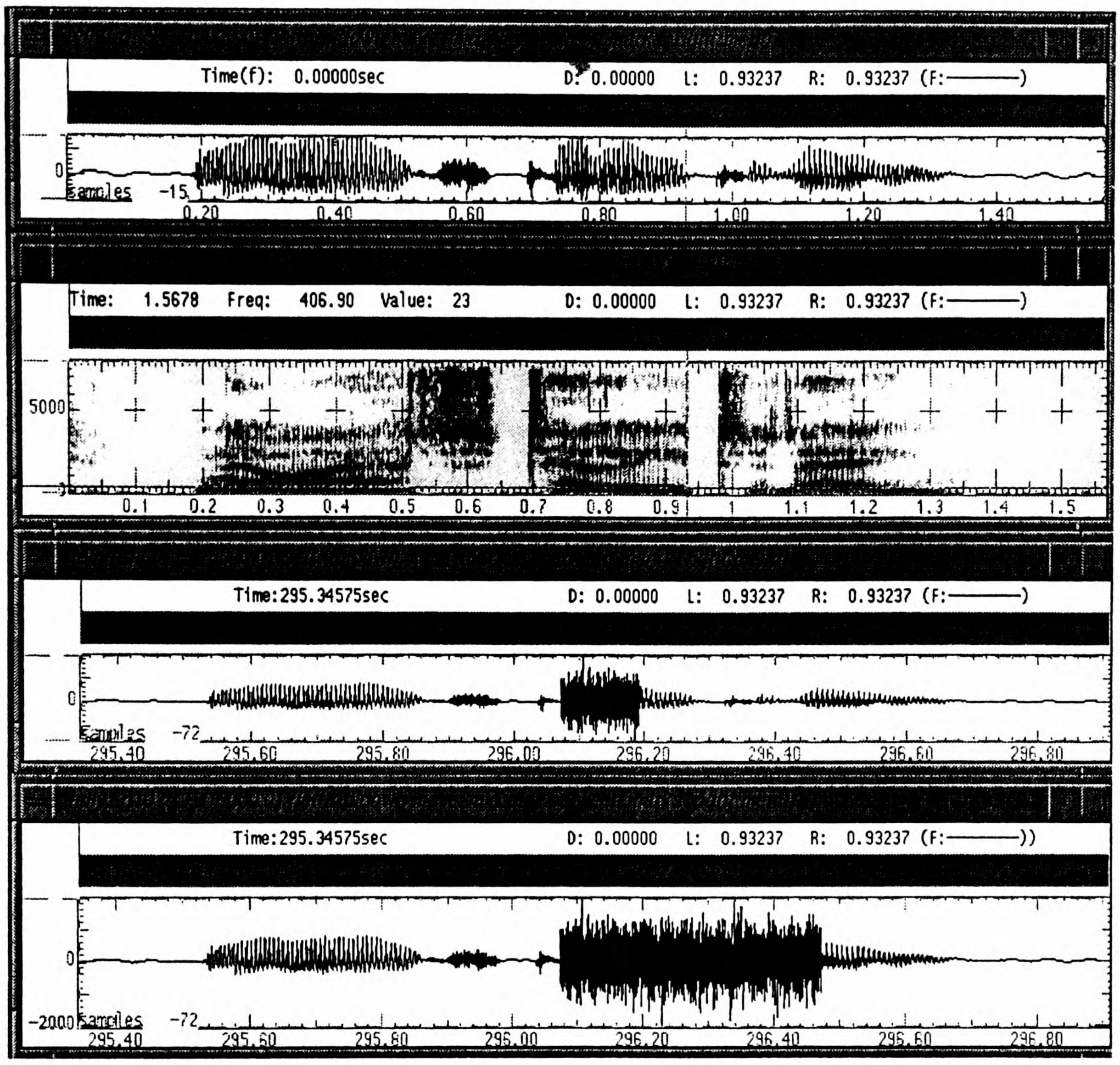
The RP data were recorded at the Oxford University Phonetics Laboratory in a sound insulated booth using a Sony DTC 1000ES PRO DAT recorder; the ME data in a quiet office at Manchester University, using a portable Sony TCD-D10 DAT recorder. An AKG C451E microphone and AKG B46E pre-amplifier were used. The data were then digitally transferred to disk (using a Silicon Graphics DAT drive) and processed using *Waves* speech processing software (Entropic Research Laboratory Inc., Washington DC) on a Silicon Graphics Indy computer (sampling frequency 16 000 Hz). In order to ensure that the stimuli used in the two parts of the experiment were comparable the average signal to noise ratio for each of the two sets of data was estimated and the ME stimuli were rescaled to a level similar to the RP recordings.

For both sets of data, each of the 72 recorded sentences was used to produce ten stimuli: sentences with a period of white noise replacing a stretch of (deleted) linguistic material. The first sentence had 40 milliseconds of the sentence replaced, subsequent sentences had an additional 40 milliseconds of noise placed immediately after the previous section of noise, thus the tenth sentence had 400 milliseconds of material replaced by noise.

The position of the noise in each sentence was fixed relative to a marker in each sound file. The markers were set by hand for each file, at a position easy to measure across sentences: the end of the vowel following the liquid. Where the vowel was followed by a voiceless consonant, the *Waves* estimate of the probability of voicing was used to determine the boundary between the vowel and the following voiceless sound. Wide-band spectrograms were used to identify the end of the vowel when it was followed by a voiced sound. These were all nasals or voiced fricatives, and were easily identified by characteristic acoustic patterns (low intensity formants for nasals; high frequency noise for fricatives). Figure 2.1 contains a waveform with reference point, displayed with wide-band spectrogram and two of the stimuli.

Once a reference point had been established for each sentence, the noise segments were generated and inserted automatically. Each began 200 ms before the reference mark and the longest ended 200 ms after it. Impressionistic perceptual assessment of the stimuli led to the expectation that the effects of /r/ and /l/ would be greater in the perseverative direction, despite earlier findings (Kelly and Local 1986; Magen 1997) that anticipatory effects were greater. Thus the bulk of the material replaced occurred after the liquid. The automatic generation of noise segments of equal size meant that in some cases the liquid was not replaced by noise; the data from these stimuli are not discussed, as identification was unsurprisingly almost 100% correct in these cases.

Figure 2.1: Waveform of the RP utterance 'No, I utter telly today', aligned with displays used to create stimuli and two stimuli. All viewed and created in *Waves*. (a) Waveform, with marker position (a vertical dashed line) for automatic generation of noise segments set at the end of the vowel following /l/ (b) Wide-band spectrogram (c) Stimulus sentence with 120 ms of noise replacing part of the speech signal (d) Stimulus sentence with 400 ms of noise replacing part of the speech signal.



The two sets of stimuli were digitally transferred to DAT tapes. The tape for each experiment consisted of ten sentences with no noise segments, followed by five pseudo-randomly selected examples containing noise segments (different versions of the same stimulus were excluded) for the familiarization run, and finally the rest of the 715 stimuli, in random order, for the experiment proper. All subjects had the same order of presentation.

2.2.1.2. Subjects and procedure

There were twelve participants, six for each part of the experiment. All subjects were right-handed monolingual speakers of the relevant dialect, who reported no known speech or hearing defects. The RP subjects (three male, three female) were first or second year psychology students at Oxford aged between 18 and 19, from south-eastern England.

Participants for the second part of the experiment were obtained by advertising in a Manchester newspaper and in sixth form colleges in the area for recent school-leavers who were born and bred in Manchester. Three male and three female subjects were used, aged between 16 and 23. The RP stimuli were obtained from a 28 year old male from Berkshire, and the Manchester stimuli from a 17 year old male, native of Manchester. All subjects were paid for their time.

The experiment was run separately for each subject, and began with the experimenter reading the subject the instructions. Subjects were told that the aim of the experiment was to determine how people identify words in a noisy environment, that they would hear words which were partly covered with noise and that they would have to indicate what word they had heard, or make a guess if they were uncertain. They put on Sennheiser HD320 headphones and were played the ten unmodified utterances over the DAT player to accustom them to the recorded voice and equipment. This was followed by instructions for using the answer book provided. Subjects were told to tick the box corresponding to the word they thought they had heard, or to make a guess if they were uncertain. They were instructed to tick the box twice if they were sure of their choice, or once if they were uncertain or guessing. This data was collected as an index of subject confidence, but proved uninteresting and shall not be discussed below. Five test examples followed to which subjects were instructed to respond. They were then asked whether they understood the task and were comfortable with it. The rest of the tape was run, and subjects were given a break of a few minutes half-way through the forty minute task.

2.2.1.3. Method of analysis

We are interested in how much material – especially material distinct from the /l/ or /r/ – subjects need to hear in order to correctly identify a word as containing an /l/ as opposed to an /r/. Thus the stimuli were divided into groups according to the amount of material deleted and replaced by noise in each utterance (as observed on *Waves* wide-band spectrograms). This was done on a segmental basis, in terms of consonants, vowels and their distance from the liquid. Although labelling stretches of acoustic signal with linguistic categories is somewhat objectionable (Repp 1981), these labels represent a first pass at the linguistic interpretation of the results. The following criteria were used: voiced and voiceless alveolar plosives were identified as having closure, burst and aspiration intervals; vowels were identified by stretches of voiced periodic signal; palato-alveolar and alveolar fricatives by frication noise; alveolar nasals by periodic sections of low intensity with low F_1. Transitions into and out of liquids (defined as starting and ending with relatively steady-state regions of periodicity) were labelled as part of the liquid, as these would be expected to provide local cues to liquid identity and were thus not of primary interest in this study. Figure 2.1 (c) is an example of a stimulus sentence where the noise segment was classified as replacing the liquid and preceding vowel. In Figure 2.1 (d), the noise segment was classified as covering the liquid and following VCVC sequence.

To allow for a finer analysis the stimuli were further subdivided, depending on where in the acoustic signal the noise segment ends. The vowels following the formant transitions out of the liquid were divided into early, mid and late portions; consonants following the liquid were divided into early and late sections. This subdivision was done by measuring the amount of the segment audible (visible in a spectrogram) after the noise segment, using *Waves*. For stops (the majority of consonants), the 'early' part consists of the stop closure and the 'late' part of the burst and aspiration. For continuant consonants the noise segment is classed as ending early in a consonant if at least 25 ms of material (frication noise or periodicity) is audible, and late if otherwise. For vowels the early category requires 50 or more ms audible periodicity, mid is between 25 and 49 ms, and late if less than 25 ms is audible.

For each category of noise segment (classified according to linguistic material replaced), the percentage of correct l/r classification was calculated.

2.2.1.4. Acoustics

The stimuli used in this perception study proved difficult to perform acoustic analysis upon: only one token of each utterance was used in the experiment and neither speaker was available for further recordings. For the RP speaker in particular, the formants were difficult to track using the Waves formant program and difficult to observe on wide-band spectrograms, partly because the vowels of interest were often voiceless. Measurements were made using 18 pole Burg spectra (Rabiner and Schafer 1978) with a 50 ms Hanning window, checked manually against wide-band spectrograms and DFT spectra. One-way ANOVAs on F_1, F_2 and F_3 measurements taken at three points in the utterance - midpoints of the schwas of *uttered* and *today* and the initial section of the vowel [ɛ] preceding the liquid (chosen because it occurred most frequently in the stimulus list) -

showed a significantly lower F_3 for the vowel before an /r/ as compared to an /l/ ($F(1,11) = 7.57$, $p < 0.02$). Other differences were not statistically significant, which is unsurprising as multiple repetitions of the stimuli sentences are needed to estimate inherent versus meaningful variation. The ME data showed no statistically significant differences in formant frequencies of the same vowels in /l/ as compared to /r/ words.

To investigate the acoustics further, a small acoustic study was carried out using six RP speakers. Different stimulus words and a different carrier phrase were used, as many of the schwas in *today* in the perception experiment stimuli were voiceless and very short. In order to investigate the role of surrounding context, six monosyllabic word pairs were used: *lip/rip, light/right, lob/rob, lag/rag, lay/ray, look/rook.* Five repetitions of each were obtained in the frame sentence 'Have you uttered a — again?', recorded amidst other, distractor, sentences.

A repeated measures MANOVA on the first three formants of the schwas in *uttered* and *again* with liquid and word pair as factors showed significance for both liquid (Pillai's test = 0.217, $F(6,338)$, $p < 0.001$) and word pair (Pillai's test = 1.295, $F(30,1710)$, $p < 0.001$). There was no significant interaction between liquid and word pair.

Univariate ANOVAs for each subject showed that the non-local difference between liquids was most consistently due to differences in F_2 with some differences in F_1 of the schwa of *uttered.* F_2 was significantly higher in /r/ than in /l/ contexts for four out of the six speakers ($F(1,48) = 12.43, 8.96, 111.383, 5.14$; $p < 0.03$). This is an unexpected result: prevocalically /r/ is expected to have dark resonance and thus a lower F_2 than a clear /r/, which has a slightly lower F_2 than a clear /l/ (Tunley 1999). The schwa preceding an /r/ was therefore expected to show F_2 differences in the same direction, *i.e.* lower F_2. Nevertheless the distinction was significant for four speakers, and a fifth exhibited a non-significant difference in the same direction. For the sixth speaker liquid was not a significant factor at all. F_1 was also significantly different for two of those four speakers: higher in /r/ contexts, as expected ($F(1,48) = 10.51, 9.94$; $p < 0.002$). This suggests that the primary cue for long-domain resonance is F_2, with some speakers also producing F_1 differences. For the first vowel of *again* only two significant differences were found, (in F_2 and F_3 of different speakers), perhaps because of the following velar stop. The F_2 difference is in accordance with results reported by Rosner and Pickering

(1994), who discuss coarticulatory effects on vowel production and perception. They note that immediate context primarily affects a vowel's F_2. Distant effects, which are relatively less studied, are weaker than those of immediate context, but would be expected to be similar.

2.2.2. Results

The data were analysed in terms of the linguistic material completely replaced by noise. A noise segment which begins in the consonant preceding the liquid and ends in the first consonant after the liquid is labelled as replacing the sequence VlV or VrV. Similarly, noise segments beginning in the consonant preceding the liquid which end early in the following vowel are labelled Vl or Vr, those ending in the middle of the following vowel as VlV_m or VrV_m, and those ending late in the vowel as VlV_l or VrV_l. Noise segments ending late in a consonant (for stops these replace the period of closure and leave at least some of the burst audible) are denoted as replacing C_m for the relevant consonant.

2.2.2.1. Percentages of correct identification

Percentages of correct identification were calculated for /l/ and /r/ stimuli separately, for two categories of noise segments: those beginning in the consonant or vowel before the liquid. The full results are contained in Table 2.2.

Table 2.2: Results for the first perception experiment. RP and ME subjects' percent correct scores for /l/ and /r/ stimuli by material replaced by noise. Tables show percent correct scores for each subject for each category. Dark shaded cells correspond to scores of 65% and over, light shaded cells correspond to scores of 35% and below.

Table 2.2 (a) RP % correct, /l/ stimuli, vowel before the liquid audible

Subject	l	lV_m	lV_l	lV	lVC_m	lVC	lVCV	$lVCVC_m$	lVCVC
SG	100	100	100	77	73	50	75	80	57
CR	90	67	67	46	67	50	83	80	43
JC	100	78	100	62	80	75	75	60	71
LR	90	67	100	54	53	75	67	80	14
MI	90	67	100	54	73	75	67	100	29
SC	90	89	83	92	87	75	58	60	57

Table 2.2 (b) RP % correct, /l/ stimuli, vowel before the liquid not audible

Subject	Vl	VlV$_m$	VlV$_l$	VlV	VlVC$_m$	VlVC	VlVCV	VlVCVC$_m$	VlVCVC
SG	79	33	29	41	42	45	30	50	60
CR	79	58	71	29	42	73	40	50	27
JC	79	83	57	47	75	91	30	75	33
LR	93	33	71	65	75	55	20	50	47
MI	86	33	57	41	50	55	50	50	40
SC	86	58	57	65	58	64	50	50	53

Table 2.2 (c) RP % correct, /r/ stimuli, vowel before the liquid audible

Subject	r	rV$_m$	rV$_l$	rV	rVC$_m$	rVC	rVCV	rVCVC$_m$	rVCVC
SG	100	83	67	92	78	67	80	100	71
CR	78	100	67	62	67	83	100	80	71
JC	100	100	67	69	67	67	80	100	57
LR	100	83	50	92	83	67	80	80	43
MI	100	83	67	85	72	83	100	60	43
SC	100	100	50	92	72	83	100	80	71

Table 2.2 (d) RP % correct, /r/ stimuli, vowel before the liquid not audible

Subject	Vr	VrV$_m$	VrV$_l$	VrV	VrVC$_m$	VrVC	VrVCV	VrVCVC$_m$	VrVCVC
SG	100	86	75	83	88	71	69	83	70
CR	67	100	88	92	75	57	77	83	70
JC	100	86	88	75	69	71	77	100	70
LR	67	71	75	58	69	86	77	100	70
MI	100	100	100	75	88	71	62	83	60
SC	100	86	100	83	75	86	62	50	80

Table 2.2 (e) ME % correct, /l/ stimuli, vowel before the liquid audible

Subject	l	lV$_m$	lV$_l$	lV	lVC$_m$	lVC	lVCV	lVCVC$_m$	lVCVC
SW	88	70	75	78	77	50	18	33	67
LD	94	90	50	78	67	67	55	67	83
DW	100	100	88	94	83	92	73	100	83
BJ	88	70	88	72	77	50	73	67	100
SD	88	70	75	94	70	75	82	100	50
SN	100	80	50	72	77	83	36	100	50

Table 2.2 (f) ME % correct, /l/ stimuli, vowel before the liquid not audible

Subject	Vl	VlV_m	VlV_l	VlV	$VlVC_m$	VlVC	VlVCV	$VlVCVC_m$	VlVCVC
SW	100	57	0	58	55	25	25	83	33
LD	100	57	67	17	44	100	50	67	0
DW	100	43	0	75	67	50	50	33	100
BJ	100	57	0	67	55	50	100	50	33
SD	0	57	0	75	44	25	50	83	33
SN	100	86	100	67	44	25	50	67	0

Table 2.2 (g) ME % correct, /r/ stimuli, vowel before the liquid audible

Subject	r	rV_m	rV_l	rV	rVC_m	rVC	rVCV	$rVCVC_m$	rVCVC
SW	78	86	64	33	54	57	21	67	38
LD	67	86	55	56	73	57	50	67	38
DW	67	100	45	44	50	29	43	67	50
BJ	78	100	36	56	65	29	50	67	38
SD	67	86	64	67	50	71	50	67	50
SN	67	100	64	56	65	43	64	67	50

Table 2.2 (h) ME % correct, /r/ stimuli, vowel before the liquid not audible

Subject	Vr	VrV_m	VrV_l	VrV	$VrVC_m$	VrVC	VrVCV	$VrVCVC_m$	VrVCVC
SW	100	33	25	36	67	29	33	80	0
LD	100	67	25	64	47	57	56	80	33
DW	100	67	50	45	33	43	44	80	0
BJ	100	67	75	64	67	57	56	50	33
SD	100	100	50	73	60	57	78	80	67
SN	50	67	25	64	47	29	56	80	67

One-sample *t*-tests were used to test for noise categories in which mean identification was significantly greater than 50% across speakers. The results of these tests (given in Table 2.3) show that for RP speakers (Table 2.3(b)), the liquid /r/ was correctly identified by a majority of speakers for all except two noise categories (those replacing rV_m and rVCVC). This result shows that cues to the identification of /r/ (in this RP speaker's speech) are of greater extent than any of the stretches of linguistic material replaced by noise. The /l/ stimuli show a quite different pattern (Table 2.3(a)). /l/ words are correctly identified for all except the longest category of noise (replacing lVCVC) by a majority of speakers, when the vowel before the liquid is audible. However, when this vowel is obscured, the majority of speakers are not able to identify /l/ words correctly. In only two categories, Vl and VlVC, do we find correct identification, and of these, only the result for Vl is highly significant. This is the category in which the whole of the vowel following the liquid is audible, one in which one would expect to find correct identification of the liquid, due to the presence of transitional cues. Thus cues to identification of /l/ stimuli appear to be relatively local.

For ME speakers we find a similar pattern for /l/ stimuli (in Table 2.3(c)): identification is on the whole correct while the vowel before the /l/ is audible, but is no better than chance for the majority of speakers when this vowel is not audible. For the /r/ stimuli we find a different pattern of results, as displayed in Table 2.3(d). There is some correct identification (r, rV_m, rVC_m, Vr and strikingly $VrVCVC_m$) but on the whole identification is poor.

Table 2.3: Significant results of one-sample (one-tailed) *t*-test for RP and ME subjects testing whether the mean scores are greater than 50%, by liquid and material replaced by noise.

Table 2.3 (a) Dialect = RP, liquid = /l/

	l	lV_m	lV_l	lV	lVC_m	lVC	lVCV	$lVCVC_m$	lVCVC
$t(5) =$	20.56	7.36	4.93	2.02	4.67	3.16	5.88	4.34	−0.57
$p <$	0.001	0.001	0.002	0.050	0.003	0.013	0.001	0.004	ns

	Vl	VlV_m	VlV_l	VlV	$VlVC_m$	VlVC	VlVCV	$VlVCVC_m$	VlVCVC
$t(5) =$	14.43	1.12	−0.40	−0.34	1.13	2.08	−2.70	1.00	−1.32
$p <$	0.001	ns	ns	ns	ns	0.047	ns	ns	ns

Table 2.3 (b) Dialect = RP, liquid = /r/

	r	rV_m	rV_l	rV	rVC_m	rVC	rVCV	$rVCVC_m$	rVCVC
$t(5) =$	12.64	1.83	3.16	5.92	9.00	6.99	8.94	5.42	1.66
$p <$	0.001	ns	0.013	0.001	0.001	0.001	0.001	0.002	ns

	Vr	VrV_m	VrV_l	VrV	$VrVC_m$	VrVC	VrVCV	$VrVCVC_m$	VrVCVC
$t(5) =$	5.61	8.62	8.25	5.88	7.71	5.28	6.85	4.45	7.75
$p <$	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.004	0.001

Table 2.3 (c) Dialect = ME, liquid = /l/

	l	lV_m	lV_l	lV	lVC_m	lVC	lVCV	$lVCVC_m$	lVCVC
$t(5) =$	17.86	5.81	2.98	7.55	10.74	2.77	0.61	2.50	2.70
$p <$	0.001	0.001	0.002	0.001	0.001	0.019	ns	0.027	0.022

	Vl	VlV_m	VlV_l	VlV	$VlVC_m$	VlVC	VlVCV	$VlVCVC_m$	VlVCVC
$t(5) =$	2.00	1.65	−1.22	1.10	0.40	−0.35	0.415	1.74	−1.13
$p <$	ns	ns	ns	ns	ns	ns	ns	ns	ns

Table 2.3 (d) Dialect = ME, liquid = /r/

	r	rV_m	rV_l	rV	rVC_m	rVC	rVCV	$rVCVC_m$	rVCVC
$t(5) =$	8.91	13.74	0.96	0.42	2.44	−0.34	−0.63	*	−2.24
$p <$	0.001	0.001	ns	ns	0.029	ns	ns	*	ns

	Vr	VrV_m	VrV_l	VrV	$VrVC_m$	VrVC	VrVCV	$VrVCVC_m$	VrVCVC
$t(5) =$	5.00	1.95	−1.00	1.34	0.64	−0.83	0.63	5.00	−1.36
$p <$	0.002	ns	ns	ns	ns	ns	ns	0.002	ns

* not calculated, all scores equal (67%).

Table 2.4 displays the results of the non-parametric repeated measures test (chosen because of the small sample size and unequal variance), Kendall's *W* (Meddis

1984), calculated for each liquid and category of noise for each dialect group. The results show that in all except two cases, there is a significant difference between percent correct scores of different noise categories, divided by liquid and the segment the noise begins in. The two cases of non-significant difference have straightforward interpretations. In the first case, RP /r/ stimuli with the vowel before the /r/ inaudible (Table 2.4 (b)), identification was correct for all noise categories as described above, thus there was no significant difference between noise categories. In the second case, the ME /l/ stimuli with the vowel before the /l/ obscured (Table 2.4 (c)), identification was uniformly no better than chance, no matter how much other linguistic material was heard.

Table 2.4: Results of Kendall's *W* test, by segment noise begins in and liquid (Significance level set at 5%).

Table 2.4 (a) Dialect = RP, liquid = /l/

	vowel before /l/ audible	vowel before /l/ not audible
$W(8) =$	0.517	0.501
$p <$	0.002	0.002

Table 2.4 (b) Dialect = RP, liquid = /r/

	vowel before /r/ audible	vowel before /r/ not audible
$W(8) =$	0.521	0.306
$p <$	0.002	ns

Table 2.4 (c) Dialect = ME, liquid = /l/

	vowel before /l/ audible	vowel before /l/ not audible
$W(8) =$	0.324	0.241
$p <$	0.049	ns

Table 2.4 (d) Dialect = ME, liquid = /r/

	Vowel before /r/ audible	vowel before /r/ not audible
$W(8) =$	0.687	0.504
$p <$	0.001	0.002

In order to explore the significant differences pointed to by the Kendall's *W* test, a non-parametric 2 related samples test (Wilcoxon sign test) was conducted. This test is based on ranking of scores and is suitable for repeated measures data. As we expect identification to get progressively worse as more linguistic material is replaced by noise, only differences between responses to consecutive categories of noise were tested. Significant differences between other categories of noise would be difficult to interpret, except in the light of differences between consecutive categories. Table 2.5 contains the results of this analysis. The tests show relatively few differences between responses to adjacent noise segments and support the results of the one-sample tests which showed few differences between noise segments within stimulus type. For example, in Table 2.5 (b), category Vl has significantly higher scores than VlV_m. Only one other set of adjacent categories differs, performance decreases slightly from VlVC to VlVCV. This reflects the finding that for the majority of RP speakers there proved to be no correct identification for all noise segments after Vl. As the other results add little to the one-sample analysis they will not be discussed further.

Table 2.5: Results of Wilcoxon sign test for % correct scores, by segment in which noise begins and liquid for RP and ME. (Significance level set at 5%, two-tailed test).

Table 2.5 (a) Dialect = RP, liquid = /l/, vowel before liquid audible

	$lV_m - l$	$lV_l - lV_m$	$lV - lV_l$	$lVC_m - lV$	$lVC - lVC_m$	$lVCV - lVC$	$lVCVC_m - lVCV$	$lVCVC - lVCVC_m$
Z	-2.060[a]	-1.473[b]	-1.997[a]	-0.943[b]	-0.943[a]	-0.406[b]	-0.743[b]	-1.782[a]
Asymp. Sig.	0.039	ns	0.046	ns	ns	ns	ns	ns

a Based on positive ranks.
b Based on negative ranks.

Table 2.5 (b) Dialect = RP, liquid = /l/, vowel before liquid not audible

	$VlV_m - Vl$	$VlV_l - VlV_m$	$VlV - VlV_l$	$VlVC_m - VlV$	$VlVC - VlVC_m$	$VlVCV - VlVC$	$VlVCVC_m - VlVCV$	$VlVCVC - VlVCVC_m$
Z	-1.992[a]	-0.524[b]	-0.943[a]	-1.782[b]	-1.153[b]	-2.201[a]	-1.826[b]	-1.160[a]
Asymp. Sig.	0.046	ns	ns	ns	ns	0.028	ns	ns

a Based on positive ranks.
b Based on negative ranks.

Table 2.5 (c) Dialect = RP, liquid = /r/, vowel before liquid audible

	$rV_m - r$	$rV_l - rV_m$	$rV - rV_l$	$rVC_m - rV$	$rVC - rVC_m$	$rVCV - rVC$	$rVCVC_m - rVCV$	$rVCVC - rVCVC_m$
Z	-0.736[a]	-0.954[a]	-1.787[b]	-1.782[a]	-0.276[b]	-2.251[b]	-0.707[a]	-2.207[a]
Asymp. Sig.	ns	ns	ns	ns	ns	0.024	ns	0.027

a Based on positive ranks.
b Based on negative ranks.

Table 2.5 (d) Dialect = ME, liquid = /l/, vowel before liquid audible

	lV_m – l	lV_l – lV_m	lV – lV_l	lVC_m – lV	lVC – lVC_m	lVCV – lVC	$lVCVC_m$ – lVCV	lVCVC – $lVCVC_m$
Z	–2.060[a]	–0.736[b]	–1.572[b]	–1.160[a]	–0.406[a]	–1.153[a]	–1.992[b]	–0.526[a]
Asymp. Sig.	0.039	ns	ns	ns	ns	ns	0.046	ns

a Based on positive ranks.
b Based on negative ranks.

Table 2.5 (e) Dialect = ME, liquid = /r/, vowel before liquid audible

	rV_m – r	rV_l – rV_m	rV – rV_l	rVC_m – rV	rVC – rVC_m	rVCV – rVC	$rVCVC_m$ – rVCV	rVCVC – $rVCVC_m$
Z	–2.214[b]	–2.207[a]	–0.210[a]	–1.265[b]	–1.261[a]	–0.106[a]	–2.226[b]	–2.251[a]
Asymp. Sig.	0.027	0.027	ns	ns	ns	ns	0.026	0.024

a Based on positive ranks.
b Based on negative ranks.

Table 2.5 (f) Dialect = ME, liquid = /r/, vowel before liquid not audible

	VrV_m – Vr	VrV_l – VrV_m	VrV – VrV_l	$VrVC_m$ – VrV	VrVC – $VrVC_m$	VrVCV – VrVC	$VrVCVC_m$ – VrVCV	VrVCVC – $VrVCVC_m$
Z	–1.786[a]	–1.897[a]	–1.476[b]	–0.736[a]	–0.954[a]	–1.378[b]	–1.787[b]	–2.214[a]
Asymp. Sig.	ns	ns	ns	ns	ns	ns	ns	0.027

a Based on positive ranks.
b Based on negative ranks.

All of the results suggests that /r/ resonances, or cues to the identification of /r/, may have a longer domain than corresponding /l/ cues. The strongest coarticulatory cues to /l/ identification appear to be relatively local, in the vowel preceding the liquid.

The possibility of response bias as a result of word or phoneme frequency was investigated. If either phoneme occurred more frequently than the other in the lexicon, then subjects (from both dialect groups) might be expected to show a response bias towards that phoneme. This would result in higher percentages of correct identification of this phoneme, which should be consistent across dialect groups. This was not the case (as is discussed in 2.2.2.2) and difference in identification of /l/ words and /r/ words seems unlikely to be related to word or phoneme frequency. /l/ and /r/ have very similar token frequencies: a search of the Kucera-Francis dictionary (Kucera and Francis 1967) produced token frequencies for /l/ of 139242 and /r/ of 140620 per million. Counts of frequency of the /l/ and /r/ words used in the experiment and found in the Kucera-Francis dictionary (65% of the stimuli) showed very little difference overall (means of 26 and 11 per million for /l/ and /r/ words respectively, medians of 8 and 7.5).

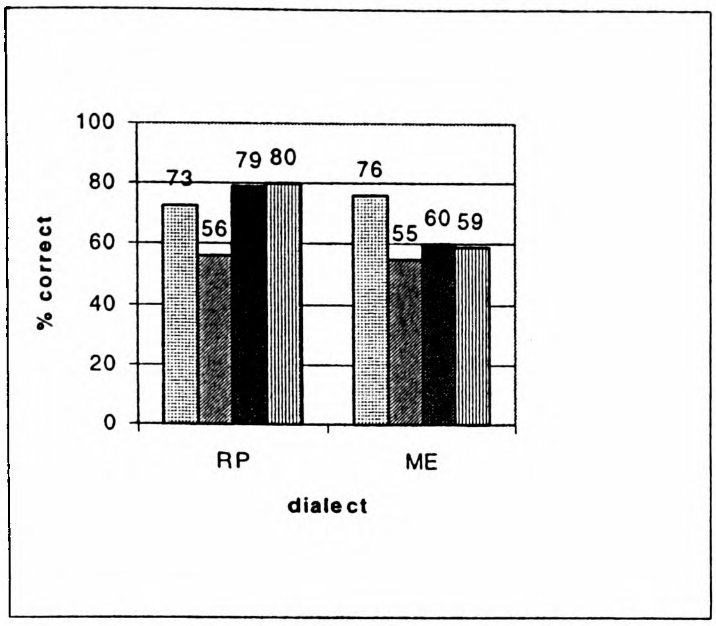
2.2.2.2. Comparison of the RP and ME results

Speakers of both dialects exhibit the same overall performance in terms of percentage of stimuli correctly identified (between 70% and 80% of all stimuli, including those with some part of the liquid audible), but there are differences in the pattern of correct identification, as mentioned above: RP subjects made correct identifications for a wider range of noise categories. Both groups showed a remarkable long-domain effect: correct identification when the noise obscured (V)rVCVC$_m$ and lVCVC$_m$.

The percent correct results show an interesting difference between the dialect groups: correct identification of /r/ stimuli differs significantly, and there is no obvious response bias towards /r/. In only one /l/ noise category (VlVCV for the RP speakers) is there a significant /r/ response. In all other cases there is no significant result, or a significant /l/ response. The mean score for /r/ identification for ME speakers is 60% correct whereas for RP speakers it is 79%. Means for the /l/ data are 65% correct for both groups. Interestingly, but possibly coincidentally, both dialect groups have better recognition of the dark liquid (RP /r/ and ME /l/) in their system. The mean percent correct results are summarised in Figure 2.2 (a) and (b). Note that presence or absence of the vowel before an /r/ does not cause any difference in overall percentage of correct identification.

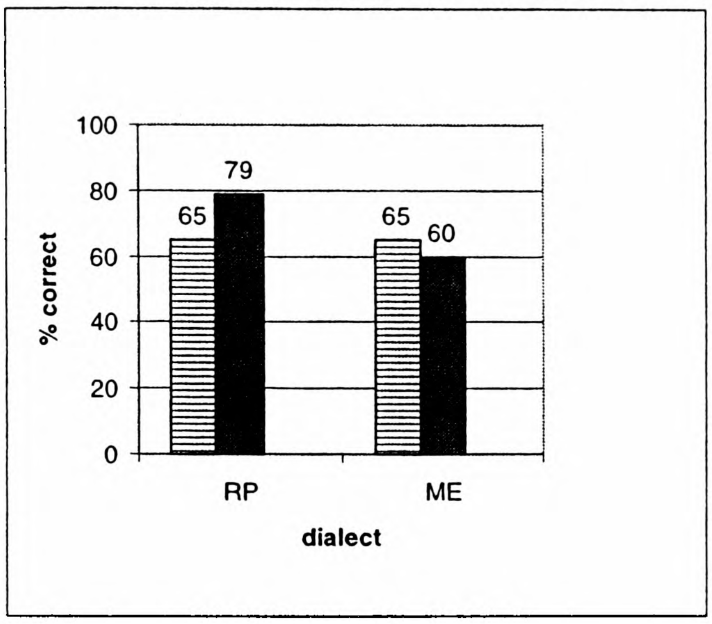
Differences between the experimental conditions and subject pool may have resulted in the poorer performance (in terms of correct identification) of the Manchester speakers. They were on average younger than the RP subjects and were unfamiliar with experimental procedures. The RP subjects were all first or second year psychology students, some of whom had participated in experiments before, and were more familiar with the general aims and methods of experimental psychology. Nevertheless, both subject groups show an ability to discriminate between /l/ and /r/ words when chunks of linguistic material including the liquid are replaced by noise.

Figure 2.2: Mean percent correct identification scores for RP and ME subjects.



(a) All words grouped by liquid and presence or absence of the vowel before the liquid.

/l/ words, V audible
/l/ words, V obliterated
/r/ words, V audible
/r/ words, V obliterated



(b) Words grouped by liquid.

/l/ words
/r/ words

The two dialects, RP and ME, were chosen to investigate whether the domain of resonance differs between dialects with contrasting resonance distinctions. The results of this experiment do not provide evidence for differing domains and are in fact consistent with the hypothesis that both dialects have similar long distance perseverative and anticipatory resonance distinctions. Most interesting is the consistent difference between the /l/ and /r/ stimuli: /r/ stimuli were identified despite absence of the vowel before the /r/, whereas /l/ stimuli were not. This finding is clearly not related to the resonance of the sounds, as it occurs in both dialects, but must be attributed to other properties of their articulation and/or acoustics. It appears that /r/ has stronger anticipatory cues than /l/. This finding is confirmed in experiment 2.

2.2.2.3. Acoustics

It is difficult to relate the results of the small acoustic study (2.2.1.4) to the perceptual experiment, as different stimuli and subjects were employed. Furthermore, the acoustic differences associated with resonance distinctions appear to be extremely subtle, although statistically significant. The acoustic results presented in 2.2.1.4 above show robust anticipatory, not perseverative, resonance distinctions. This suggests an explanation of the long-domain effect found in the perceptual study: cues to perception of /r/ are present earlier in the utterance than we had expected when designing the stimuli. Note, however, that anticipatory cues alone cannot have been sufficient for correct identification of /l/ for both dialects and /r/ for ME, because, when material after the liquid is obliterated, presence of the preceding vowel is not sufficient for correct identification of the liquid. Further detailed acoustic work is required to establish the nature and extent of the resonance distinction; this is attempted in experiment 3 (chapter 3).

2.2.3. Conclusions

This study was designed to test whether listeners are able to use long-domain resonance patterns for speech perception. It demonstrates that subjects are correctly able to identify /l/ and /r/ sounds when some surrounding material is deleted, and /r/ sounds when the noise segment covers a long stretch of material ($VrVCVC_m$). Correct identification of /l/ and /r/ sounds differ. Identification of /r/ sounds is robust over a longer domain than /l/ sounds in both dialects, most strikingly for the RP group. This suggests that the articulatory nature of these sounds and their secondary articulations needs to be further investigated. One possibility is that labialisation of the /r/ and surrounding stretches in RP speech may facilitate identification.

The results of this experiment provide support for Kelly and Local's hypotheses about the long domain of resonance and its potential use in speech perception. Long-domain resonance cues provide information which some speakers can use as cues to perception. This confirms Whalen's (1990) hypothesis that coarticulation can serve a communicative function. The results of this experiment are also consistent with Warren's view of phoneme restoration as a process of auditory induction. Listeners were able to identify /l/s and /r/s in so-called minimal pairs on the basis of acoustic information alone: no semantic contextual information was present. Thus acoustic cues to the identity of the excised sound must be present over long domains. The results suggest that in many of the illusory cases acoustic information about the missing sound is present in the remaining speech signal, so in some sense the phoneme is not absent at all.

Several questions remain to be answered. Phonetic variability between subjects is a commonplace phenomenon, but it might be interesting to explore the differences between subjects. The most challenging linguistic question which arises is why perception is sometimes better for longer noise segments, where more material is replaced by noise, than for shorter ones, particularly as all of the material audible in the longer case can be heard in the shorter case. A possible hypothesis is that conflicting cues are present in material which is subsequently replaced by the longer noise segment, *i.e.* that the resonance of the liquid carries across shorter domain effects of intervening segments.

Although segment-based models of phonology generally allow for overlapping of information, this is usually restricted to adjacent sounds (*e.g.* Nearey 1990). While such models adequately describe the results of some experimental data, they do not capture the full range of coarticulatory behaviour. The results of this experiment show that a liquid may exert long-domain coarticulatory influence on up to two syllables remote from it. The preliminary acoustic study shows that long-domain resonance is not a necessary feature of production, as differences attributed to long-domain coarticulation are not significant for all speakers studied. Nevertheless it proves to be a possible, and in fact natural, feature of speech (Hawkins and Slater 1994) which, when present, is perceptually available to listeners.

2.3. Experiment Two

Given some of the shortcomings of experiment one, the second experiment was designed to address two related questions: the perceptual relevance of long-domain cues, and the relative importance of local and long-domain cues. Articulatory and acoustic data were recorded from a male southern British English speaker (EB). The articulatory data, which are discussed in chapters 3 and 4, demonstrate differences in articulation associated with the distinction between /l/ and /r/. These distinctions are found in segments adjacent and non-adjacent to the liquid. This perceptual experiment was run using the acoustic recordings corresponding to the articulatory data from EB.

Two experimental techniques were used to modify EB's utterances to create the experimental stimuli: cross-splicing (*e.g.* Repp 1985, Whalen 1991) and a modification of the gating paradigm (Grosjean 1996). The modified gating paradigm, or progressive replacement paradigm, was used in the first perceptual experiment: varying stretches of the stimuli (including the liquid) were replaced by noise. This technique was used to investigate how much of a stimulus subjects needed to hear in order to identify it correctly. The splicing technique was used to create cross-spliced stimuli, in which /l/s were spliced into /r/ sentences, replacing the original liquid, and vice versa. Cross-spliced stimuli were created to investigate the relative importance of local and long-domain cues to an /l/ and /r/. The cross-spliced stimuli contain local cues to one liquid, and long-domain cues to the other, thus they contain conflicting cues to the stimulus. They were designed to investigate the relative perceptual salience of local and long-domain cues. As a control, spliced stimuli were also created. These had the same liquid (taken from another token of the utterance) spliced in place of the original liquid. These stimuli were used to investigate how much of the stimulus subjects needed to hear in order to identify the liquid correctly. Discontinuities in stimuli introduced by splicing might lead to poorer identification of stimuli by subjects. Spliced stimuli were thus used rather than naturally occurring stimuli, so that responses to these stimuli would be more comparable with responses to the cross-spliced stimuli. Subjects were required to indicate which liquid they had heard, by picking one word of the relevant minimal pair.

2.3.1. Experimental design

2.3.1.1. Aims

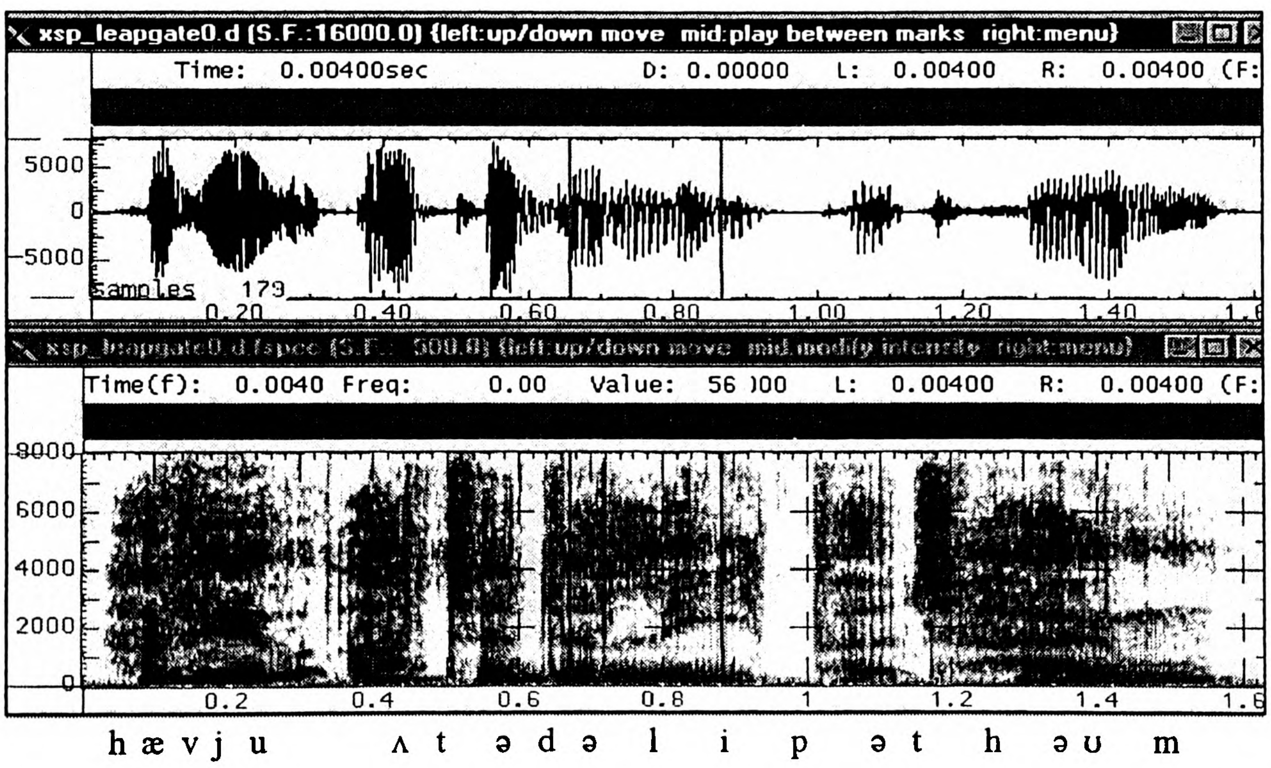
The aims of the experiment were twofold: to check whether subjects could identify /l/s and /r/s from long-domain cues alone, and to investigate the relative importance of local and long-domain cues.

2.3.1.2. Stimuli

The experimental stimuli were prepared from the set of stimuli produced by speaker EB during the articulatory experiment (experiment 3). The frame sentence and word pairs were designed for this experiment and their choice is discussed in the experimental method section of chapter 3. Twelve tokens of the frame sentence 'Have you uttered a — at home?' were chosen from which to create the stimuli, one for each of the words *leap, reap, lip, rip, lap, wrap, lope, rope, lobe, robe, lob, rob*. Sentences with similar intonation patterns (judged impressionistically) and voiced schwas (judged with the aid of wide-band spectrograms) were chosen.

A spliced and a cross-spliced stimulus were created from each utterance. The splicing was done by hand, using waveform and wide-band spectrogram displays created in *Waves*. In order to minimise discontinuities, the original liquid and at least half (in some cases considerably more) of both of its surrounding vowels were replaced by a similar section from another utterance. For the spliced stimuli, these sections came from another token of the same utterance. For the cross-spliced stimuli, a section from the corresponding minimal pair was spliced in. Splicing was done at zero-crossings, and waveforms and spectrograms were used in an attempt to match the signals as closely as possible. The stretch of signal up to the initial splice point, the material to be spliced in and the material after the final splice point were concatenated: there was no alteration of the signal. This simple technique resulted in some stimuli with audible discontinuities, however these were informally judged to be minor (by the author and two other listeners), and preferable to stimuli created by artificially altering transitions. Figure 2.3 shows a waveform and wide-band spectrogram, together with segmental labelling, of one of the cross-spliced stimuli. The utterance was ‘Have you uttered a reap at home?’ with an /l/ and parts of the adjacent vowels spliced in place of the /r/ and its surroundings.

Figure 2.3: Waveform and wide-band spectrogram of a cross-spliced stimulus, the utterance ‘Have you uttered a reap at home?’ with an /l/ and parts of the adjacent vowels spliced in. Segmental labelling is included below the spectrogram and the vertical cursors mark the spliced section.



From the spliced and cross-spliced stimuli, stimuli with sections of speech replaced by noise were created. These will be referred to as gated stimuli, and the sections of noise as gates. The noise used for gating was created using the *Waves* ‘testsd’ function. Six gate lengths were chosen, from gate 0 which has no linguistic material replaced by noise, to gate 5 in which the stretch replaced by noise is the entire CVC word beginning with the liquid, and the preceding syllable. The gate design is displayed in Table 2.6.

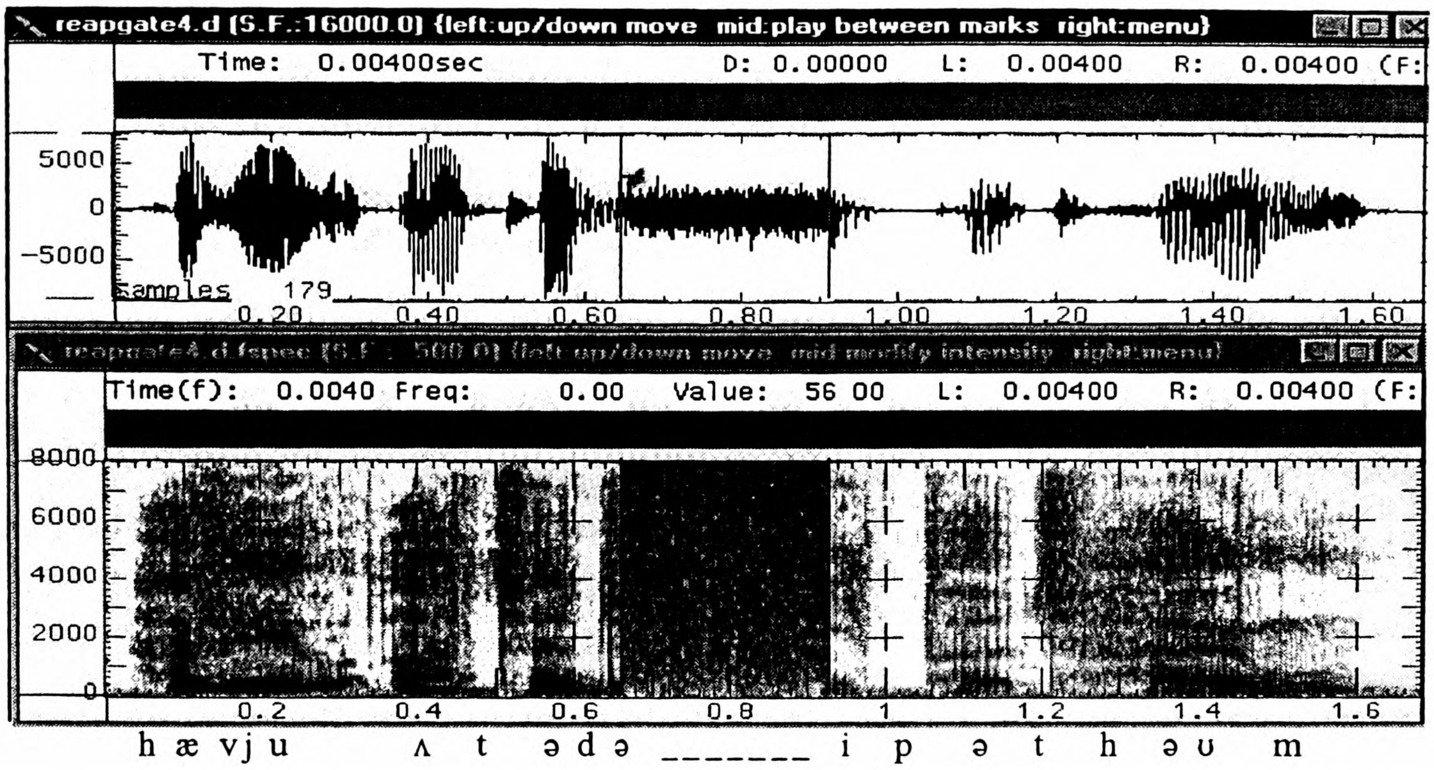
Table 2.6: Design of gates. The linguistic material replaced by each gate is shown, where LIQ = liquid, V = vowel, C = consonant. ^{spl} indicates that a gate begins or ends at the splice point, and thus does not cover the whole of the vowel (ə or V).

Gate Number	Gate 0	Gate 1	Gate 2	Gate 3	Gate 4	Gate 5
Material replaced by gate	none	LIQ	^{spl}əLIQ	LIQV^{spl}	^{spl}əLIQV^{spl}	dəLIQVC

Identification of the speech section to be replaced by noise was generally straightforward. Gate 5 was the most straightforward: the consonants defining the gate endpoints are both stop consonants. Gating began at the end of the regular periodicity of the vowel preceding the /d/ and ended after the burst of C (either /p/ or /b/). Gate 4, and the splice-point defined ends of gates 2 and 3, were defined by the known splice points (these points were noted during the splicing procedure). For this speaker, segmentation of /l/ proved simple: there were noticeable discontinuities in the wide-band spectrograms and discontinuities and loss of energy in the waveform. The /l/ was defined as the stretch between these discontinuities. /r/ segmentation was more problematic. The solution adopted was to define /r/ as the section between the midpoints of the transitions into and out of the segment, following the speech segmentation criteria for the SCRIBE project, a speech corpus labelling project (Hieronymus *et al.* 1990). Once each stretch of material to be gated out had been identified, a noise signal of the same duration was created and used to replace the relevant linguistic material. The initial stretch of the stimulus (up to the initial gate point), the noise and the final stretch (following the final gate point) were concatenated to create the experimental stimuli.

Stimuli with gates 0 to 3 were created for all the spliced and cross-spliced stimuli, resulting in 8 stimuli for each of the 12 original utterances, 96 stimuli in total. Gates 4 and 5 were created only once for each utterance, as the replacement of the spliced material by the gate meant that splicing was irrelevant for these stimuli. The total number of stimuli was 120, 10 stimuli for each of the 12 original utterances. Figure 2.4, included to illustrate the gating, contains the *reap* stimulus with gate 4 applied.

Figure 2.4: Waveform and wide-band spectrogram of a gate 4 stimulus, the utterance 'Have you uttered a reap at home?' with the liquid and most of the adjacent vowels replaced by noise. Segmental labelling is included below the spectrogram, and red cursors mark the noise section.



The stimuli were then converted to Microsoft audio (*wav*) files, so that they could be played from a PC. For the experiment, 8 randomisations of the 120 stimuli were created, resulting in a list of 960 items. These were used to create a control file, containing the randomised lists of sound files and possible responses (*e.g. leap* or *reap*).

2.3.1.3. Subjects

Ten subjects participated in the experiment, five male and five female. All were right-handed, monolingual speakers of standard Southern British English, who reported no speech and hearing difficulties. Their ages ranged from 18 to 24. Subjects were paid for their time and were recruited personally in Oxford.

2.3.1.4. Experimental procedure

The experimental stimuli were presented to subjects using a stimulus presentation tool developed in the Oxford University Phonetics Laboratory by A. Slater. Subjects were seated in a sound-treated room in the Oxford University Phonetics Laboratory, facing a VDU. They were given a pair of Yamaha HP3 headphones, through which the stimuli were presented, and a PC keyboard on which to respond. The stimulus tool used the control file to run the experiment. Sound files were played via a Soundblaster 16 digital-analog card installed in the PC on which the presentation tool was run. As each file was played, the 2 possible responses were displayed on a screen and the subject's choice recorded. The side of presentation (left or right) of responses on the screen was randomised, to minimise response bias. The two keys to be used for responses were clearly identified by labels.

The experiment began with a short practice run consisting of 10 stimuli. The main run consisted of 960 stimuli. The task took approximately an hour to complete, varying slightly according to the speed with which the subject responded. The experiment was driven by the subject's response time and controlled by some timing parameters. For each stimulus a screen appeared displaying the text "Please listen now ...". There was then a 750 millisecond pause followed by a screen displaying the frame sentence and 2 response options. 550 milliseconds later the relevant sound file was played. If subjects did not respond within 2 seconds, they were prompted to do so by a message on the screen. Once the response had been recorded, there was a 200 millisecond gap before the next cycle began.

Subjects were told that the aim of the experiment was to discover how well people could hear words under noisy conditions. They were instructed to respond on the keyboard by pressing the key corresponding to the word they had heard, or to make a guess if they were unsure.

2.3.2. Results

For the analysis, percent correct identification scores were calculated for each subject for each category of gate. The 6 gate categories (gates 0 to 5) were divided into the 4 gates created from spliced stimuli (spliced stimuli, gates 0 to 3), the 4 gates created from cross-spliced stimuli (cross-spliced stimuli, gates 0 to 3) and gates 4 and 5 (for which splicing was irrelevant as so much material was gated out). As the previous perception experiment suggested that /r/s were in general identified better than /l/s, the percent correct scores were calculated separately for /l/ and /r/ stimuli.

The percent correct results for /l/ and /r/ stimuli are presented in Table 2.7. Percent correct means percentage of correct identification of the original liquid in the spliced and unspliced stimuli, and of the spliced-in liquid in the cross-spliced stimuli. The score thus refers to correct identification of local cues in the cross-spliced case.

Table 2.7: Percent correct results for all subjects, divided by stimulus type.

	SPLICED STIMULI				CROSS-SPLICED STIMULI				UNSPLICED STIMULI	
/r/ stimuli	Gate 0	Gate 1	Gate 2	Gate 3	Gate 0	Gate 1	Gate 2	Gate 3	Gate 4	Gate 5
JD	100	100	98	100	100	83	38	50	88	73
MS	100	100	92	100	100	44	31	35	94	40
JC	100	100	100	100	98	85	46	75	98	54
GI	100	100	98	100	100	85	44	50	94	63
MF	100	100	94	92	98	88	42	69	81	56
MR	100	100	100	98	100	90	60	60	75	44
LM	100	100	94	100	100	56	33	13	94	79
SR	100	100	98	98	100	85	60	52	77	56
AH	100	100	94	100	100	63	21	42	88	58
MM	100	100	92	92	100	71	56	46	60	50
/l/ stimuli	Gate 0	Gate 1	Gate 2	Gate 3	Gate 0	Gate 1	Gate 2	Gate 3	Gate 4	Gate 5
JD	100	83	83	38	100	100	98	100	40	44
MS	100	50	42	48	100	100	96	100	17	58
JC	100	98	77	75	100	98	100	98	58	44
GI	100	81	75	48	100	100	100	100	46	35
MF	98	92	71	71	100	98	94	98	42	56
MR	100	100	92	65	100	98	96	100	75	46
LM	100	65	56	46	100	100	96	98	15	25
SR	100	92	83	65	100	100	96	98	56	42
AH	98	85	63	54	98	100	92	96	29	46
MM	100	79	67	58	100	100	90	100	48	58

For each gate, each subject responded 48 times (8 times for each of the 6 word pairs). The distribution of responses is not always normal: in some cases (*e.g.* Gate 0), responses are all at around 100%. As the data was collected by repeatedly exposing subjects to the same stimuli, a non-parametric repeated measures analysis, Kendall's W was chosen. The result, $W(19) = 0.843$, $p < 0.001$, confirms that there are significant differences between subjects' responses to different gates. To test whether generalisations can be made across subjects, *i.e.* whether subjects as a group have scores significantly different from 50%, a one-sample t-test was calculated for each gate. (For the gates of interest, *i.e.* where scores diverged noticeably from 100%, a Kolmogorov-Smirnov test showed no departures from normality significant at the 1% level). Table 2.8 gives the results of the t-test. A significant result ($p < 0.05$) for any gate is interpreted as demonstrating that the subjects as a group correctly identified the liquid. Missing values result when the standard deviation for that sample is 0 - identification for all subjects is 100% correct. Note that for the cross-spliced stimuli, correct identification refers to identification of the spliced-in liquid, on the basis of local cues.

Table 2.8: Results of one-sample *t*-test, testing whether subjects' scores differ as a group from 50% for each gate.

/r/ stimuli	Spliced gates				Cross-spliced (spliced in /l/)				No splicing	
	gate 0	gate 1	gate 2	gate 3	gate 0	gate 1	gate 2	gate 3	gate 4	gate 5
$t(9) =$	-	-	46.52	46.48	186	5.01	−1.68	−0.15	9.42	1.93
$p <$	-	-	0.001	0.001	0.001	0.001	0.126	0.888	0.001	0.086

/l/ stimuli	Spliced gates				Cross-spliced (spliced in /r/)				No splicing	
	gate 0	gate 1	gate 2	gate 3	gate 0	gate 1	gate 2	gate 3	gate 4	gate 5
$t(9) =$	186	6.64	4.26	2.50	249	28.73	27.83	7.01	−0.81	−1.52
$p <$	0.001	0.001	0.002	0.034	0.001	0.001	0.001	0.001	0.439	0.164

The *t*-tests show that subjects identified liquids correctly for many of the gates. Identification was very good for all of the spliced gates, except gate 3 of the /l/ stimuli, although this result is significant for $\alpha = 0.05$. For the cross-spliced stimuli, excellent identification is found for gates 0 and 1. Only the stimuli with spliced in /r/ (*i.e.* /l/ stimuli) have accurate identification for the group of subjects at gates 2 and 3. The unspliced stimuli, designed to test subjects' use of non-local cues, show that subjects as a group did not identify liquids from non-local cues in this experiment, except for /r/ stimuli at gate 4, where small portions of the vowels surrounding the gated-out /r/ are audible.

The non-parametric Wilcoxon test for 2-related samples was used to explore whether responses to any two gates differ. This test was performed for gates 2 and 3, where identification was not uniformly good for all subjects, to enable comparison of subjects' performance at these gates. For the /r/ stimuli, gates 2 and 3 of the cross-spliced stimuli do not differ significantly ($Z = -1.125$, $p < 0.260$). Gates 2 and 3 of the spliced /l/ stimuli show a significant difference ($Z = -2.138$, $p < 0.033$), which suggests that identification at gate 3 is worse than that at gate 2. Gate 3 of the /l/ spliced stimuli differs significantly from gates 2 and 3 of the /r/ cross-spliced stimuli ($Z = -2.494$ and -2.807, $p < 0.013$ and 0.005 respectively). These results suggest that the subjects as a group identified the liquid more accurately for gate 3 of the /l/ spliced stimuli than for gates 2 and 3 of the /r/ cross-spliced stimuli.

There are significant Spearman correlations (α = 0.01) between responses for several gates. Correlations of interest include a significant negative correlation (ρ = –0.801, $p < 0.005$) between the responses to /r/ stimuli for gate 5 and /l/ stimuli for gate 5. This demonstrates a response bias for subjects: those with higher correct responses for the /r/ stimuli had lower correct /l/ responses (*i.e.* more /r/ responses) for the /l/ stimuli. Gates 1 and 3 of the /l/ spliced stimuli are significantly positively correlated with gate 3 of the cross-spliced /r/ stimuli (ρ = 0.068 and 0.780, $p < 0.010$ and $p < 0.008$). This shows that subjects who could identify /l/s well at gates 1 and 3 of the spliced stimuli identified them better at gate 3 of the cross-spliced stimuli. This suggests that they were better able to make use of the anticipatory cues present in the /ə/. The responses to spliced /l/ stimuli are highly positively correlated with each other (α = 0.01), suggesting that some subjects were better at identifying /l/s than others. Similarly, the responses to cross-spliced /r/ stimuli (those with an /l/ spliced in) are correlated, with gate 1 and gate 3 significant at the 1% level. For the /r/ stimuli, responses are generally very consistent (very high correct identification scores) between subjects, and correlations were therefore not examined for such gates.

In order to check whether stimulus affected subject responses, the percent correct identification for each of the 120 stimuli was calculated, for each subject. One-sample *t*-tests, testing for a significant difference in the sample from 50% (α = 0.05), showed that the majority of the stimuli (82) were correctly identified by subjects as a group. A further 29 showed no significant difference from 50%, and 9 showed a significant difference in the other direction: subject responses were significantly below 50%. The group of non-significant stimuli contains all 12 stimuli with gate 5 and 5 stimuli from gate 4 (all /l/ stimuli). It also contains 4 spliced stimuli (2 from gate 2 and 1 each from gates 1 and 3, again all /l/ stimuli), and 7 cross-spliced stimuli (2 from gates 2 and 3, and 3 from gate 1, all /r/ stimuli with spliced-in /l/s). The 9 stimuli with responses significantly lower than 50% were all /l/ stimuli or cross-spliced /r/ stimuli containing an /l/. They were consistently identified as /r/ stimuli. In the case of the 6 cross-spliced stimuli (*rope* gate 1, *rob, reap* and *wrap* gate 2, *rip* and *rope* gate 3), this may be due to the presence of stronger cues for the /r/. The remaining three stimuli were *lap* gate 4 and spliced gate 2, and spliced gate 3 of *lope*. The small number of stimuli in this category suggests that no importance should be attached to the significant result of the *t*-test, particularly given the large number of tests performed.

2.3.3. Discussion and conclusions

For gate 0, in which no material was replaced by the gate, subjects scored from 98% to 100% correct identification. This shows that the task of identification was performed well by all subjects. The splicing and cross-splicing did not affect identification for this gate and the spliced-in liquid was identified regardless of the long-domain context: local information clearly outweighed any more distant cues. There are several other sets of stimuli where all subjects identified the liquid correctly at above 90%. These are gates 1, 2 and 3 of the spliced /r/ stimuli and the cross-spliced /l/ stimuli. Subjects respond /r/ consistently for all of these gates, to both the spliced /r/ stimuli and the cross-spliced /l/ stimuli, in which an /r/ is spliced in. Again, the local information outweighs long-distance cues, and subjects consistently identify the /r/.

Gates 1, 2 and 3 of the spliced /l/ stimuli and cross-spliced /r/ stimuli show a different pattern of responses. This is one of the clearest results of this experiment: subjects respond differently to /l/ and /r/ stimuli. Identification of /r/ is good when any of the adjacent vocalic material is present (gates 0, 1, 2, 3 and 4), regardless of whether the stimulus had an original /r/ or an /r/ spliced into an /l/ context. Identification of /l/ is noticeably worse, particularly in the cross-spliced case. For the spliced /l/ stimuli, identification is correct for subjects as a group for all gates (0-3), as shown by the *t*-tests. The cross-spliced stimuli (/r/ stimuli with /l/s spliced in) are not correctly identified as /l/s (or as /r/s), except in the case of gates 0 and 1. In gate 0 the spliced-in /l/ is audible, and in gate 1 the /l/ is replaced by noise but local information from both adjacent vowels is audible. The cross-spliced /l/ is not consistently identified in stimuli with gates 2 and 3, which have the liquid and one of the stretches of adjacent vowel replaced by noise (^{spl}əLIQ or LIQV^{spl}), yet subjects had no difficulty identifying cross-spliced /r/s for gates 2 and 3 of the /l/ stimuli. This suggests that cues to an /l/ may be weaker or not always strongly present in adjacent stretches of the speech signal, whereas cues to an /r/ are spread further through the speech signal and are stronger.

For the cross-spliced /r/ stimuli, longer-distance cues from an original /r/ and local adjacent cues from the spliced-in VlV sequence appear to conflict, as neither the cross-spliced /l/ nor the original /r/ are consistently identified. All of the information present at gates 2 and 3 of the cross-spliced /r/ stimuli, with the exception of an additional portion of spliced-in vowel from the /l/ context, is present for gate 4 of the /r/ stimuli, which removes ^{spl}əLIQV^{spl}. For gate 4 identification of the /r/ is fairly good and generalizable across subjects. Thus in gates 2 and 3 of the cross-spliced stimuli there is enough information in the speech signal for subjects to identify an /r/, and yet they fail to do so. This must be because of the presence of conflicting cues for an /l/, which are however not strong enough to override the /r/ cues. Note that for the spliced /l/ stimuli, where an /l/ is spliced into an /l/ context and no conflicting cues are present, identification of the /l/ is above chance for both gates 2 and 3. For gate 4 of the /l/ stimuli, identification of the /l/ was not above chance, in contrast to the /r/ stimuli where identification was significantly above chance. This suggests an explanation for the discrepancy between responses to cross-spliced /l/ and /r/ stimuli for gates 2 and 3. In the /l/ context, cross-spliced /r/s are identified well above chance because information from the longer-distance /l/ cues is not perceptually available to these subjects. In the /r/ context, conflicting cues are present. The experiment was designed to investigate the relative strength in terms of perceptual usefulness of local *vs.* long-distance cues. The results suggest that local cues are in general more important in identifying liquids, although more remote cues may affect perception as evidenced by responses to the cross-spliced /r/ stimuli.

The relative strength of anticipatory *vs.* perseverative cues can be examined by comparing responses to gates 2 and 3. As identification of /r/ was always very good, the question can only be addressed for the /l/ stimuli: for /r/ stimuli, cues in the vowel preceding and following the liquid are sufficient to allow near perfect identification of the liquid, and their relative strength cannot be evaluated. For (spliced) /l/ stimuli, responses to gates 2 and 3 differ. The /l/ is identified more consistently for gate 2 (which removes ^{spl}əLIQ) than for gate 3 (which removes LIQV^{spl}), as shown by the Wilcoxon test reported above. This result suggests that subjects found perseverative cues in the vowel following the liquid, which are not removed by gate 2, more useful for identification than anticipatory cues. A possible explanation for this result is that the following (stressed) vowel is generally longer than the ə, and subjects hear a longer stretch of an adjacent vowel in gate 2. The acoustic measurements for both sets of vowels reported in chapter 3, while showing significant differences in different liquid contexts, do not suggest that there is any difference in the magnitude of acoustic cues. Nevertheless, subjects identify an /l/ better for gate 2. The significant positive correlation between results at gate 3 of the /l/ spliced stimuli and gate 3 of the /r/ cross-spliced stimuli suggest that some subjects are better able to make use of the anticipatory cues to an /l/. Those subjects who make use of the cues to an /l/ in the anticipatory schwa, also do so in the cross-spliced /r/ context where they have higher /l/ identification rates.

Gates 4 (^{spl}əLIQV^{spl}) and 5 (dəLIQVC) were designed to investigate the use of long-domain differences between /l/ and /r/, which are present in articulation and in the acoustic signal (shown in chapter 4). Identification of the liquid was only significantly above chance for gate 4 of the /r/ stimuli. Although /r/ was identified at above chance at gate 4, /l/ was not. This suggests that the presence of coarticulatory cues are useful but not necessary for the identification of /r/: subjects, when faced with the absence of these cues in the /l/ stimuli, responded with both /l/ and /r/. This result was surprising, but perhaps due to several factors: the nature of the speech used, the experimental task, an induced response bias and considerable variability between subjects. The speech stimuli were created from recordings made by a subject wearing an EPG artificial palate and with 7 EMA coils attached to his articulators. This extra equipment may well have produced slightly unnatural sounding speech. The stimulus preparation technique was deliberately fairly unsophisticated, in an attempt to keep the stimuli as natural as possible and to avoid the possibility of an experimental effect caused by the experimenter's modification of formant frequency patterns. This resulted in some pitch discontinuities and clicks in the stimuli. More importantly, a strong response bias towards /r/ was induced in some subjects, noticeably LM and MS. As /r/s were more easily identified, subjects responded /r/ more often than /l/ for most gates, and some subjects appear to have carried this response bias over to gate 4 and 5. Gate 5 of the /r/ stimuli is negatively correlated with gate 5 of the /l/ stimuli, showing that subjects were responding at close to chance with a response bias towards /r/. There is also considerable variance in the sample, showing that behaviour is quite varied between individuals. The standard deviation ranges from 0 to 19% at the different gates. To estimate the population mean with error no greater than 5%, a sample size of up to 50 subjects would be needed for the gates with the most variation (Woods *et al.* 1986, 107-8). The sample in this experiment is probably too small to find significant results when responses vary to this extent. Although running the study for 50 subjects might produce a significant result, the current study shows that there are great individual differences and suggests that the perceptual relevance of long-domain cues is small, and that they are not used extensively by hearers in speech recognition. Different or subtler experimental techniques might be more worth developing to pursue this line of research.

The results of this experiment show that /r/ has much stronger, or more perceptually available cues, distributed over a greater stretch of the speech signal than /l/. Local cues to a segment's identity, found in adjacent stretches of vowel, generally outweighed any more distant cues. However, relatively distant /r/ cues conflicted with local cues to an /l/, causing subjects to be unable to identify the missing liquid consistently. Listeners are able to use coarticulatory information present in the vowels adjacent to a liquid to identify that liquid.

2.4. Summary and conclusions

2.4.1. Perceptual effects

The two experiments described in this chapter demonstrate that the coarticulatory effects of liquids are available to listeners for use in speech perception. Experiment one shows that listeners can correctly identify /l/ and /r/ sounds when some surrounding material is deleted, and /r/ sounds when the noise segment covers a long stretch of material ($VrVCVC_m$). Correct identification of /l/ and /r/ sounds differ in domain, with identification of /r/ sounds occurring over a longer domain than /l/ sounds. Experiment two confirms that listeners are able to use coarticulatory information present in the vowels adjacent to a liquid to identify that liquid. It also confirms that subjects respond differently to /l/ and /r/ stimuli. /r/ has much stronger, or more perceptually available cues, distributed over a greater stretch of the speech signal than /l/. Local cues to a segment's identity, found in adjacent stretches of vowel, generally outweigh any more distant cues, although more remote cues may affect perception.

2.4.2. Implications for theories of word recognition

The experiments described in this chapter are concerned with the perception of liquids in real words: subjects were asked to identify which of two words (differing only in the liquid they contained) they had heard. Thus the task is at one level a word recognition task. To the extent that these experiments test word recognition, they have clear implications for theories of spoken word recognition. Most theories of word recognition assume that some input, possibly a string of phonemes, is provided by early processes of speech perception and compared to the mental lexicon until a best match is found (Goldinger *et al.* 1996). The results of these experiments show that an adequate theory must allow for the effects of non-adjacent phonetic segments and fine coarticulatory detail in word recognition. A phonemic representation is insufficient.

The experimental results provide support for a non-segmental theory of word recognition such as that outlined by Nguyen and Hawkins (*cf.* Hawkins 1995). Hawkins (1995) argues for a non-segmental model of spoken word recognition, stating that phonemes are not essential for word recognition as perception takes place mainly by reference to non-segmental phonetic structure. They propose that words are recognised by mapping a fine-grained (non-segmental) auditory representation of the acoustic input directly onto a (non-segmental) lexicon. In Nguyen and Hawkins' model (Nguyen and Hawkins 1999), both short-domain and long-domain phonetic information are central to word recognition. They argue that long-domain information has a central role to play in word recognition for at least three reasons: it may speed up lexical access, may allow backtracking and may make recognition more resistant to noise.

A non-segmental model is consistent with an exemplar theory of perception in which the auditory properties of an item are compared with those of the exemplars. Johnson (1997) advocates an exemplar theory in which information about individual talkers is stored in exemplars. This model is easily extended to store fine-grained coarticulatory detail. Hawkins and Nguyen point out that a non-segmental approach is also compatible with well-established theories of word recognition such as Cohort theory (*e.g.* Marslen-Wilson and Warren 1994) and LAFS (Klatt 1979), both of which do not require an intermediate level of phonemic representation.

Finally, a theory of speech perception which involves the matching of detailed auditory representations to a lexicon is entirely consistent with recent neurolinguistic and psycholinguistic research (*e.g.* Goldinger 1997, Johnson 1997). Coleman (1998a), reviewing the neuroscience literature, concludes that the mental lexicon is very probably an auditory one. This topic will be returned to in chapter 6.

CHAPTER THREE

LOCAL COARTICULATORY EFFECTS OF /l/ AND /r/

3.1. Introduction

In order to relate acoustic and perceptual data on the coarticulatory effects of English liquids to claims about secondary articulation, an articulatory study was conducted. The techniques used were electromagnetic articulography (EMA) and electropalatography (EPG), which will be described below. The experiment was run at Queen Margaret University College, Edinburgh (QMC) and the three subjects were speakers of standard Southern British English. This chapter describes the experimental method, acoustic results, and articulatory results for segments adjacent to the liquid, *i.e.* local coarticulatory effects. Chapter 4 considers the extent of coarticulation and shows that it extends well beyond adjacent segments.

3.1.1. EMA

Electromagnetic articulography provides data on articulator movement in the midsagittal plane. Transmitter coils placed on a helmet worn by subjects produce alternating magnetic fields. These fields induce an alternating signal in small insulated receiver (transducer) coils, which are attached (with dental adhesive or sutures) to articulatory structures of interest in the subject's midsagittal plane. The induced voltage is known to be approximately inversely proportional to the cube of the distance between the transmitter and receiver, but needs to be determined empirically (Hoole 1996). This allows calculation of the position of the receiver coils. Movement of the receiver coils off the midsagittal plane causes measurement error, but this can be largely eliminated by using three transmitter coils with different frequencies, such as in the Carstens AG100 system used for this study (Gracco 1995; Perkell *et al.* 1992; Schönle *et al.* 1987).

Accuracy and sources of measurement error have been evaluated for at least two of the systems available: movetrack (Branderud *et al.* 1993; Nguyen and Marchal

1993; Perkell *et al.* 1992) and the Carstens AG100 (Alfonso *et al.* 1993; Honda and Kaburagi 1993; Hoole 1993; Perkell *et al.* 1992; Tuller *et al.* 1990). Accuracy, with adequate calibration procedures, has been shown to be within the order of 0.5 mm for likely speech movements with the AG100 system (Hoole 1996). Furthermore, relative accuracy (differences in placement of a particular articulator) is generally better than absolute accuracy (Hoole and Nguyen 1999).

The reliability of EMA during speaking sequences has been investigated, by Horn *et al.* (1997), using a set of German syllables of the form /aCa/ produced by 31 speakers. For each speaker, three coils were placed on the front part of the tongue (on the tip, 10 mm and 20 mm behind the tip). They showed that distances and angles measured were in general highly reliable across speakers (Horn *et al.* 1997). The influence of pellet markers on acoustic and perceptual measures have been examined, and found to differ with speaker, with no consistent results across a sample of 57 speakers (Weismer and Bunton 1999). The reliability of simultaneous EMA and EPG data collection has been documented by Rouco and Recasens (1996). They show that EMA data is not significantly affected by the use of the EPG system. Tongue-palate contacts were affected in their study by the presence of EMA coils, which interfered with the contact pattern. This suggests that EPG data acquired simultaneously with EMA data should be used cautiously. Fitzpatrick and Ní Chasaide (2000) show that EMA and EPG data provide complementary data on tongue constriction location, and that EMA provides a more accurate estimation of tongue height than EPG.

The relationship between EMA data and acoustics has been investigated by Maurer *et al.* (1993). They report that different articulatory positions can give rise to similar vowel qualities and conversely that similar articulatory positions can give rise to different vowel qualities. Their study used five coils, two placed on the lips, two on the tongue and one on the jaw. The speech material was spontaneous repetition of (German) words with five vowel qualities and production of the isolated vowels at different levels of F_0. The two subjects were also instructed to produce the vowels whilst trying to move the articulators as much as possible without altering vowel quality. This task produced more extreme articulatory movements, compared with the spontaneous speech results. The fact that movement trajectories for one vowel overlapped with another, despite relatively constant Fourier spectra, led the authors to conclude that similar articulatory positions may correspond to different vowels. The nature of the tasks and the limitations of EMA may well account for this conclusion, although the theory of distinctive regions and modes (Mrayati *et al.* 1988) predicts similar results.

A more recent study by Hogden *et al.* (1996) shows that articulator positions measured by EMA can be very accurately recovered from acoustic data. A data set (/gV_1V_2g/ data for all combinations of 9 Swedish vowels and one English vowel, produced by one speaker) was used to create look-up tables for estimating articulator positions from acoustics. They achieved 94% correlations between predicted and actual tongue positions on two further sets of the same data, with root-mean-squared errors of less than 2mm (Hogden *et al.* 1996). On a smaller scale, the spectral shapes of the French fricatives [s] and [ʃ] have been successfully regenerated from articulatory parameters extracted from EMA data (Nguyen *et al.* 1994). EMA has also been used to investigate Steven's quantal theory (Stevens 1989) for a series of palatal and velar stops (Fitzpatrick and Ní Chasaide 1993) and the existence of trading relations between the tongue and the lips in the production of /u:/ (Perkell *et al.* 1993).

EMA is an effective tool for investigating coarticulation, in particular lingual coarticulation (Hoole and Nguyen 1999). Studies of local coarticulatory effects include coarticulation of velar consonants with adjacent vowels by German speakers (Mooshammer and Hoole 1993; Mooshammer *et al.* 1995), coarticulatory effects on German /r/ (Mooshammer and Schiller 1996) and variability in German vowels (Hoole and Kühnert 1995). Lingual coarticulation in English /kl/ clusters (Hardcastle *et al.* 1996) and fricatives (Hoole *et al.* 1993) has also been investigated, as have alveolar-velar assimilations in German and English (Kühnert 1993; Pompino-Marschall 1991). One study provides evidence of longer domain coarticulation: in a study of anticipatory coarticulation in German speakers, lip rounding in anticipation of a rounded vowel was found to begin shortly (10 ms) before or during the syllable preceding that containing the rounded vowel (Katz *et al.* 1990). This provides articulatory evidence for earlier acoustic findings of vowel-to-vowel coarticulation in VCV utterances (Öhman 1966) and anticipatory lip rounding (Lubker and Gay 1982).

3.1.2. EPG

Electropalatography (EPG) is a technique for measuring contact between the tongue and the palate which shows both the shape of tongue-palate contact and changing contact patterns over time. A thin acrylic pseudo-palate is constructed from a palatal cast of the speaker concerned. In the Reading EPG system used for this experiment the palate contains 62 electrodes (contacts), distributed in rows across the palate, with 6 contacts in the most anterior row and 8 contacts per row for the remaining 7 rows (Hardcastle *et al.* 1989). Thin wires are attached to each electrode, and these are wound behind the back molars on each side of the mouth and led out through the corners of the mouth. An undetectable direct current runs through these wires. A hand-held ground electrode is used to complete the circuit. When contact is made between the tongue and an electrode in the palate, this closes the circuit and the contact is registered. EPG can be used to study spontaneous speech as sampling can take place frequently: the equipment for this experiment had a sampling rate of 200 Hz. EPG has been extensively used in studies of normal speech and in speech therapy (Hardcastle and Gibbon 1997) and several techniques have been described for data reduction (Hardcastle *et al.* 1991). As the technique is better known than EMA no attempt to review the literature is made here.

3.2. Experiment Three

3.2.1. Experimental design

3.2.1.1. Aim

The aim of this experiment was to discover articulatory distinctions correlated with the difference between /l/ and /r/, and their extent. A further aim was to investigate the relationship between articulatory and acoustic data, and to relate the data to the results of perceptual studies of recognition of /l/ and /r/.

3.2.1.2. Stimuli

The stimuli for the experiment were chosen on the basis of prior acoustic work, the pilot production study conducted on 6 speakers, described in Chapter 2. All words were produced in a frame sentence, 'Have you uttered a — at home?'. The previous acoustic study suggested that using alveolar stops was preferable to using velars which exerted a strong local influence on adjacent vowels and obliterated subtler resonance distinctions (hence the replacement of 'again' of earlier stimulus sets with 'at home'). 6 l/r pairs were chosen, all English words: *leap/reap*, *lip/rip*, *lap/wrap*, *lope/rope*, *lobe/robe*, *lob/rob*. Twelve repeats of each stimulus sentence were randomised, giving 144 stimulus sentences. One of a set of filler sentences was placed after each stimulus sentence, to reduce the likelihood of interference of neighbouring, similar stimuli. The filler sentences were also included in an attempt to hide the real object of the study from the subjects; the EMA and EPG data from them was not recorded, though the subjects were unaware of this. The capture time required by the software, which does some online processing of the EMA data, was long enough to allow the subject to read the filler sentence while the data was saved and rest before the next recording. The EMA experimental paradigm places a limit on the number of sentences that can be recorded in a session, as the adhesive weakens and the coils fall off after an unpredictable period of time. The collection of the 144 stimuli described above in one session was just within the limits of the experimental technique at the time of the first recording. Data capture was subsequently speeded up by modifications to the software involved, so that the second and third recording sessions were easily completed.

3.2.1.3. Subjects

The subjects of this experiment were right-handed males, EB, MB and AS aged 21, 20 and 31 respectively, who reported no speech or hearing defects. They were students at Oxford, had no training in phonetics and participated in the experiment voluntarily. Subjects were screened for potential problems such as claustrophobia and a severe gag reflex, and were well informed about the experimental technique and shown a video of EMA data collection produced at QMC. The QMC ethics committee approved the experiment and method of subject selection.

3.2.2. Experimental facility and equipment

The speech production research facility at QMC in Edinburgh consists of an acoustically damped studio and control room, separated by quadruple glazing. (For detailed specifications see the Report on the Speech Production Research Facility (Wrench 1998)). The primary research instruments are a Carstens AG100 Articulograph and a Reading EPG2 Electropalatograph system. Separate personal computers are used for the different pieces of recording instrumentation. One computer initiates recording by transmitting starting signals to the others via a serial port.

Recordings were made on different occasions for each subject, resulting in some slight differences in the recording set-up. For subject EB an audio signal was recorded using a Shure 849 electret microphone directly connected to a SoundBlaster analogue-to-digital converter in a PC and digitised with 16 bit resolution at 16 kHz. For subjects MB and AS a Soundlab G162 condensor microphone was used. The EPG sampling rate is approximately 200 Hz, the EMA sampling rate 500 Hz. The audio, EPG and EMA signals are synchronised to within the 5 ms accuracy imposed by the EPG sampling rate. The Articulograph electronics unit, which is situated in the studio, produces a hum at around 160-200 Hz, which is audible in the first recording, that of EB. For this reason, a second audio recording was made simultaneously with the first for EB with an Audio Technica AT803B electret lapel microphone attached to the helmet worn by the subject, an Alice Mic-Amp-Pack 2 preamplifier and a Sony DTC690 DAT player. This recording was used for the subsequent acoustic analysis; all acoustic interpretation of articulatory data is based on the synchronised recordings made using the PC, including the first audio recording, as the quality of that recording was more than adequate to identify acoustic landmarks. For subjects MB and AS only one audio recording was made as for these experiments the Articulograph electronics unit had been placed in a box which muffled the noise.

A plastic T-bar was made to approximate each subject's occlusal plane (as described in Hoole (1996) and Westbury (1991; 1994), and to determine a co-ordinate system in which to situate the EMA data. The T-bar was marked (using a cast of the subject's upper teeth) so that coils could be placed at the junction of the front incisors and on the midline at the back of the hard palate. Each subject had different dentition, which made it impracticable to use teeth as anatomical landmarks for establishing a co-ordinate axis system which would be comparable across subjects.

3.2.3. Procedure

Prior to each experiment the sensor coils were sterilised and coated in latex, the Articulograph was calibrated and the other equipment was set up. The subject was seated beneath the helmet suspension point and was asked to insert the artificial palate. The EMA coils were then glued into place: on the bridge of the nose, the gum above the upper incisors, the lower gum, the upper and lower lips and three on the tongue. The coils on the tongue were placed approximately as follows: the 'tongue tip coil' 1cm back from the tip of the tongue (on the midline), the 'tongue mid' coil

2.5cm further back and the 'tongue back' coil a further 2cm back, opposite the soft palate. After attachment of the coils, the helmet was placed on the subject and positioned so that the coils lay in the midsagittal plane.

The stimuli were printed on 9 sheets of A4 paper. The experimenter and subject each had a copy, and the subject held the relevant stimulus sheet in his hand. Each stimulus sentence was recorded individually. The equipment operator signalled to the experimenter for recording to begin and the subject was instructed to read two sentences at a time: a stimulus sentence and a filler sentence. The subject was not informed which sentences were of interest and was under the impression that all sentences were recorded, although only the stimuli were recorded.

During subject EB's recording session, after 55 sentences had been recorded, the equipment operator became concerned that the tongue tip coil was not properly attached. It appeared that the coil had rotated sideways slightly and it was therefore re-attached. This affected the data used in analysis as discussed below. Data from subject MB had a different problem: due to an underpowered neck transmitter the reference coils fell into an area of instability in the magnetic field and had to be smoothed and corrected, as is described in 3.3 below. The weak transmitter signal resulted in greater fluctuation in coils placed far from the neck transmitter. For subject AS, two coils were placed on the bridge of the nose, in an attempt to avoid the problem found during MB's recording session. One of these coils (placed slightly lower on the nose) proved more reliable, although it was still affected by the discontinuity in measurement. The upper incisor reference coil in this data set did not fall into the area of instability.

After all of the stimuli had been read, the T-bar (with two coils attached in the marked positions) was inserted into the subject's mouth and the position of the coils recorded. Positioning of the T-bar proved difficult for EB, as the subject had an overbite, so he was instructed to push the T-bar up against his upper teeth with his tongue. For subject MB the rear coil position was not recorded due to (accidental) reversal of polarity of the coil, and the position is estimated based on the position of the mid tongue coil and the thickness of the T-bar. Finally traces of the subjects' palates were made. For EB, a coil was glued to the subject's right index finger and he was instructed to draw it along his palate. Due to head movement during this task and the subject's difficulty in keeping the coil correctly aligned, the task was repeated several times until a satisfactory palate trace was obtained. For MB and AS, a flexible plastic rod with a coil attached to it was used to trace the palate. This produced much less head movement and more reliable palate traces.

3.3. Data analysis

3.3.1. Post-processing of the data

Immediate post-processing of EB's data was done at QMC by Alan Wrench, the research officer. Following QMC protocol, *Tailor* and *Emalyse*, programs produced by Carstens, were used for post-processing and checking the EMA data. The reference coil positions were smoothed by filtering, and the data was then corrected for head movement. Head movement correction translates and rotates each frame to ensure that the reference coils are placed at the same geometric co-ordinates. This corrects for head movement backwards, forwards, up and down relative to the helmet. The algorithm may fail if the reference coils do not remain a fixed distance apart, which may happen as a result of movement of the skin under the bridge of the nose reference coil.

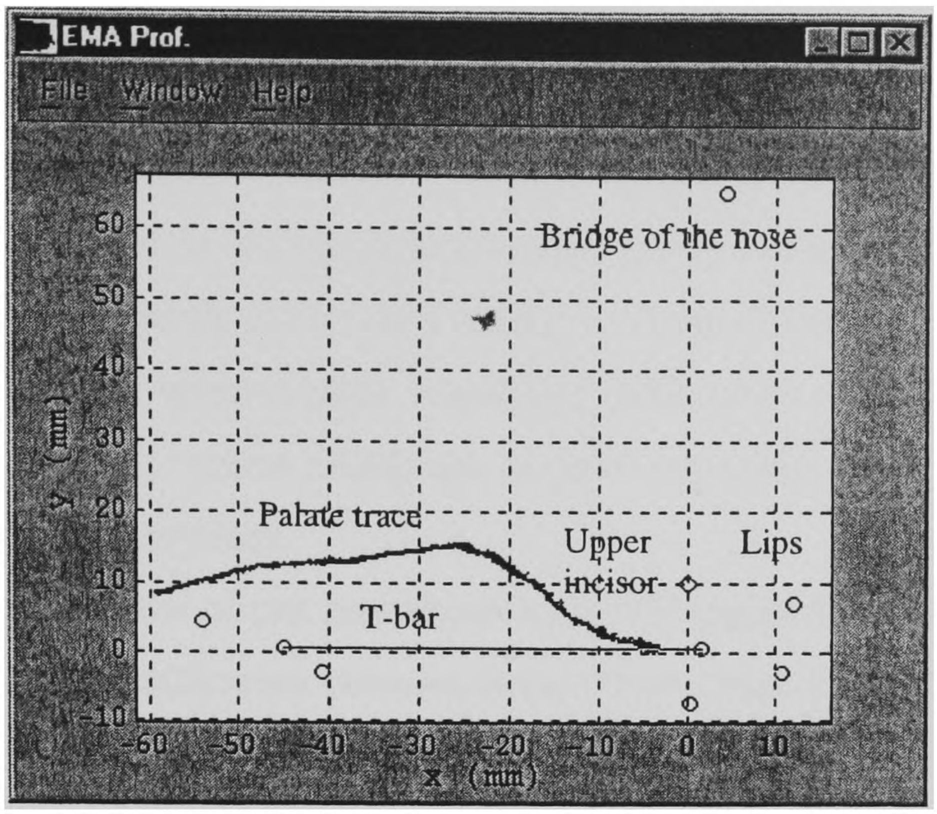
During post-processing, it became apparent that the receiver coils on EB's tongue tip had been malfunctioning until the coil was reattached. In some utterances there were large spikes in the data where the positional data for the tongue tip coil had registered falsely. These errors appear to have been triggered by bilabial closure, which put pressure on the connecting wire from the tongue tip coil and rotated the coil. Although these errors were obvious and easy to spot, and could have been compensated for by smoothing, all data containing the errors were discarded. This left 7 repeats of each stimulus sentence for analysis in EB's data set.

Twelve repeats were obtained and used for MB and AS. Post-processing of this data was done by the author in Oxford, as the reference coil data needed to be adjusted. For subject AS, the horizontal position of the bridge of the nose coil required adjustment. The instability in the measurement field caused large spikes to appear in the data. These were easily visible and were removed automatically with a *Matlab* function written for this purpose. Differences due to spikes were all found to be larger than 0.23 mm. Traces were sequentially examined by the function, and when a difference in adjacent samples of more than 0.23 mm was found, the second sample was set to the value of the first, and the spike thus removed. At the first spike in the data, a step down of about 1.5 mm systematically occurred. This was corrected by setting the sample value after the step to the average of the ten preceding and ten following samples. The correction required was obvious on inspection of the unaffected data (63 out of 144 utterances were affected). All of the coils were smoothed using a 10 point moving average smoothing algorithm. Head correction was performed, again using *Matlab* routines written by the author. The head correction algorithm was one based on Carstens' procedure: the midpoints on the lines between the two reference coils were aligned for each sample and the data was rotated to make the lines congruent. In most cases the rotation required was so small that *Matlab* did not perform rotation (rounding of the trigonometric functions gave the angular rotation required as 0).

For subject MB, both the bridge of the nose and upper incisor coil positions were processed to remove noise. Low pass filtering was the technique chosen, as there was some periodicity in the noise. 7th order elliptical filters were used to filter out all frequencies above 3 Hz. This preserved the overall shape of the signal, but removed most of the noise. The data was corrected for head movement as described for AS.

For all subjects the coil co-ordinates were then rotated so that the T-bar, which represented the occlusal plane lay parallel to the x-axis. The data was translated so that the origin of the co-ordinate system lay 10mm (for EB and MB) and 12mm (for AS) below the mean position of the upper incisor coil for each utterance. These distances are the length of the central maxillary incisors as measured from a dental cast. This point was estimated to be at the junction of the central-maxillary incisor diastema and the incisor's exposed tips. The x data was then reflected around the x-axis to comply fully with the descriptive standard proposed by Westbury (Westbury 1994). Figure 3.1 shows the palate trace (curved line) and T-bar trace (straight line), together with positions of coils (open circles) for the upper and lower lips and incisors, tongue middle and tongue back, upper and lower lips and bridge of the nose taken from the T-bar trace recording for subject EB. The origin of the co-ordinate system is slightly posterior to the front coil on the T-bar, which is approximately equidistant from the upper and lower incisor coils in this articulatory configuration.

Figure 3.1: Palate trace and T-bar trace: subject EB.



3.3.2. Statistical Analysis: multivariate general linear models

Multivariate general linear models (constructed in *SPSS*) are used throughout the analysis of the EMA and acoustic data. Multivariate tests, as opposed to univariate ANOVAs, take into account the correlations between variables (Chatfield and Collins 1980, 149). As the articulatory and acoustic variables are intercorrelated, multivariate models are the appropriate choice. Measurements are assumed to be independent, *i.e.* the error term in these models is assumed to be a vector of independent errors, and a repeated measures model is not used.

Use of multivariate general linear models (multivariate GLMs) requires certain assumptions about the data to be met: multivariate normal distribution of errors and equality of the variance-covariance matrices. These assumptions correspond to normality of distribution and equality of variance required for univariate ANOVAs. Tests for multivariate normality are complex, and not provided by *SPSS* or any of the standard statistical packages. However, most multivariate analysis techniques are relatively unaffected by non-normality, provided that the data is fairly symmetric, and symmetry can be examined by treating each variable separately (Krzanowski 1988, 211-215). Thus, in this study, the normality of each variable is examined separately, by means of one-sample Kolmogorov-Smirnov tests for the normal distribution. When these tests show significant departures from normality, alternative analyses of the data are pursued. Equality of the variance-covariance matrices can be tested by using Box's *M* test in *SPSS*. This test cannot always be calculated for small samples, and is sensitive to multivariate normality. When Box's *M* test cannot be calculated, Levene's test for equality of error variance, which tests for homogeneity of variance for each variable across all combinations of levels of factors, is used as a best approximation (*SPSS* online help, version 7.5.1 December 1996). The result of the data failing to meet the assumptions is one of conservativity: the test under-reports significant differences instead of producing spurious results. For this reason GLMs are occasionally constructed for data which does not meet these assumptions.

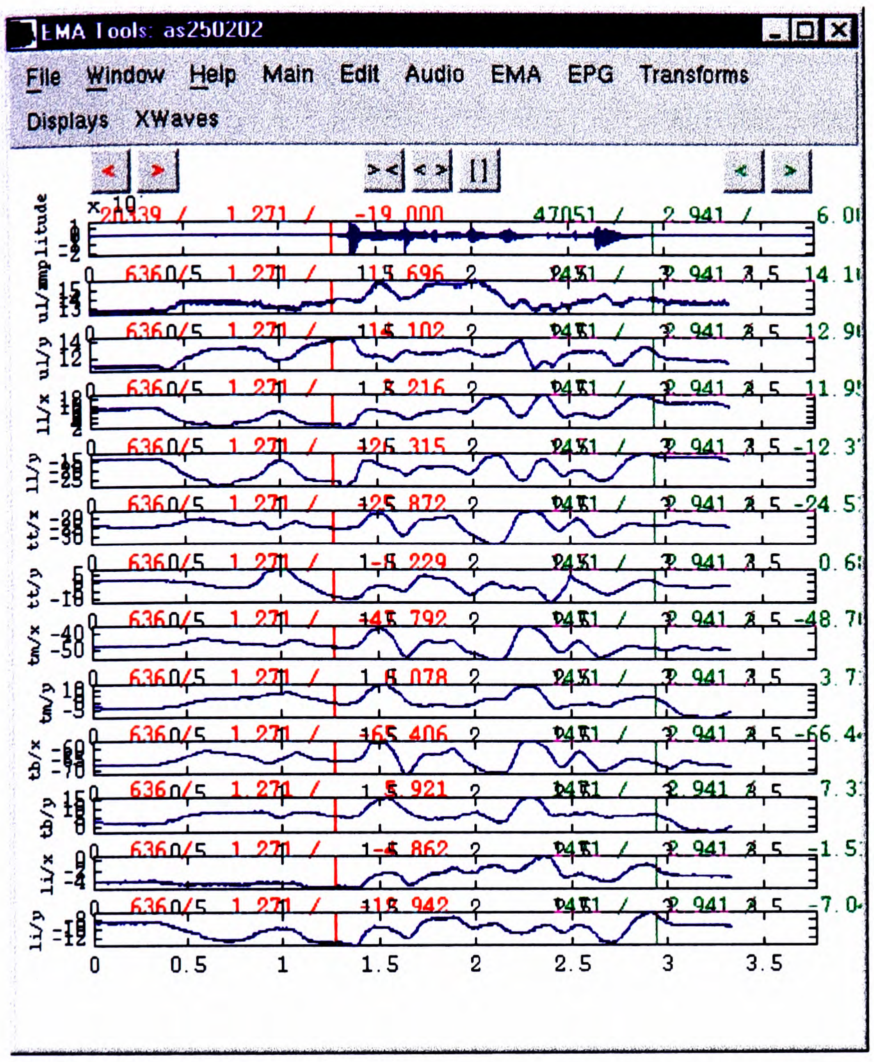
There are a number of test procedures used with multivariate GLMs and there appears to be very little difference between them. Pillai's trace has been claimed to be slightly more robust than the others (Chatfield and Collins 1980, 146-148), and is thus used in this study.

3.3.3. Analysis tools

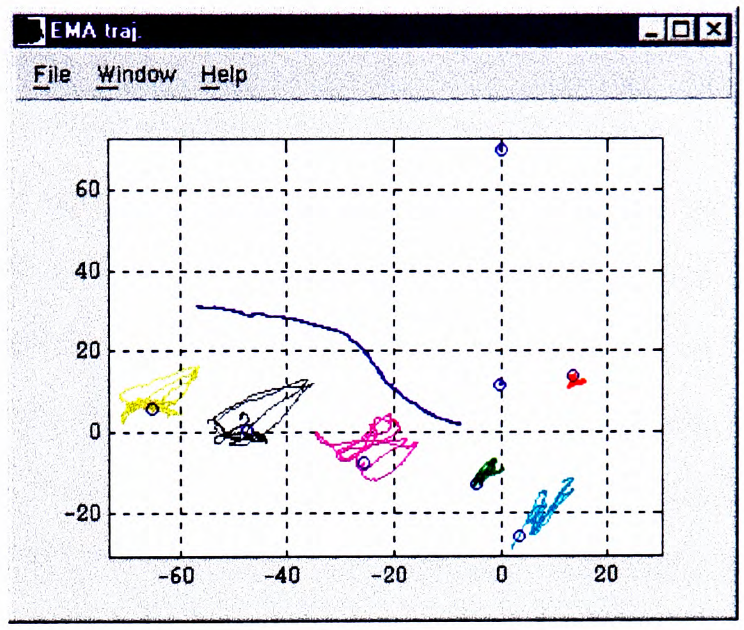
Analysis of all data was performed using *Matlab*, a technical computing environment for numeric computation, signal processing and visualisation. A set of *Matlab* programs, developed by Noël Nguyen (Nguyen 1996) was used, modified as necessary by the author. Most of the processing was done using additional programs written in *Matlab* by the author and run on a Silicon Graphics Indy computer. Figure 3.2 contains some of the displays available for viewing the articulatory and acoustic data (taken from AS's data): (a) shows an acoustic waveform (the top trace) and the 12 EMA traces, *x* and *y* traces for each of the 6 coils (b) the EMA coil movement trajectories between the left and right cursor positions in (a), (c) the EPG contact data at the left and right cursor positions and (d) the EMA coil positions at the cursors, with the 3 tongue coils joined by lines.

Figure 3.2: Audio, EPG and EMA data displays.

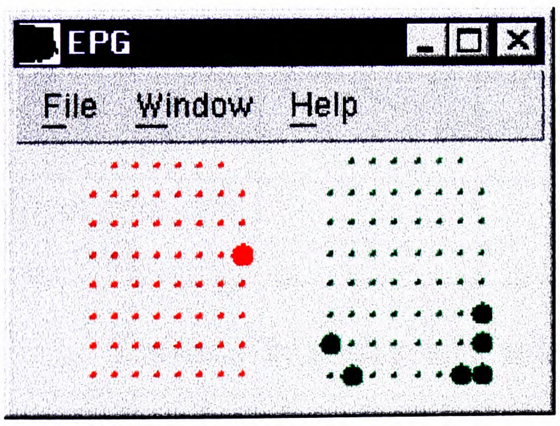
(a) Acoustic and EMA window



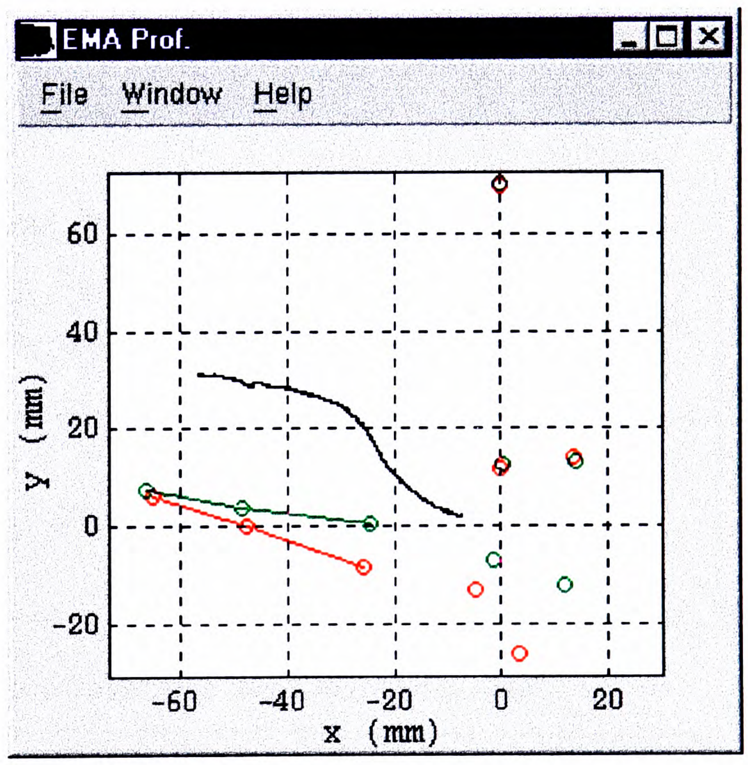
(b) EMA trajectory window



(c) EPG pattern window



(d) EMA profile window



3.3.4. Reliability of the EMA data

The first check made on the EMA data was on the reliability of cross-utterance comparisons, or the stability of reference coils from one utterance to the next. Preliminary inspection of the data suggested that there were some inconsistencies from one utterance to the next. Within each utterance there was very little variability in reference coil positions due to the effectiveness of the head correction algorithms used. However, the assumption of no head movement and fixed distance between the two reference coils (bridge of the nose and upper incisor coils) was not met, and is probably never met under experimental conditions. This resulted in slightly different positioning of the reference coils from one utterance to the next, as the angle and distance between the coils differed occasionally.

To check whether this constituted a serious problem for analysis, the mean *x* and *y* positions of the bridge of the nose and upper incisor coils were calculated for each utterance, as was the Euclidean distance and the angle between the coils. From this data, the mean positions, Euclidean distance and angle between the coils were calculated across all utterances, together with standard deviations. The results are given in Table 3.1, Table 3.2 and Table 3.3. Note that the small standard deviations associated with the upper incisor coil are a result of the use of this coil as a reference for the co-ordinate system. The reference coils were deemed sufficiently stable relative to each other for analysis to proceed. This data supports claims of positional accuracy to within 0.5 mm for EB. For MB and AS, for whom there were problems with the reference coil data, the reference coil accuracy is probably not a good indication of the accuracy of the other coil positions.

Table 3.1: Reference coil data: mean (standard deviation) for EB.

	x position, in mm	y position, in mm	Euclidean distance between the coils, in mm	angle between the coils, in degrees
Bridge of the nose coil	4.79 (0.08)	63.22 (0.29)	53.44 (0.28)	5.15 (0.11)
Upper incisor coil	0 (0)	10 (0)		

Table 3.2: Reference coil data: mean (standard deviation) for MB.

	x position, in mm	y position, in mm	Euclidean distance between the coils, in mm	angle between the coils, in degrees
Bridge of the nose coil	0.5 (0.31)	79.29 (0.12)	69.23 (0.17)	–0.41 (0.48)
Upper incisor coil	0.01 (0.28)	10.06 (0.11)		

Table 3.3: Reference coil data: mean (standard deviation) for AS.

	x position, in mm	y position, in mm	Euclidean distance between the coils, in mm	angle between the coils, in degrees
Bridge of the nose coil	0.13 (1.55)	70.08 (0.25)	58.10 (0.20)	–0.13 (1.53)
Upper incisor coil	0 (0)	12 (0)		

3.4. Acoustic measurements

3.4.1. Preliminary acoustic measurements

Preliminary acoustic investigation of the subjects' speech showed that they have the resonance contrasts of interest and make long distance distinctions in minimal pair sentences: statistically significant differences in formant values were found in the schwas indicated by boldface orthography in 'Have you utt**ered** a — **at** home?', depending on liquid beginning the word placed in the slot '—'. The schwa in *uttered* is labelled 'schwa 1', the schwa in *at* 'schwa 2'.

Before the articulatory experiment was conducted, each subject was recorded producing a set of test sentences: ten (EB) or twelve (MB and AS) of each of the twelve minimal pair words *leap*, *reap*, *lip*, *rip*, *lap*, *wrap*, *lope*, *rope*, *lobe*, *robe*, *lob* and *rob* in the frame sentence. Measurements of formant frequencies were made at the temporal midpoint of each vowel using 18 pole Burg spectra with a 50 ms Hanning window. Formant frequency values were checked against values read from 18 pole 50 ms discrete Fourier transform (DFT) spectra and wide-band spectrograms created in *Waves*.

For each subject, a multivariate general linear model was constructed in *SPSS* with liquid and word pair as factors and the formant frequencies as dependent variables. For subject EB the model is slightly problematic, as the assumption of normality required for a multivariate GLM does not hold: the distribution of F_1 of schwa 1 (the schwa in *uttered*) is significantly different from normal (one-sample Kolmogorov-Smirnov test gives $Z = 1.681$, $p < 0.007$). Nevertheless, Box's test of equivalence of covariance matrices shows a non-significant result ($\alpha = 0.01$). Both liquid and word pair were highly significant factors (Pillai's trace: $F(6,103) = 16.520$, $p < 0.001$ and $F(30,535) = 8.977$, $p < 0.001$ respectively) and there was a significant interaction between word pair and liquid ($F(30,535) = 1.750$, $p < 0.009$). The interaction was largely due to F_2 differences in both schwas and F_3 differences in schwa 1.

Consequently, independent samples *t*-tests were conducted on each word pair separately to explore the interaction. (These results were all replicated at the 5% level using the Mann-Whitney non-parametric test, confirming the robustness of the *t*-test under these conditions). All pairs had a significant difference ($p < 0.05$) between /l/ and /r/ contexts, but these were found in different formants and not always in both schwas. The significant results are displayed in Table 3.4 below. The significant differences are all in the second and third formants: F_1 values appear unaffected by resonance context. There are a few differences in schwa 2, but these are less consistent than the differences in schwa 1. Mean formant frequencies for each word pair are displayed in Figure 3.3 and Figure 3.4. In general, F_2 is lower and F_3 higher for /l/ contexts.

Table 3.4: Significant ($p < 0.05$) results of independent samples *t*-tests on acoustic data of subject EB.

Word Pair	Schwa 1 F_2	Schwa 1 F_3	Schwa 2 F_2	Schwa 2 F_3
leap/reap		$t(18) = 4.391$ $p < 0.001$		
lip/rip	$t(18) = -4.397$ $p < 0.001$			
lap/wrap	$t(18) = -2.557$ $p < 0.02$	$t(18) = 2.541$ $p < 0.02$		
lope/rope	$t(18) = -3.363$ $p < 0.003$	$t(18) = 2.131$ $p < 0.047$	$t(18) = -3.025$ $p < 0.007$	
lobe/robe	$t(18) = -4.379$ $p < 0.001$			$t(18) = 2.446$ $p < 0.025$
lob/rob	$t(18) = -4.215$ $p < 0.001$		$t(18) = -2.625$ $p < 0.017$	

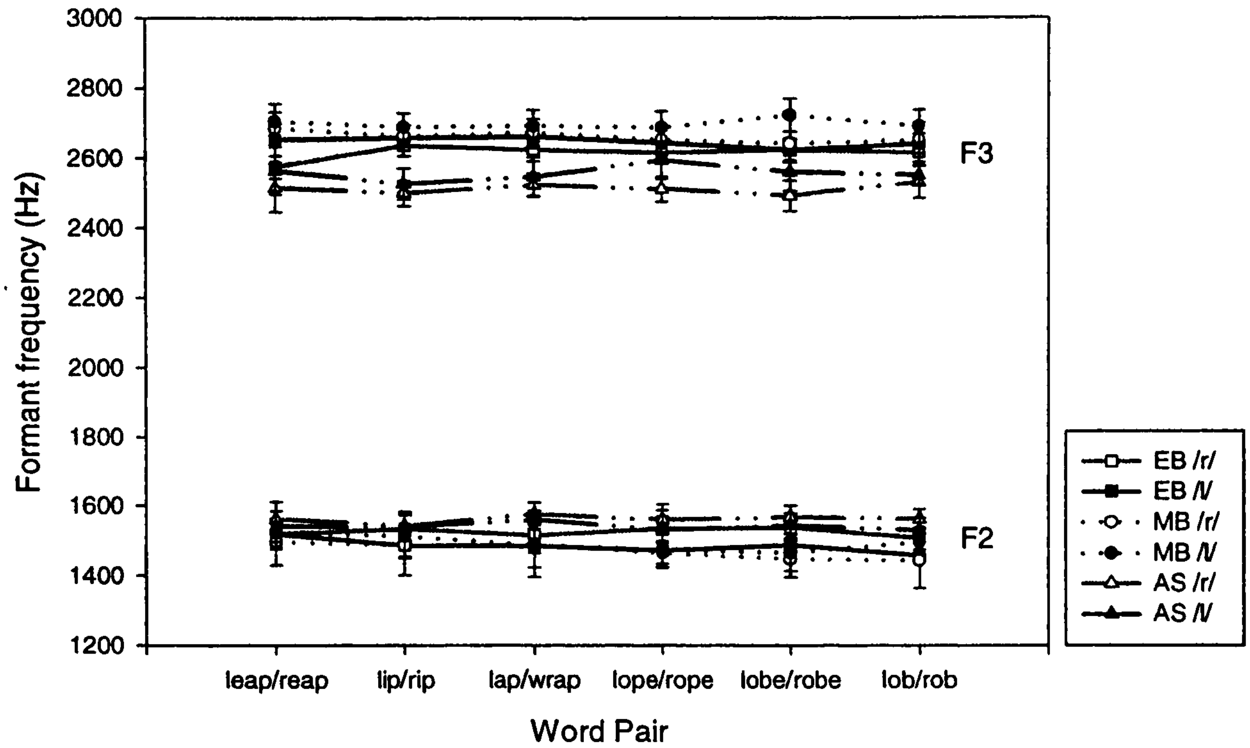
There are no departures from normality for subject MB's acoustic data, and Box's test is not significant. In the multivariate GLM, both factors liquid and word pair were highly significant (Pillai's trace: $F(6,125) = 5.463$, $p < 0.001$ and $F(30, 645) = 6.034$, $p < 0.001$) with no interaction between them. The differences in liquid are a result of differences in F_2 and F_3 of schwa 1.

The GLM for subject AS is unproblematic, with no significant departures from normality and a non-significant result for Box's test. Only the factor liquid was highly significant for this subject (Pillai's trace: $F(6,94) = 9.620$, $p < 0.001$), with significant differences in F_2 and F_3 of schwa 1 and F_3 of schwa 2.

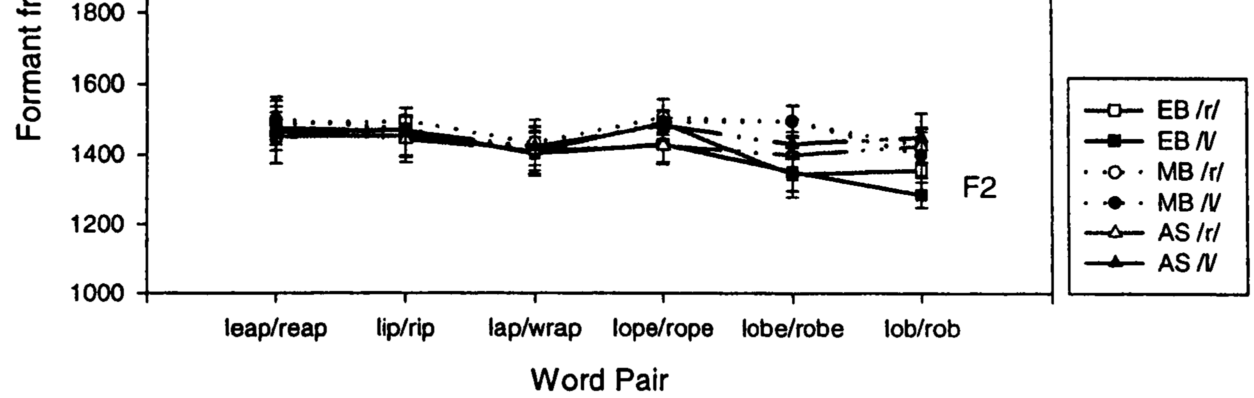
Figure 3.3 (a) shows mean F_2 and F_3 calculated for each word pair for schwa 1, for all three speakers. Figure 3.3 (b) shows the same data for schwa 2. Note that for each speaker, F_3 is lower in the /r/ context than the /l/ context, whereas F_2 results are more varied.

Figure 3.3: Mean formant frequencies for schwas (indicated by capitals) in the frame sentence 'Have you uttEREd a — At home?', for all three subjects. /l/ and /r/ stimuli shown separately.

(a) Schwa in *uttered* (schwa 1)



(b) Schwa in *at* (schwa 2)



3.4.2. The effect of EPG

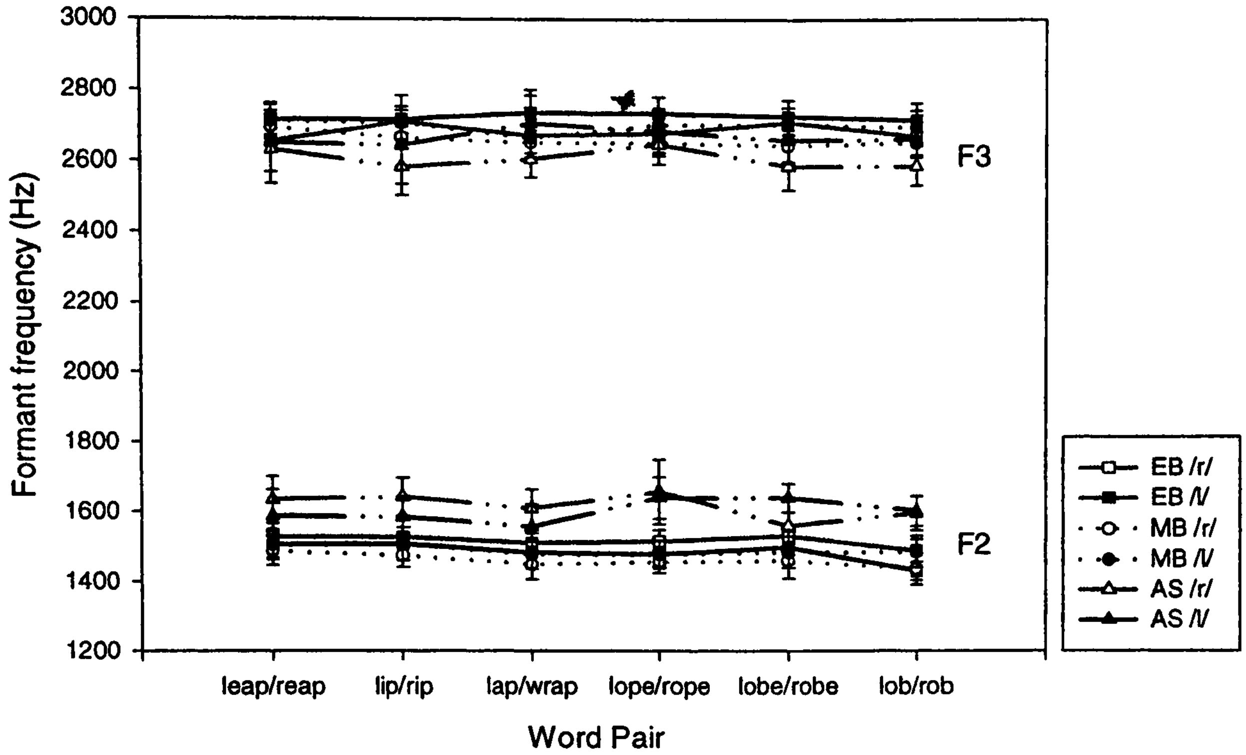
To examine the effect of EPG on acoustic output, the subjects' speech was recorded with their artificial palates inserted, and was analysed in the same way. Twelve repeats of the same test sentences were used (ten for EB). This procedure was intended as a check, to make sure that the subtle distinctions found in the preliminary study would not be lost when the palates were worn.

Multivariate GLMs were constructed for all subjects, as for the previous data sets. For subjects EB and AS the assumptions behind the model were met, as tested by Box's test and one-sample Kolmogorov-Smirnov tests. Subject MB's data gave a significant result for Box's test ($F(231, 13582) = 1.289$, $p < 0.002$), rendering use of the model problematic. For all subjects, liquid and word pair were significant factors (EB: $F(6,103) = 13.231$, $p < 0.001$ and $F(30,535) = 9.367$, $p < 0.001$; MB: $F(6,125) = 7.684$, $p < 0.001$ and $F(30,645) = 5.860$, $p < 0.001$; AS: $F(6,85) = 5.727$, $p < 0.001$ and $F(30,445) = 2.328$, $p < 0.001$). There were no interactions between word pair and liquid. The significance of the factor liquid is due to highly significant differences in F_3 of both schwas for all subjects, and F_2 of schwa 1 for EB and MB. As the model for MB was problematic, a one-way ANOVA was conducted for all variables with liquid as factor. (Levene's test for homogeneity of error variance showed no significant results and the data is normally distributed as mentioned above). The ANOVA produced the same significant results for the factor liquid as the GLM and is thus not reported further here.

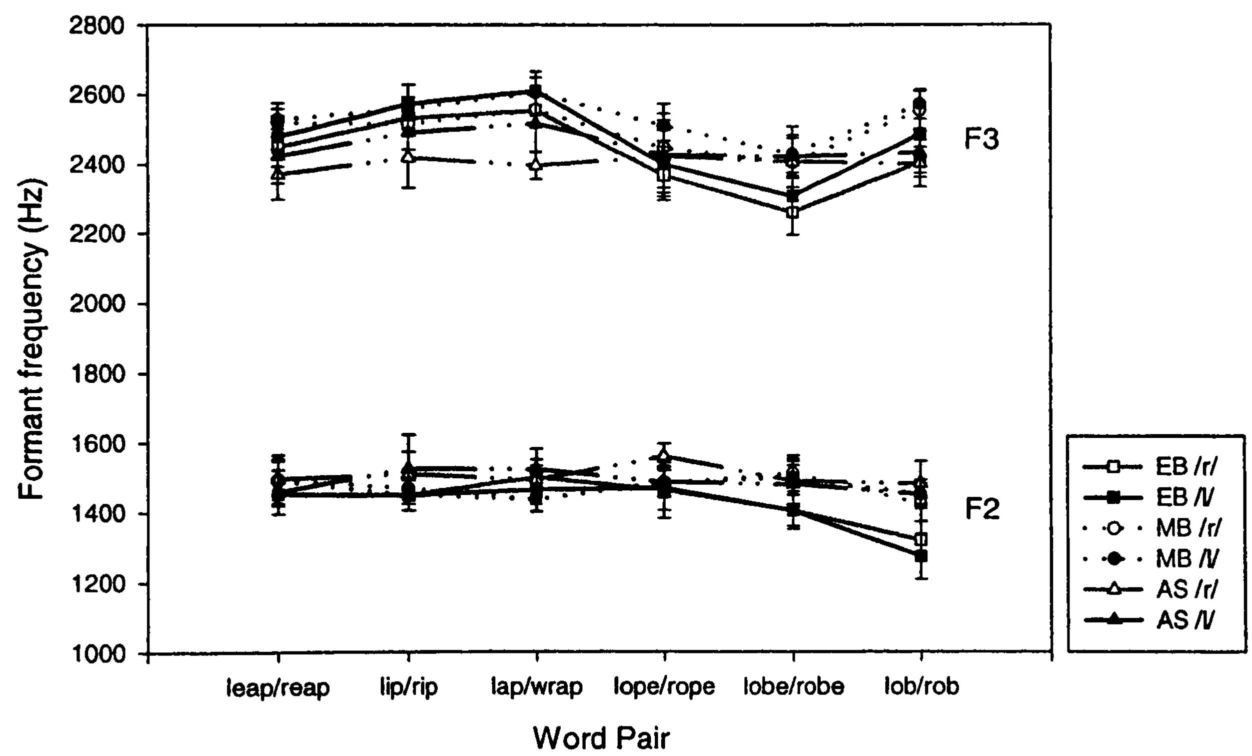
Mean F_2 and F_3 for schwa 1 and 2 in each /l/ and /r/ word context are plotted in Figure 3.4 for each subject separately. Although there seems to be some interference from the palate, all subjects made the distinctions of interest quite clearly, with obviously lower F_3 in the /r/ contexts. This result justified the decision to conduct the EMA experiment simultaneously with EPG.

Figure 3.4: Mean formant frequencies for schwas (indicated by capitals) in the frame sentence 'Have you uttEREd a — At home?', for all three subjects wearing artificial palates. /l/ and /r/ stimuli shown separately.

(a) Schwa in *uttered* (schwa 1)



(b) Schwa in *at* (schwa 2)



3.4.3. The effect of EMA and EPG

The acoustic data recorded during the EMA and EPG sessions were analysed to check whether the articulatory instrumentation had affected the acoustic output significantly, and to see whether long-distance coarticulatory differences between /l/ and /r/ contexts were present during this experimental session. For EB two acoustic recordings had been made. The DAT recording, although not synchronised with the EMA and EPG recordings, was used for this analysis as the higher quality made measurement less prone to error (and easier). For MB and AS only one acoustic recording was made, which was synchronised with the articulatory data.

The formant frequency data from each subject was analysed separately. For EB and MB, one-sample Kolmogorov-Smirnov tests for the normal distribution showed no significantly non-normal distributions and Box's test was non-significant ($\alpha = 0.01$). AS's data showed two significant departures from normality: schwa 1 F_3 ($Z = 1.702$, $p < 0.006$) and schwa 2 F_1 ($Z = 1.643$, $p < 0.009$), but the result of Box's test was not significant. Multivariate GLMs with word and liquid as factors and the formant frequencies as dependent variables were constructed for all three subjects.

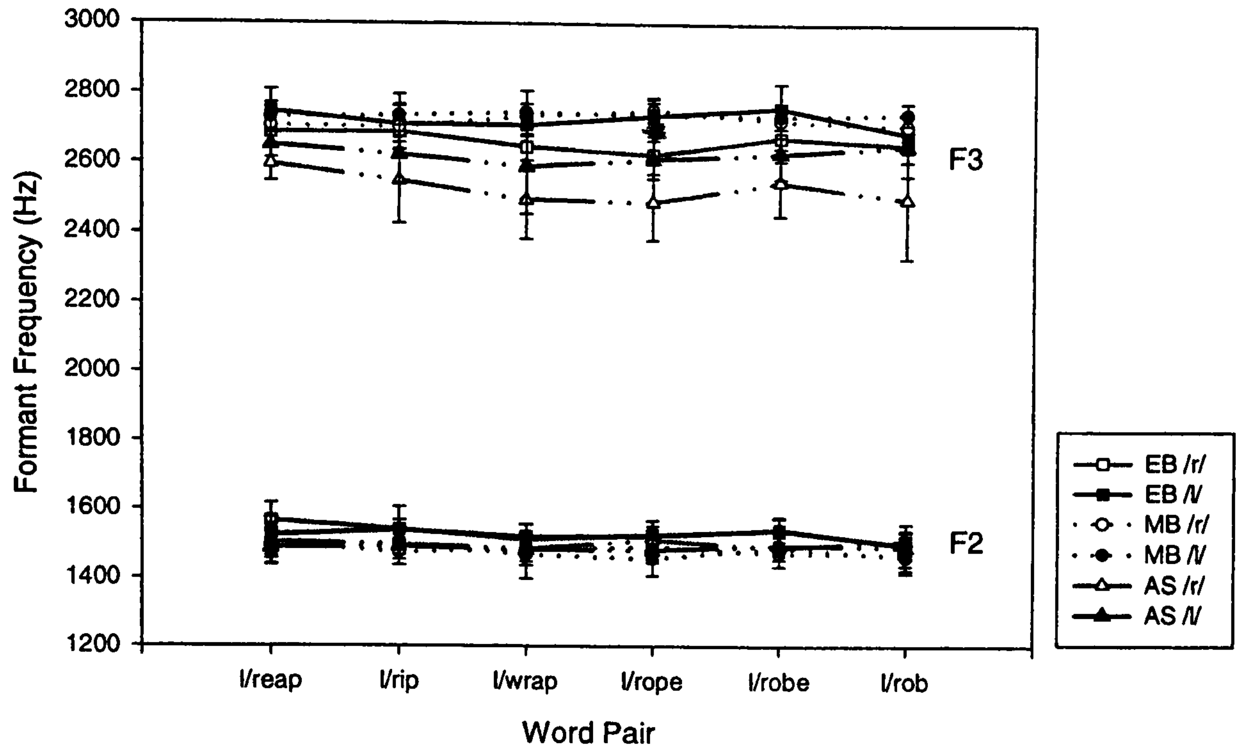
For all subjects liquid and word pair were highly significant factors (Pillai's trace: $F(6, 103) = 6.884$, $p < 0.001$ and $F(30, 535) = 10.264$, $p < 0.001$ for EB; $F(6, 125) = 8.12$, $p < 0.001$ and $F(30, 645) = 7.376$, $p < 0.001$ for MB and $F(6, 83) = 7.914$, $p < 0.001$ and $F(30, 435) = 4.411$, $p < 0.001$ for AS). MB was the only subject to show a significant interaction between the two factors liquid and word pair ($F(30, 645) = 1.821$, $p < 0.005$).This interaction was largely due to F_3 of schwa 2.

The significance of the factor liquid is due to differences in F_3 of schwa 1 for EB, and F_3 of schwa 1 and schwa 2 for MB and AS. The experimental environment caused loss some of the perseverative differences that were found in previous recordings (see Table 3.4 and Figure 3.3 above) for EB and confirms that these are less consistent for this speaker. AS and MB show both perseverative and anticipatory differences, as in their pre-tests.

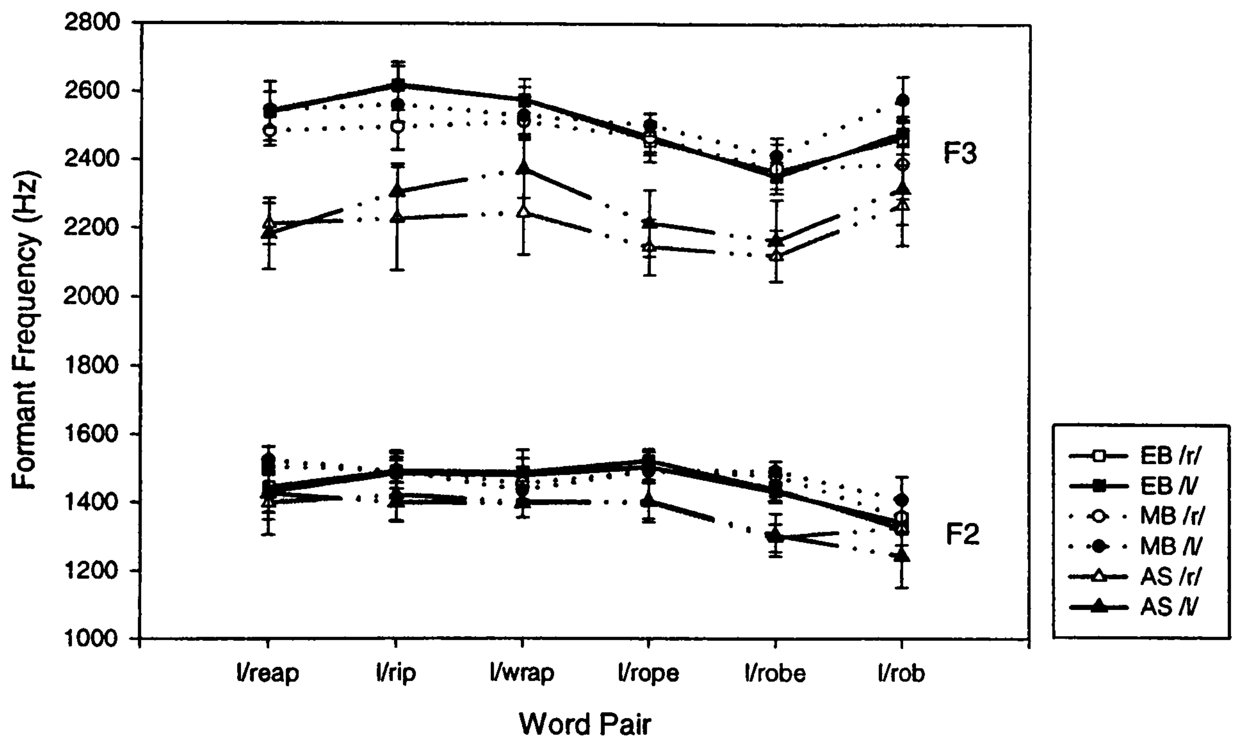
The means for all speakers are displayed in Figure 3.5 (a) and (b) for schwa 1 and 2 respectively. Note that mean F_3 is lower for most speakers in most word pairs in the /r/ context.

Figure 3.5: Mean formant frequencies for schwas (indicated by capitals) in the frame sentence 'Have you uttEREd a — At home?', for all three subjects with EMA coils and artificial palates inserted. /l/ and /r/ stimuli shown separately.

(a) Schwa in *uttered* (schwa 1)



(b) Schwa in *at* (schwa 2)



The significant coarticulatory differences in F_3, which have proved fairly robust across recording conditions for all subjects, require examination of the EMA and EPG data in search of an articulatory explanation. This also necessitates an examination of the relationship between EMA and acoustic data.

3.5. Local coarticulatory effects

The phrase 'local coarticulatory effects' is used throughout to refer to effects found in vowels adjacent to the liquid. Coarticulatory effects found in non-adjacent segments will be referred to as 'long-distance' or 'long-domain', and will be discussed in chapter four. Both the vowel before and after the liquid were examined for coarticulatory effects. These are discussed separately under the headings anticipatory and perseverative coarticulation. The results from each subject are presented in turn, as individual differences in anatomy and articulation argue against pooling of the data.

3.5.1. Anticipatory effects

To examine local anticipatory coarticulation, EMA and acoustic measurements were taken in the schwa before the liquid, the vowel of *a* in 'Have you uttered a — at home?'. The measurement point was defined as the midpoint of the vowel periodicity in the waveform, as observed from wide band spectrograms and the waveform. Initial segmentation of the schwa was relatively straightforward as all vowels were preceded by an alveolar plosive. In some cases this plosive was not realised with full closure, but the dip in amplitude was present. Segmenting the schwa from a following /r/ was difficult, so a conservative approach was used, counting transitions into the /r/ as part of the liquid. This procedure resulted in measurements being taken in the steady-state portion of the vowels.

At each measurement point, EMA positional data for the upper lip, lower lip, lower incisor, tongue tip, tongue mid and tongue back were collected in the form of (x, y) co-ordinate pairs. F_1, F_2 and F_3 were measured at the stipulated time-point using 18 pole DFT and Burg spectra with 0.025s Hanning and Hamming windows respectively, created in *Waves*. The measurements were imported into the graphical package *Sigmaplot* and the statistical package *SPSS* for display and analysis.

The EMA data collected is displayed in Figure 3.6 (EB), Figure 3.7 (MB) and Figure 3.8 (AS), and the acoustic data in Figure 3.9 (EB), Figure 3.10 (MB) and Figure 3.11 (AS), separated according to the following liquid.

Figure 3.6: Articulatory data, /ə/ before the liquid: subject EB.

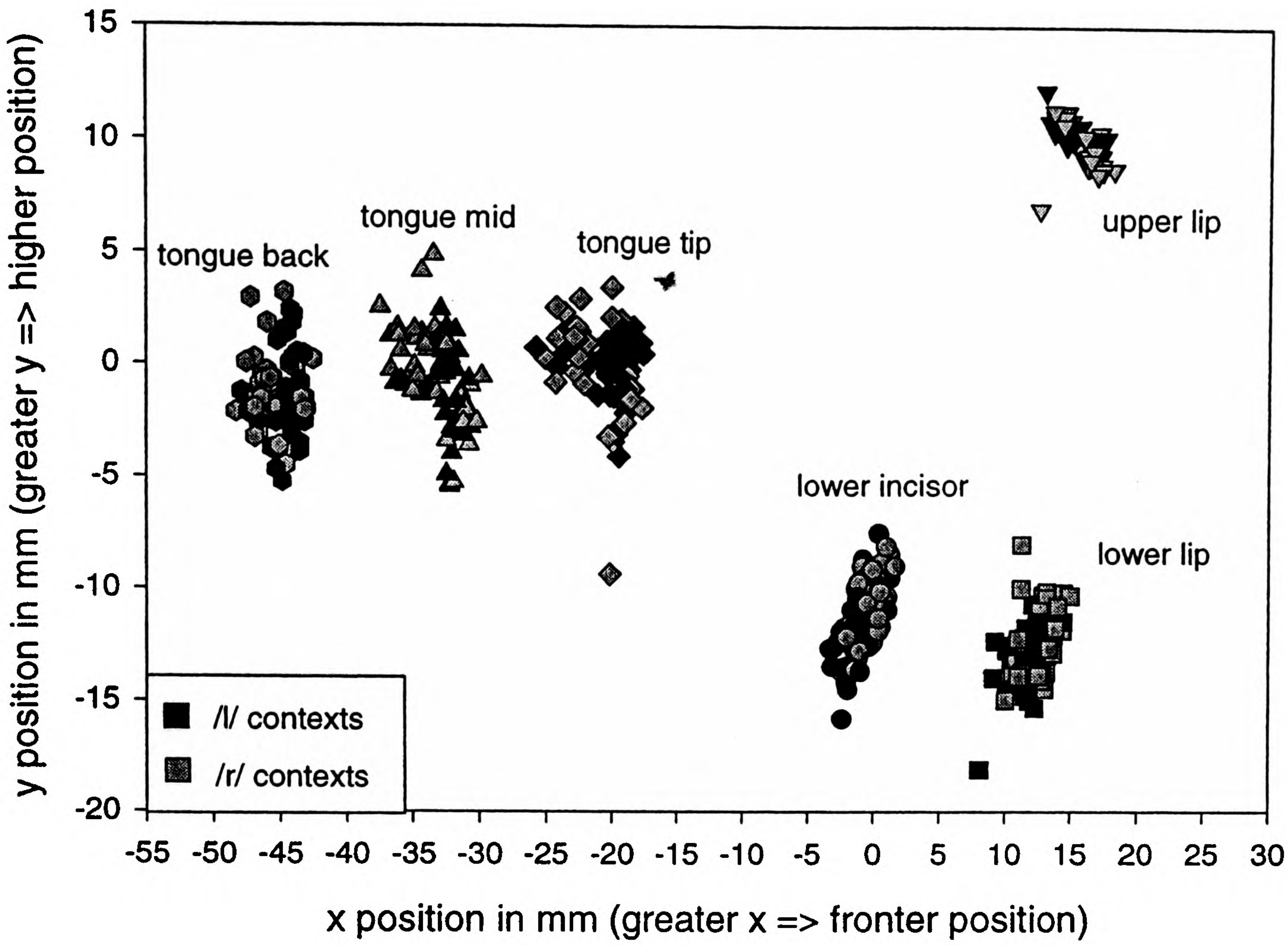


Figure 3.7: Articulatory data, /ə/ before the liquid: subject MB.

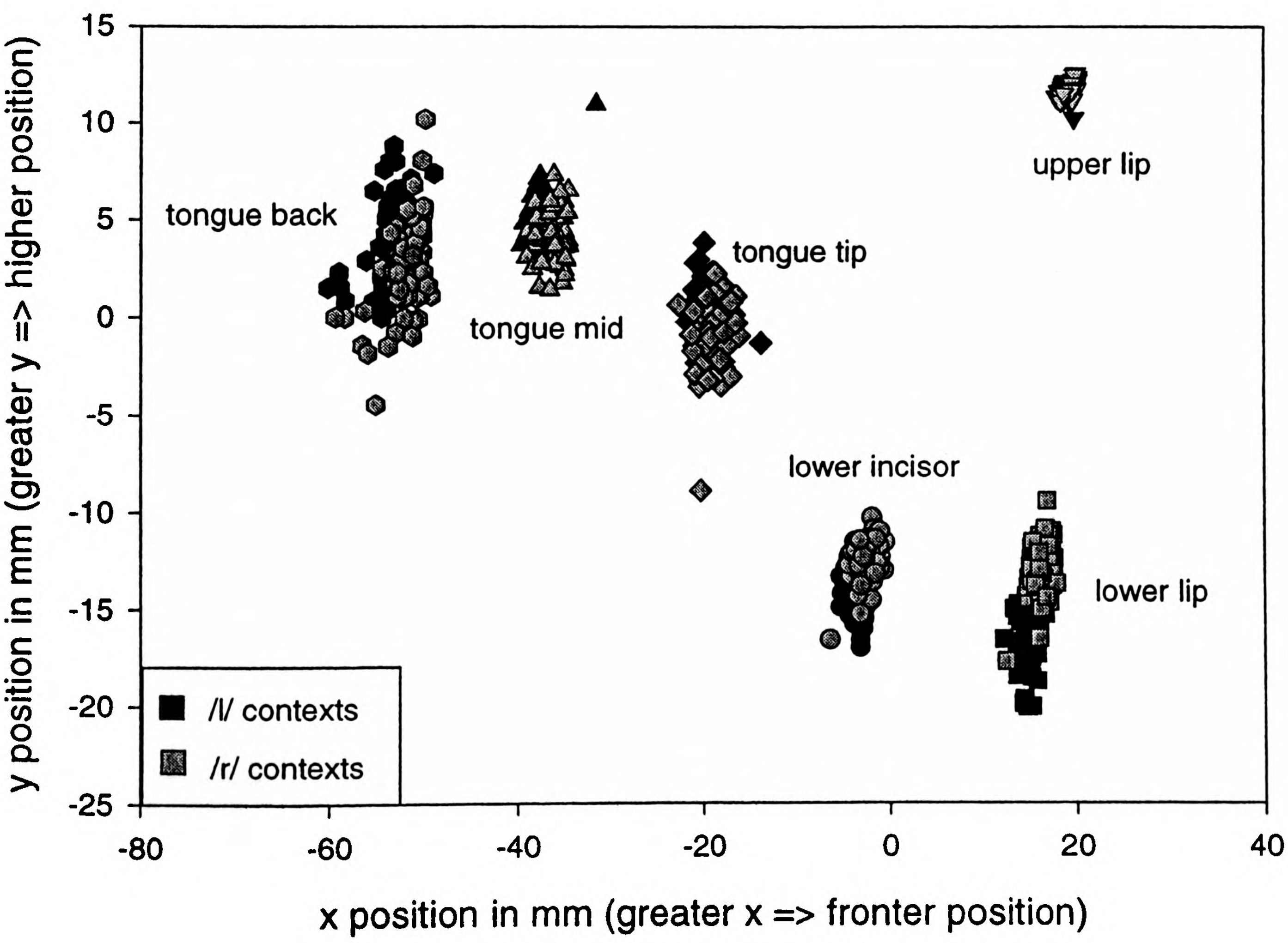


Figure 3.8: Articulatory data, /ə/ before the liquid: subject AS.

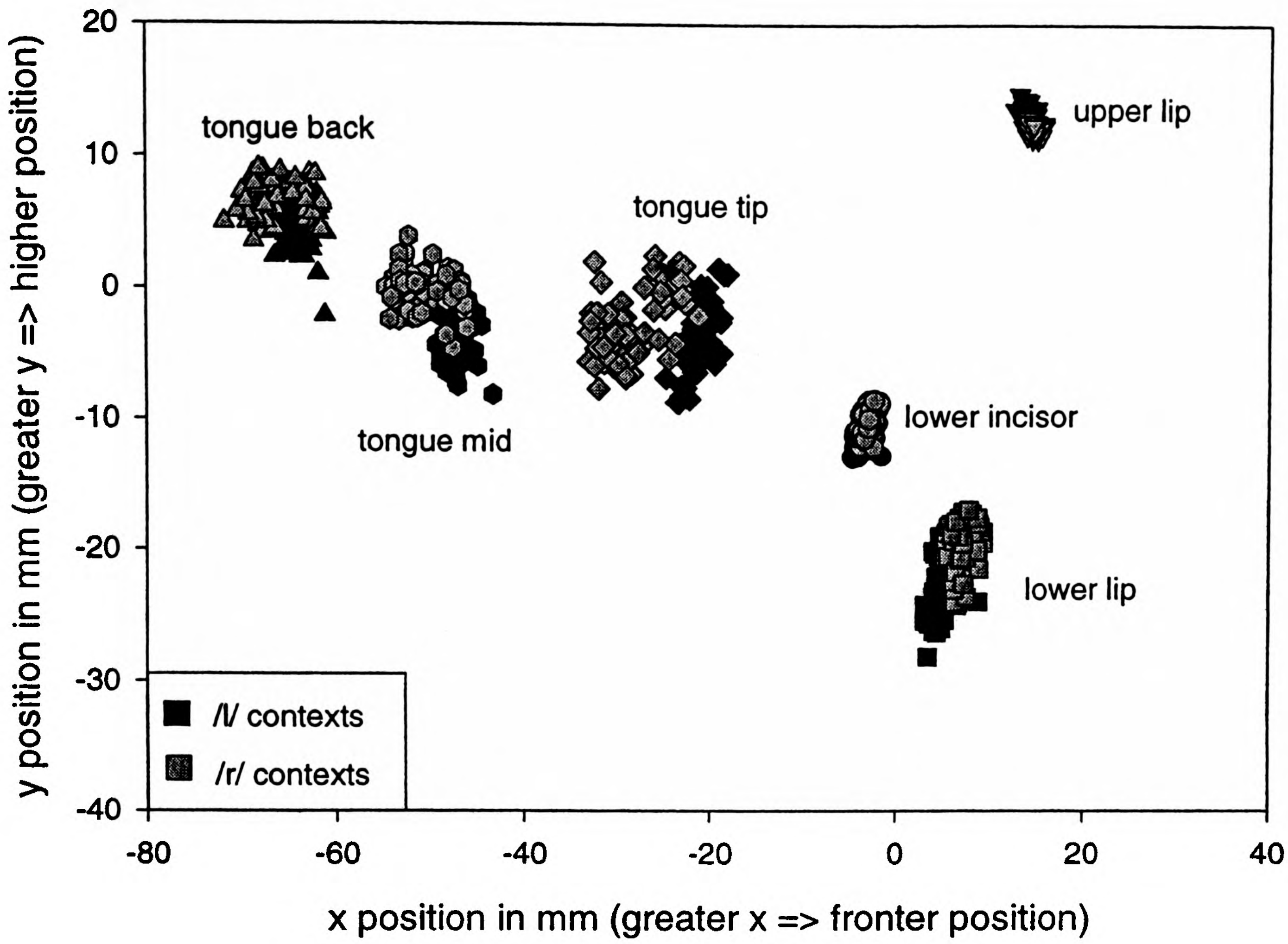


Figure 3.9: Acoustic data, /ə/ before the liquid: subject EB.

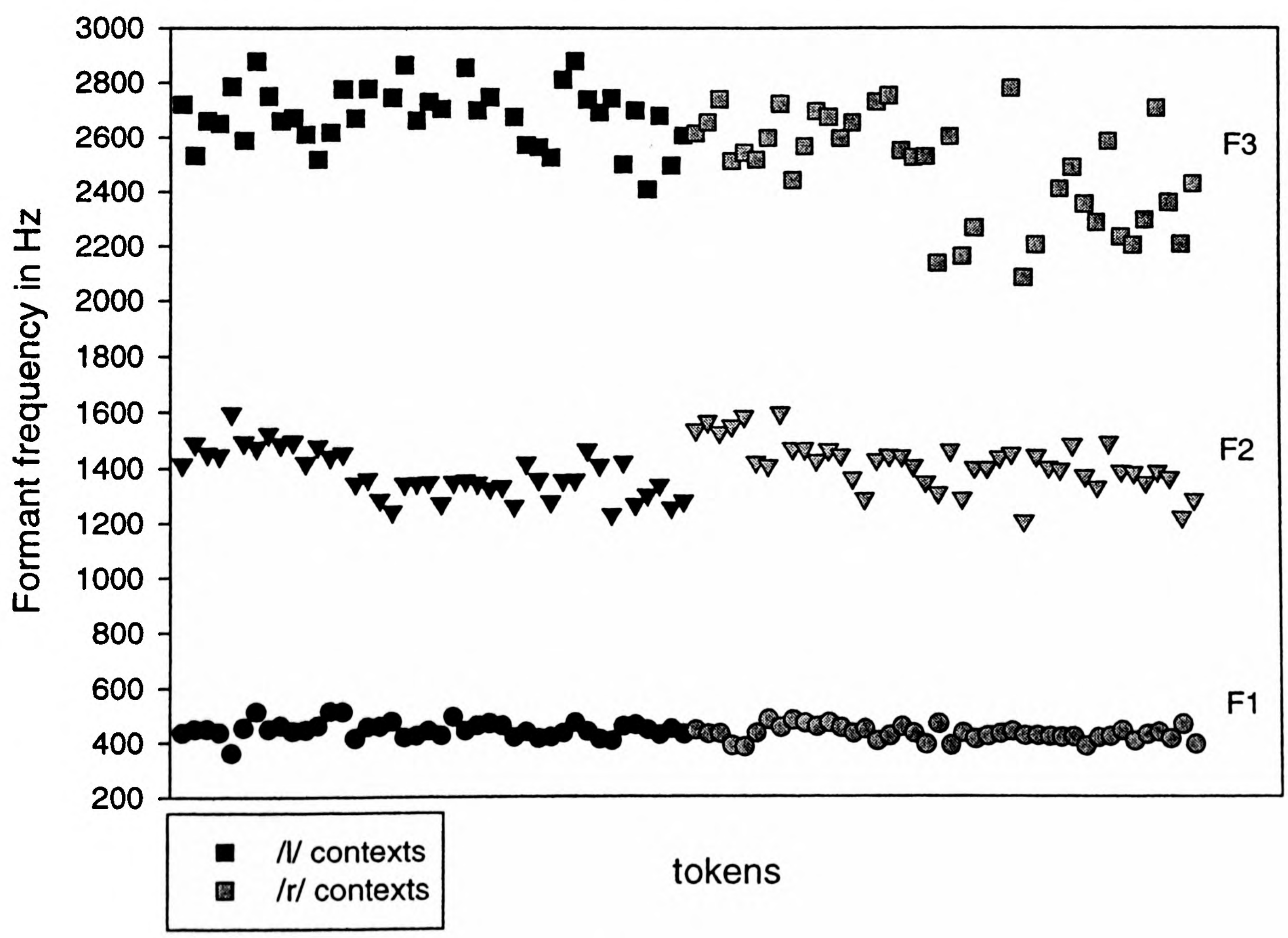


Figure 3.10: Acoustic data, /ə/ before the liquid: subject MB.

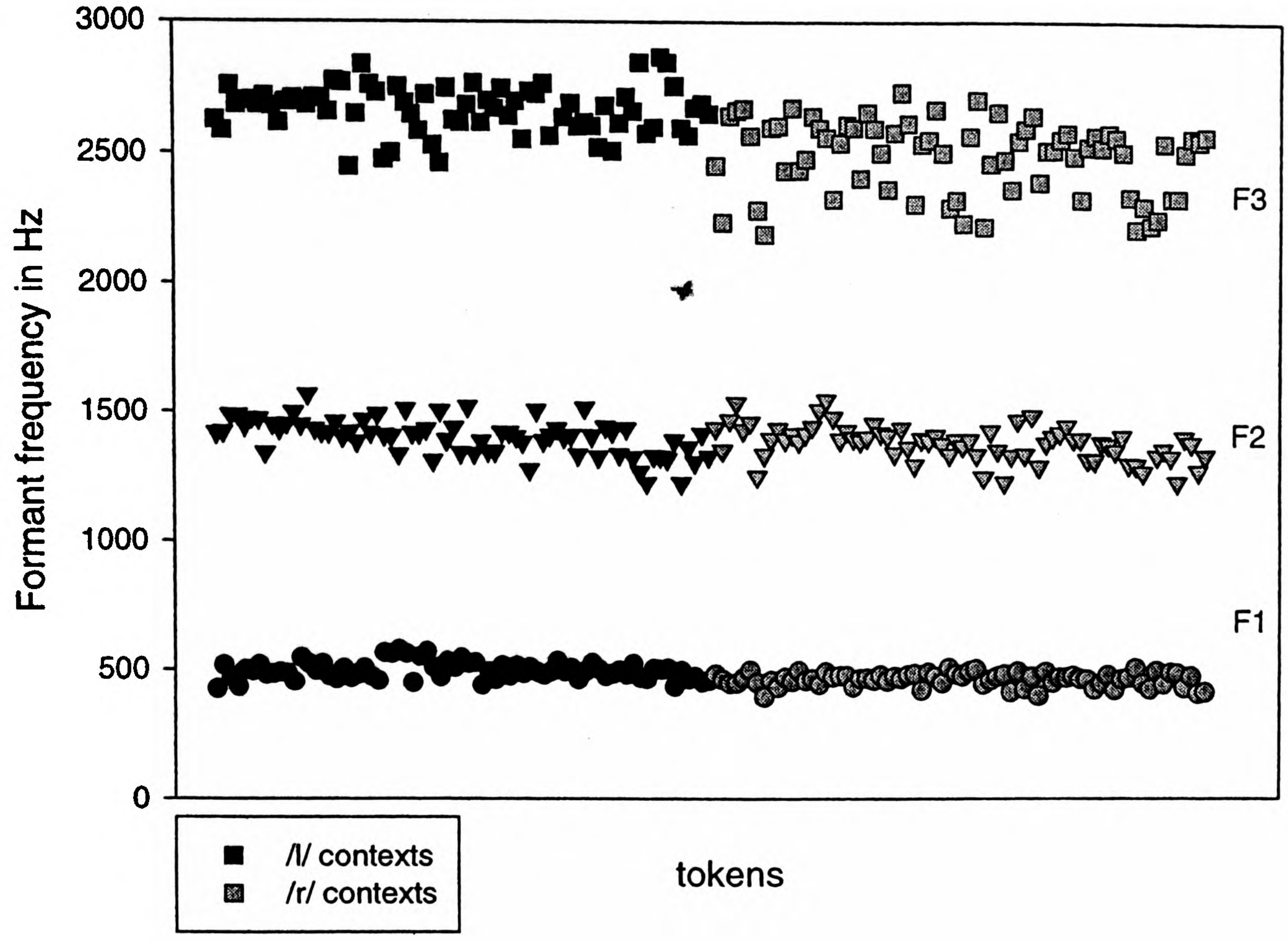
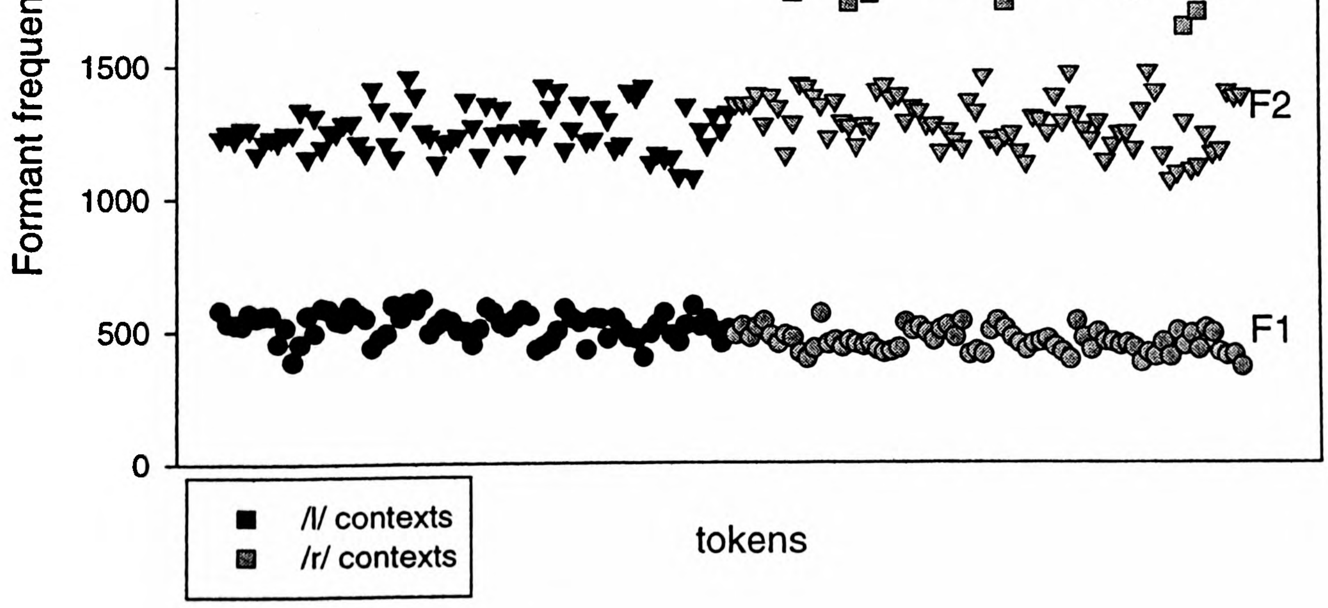


Figure 3.11: Acoustic data, /ə/ before the liquid: subject AS.



A multivariate GLM of the articulatory and acoustic data was constructed for each subject, with word pair and liquid as factors. Tests for normality and equality of variance were performed for all data. Only significant results ($\alpha = 0.01$) are reported below.

For subject EB the variance of F_3 and upper lip *x* are not equal ($F(11,65) = 2.605$, $p < 0.008$ and $F(11,65) = 3.582$, $p < 0.001$). Both factors liquid and word pair were highly significant (Pillai's test: $F(15,51) = 45.557$, $p < 0.001$ and $F(75,275) = 2.455$, $p < 0.001$ respectively) and there was no interaction between the two factors. The factor liquid was significant ($\alpha = 0.01$) for all of the acoustic and articulatory variables except tongue back *y* and F_2. Overall the model suggests that there are strong coarticulatory effects on acoustic output and place of articulation for a schwa preceding a liquid. Non-parametric two-sample Kolmogorov-Smirnov tests produced no additional or contradictory results: all variables except tongue back *y* and F_2 are significant ($\alpha = 0.01$). Means, standard deviations and differences between the means for the significant EMA and acoustic data are given in Table 3.5. Abbreviations are used for the EMA data labels: LIX indicates lower incisor *x* data, LIY lower incisor *y* data, *etc.* TT, TM and TB stand for tongue tip, mid and back respectively, and LL and UL for lower and upper lip.

In /r/ contexts, F_1 and F_3 are lower than in /l/ contexts. In /r/ contexts we find higher means for lower incisor *x* and *y*, lower lip *x* and *y* and upper lip *x* and a lower mean for upper lip *y*. These differences imply lip protrusion, a narrower lip aperture and thus rounding in the /r/ context. The tongue data, with the exception of tongue back *y* which was not significant, has lower *x* values and higher *y* values in the /r/ context, suggesting backing and raising of the tongue in the /r/ as opposed to the /l/ context. These differences are evidence of anticipatory coarticulation, or preparation for the following /r/ or /l/.

Table 3.5: Means, standard deviations and difference between means, EMA data (in mm) and acoustic data (in Hz), /ə/ before the liquid: subject EB.

	LIQUID	N	Mean	Std. Deviation	Difference between means
LIX	r	42	-.0255	.9543	1.05
	l	42	-1.0769	1.1499	
LIY	r	42	-10.5193	1.1561	1.30
	l	42	-11.8150	1.6019	
LLX	r	42	13.0888	1.1735	1.37
	l	42	11.7155	1.2659	
LLY	r	42	-11.7402	1.2713	1.42
	l	42	-13.1645	1.4196	
TBX	r	42	-46.1112	1.1414	-2.04
	l	42	-44.0726	.6599	
TMX	r	42	-34.6874	1.4002	-2.86
	l	42	-31.8248	.7963	
TMY	r	42	.2907	1.7220	1.61
	l	42	-1.3234	1.8943	
TTX	r	42	-22.7481	1.5775	-3.56
	l	42	-19.1833	.8323	
TTY	r	42	.5981	1.9002	1.36
	l	42	-.7636	1.4430	
ULX	r	42	16.4662	1.0163	1.77
	l	42	14.6952	.8236	
ULY	r	42	9.3821	.6976	-1.09
	l	42	10.4760	.5737	
F1	r	42	425.0714	26.4357	-18.40
	l	42	443.4762	29.8600	
F3	r	38	2482.6053	199.8074	-195.16
	l	39	2677.7692	112.6334	

Subject MB's data showed unequal variance of errors for F_3 in Levene's test ($F(11,130) = 5.02$, $p < 0.001$). Both factors liquid and word pair are significant (Pillai's test: $F(15,116) = 35.528$, $p < 0.001$ and $F(75,600) = 2.145$, $p < 0.001$ respectively) with no interaction between them. The factor liquid was significant ($\alpha = 0.01$) for all of the acoustic and articulatory variables except tongue tip x, upper lip y and F_2. As for EB, we find evidence of strong coarticulatory effects in the vowel before the liquid.

Means, standard deviations and differences between the means for the significant EMA and acoustic data are given in Table 3.6. Again we find lower formant frequencies in the /r/ context. Fronter and higher lower lip and incisor position and fronter upper lip position, indicative of lip rounding and protrusion are also found in the /r/ context. However, interestingly, we find lower and fronter tongue position (lower y and higher x) in the /r/ context for this subject.

Table 3.6: Means, standard deviations and difference between means, EMA data (in mm) and acoustic data (in Hz), /ə/ before the liquid: subject MB.

	LIQUID	N	Mean	Std. Deviation	Difference between means
LIX	r	72	-2.4763	1.0213	.88
	l	72	-3.3567	.7553	
LIY	r	72	-12.4789	1.0651	1.55
	l	72	-14.0260	1.0813	
LLX	r	72	16.2887	.9488	1.35
	l	72	14.9366	.7909	
LLY	r	72	-13.0507	1.2527	3.59
	l	72	-16.6408	1.5277	
TBX	r	72	-52.1187	1.9067	1.14
	l	72	-53.2546	1.9325	
TBY	r	72	2.4628	2.4386	-1.73
	l	72	4.1940	2.0343	
TMX	r	72	-36.4619	1.1482	1.01
	l	72	-37.4747	1.2531	
TMY	r	72	4.1458	1.1915	-.66
	l	72	4.8076	1.2343	
TTY	r	72	-.9709	1.6938	-1.16
	l	72	.1892	1.1701	
ULX	r	72	19.4161	.4555	.39
	l	72	19.0276	.5275	
F1	r	72	458.8611	25.8878	-33.76
	l	72	492.6250	33.5162	
F3	r	72	2488.2639	142.3021	-178.91
	l	71	2667.1690	90.1389	

For subject AS two variables, tongue tip *x* and tongue mid *x*, are not normally distributed ($Z = 1.961$, $p < 0.001$ and $Z = 2.108$, $p < 0.001$). Inspection of histograms shows that this is because both variables fall into two normal distributions with a clear separation between the /l/ and /r/ data. There are also significant departures from equality of variance for tongue tip, mid and back *x* ($F(11,132) = 7.588$, 3.992 and 4.509 respectively, $p < 0.001$). The GLM shows both liquid and word pair significant (Pillai's test: $F(15,118) = 101.929$, $p < 0.001$ and $F(75,610) = 1.938$, $p < 0.001$ respectively), with an interaction between the two factors ($F(75,610) = 1.631$, $p < 0.001$). All variables except F_2 and tongue tip *y* were significantly different with respect to the factor liquid. The interaction term was only significant for upper lip *x* and *y*. Because of the interaction, non-parametric Kolmogorov-Smirnov tests for two independent samples were conducted for the data split by word pair. These tests show that most of the variables are significant for each word pair, but that there are some differences between pairs.

Table 3.7 contains the means, standard deviations around the means and difference between the means for AS's data. The lip and lower incisor data show evidence of rounding in the /r/ context as expected: the lower lip and incisor are protruded (greater *x*) and raised (greater *y*) and the upper lip is protruded and lowered (greater *x* and lower *y*) in the /r/ context. The tongue is retracted and raised (lower *x* and greater *y*) and F_1 and F_3 are lowered in the /r/ context. This is the pattern that was found for subject EB.

Table 3.7: Means, standard deviations and difference between means, EMA data (in mm) and acoustic data (in Hz), /ə/ before the liquid: subject AS.

	LIQUID	Mean	N	Std. Deviation	Difference between means
LIX	r	-2.6916	72	.5032	.93
	l	-3.6225	72	.5336	
LIY	r	-10.1894	72	.9788	1.08
	l	-11.2722	72	1.0616	
LLX	r	7.3469	72	1.0162	2.64
	l	4.7021	72	.8820	
LLY	r	-19.7601	72	1.8042	3.73
	l	-23.4949	72	2.2221	
TBX	r	-66.1924	72	2.7081	-2.18
	l	-64.0097	72	1.3255	
TBY	r	6.3249	72	1.2864	1.98
	l	4.3498	72	1.5591	
TMX	r	-50.8271	72	2.4415	-3.59
	l	-47.2405	72	1.1842	
TMY	r	-8.63E-02	72	1.4105	4.18
	l	-4.2687	72	1.6378	
TTX	r	-28.7818	72	3.2169	-7.75
	l	-21.0339	72	1.2355	
TTY	r	-2.7688	72	2.5504	1.05
	l	-3.8213	72	2.3010	
ULX	r	14.7671	72	.5244	.76
	l	14.0054	72	.4908	
ULY	r	12.1814	72	.3754	-1.30
	l	13.4772	72	.5114	
F1	r	455.7778	72	43.3584	-64.56
	l	520.3333	72	51.9978	
F2	r	1285.1250	72	96.8489	28.97
	l	1256.1528	72	84.4126	
F3	r	1951.5694	72	126.3547	-432.72
	l	2384.2917	72	153.2642	

To explore the relationship between the articulatory and acoustic data, Spearman correlations were calculated for the three formants and the articulatory data for each subject. Only significant results ($\alpha = 0.01$) are discussed below.

For EB, F_1 is negatively correlated with lower lip and lower incisor x and y, and positively correlated with upper lip y. Thus F_1 decreases with protrusion and raising of the jaw and lower lip and lowering of the upper lip, *i.e.* the decrease in F_1 is correlated with lip rounding. F_2 is positively correlated with jaw, lower lip and tongue raising (y data). The increase of F_2 with tongue and jaw raising corresponds to the marginally significantly higher F_2 found preceding an /r/ (in this data set). F_3 has significant positive correlations with tongue tip, mid and back x, and upper lip y. It is negatively correlated with upper lip x and lower lip y. I interpret these correlations as follows: F_3 lowers with tongue retraction and lip rounding (protrusion and lowering of the upper lip, raising of the lower lip). This corresponds exactly to the articulatory behaviour expected for the /r/ context in which F_3 is lower. F_3 lowering is an expected outcome of lip rounding. Partial correlations between the acoustic and articulatory data controlling for variation in word pair and liquid were also calculated as a check on the above results. Only results for $p < 0.001$ are reported. As found above, F_2 is positively correlated with lower incisor y ($r = 0.3786$). Controlling for variation in liquid and word pair removes all other correlations, which suggests that these correlations can indeed be partially attributed to liquid context.

For MB, F_1 is negatively correlated with lower lip x and y and lower incisor y, which means that F_1 decreases with lip-rounding, as seen for EB. F_2 is positively correlated with tongue tip x and y and negatively correlated with upper lip x. Thus F_2 increases with fronting and raising of the tongue tip and retraction of the upper lip. F_3 is negatively correlated with upper lip x, lower lip and lower incisor x and y. It is positively correlated with tongue tip and back y. Thus, F_3 decreases with lip rounding and lowering of the tongue for this subject. This is the articulatory behaviour we see in his /r/ productions. Partial correlations controlling for variation in liquid and word pair were calculated, and the following significant ($\alpha = 0.001$) correlations were found: F_2 and tongue back x and y ($r = 0.3388$ and 0.2702 respectively); F_3 and tongue mid and tip x ($r = 0.2795$ and 0.2695 respectively).

AS's data has many strong correlations ($\alpha = 0.01$) between acoustic and articulatory variables. F_1 is negatively correlated with lower incisor and lip x and y, tongue back, mid and tip *y*. It is positively correlated with upper lip *y* and tongue tip *x*. F_1 decreases with lip rounding, tongue raising and retraction of the tongue tip, the articulatory behaviour for /r/ for this subject. F_2 is positively correlated with lower incisor *x* and *y*, lower lip *y*, tongue back *x*, tongue mid *x* and *y* and tongue tip *y*. Thus F_2 increases with lower lip raising, jaw and tongue raising and fronting. F_3 is negatively correlated with lower lip and incisor *x* and *y*, upper lip *x*, tongue back and mid *y*. It is positively correlated with upper lip *y*, tongue tip, mid and back *x*. Thus F_3 decreases with lip rounding and tongue raising and retraction, as does F_1.

Partial correlations (controlling for variation due to liquid and word pair) still show strong correlations between many of the articulatory and acoustic variables. F_1 is negatively correlated with lower incisor *x* and *y*, lower lip *y*, tongue tip, mid and back *x* and tongue tip *y*. It is positively correlated with upper lip *x*. F_1 increases with tongue backing, upper lip protrusion and lower lip, incisor and jaw raising. F_2 is positively correlated with lower lip and incisor *y*, tongue tip, mid and back *x* and tongue tip *y*. It thus increases with fronting of the tongue and raising of the jaw, lower lip and tongue tip. F_3 is positively correlated with tongue tip, mid and back *x* and thus decreases with tongue retraction, once variation due to liquid and word pair are controlled for. This particular result confirms that the F_3 correlation with lip position and tongue height is at least partly attributable to variation caused by the liquid.

3.5.2. Perseverative effects

To examine local perseverative coarticulation, the monophthongs from the set of minimal pairs were measured: /i ɪ æ ɒ/ from the words *leap*, *reap*, *lip*, *rip*, *lap*, *wrap*, *lob* and *rob*, recorded in the frame sentence 'Have you uttered a — at home?'. Measurements were made at the temporal midpoint of the voiced part of the vowel of interest. The articulatory data at each measurement point were looked up automatically in the EMA data matrices. Positional data for the upper lip, lower lip, lower incisor, tongue tip, tongue body and tongue back were collected in the form of (*x*, *y*) co-ordinate pairs. Acoustic measurements were made using *Waves* as described in 3.5.1 above.

For EB a subset of the stimuli was chosen: the monophthongs /ɪ æ ɒ/. Extensive devoicing of the high vowel /i/ determined its exclusion from the set of vowels measured. The acoustic measurements were made using 22 pole DFT and

Burg spectra. The large number of poles was required to pick out all three formants, as many of the vowels were nasalised and displayed prominent nasal formants. F_3 proved unmeasurable for /ɒ/ and for two tokens of /æ/ for this subject.

The articulatory data for EB is presented in Figure 3.12, MB in Figure 3.13 and Figure 3.14. Different colours represent the word in which the measurement was taken, and different symbols the different articulatory measurement points (*e.g.* upper lip *vs.* lower lip). The acoustic data (F_1, F_2 and F_3 measurements) corresponding to the articulatory data are presented graphically in Figure 3.15, Figure 3.16 and Figure 3.17 for EB, MB and AS respectively. Data for each word is shown in a different colour, with tokens of vowels of the same category adjacent to each other.

Figure 3.12: Articulatory data, vowel after the liquid: subject EB.

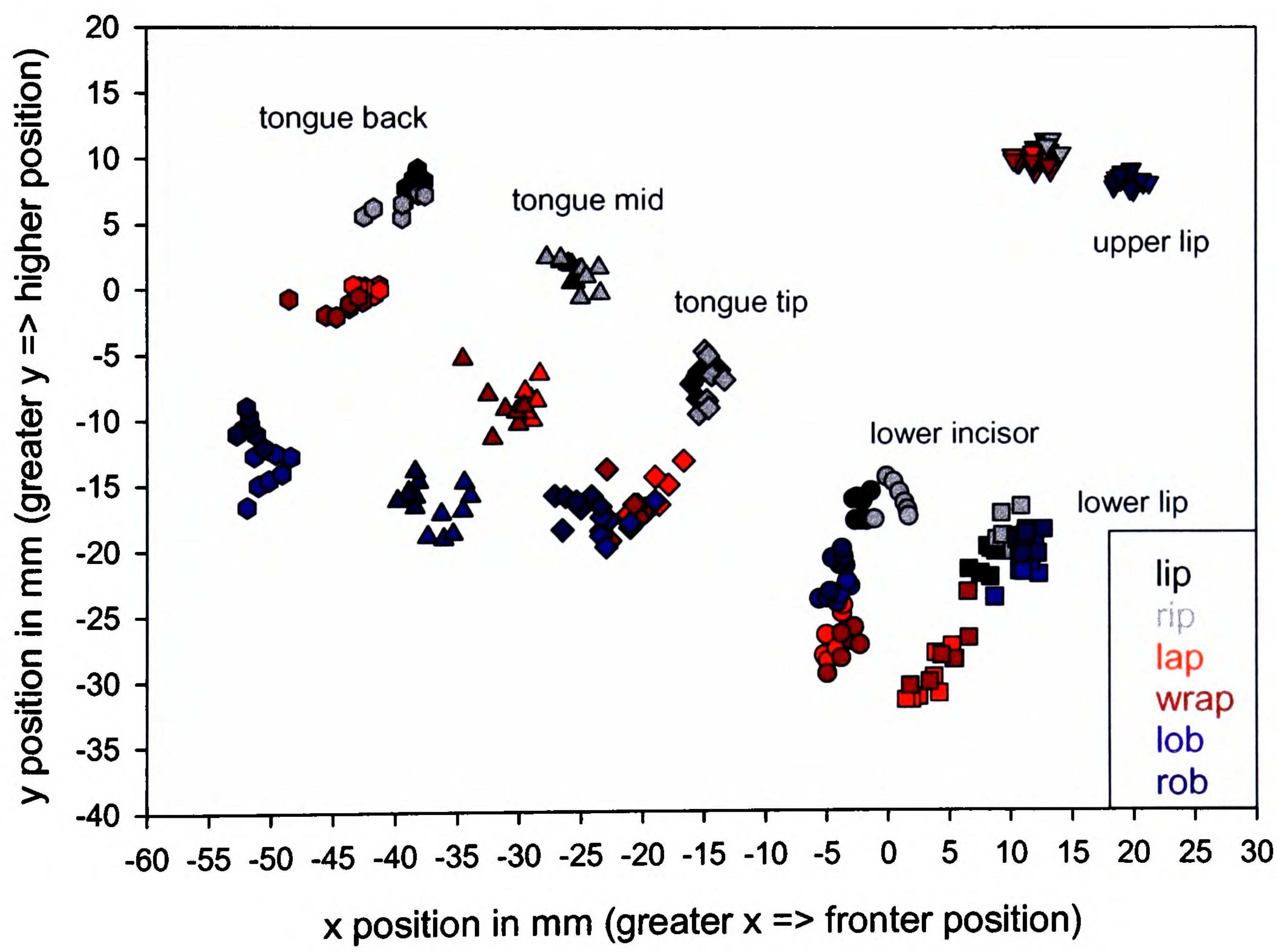


Figure 3.13: Articulatory data, vowel after the liquid: subject MB.

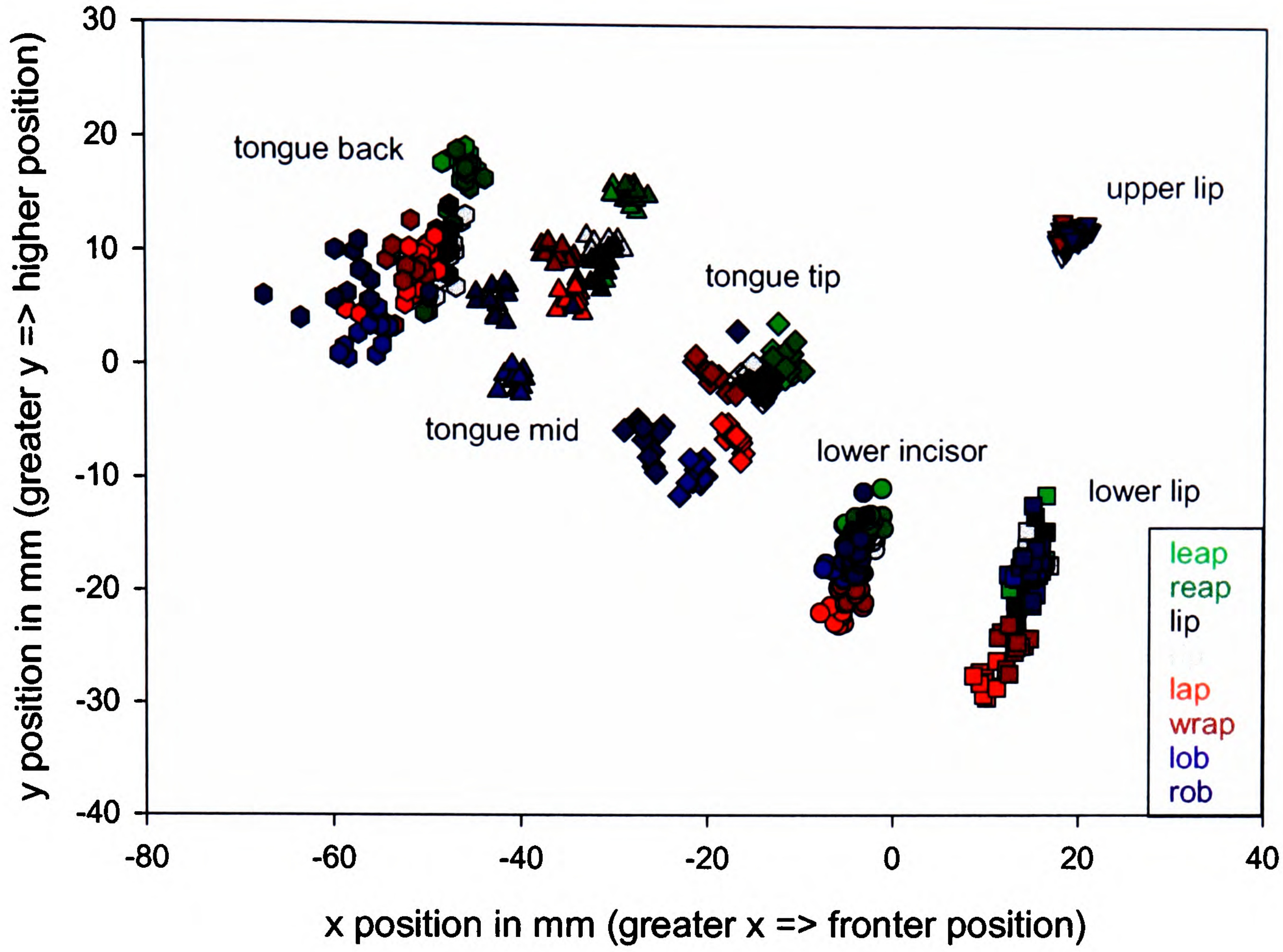


Figure 3.14: Articulatory data, vowel after the liquid: subject AS.

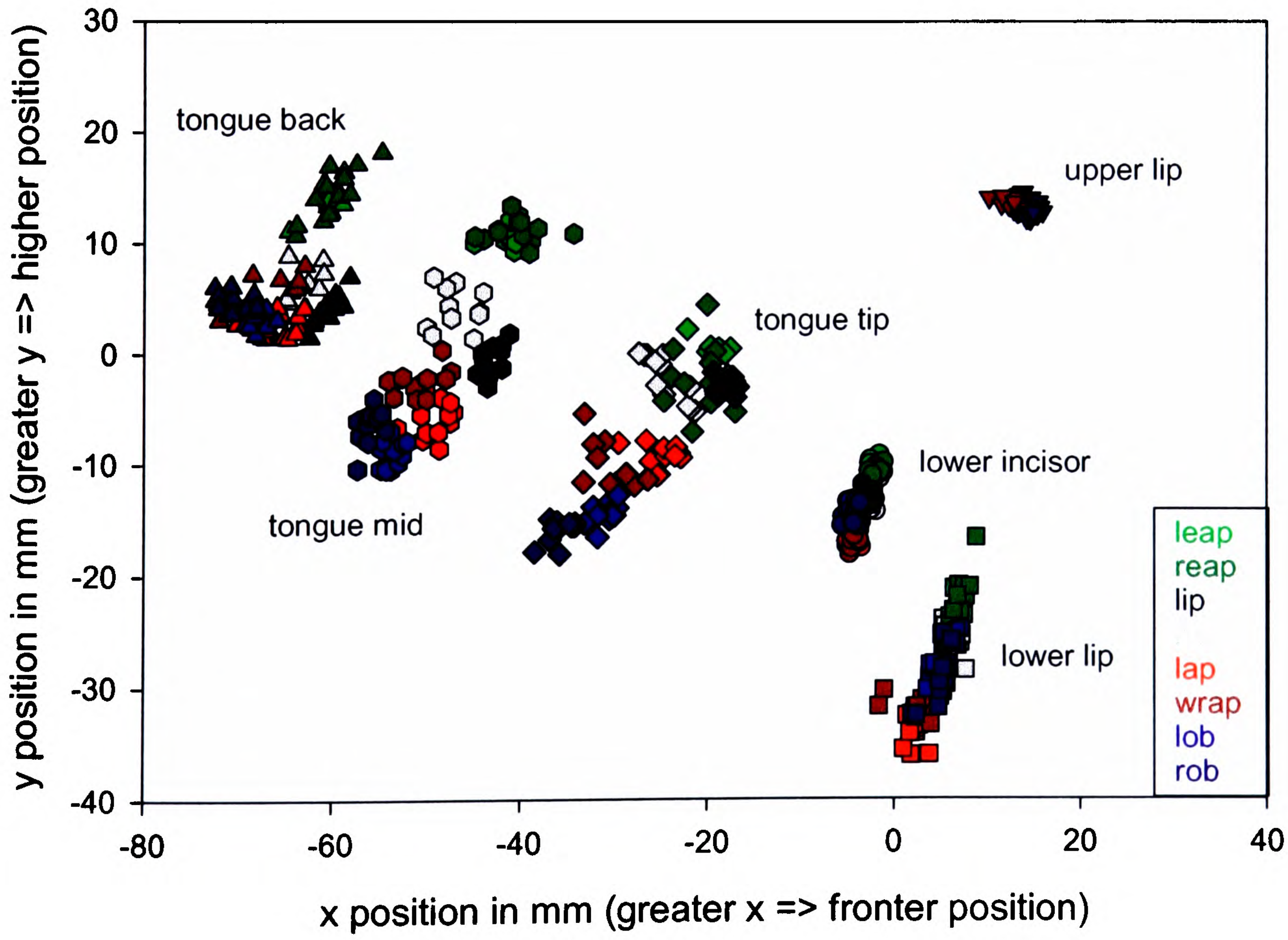


Figure 3.15: Acoustic data, vowel after the liquid: subject EB.

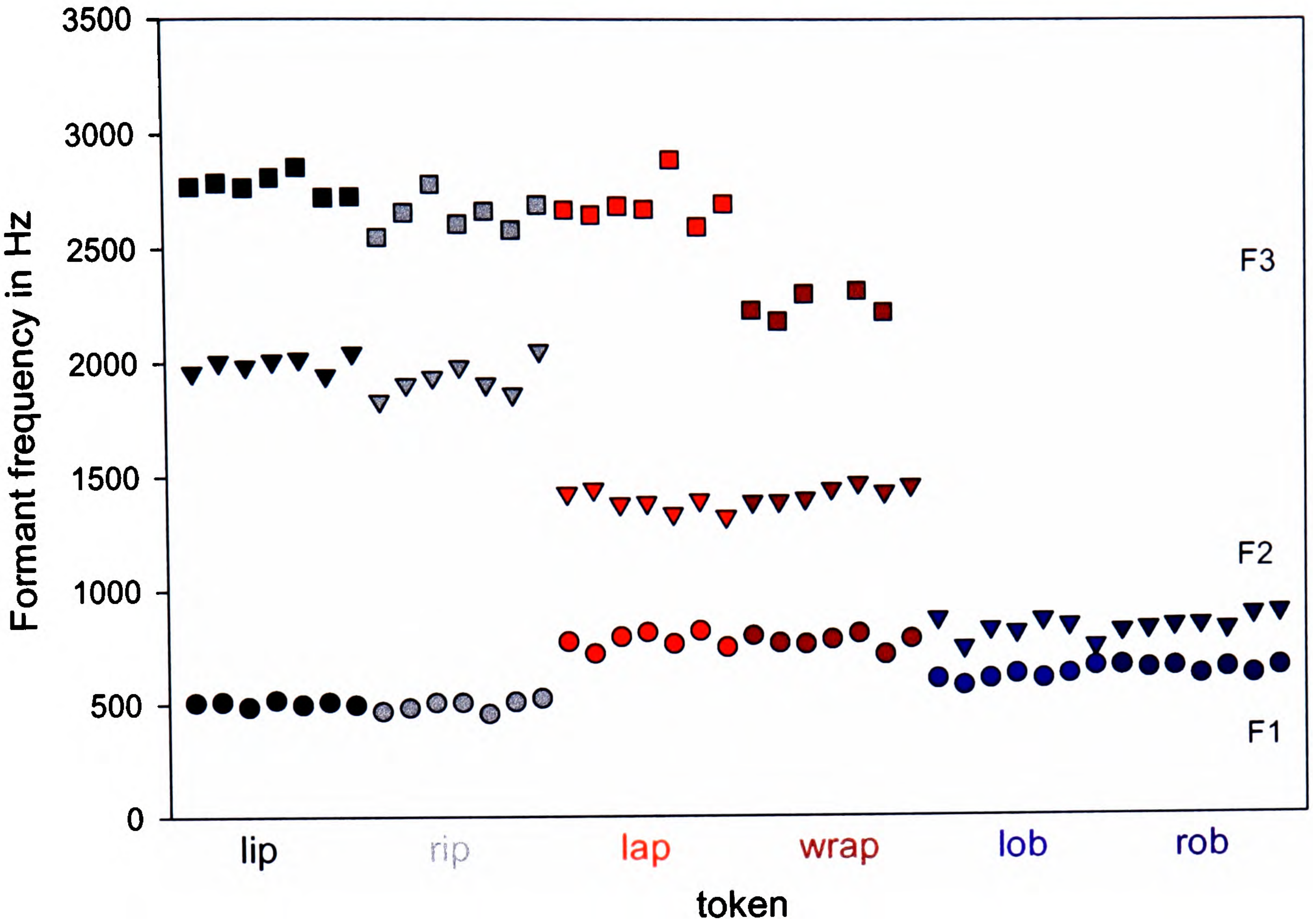


Figure 3.16: Acoustic data, vowel after the liquid: subject MB.

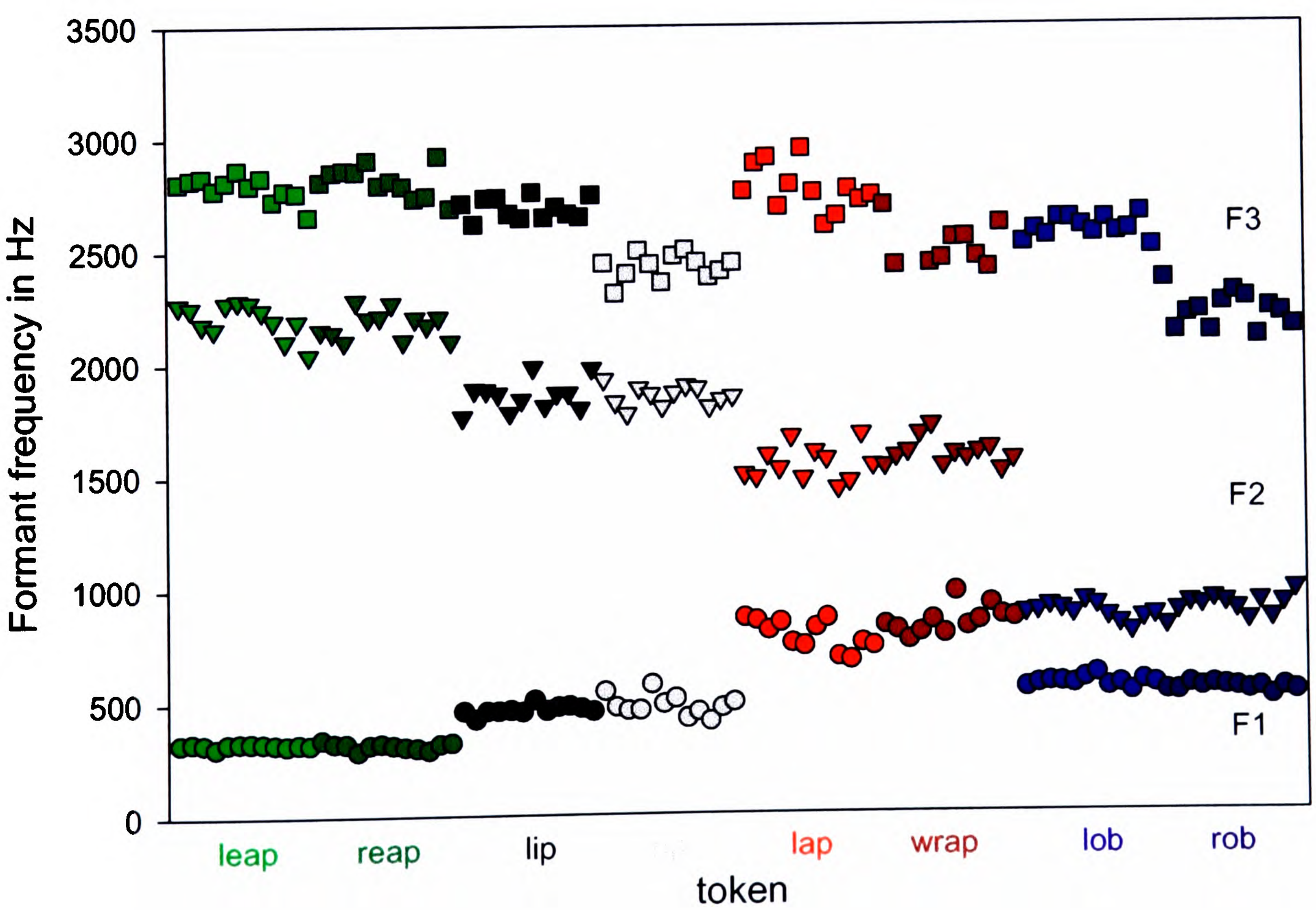
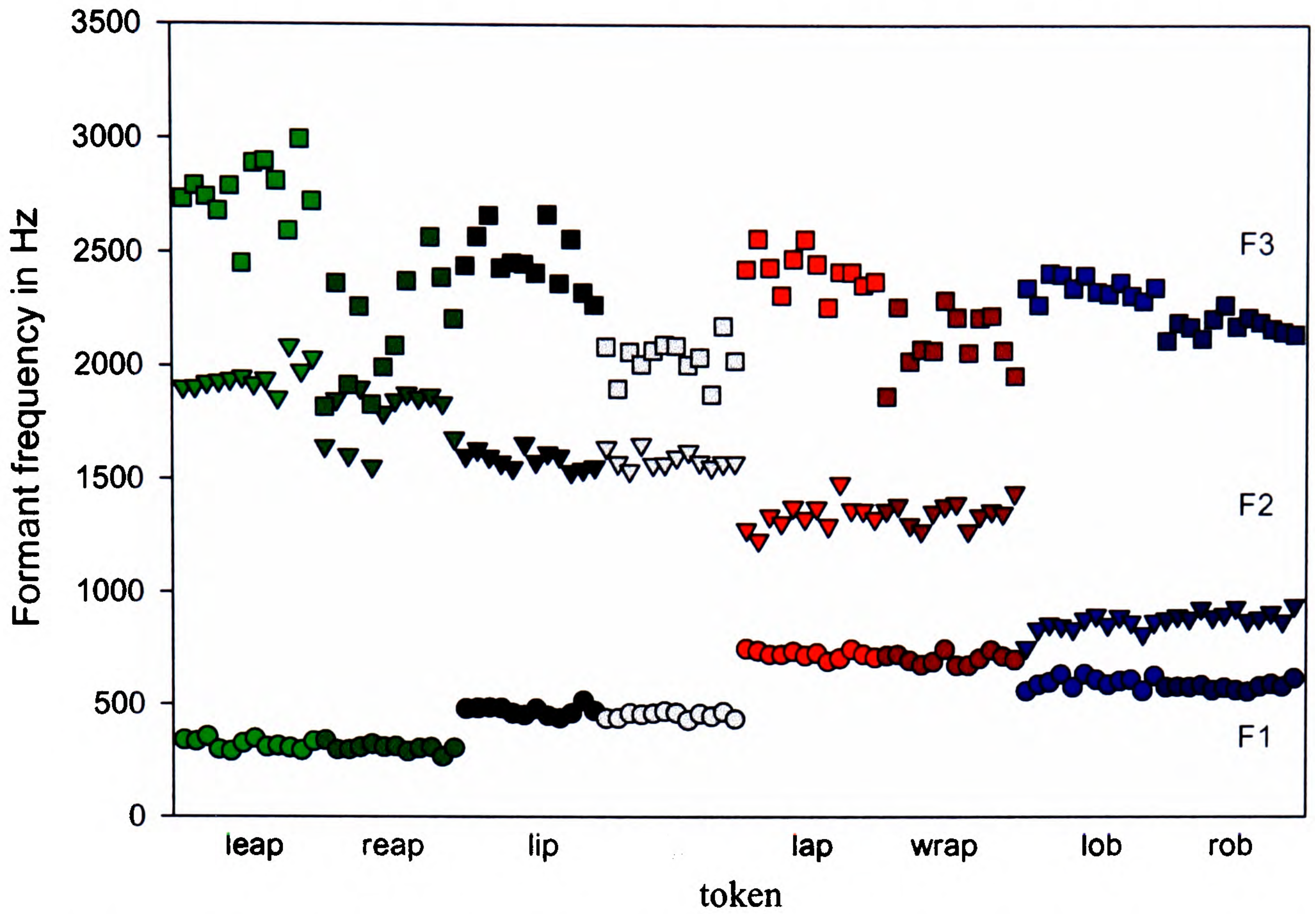


Figure 3.17: Acoustic data, vowel after the liquid: subject AS.



Inspection of the articulatory and acoustic data shows clear groupings on the basis of word pair and liquid for all subjects. Multivariate general linear models with liquid and word pair as factors were constructed in *SPSS* to investigate this question.

For EB, acoustic data was not included in the GLM as the missing F_3 measurements would result in the data for those vowels not being modelled. One-sample Kolmogorov-Smirnov tests showed departure from normality in one parameter at the 1% level: upper lip x ($Z = 1.620$, $p < 0.01$). Levene's test for equality of error variances showed a significant result at this level: tongue mid x ($F(5,36) = 3.737$, $p < 0.008$). Both the factors liquid and word pair and the interaction between them are highly significant (Pillai's test: $F(12,25) = 16.834$, $p < 0.001$; $F(24,52) = 88.224$, $p < 0.001$; $F(24,52) = 5.941$, $p < 0.001$). The factor liquid was significant for tongue tip, mid and back x, lower lip and lower incisor x and y. Post-hoc tests on the factor word pair were conducted using Tamhane's *T2*, a conservative test which does not require equality of variance. The post-hoc tests show significant differences ($p < 0.01$) between all three vowels in all but a few variables. For lower incisor x and tongue tip y, /æ/ and /ɒ/ are not significantly different; /ɪ/ and /ɒ/ are not significantly different in lower lip y.

To explore the interaction between liquid and word pair, and to test for differences in formant frequencies associated with liquid context, independent samples *t*-tests were performed on each variable for each word pair with liquid as factor. Table 3.8 lists the variables which differed significantly with liquid context and their significance levels. Levene's test for equal variances showed no significant results, so the table contains the values of the test when equal variances are assumed.

Table 3.8: Independent samples t-test for EMA and acoustic data by word pair, liquid as factor: subject EB.

		Equal variances assumed		
		t-test for Equality of Means		
WORDPAIR		t	df	Sig. (2-tailed)
lip/rip	LIX	6.703	12	.000
	LLX	4.261	12	.001
	LLY	2.977	12	.012
	TBY	-5.518	12	.000
	F2	-2.211	12	.047
	F3	-3.762	12	.003
lap/wrap	LIX	2.734	12	.018
	TBX	-2.937	12	.012
	TBY	-4.956	12	.000
	TMX	-3.568	12	.004
	TTX	-2.682	12	.020
	F3	-9.504	10	.000
lob/rob	LIY	5.059	12	.000
	LLY	3.440	12	.005
	TBX	-2.676	12	.020
	TBY	5.080	12	.000
	TMX	-6.420	12	.000
	TMY	2.587	12	.024
	TTX	-3.901	12	.002
	TTY	2.797	12	.016
	F1	2.594	12	.023

The differences in means for these significant variables are shown in Table 3.9. Articulatory differences range from 1 to 3.5 mm. For /ɪ/ the lower incisor is fronter, the lower lip fronter and higher and the tongue back lower in /r/ contexts. For /æ/ (the *lap/wrap* pair) the lower incisor is fronter, the whole of the tongue backer and the tongue back lower for /r/ contexts. The vowel /ɒ/ (*lob/rob* pair) has the whole tongue backer and higher and the lower lip and lower incisor higher for /r/ contexts.

Table 3.9: Means, standard deviations, and difference in means for EMA data (in mm) and acoustic data (in Hz), vowel after the liquid: subject EB.

WORDPAIR		LIQUID	N	Mean	Std. Deviation	Difference in means
lip/rip	LIX	r	7	.6829	1.0273	2.90
		l	7	-2.2214	.5090	
	LLX	r	7	9.9271	.7844	1.83
		l	7	8.0971	.8220	
	LLY	r	7	-18.7914	1.3594	1.94
		l	7	-20.7286	1.0559	
	TBY	r	7	6.4229	.7129	-1.91
		l	7	8.3314	.5738	
	F2	r	7	1929.2857	73.6856	-68.00
		l	7	1997.2857	34.5191	
	F3	r	7	2650.0000	77.2528	-128.00
		l	7	2778.0000	46.2133	
lap/wrap	LIX	r	7	-3.4357	.8602	1.12
		l	7	-4.5514	.6522	
	TBX	r	7	-44.5800	2.0365	-2.44
		l	7	-42.1429	.8207	
	TBY	r	7	-1.2429	.5997	-1.26
		l	7	1.714E-02	.3048	
	TMX	r	7	-31.4886	1.7488	-2.43
		l	7	-29.0543	.4475	
	TTX	r	7	-21.0986	1.2111	-1.99
		l	7	-19.1043	1.5507	
	F3	r	5	2252.2000	56.4907	-443.23
		l	7	2695.4286	91.9019	
lob/rob	LIY	r	7	-21.1671	1.0373	2.26
		l	7	-23.4229	.5617	
	LLY	r	7	-19.3614	.8191	2.07
		l	7	-21.4314	1.3654	
	TBX	r	7	-51.7229	.7221	-1.47
		l	7	-50.2500	1.2647	
	TBY	r	7	-10.5757	1.0145	3.47
		l	7	-14.0457	1.4957	
	TMX	r	7	-38.7471	.5761	-3.29
		l	7	-35.4586	1.2266	
	TMY	r	7	-15.4800	.9136	1.85
		l	7	-17.3257	1.6517	
	TTX	r	7	-25.4171	1.3333	-3.15
		l	7	-22.2643	1.6718	
	TTY	r	7	-16.5357	.9374	1.55
		l	7	-18.0843	1.1258	
	F1	r	7	643.8571	16.9846	30.14
		l	7	613.7143	25.6236	

These differences appear to correspond to differences in the articulation of the preceding liquid: EB produces /r/ with a strongly retracted tongue in comparison to /l/, and the tongue is backer after /r/. The lip and lower incisor differences point to perseverative lip rounding after the /r/: in particular a fronter and/or higher lower lip suggests lip rounding. The tongue height differences require a more complicated explanation. Tongue back is lower in /r/ contexts for /ɪ/ and /æ/, but together with the rest of the tongue, is higher for /ɒ/. Inspection of the articulatory data in Figure 3.12 shows a trend for the rest of the tongue data to follow the tongue back data. A possible explanation is that the articulatory configuration for an /r/ has lower tongue height than /ɪ/ and /æ/, but higher tongue height than the low vowel /ɒ/. If the tongue height of /r/ affects the tongue height of the following vowel, these differences are as would be expected. This explanation assumes that tongue back height is relatively free for /l/, or that /l/ is strongly coarticulated with the following vowel in this respect, whereas /r/ is less influenced. This interpretation supports the DAC model of Recasens *et al.* (1997) in which a clear /l/ is not strongly constrained and is therefore more strongly coarticulated than a segment such as British English /r/ which has velarisation.

The means for the acoustic data show that F_3 is lower in /r/ than /l/ contexts, by as much as 443 Hz for /æ/. This might well be the result of lip-rounding associated with the /r/. The F_2 difference is one of 68 Hz, with F_2 again lower in the /r/ context, consistent with a backer tongue position for the darker liquid, /r/. The small difference of 30 Hz between the means of F_1 in /ɒ/ reflects a lower F_1 in the /l/ context.

For MB, lower lip y and tongue tip y are not normally distributed ($Z = 1.879$ and 1.833, $p < 0.002$). Levene's tests show no significant results. The GLM, which includes both articulatory and acoustic data, shows a strong interaction between the factors liquid and vowel ($F(45, 219) = 8.176, p < 0.001$). Both factors liquid and word pair are highly significant ($F(15,71) = 45.647, p < 0.001$ and $F\ (45,219) = 59.687, p < 0.001$). Post-hoc tests show that there are significant differences between the word pairs in all variables except the following: lower incisor does not differ significantly for /i/ and /ɪ/, or /æ/ and /ɒ/; lower lip x and y are the same for /i/, /ɪ/ and /ɒ/; tongue back x and upper lip x are not different for /i/ and /ɪ/; tongue back y is the same for /ɪ/, /æ/ and /ɒ/ and upper lip y is not significantly different for any pair except /i/ and /ɪ/.

To explore the interaction and look for significant differences attributable to liquid, independent samples t-tests were conducted for each word pair separately with liquid as a factor. Table 3.10 displays the variables which differed significantly with liquid context and their significance levels. Levene's test for equal variances showed only one significant result (tongue mid *x* for *leap/reap*) which did not change the pattern of results, so the table contains the values of the test when equal variances are assumed.

Table 3.10: Independent samples t-test for EMA and acoustic data by word pair, liquid as factor: subject MB.

		Equal variances assumed		
		t-test for Equality of Means		
WORDPAIR		t	df	Sig. (2-tailed)
leap/reap	LIX	2.549	22	.018
	LLX	2.567	22	.018
	TMX	2.199	22	.039
	TMY	2.103	22	.047
	ULY	2.283	22	.032
lip/rip	LIX	3.067	22	.006
	LLX	2.446	22	.023
	TMY	4.166	22	.000
	TTX	-2.930	22	.008
	F3	-12.441	22	.000
lap/wrap	LIX	4.398	22	.000
	LIY	4.829	22	.000
	LLX	8.085	22	.000
	LLY	7.388	22	.000
	TMX	-3.688	22	.001
	TMY	15.017	22	.000
	TTX	-5.699	22	.000
	TTY	14.159	22	.000
	F1	2.349	22	.028
	F3	-5.799	19	.000
lob/rob	LIX	4.984	22	.000
	LIY	2.649	22	.015
	LLY	2.441	22	.023
	TBY	5.882	22	.000
	TMY	17.417	22	.000
	TTX	-4.895	22	.000
	TTY	4.024	22	.001
	ULX	-2.407	22	.025
	F1	-3.747	22	.001
	F3	-14.399	22	.000

Table 3.11: Means, standard deviations, and difference in means for EMA data (in mm) and acoustic data (in Hz), vowel after the liquid: subject MB.

WORDPAIR		LIQUID	N	Mean	Std. Deviation	Difference in means
leap/reap	LIX	r	12	-2.4423	.7758	1.03
		l	12	-3.4703	1.1621	
	LLX	r	12	15.1848	.7625	1.09
		l	12	14.0957	1.2565	
	TMX	r	12	-28.0746	.6099	.85
		l	12	-28.9208	1.1854	
	TMY	r	12	15.6249	.3356	1.35
		l	12	14.2728	2.2023	
	ULY	r	12	12.0747	.4013	.37
		l	12	11.7038	.3945	
lip/rip	LIX	r	12	-2.7883	.7787	.91
		l	12	-3.7023	.6779	
	LLX	r	12	15.5576	.7470	.85
		l	12	14.7061	.9467	
	TMY	r	12	10.5179	.7690	1.73
		l	12	8.7858	1.2179	
	TTX	r	12	-14.9099	.9756	-1.05
		l	12	-13.8625	.7631	
	F3	r	12	2423.7500	57.0281	-266.08
		l	12	2689.8333	47.2995	
lap/wrap	LIX	r	12	-4.3326	1.0182	1.61
		l	12	-5.9427	.7559	
	LIY	r	12	-20.2923	.7263	1.67
		l	12	-21.9647	.9548	
	LLX	r	12	13.0357	.9928	2.97
		l	12	10.0644	.7968	
	LLY	r	12	-24.8278	1.2518	3.36
		l	12	-28.1888	.9572	
	TMX	r	12	-36.3049	1.2414	-1.65
		l	12	-34.6514	.9336	
	TMY	r	12	9.8381	.6381	4.09
		l	12	5.7433	.6965	
	TTX	r	12	-19.6164	1.2914	-2.44
		l	12	-17.1784	.7266	
	TTY	r	12	-.7707	.9792	5.53
		l	12	-6.3023	.9342	
	F1	r	12	838.9167	59.2950	60.92
		l	12	778.0000	67.4887	
	F3	r	9	2511.1111	96.4656	-257.22
		l	12	2768.3333	103.4815	
lob/rob	LIX	r	12	-4.0094	.7105	1.65
		l	12	-5.6576	.8986	
	LIY	r	12	-16.5201	1.8776	1.49
		l	12	-18.0095	.5182	
	LLY	r	12	-17.1663	2.0164	1.61
		l	12	-18.7748	1.0699	
	TBY	r	12	7.4374	2.1069	4.62
		l	12	2.8222	1.7173	
	TMY	r	12	5.7023	1.1612	7.01
		l	12	-1.3077	.7716	
	TTX	r	12	-25.6415	3.0147	-4.42
		l	12	-21.2240	.8263	
	TTY	r	12	-5.7727	3.2068	3.89
		l	12	-9.6632	.9656	
	ULX	r	12	19.6296	.5793	-.60
		l	12	20.2247	.6310	
	F1	r	12	533.6667	15.3109	-27.50
		l	12	561.1667	20.2970	
	F3	r	12	2210.5833	77.7390	-374.92
		l	12	2585.5000	45.7434	

In Table 3.11 the difference in means for the significant variables in each word pair are given. Articulatory differences range from 0.3 mm to 7 mm, and acoustic differences from 27 Hz to 374 Hz. Some differences in means are in the same direction for all word pairs in which they are significant: lower incisor and lower lip are more protruded and raised, tongue tip, mid and back are more raised and the tip more retracted and F_3 is lowered in the /r/ context. There are also differences in the direction in which differences in means lie: tongue mid is fronter in the /r/ context for *leap/reap* and backer for *lap/wrap* (backness is the trend for the other word pairs). This might be a result of articulatory undershoot: after an /l/ the fronter target for the /i/ is not reached, whereas after an /r/ the tongue travels further to the front position. Another different direction in means is found in F_1, which is lower in the /r/ context for the *lob/rob* pair, but higher for the *lap/wrap* pair. Interestingly, this subject showed fronter and lower tongue in the schwa just before the liquid in the /r/ context, but in the perseverative context shows tongue raising and retraction as do the other subjects. As will be seen below, in the long domain anticipatory context there were no significant differences in tongue placement in vowels for this subject, but long-domain perseverative differences in tongue placement (retraction and raising) are seen.

For AS, tongue back *y* is not normally distributed ($Z = 1.978$, $p < 0.001$). Levene's test for equality of error variances shows 8 out of 15 variables highly significant, and the GLM was thus not pursued (the model showed liquid and word pair and the interaction between them highly significant). Independent samples *t*-tests (unequal variances assumed) conducted for each word pair separately, with liquid as factor show many significant differences between /l/ and /r/ contexts. These are presented in Table 3.12 and the corresponding means in Table 3.13.

The differences in means for the significant variables show a similar pattern to those of the other subjects in most variables. The articulatory differences range from 0.4 mm (again for lip data) to over 6 mm for tongue data. Acoustic differences range from approximately 20 Hz to 600 Hz. The lower lip is raised and protruded in /r/ contexts, tongue tip, mid and back are retracted, the tip lowered and the mid and back raised and F_1 and F_3 are lowered. Lower incisor is (interestingly) retracted in the /r/ context for *leap/reap* and *lob/rob* and not significantly different for the other word pairs. Upper lip *y* is lowered, as would be expected, in the /r/ context, for *leap/reap* and *lip/rip*, but it is also retracted in these pairs and raised for *lob/rob*. It is possible that rounding on the vowel /ɒ/ is stronger following the unrounded /l/ than when it follows the rounded consonant /r/. Other data (anticipatory and perseverative measurements), suggest that lip rounding is less salient in /r/ for this subject (or that the placement of the coils was not optimal to measure rounding activity).

Table 3.12: Independent samples t-test for EMA and acoustic data by word pair, liquid as factor: subject AS.

		Equal variances not assumed		
		t-test for Equality of Means		
word pair		t	df	Sig. (2-tailed)
leap/reap	LIY	-3.387	17.908	.003
	LLX	4.688	22.000	.000
	LLY	6.995	13.593	.000
	TTX	-2.124	15.596	.050
	TTY	-2.836	11.982	.015
	ULX	-3.581	14.826	.003
	ULY	-3.028	21.272	.006
	F2	-4.301	16.328	.001
	F3	-6.834	15.630	.000
lip/rip	LLX	7.041	19.941	.000
	LLY	7.887	18.559	.000
	TBX	-4.620	17.268	.000
	TMX	-5.645	15.607	.000
	TMY	6.749	20.100	.000
	TTX	-9.697	14.853	.000
	ULX	-3.462	20.163	.002
	ULY	-2.172	18.556	.043
	F1	-2.635	18.573	.017
	F3	-9.895	19.116	.000
lap/wrap	LLY	4.755	21.913	.000
	TBY	3.943	17.034	.001
	TMX	-2.091	20.670	.049
	TMY	5.911	20.792	.000
	TTX	-5.671	20.961	.000
	F1	-2.335	19.193	.031
	F3	-6.789	19.403	.000
lob/rob	LIY	-3.868	18.673	.001
	TBX	-4.165	21.759	.000
	TBY	5.859	21.415	.000
	TMX	-5.986	21.992	.000
	TMY	4.520	19.735	.000
	TTX	-10.179	21.805	.000
	TTY	-3.660	21.747	.001
	ULY	3.042	21.527	.006
	F1	-2.442	17.719	.025
	F2	3.604	17.948	.002
	F3	-9.311	21.976	.000

Table 3.13: Means, standard deviations, and difference in means for EMA data (in mm) and acoustic data (in Hz), vowel after the liquid: subject AS.

word pair		LIQUID	N	Mean	Std. Deviation	Difference in means
leap/reap	LIY	r	12	-10.8490	.7352	-.84
		l	12	-10.0129	.4369	
	LLX	r	12	7.1910	.7858	1.51
		l	12	5.6838	.7894	
	LLY	r	12	-21.8467	2.0090	4.29
		l	12	-26.1390	.6947	
	TTX	r	12	-20.7978	2.4567	-1.66
		l	12	-19.1349	1.1495	
	TTY	r	12	-2.2312	3.0377	-2.54
		l	12	.3108	.6423	
	ULX	r	12	13.2128	.3569	-.40
		l	12	13.6135	.1512	
	ULY	r	12	13.3988	.3919	-.44
		l	12	13.8437	.3250	
	F2	r	12	1778.3333	121.3914	-169.08
		l	12	1947.4167	61.6979	
	F3	r	11	2167.0000	252.4718	-593.25
		l	12	2760.2500	144.5590	
lip/rip	LLX	r	12	6.5648	.7414	1.85
		l	12	4.7109	.5313	
	LLY	r	12	-25.2170	1.5883	4.28
		l	12	-29.4926	1.0020	
	TBX	r	12	-64.1123	2.3872	-3.65
		l	12	-60.4647	1.3351	
	TMX	r	12	-47.2010	2.2564	-4.06
		l	12	-43.1408	1.0572	
	TMY	r	12	3.9014	1.8920	4.56
		l	12	-.6577	1.3770	
	TTX	r	12	-24.4686	2.1173	-6.44
		l	12	-18.0275	.9004	
	ULX	r	12	13.4866	.3600	-.45
		l	12	13.9325	.2636	
	ULY	r	12	13.2872	.3307	-.39
		l	12	13.6759	.5243	
	F1	r	12	451.6667	13.3983	-19.08
		l	12	470.7500	21.2095	
	F3	r	12	2039.5000	83.4184	-430.92
		l	12	2470.4167	125.6875	
lap/wrap	LLY	r	12	-31.4574	1.3664	2.57
		l	12	-34.0303	1.2831	
	TBY	r	12	5.0512	1.8536	2.40
		l	12	2.6469	1.0131	
	TMX	r	12	-50.9099	2.2739	-1.73
		l	12	-49.1759	1.7545	
	TMY	r	12	-2.6021	1.2201	3.38
		l	12	-5.9818	1.5602	
	TTX	r	12	-29.9161	2.4275	-5.08
		l	12	-24.8333	1.9355	
	F1	r	12	704.5833	24.6704	-20.00
		l	12	724.5833	16.4894	
	F3	r	12	2113.6667	130.7067	-310.00
		l	12	2423.6667	89.0642	
lob/rob	LIY	r	12	-14.7199	.7207	-.95
		l	12	-13.7654	.4595	
	TBX	r	12	-70.7819	1.5595	-2.52
		l	12	-68.2593	1.4033	
	TBY	r	12	4.9527	.8687	1.92
		l	12	3.0277	.7352	
	TMX	r	12	-56.0052	.9483	-2.34
		l	12	-53.6647	.9671	
	TMY	r	12	-6.4239	1.6187	2.58
		l	12	-9.0052	1.1375	
	TTX	r	12	-35.8713	1.2689	-5.04
		l	12	-30.8313	1.1540	
	TTY	r	12	-15.9553	1.0505	-1.49
		l	12	-14.4639	.9428	
	ULY	r	12	12.8597	.3618	.42
		l	12	12.4404	.3116	
	F1	r	12	578.1667	15.1408	-21.17
		l	12	599.3333	25.9311	
	F2	r	12	900.7500	23.2267	47.17
		l	12	853.5833	38.9392	
	F3	r	12	2178.3333	43.6751	-168.83
		l	12	2347.1667	45.1479	

There are many significant Spearman correlations between acoustic and articulatory variables for all three subjects. However these are likely to be the result of the noticeably different articulatory configurations and formant patterns for the different vowels. A partial correlation analysis was thus conducted on all of the articulatory and acoustic data controlling for word pair (hence, vowel) and liquid. For EB, this showed no significant correlations between the formants and any articulatory measure. Most variability is accounted for by the combination of vowel and liquid and within these categories there is no correlation between formant frequency variation and articulatory variation as measured by EMA. MB and AS, however, perhaps because there are more measurements for these subjects, show many significant correlations ($\alpha = 0.001$) between acoustic and articulatory data. As both subjects show a similar pattern, these are discussed briefly together. F_1 is negatively correlated with all variables but upper lip x for AS, and all but upper lip y, tongue back x and tongue tip y for MB. Thus F_1 increases with retraction and lowering of the articulators in general. F_2 is positively correlated with lower incisor x and y, upper lip x, tongue mid x and y for both subjects; in addition tongue back y for AS and lower lip x and y for MB. F_2 increases with fronting and raising. F_3 is (positively) correlated with fewer variables: tongue back and mid y for AS and upper lip y for MB. F_3 increases with tongue raising for AS and decreases with upper lip lowering for MB.

3.6. Summary and conclusions

The existence of coarticulatory effects of liquids on adjacent vowels is clear: there are consistent effects of lip rounding and fairly consistent formant lowering associated with /r/. Tongue body differences vary quite markedly with subject and word pair, but in general show significant differences between /l/ and /r/ contexts.

The acoustic results for vowels following the liquid can be compared with Tunley's results (1999). Tunley found evidence of F_2 and F_3 lowering after an /r/ and some slight evidence of F_2 lowering after an /l/. The results found here are a consistently lower F_3 and F_1 after an /r/ (except for MB /æ/ where F_1 is higher after /r/). F_2 effects are few and vary with speaker. Given Tunley's results, it is not surprising that F_2 is often not significant – the coarticulatory effects of /r/ and /l/ are similar, both lower the second formant in comparison to /h/. Tunley proposes a hierarchy of strength of effects, with /ɪ/ exhibiting most coarticulation and /æ/ least.

This hierarchy is not supported by the data described here, possibly because of the effect of different following consonants. With more speakers, more consistent effects might be found and the hierarchy thus reconsidered.

CHAPTER FOUR

LONG-DOMAIN COARTICULATORY EFFECTS OF /l/ AND /r/

4.1. Introduction

This chapter contains further analysis of the experimental data described in chapter 3 (experiment 3). The experiment was primarily designed to test whether long-domain coarticulatory differences in articulation and acoustics could be found between /l/ and /r/ contexts, to test Kelly and Local's observations that liquid resonances have great extent. The previous chapter examines the local coarticulatory effects of /l/ and /r/; this chapter focuses on the long-distance effects, that is effects on non-adjacent segments. Previous experimental work on long-domain liquid resonances, conducted by Tunley (1999), showed /r/ colouring on unstressed vowels which were non-adjacent to the /r/. Tunley's work compared /r/ and /h/, and found that F_2 and F_3 exhibited rhotic resonance effects. Hawkins and Heid (2000) found long-domain differences between /l/ and /r/ contexts up to five syllables before the liquid. Their study was an acoustic one, which only considered anticipatory effects. This experiment compares /l/ and /r/, examining non-adjacent consonants and vowels for long-domain anticipatory and perseverative resonance effects, using EMA, EPG and F_1, F_2 and F_3 measurements.

4.2. The extent of coarticulatory effects

4.2.1. Time aligning

To examine the EMA data for long-domain resonance differences, the sets of EMA data for each word were aligned, so that tokens could be compared. (Data from each subject was processed separately.) A *Matlab* routine was written to find the best alignment of data from any two coils, taking into account both the x and y dimensions and using the Euclidean distance metric. The best alignment was deemed the one in which the distance (sum of Euclidean distances at each point) between the data from the two coils was at a minimum. For each of the tokens, the best alignment of each coil's data to the corresponding data of every other token was calculated. (Reference coils were not included as their stability led to alignments which differed substantially from the other coils.)

The *alignment distance* was quantified as the number of samples which one coil trace needed to be shifted to align it with the corresponding trace for the other tokens. The sign of the shift value refers to the direction in which the coil traces need to be shifted relative to each other. As the optimal shift distance varies slightly from coil to coil, an average shift was required to keep all of the articulatory data in step when two utterances were aligned. Shifting each coil trace of one utterance a slightly different distance would have resulted in articulatory data which was uninterpretable in terms of the original acoustic waveform. Thus, for each pair of tokens, the mean alignment distance over all the coils (the shift needed to align all the coils for those two tokens) was calculated. The corresponding standard deviations were calculated. A standard deviation of 3 corresponds to a shift of three samples which at the sampling rate of 500 Hz is 6 ms. In most cases the mean alignment was a good choice for aligning the data from two utterances. However, in a few cases, the standard deviations were large, *e.g.* the mean alignment distance between EB *leap* token 3 and token 4 has a standard deviation of 133, or 0.266 s. For each standard deviation above 10 samples, the shift values for each coil were inspected. In each case it turned out that there were only one or two pairs of coils which had a noticeably different alignment value and were causing the large standard deviation. The alignments were checked, and in these cases it appeared that adjacent similar peaks in one coil trace had caused the odd alignment, and that the values for the other coils were more accurate in terms of the acoustic data. As in each case there were no more than two outliers, the median alignment values were used in these cases.

For each utterance, one token (the first instance) was arbitrarily chosen as the reference token to which all the others would be aligned. They were then aligned using a purpose-written *Matlab* routine and the median alignment values (rounded up to the next whole unit). The routine aligned the remaining tokens, one by one, with the first (reference) token by shifting the data in the required direction by the alignment distance. The initial section of each non-reference token's file was either padded with zeros, or deleted, depending on the direction and distance which the token needed to be shifted relative to the reference token. As much of the initial part of each recording was silence (the time from instructing the subject to speak until the onset of phonation) this led to no loss of speech data. The aligned tokens were then truncated to be of equal length. This had the result of removing final periods of silence from the files. Finally, the longest stretch of initial zeros (introduced as padding) for any repeat was found and the files were again truncated to remove these zeros and the corresponding non-speech stretches in the other files.

The files for each word pair were then examined together. From the aligned data, an average set of traces was computed for each word. The average /l/ and /r/ data for each word pair were then used to calculate the best alignment between /l/ and /r/ utterances (the rounded up median alignment of all coils) and all of the /l/ tokens were aligned to all the /r/ tokens, by truncating the longest tokens appropriately. As all of the repeats for one utterance were now the same length, this was a simple procedure. The audio files corresponding to the first repeat - the reference token for alignment - were similarly truncated to align with the articulatory data, so that a waveform would be available for display alongside the averaged, aligned trajectories.

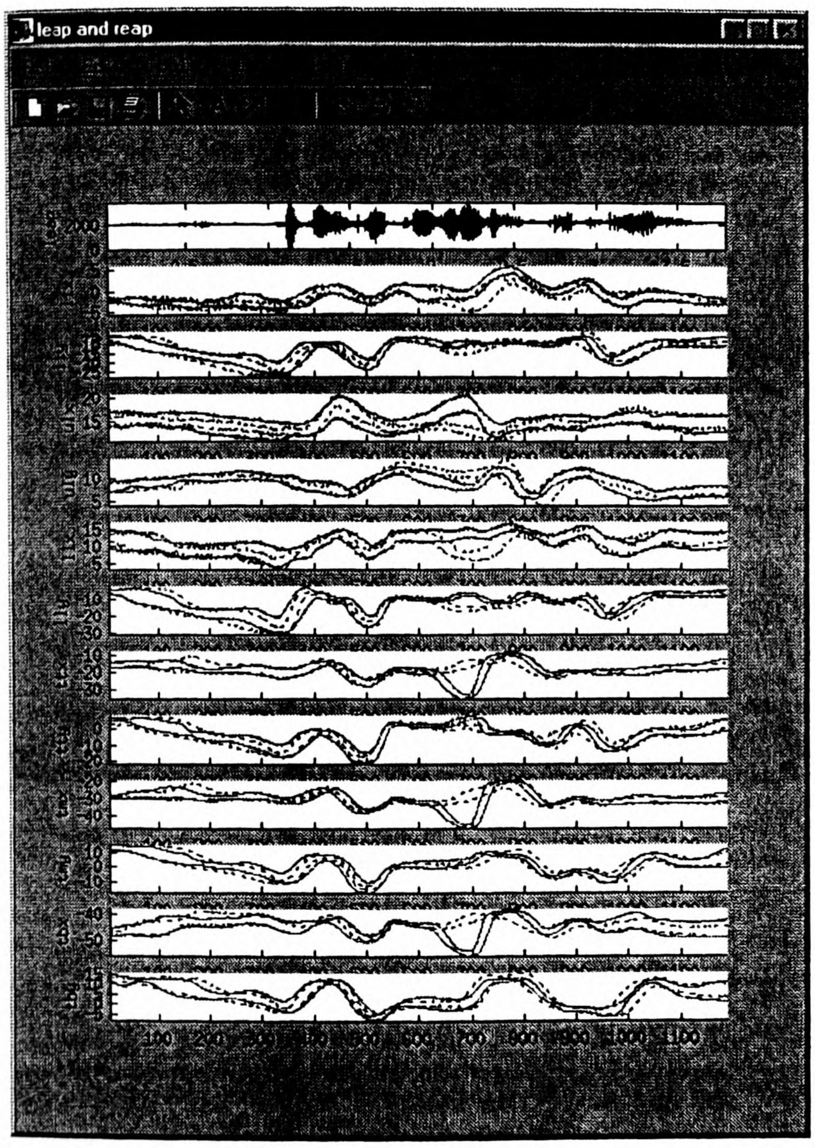
As a first measure of difference between the /l/ and /r/ sets for each word pair, 95% confidence intervals were calculated from the averaged data and aligned repeats. Confidence intervals were calculated at each sampling point *t*, using the formula:

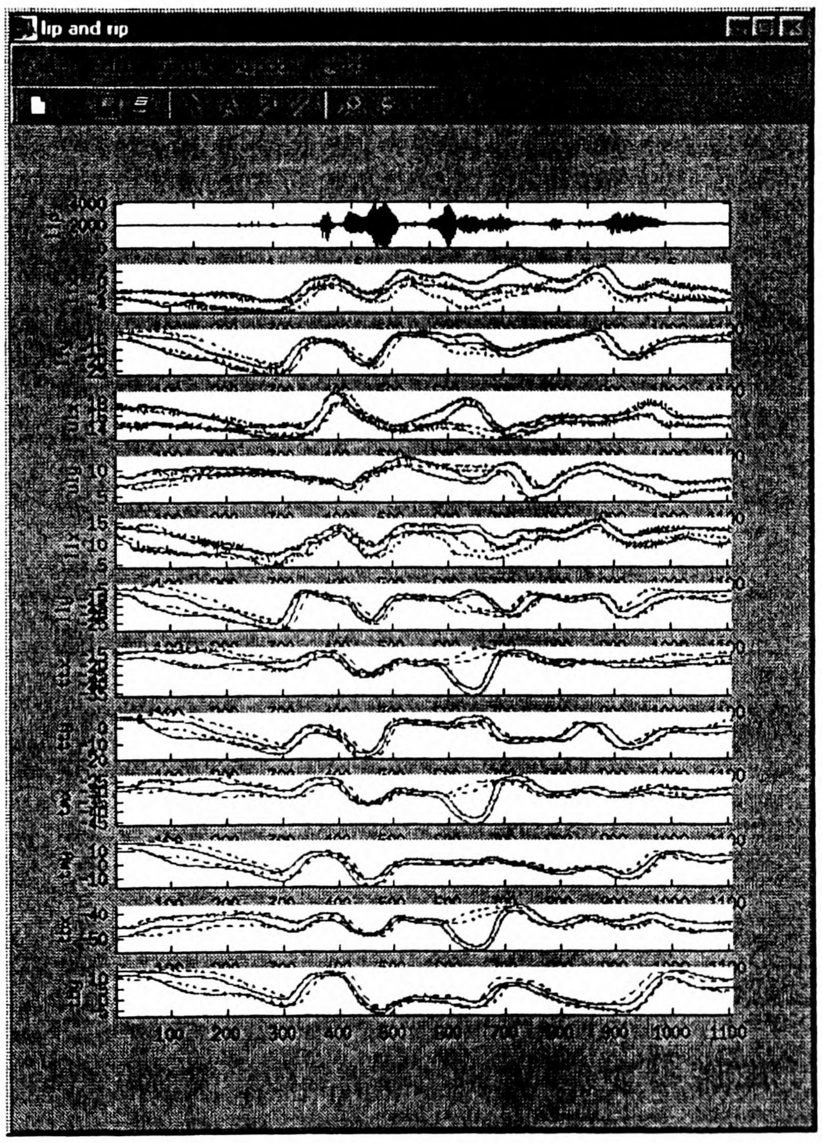
$$\text{ci(t)} = X(t) \pm \frac{2.97 * s(t)}{\sqrt{n-1}}$$

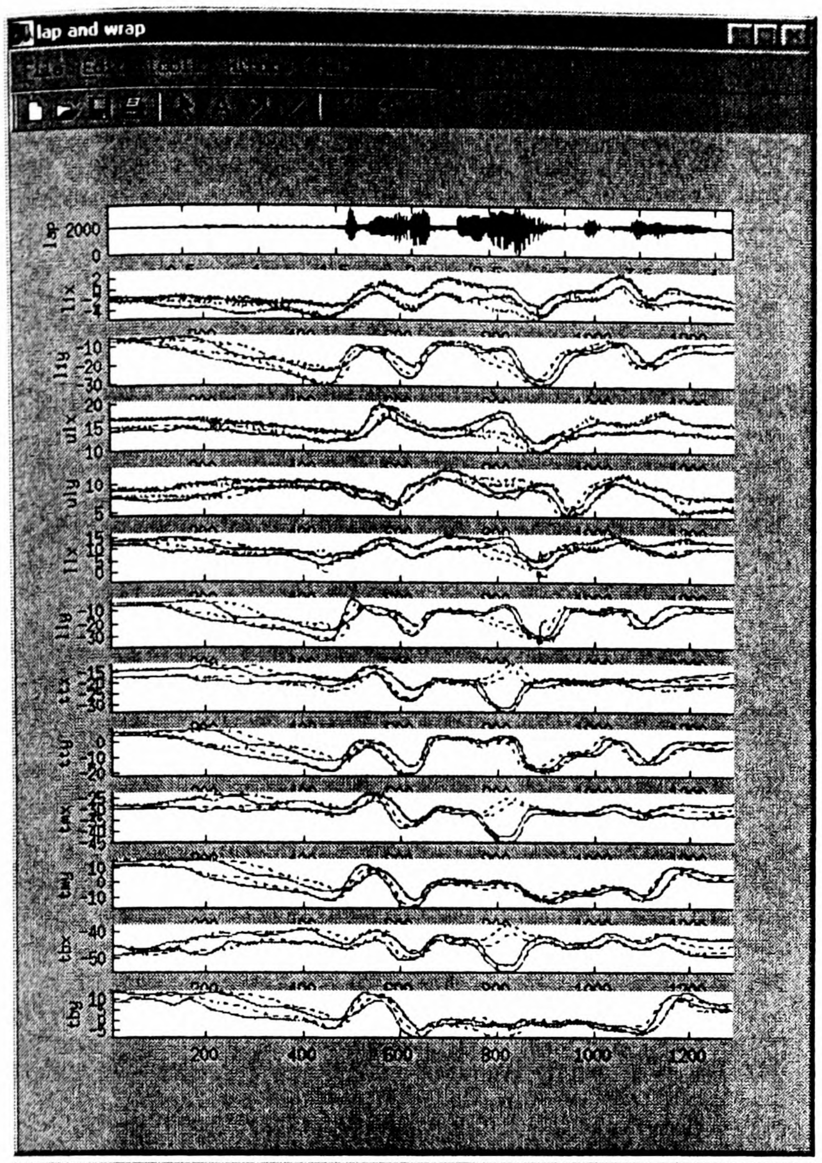
where *X* is the mean value of the sample, *s* the standard deviation and *n* the number of samples. The threshold value of 2.97 was used because of the small sample size. These confidence intervals are plotted in Figure 4.1 for EB, Figure 4.2 for MB and Figure 4.3 for AS, along with acoustic data from one token (the reference token for alignment) of the /l/ word. Inspection of these plots shows obvious differences in articulation, local to the liquid and adjacent vocalic material and occasionally spreading to earlier or later material. The liquids can be identified roughly by amplitude drops in the waveform aligned with the large differences in the articulatory data. The order of traces plotted is (from top to bottom): /l/ waveform, lower incisor *x* and *y*, upper lip *x* and *y*, lower lip *x* and *y*, tongue tip *x* and *y*, tongue mid *x* and *y* and tongue back *x* and *y*.

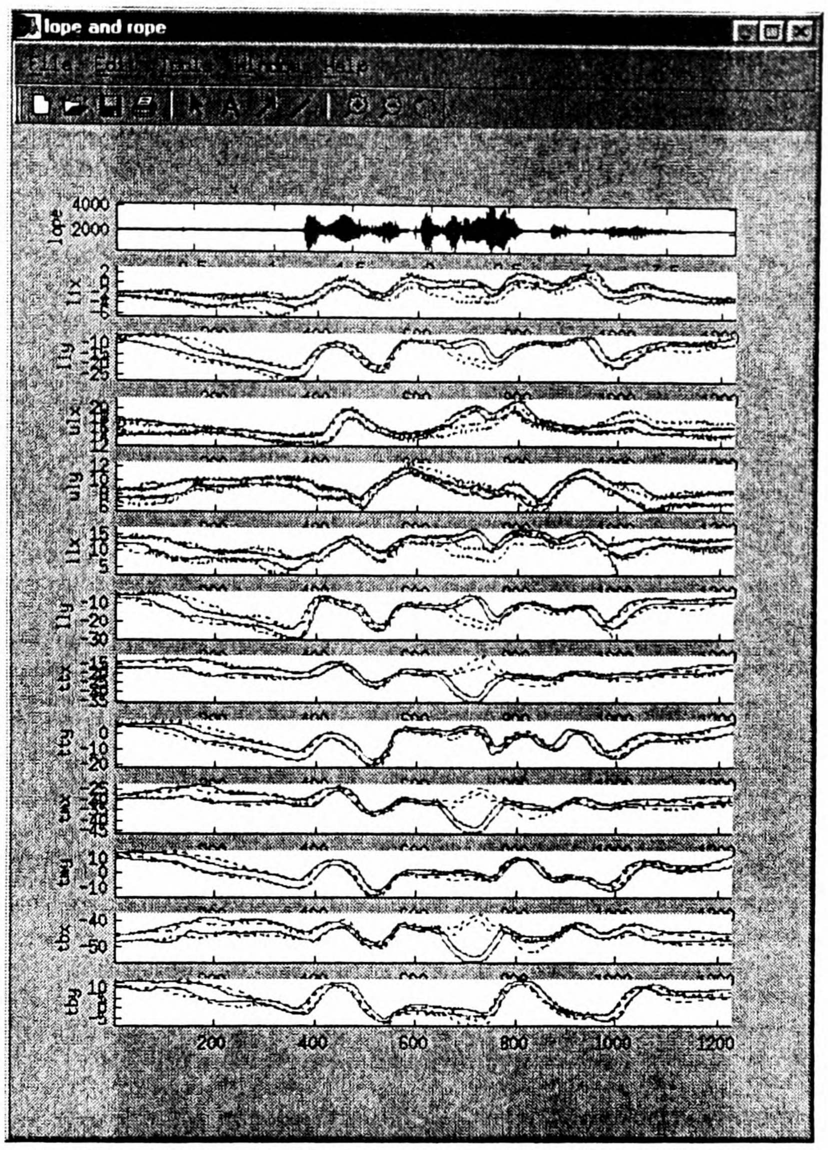
For EB, most clearly visible is retraction of the tongue in the /r/ context relative to the /l/ context (decrease in ttx, tmx and tbx), raising of the lower lip (lly) and lower incisor or jaw (liy) (increase in *y* values), lowering of the upper lip (decrease in uly values) and protrusion of both lips and jaw (increase in ulx and llx values). MB shows fewer differences in the tongue data, but clear evidence of lip rounding in the /r/ context: protrusion of the lips and jaw (higher llx, lix and ulx) and raising of the lower lip (higher lly and liy) in the /r/ context. AS shows clear retraction of the tongue (lower ttx, tmx and tbx) and some raising (higher tmy) in the /r/ context, together with fronted lips and jaw (greater lix, llx and ulx), lowered upper lip (lower uly) and raised lower lip (higher lly).

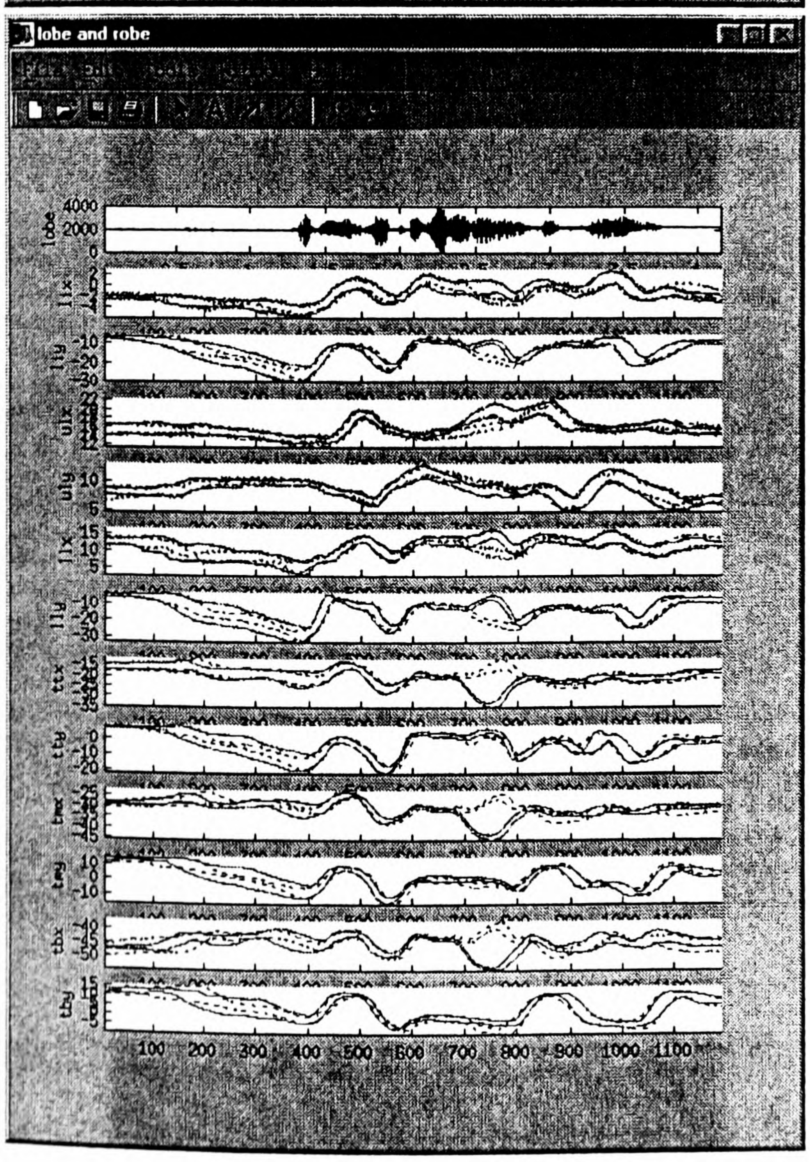
Figure 4.1: 95% confidence intervals for EMA data, subject EB. Solid lines indicate /r/ articulatory traces, dashed lines /l/.






lap and wrap


lope and rope


lobe and robe

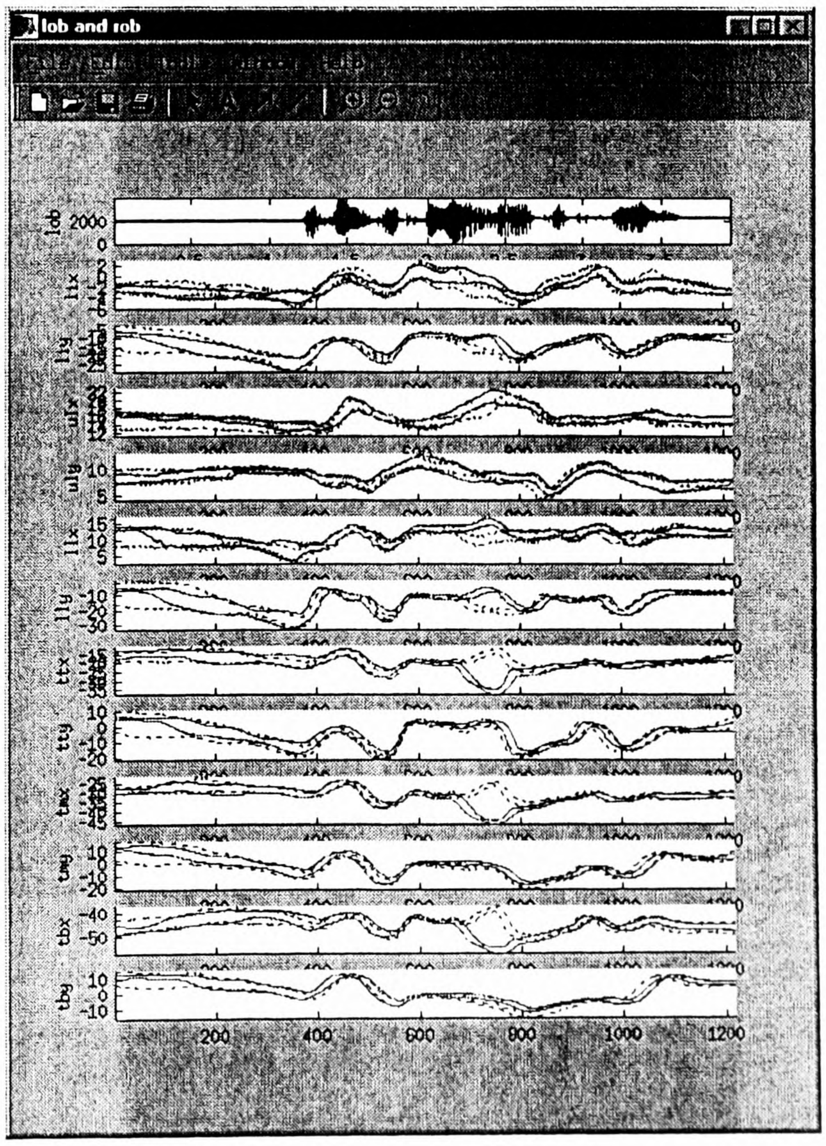
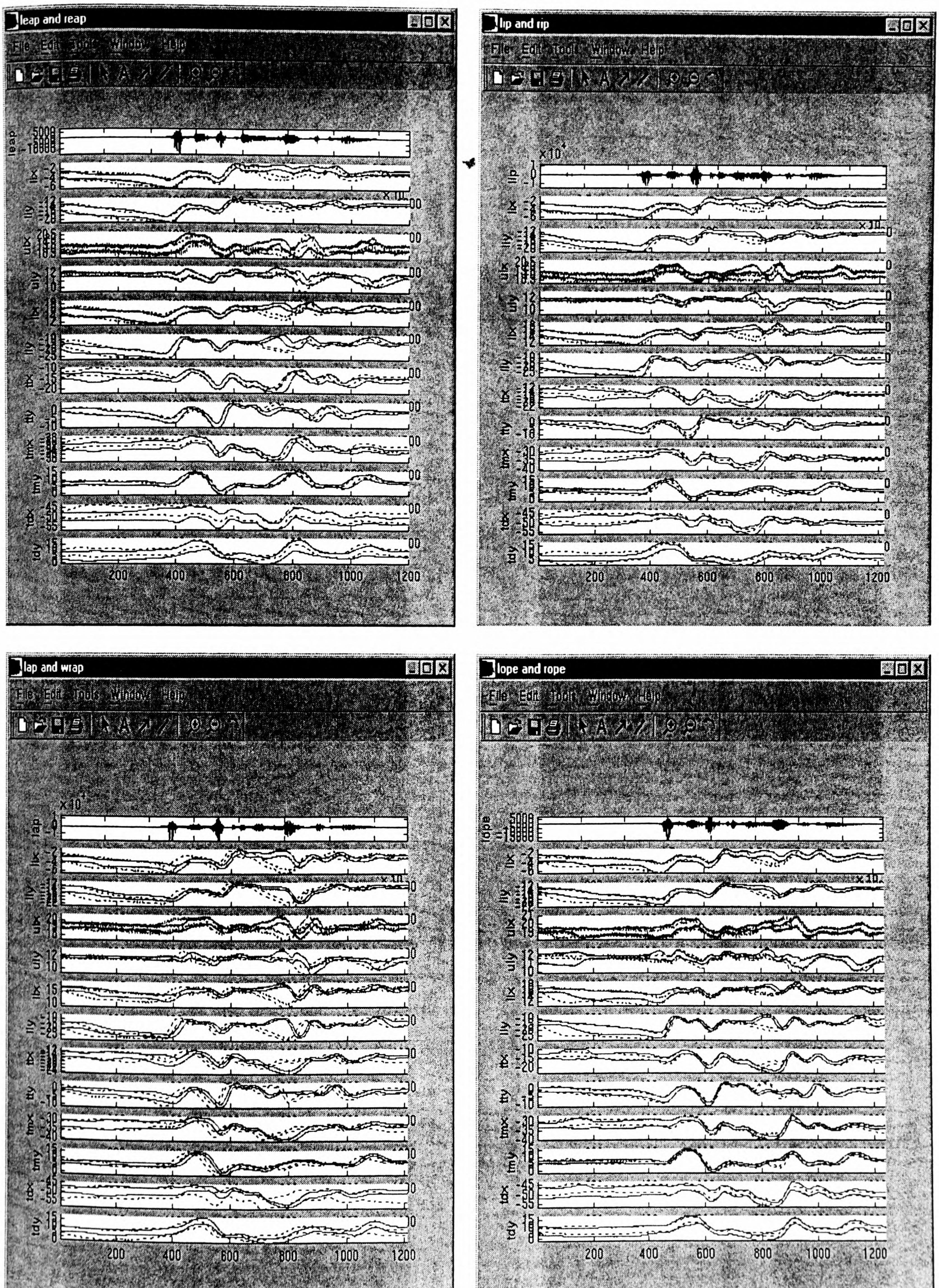

lob and rob

Figure 4.2: 95% confidence intervals for EMA data, subject MB. Solid lines indicate /r/ articulatory traces, dashed lines /l/.



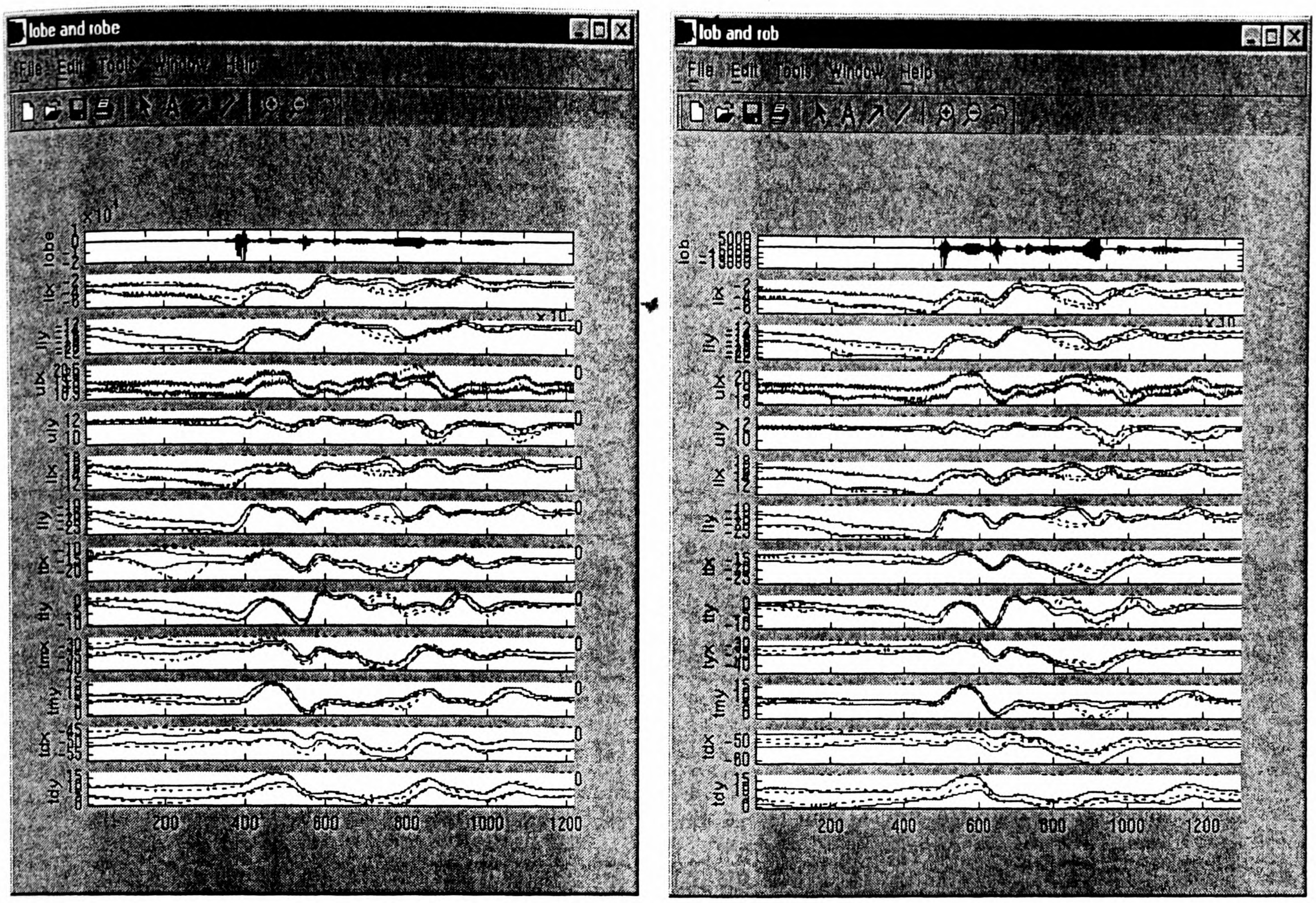
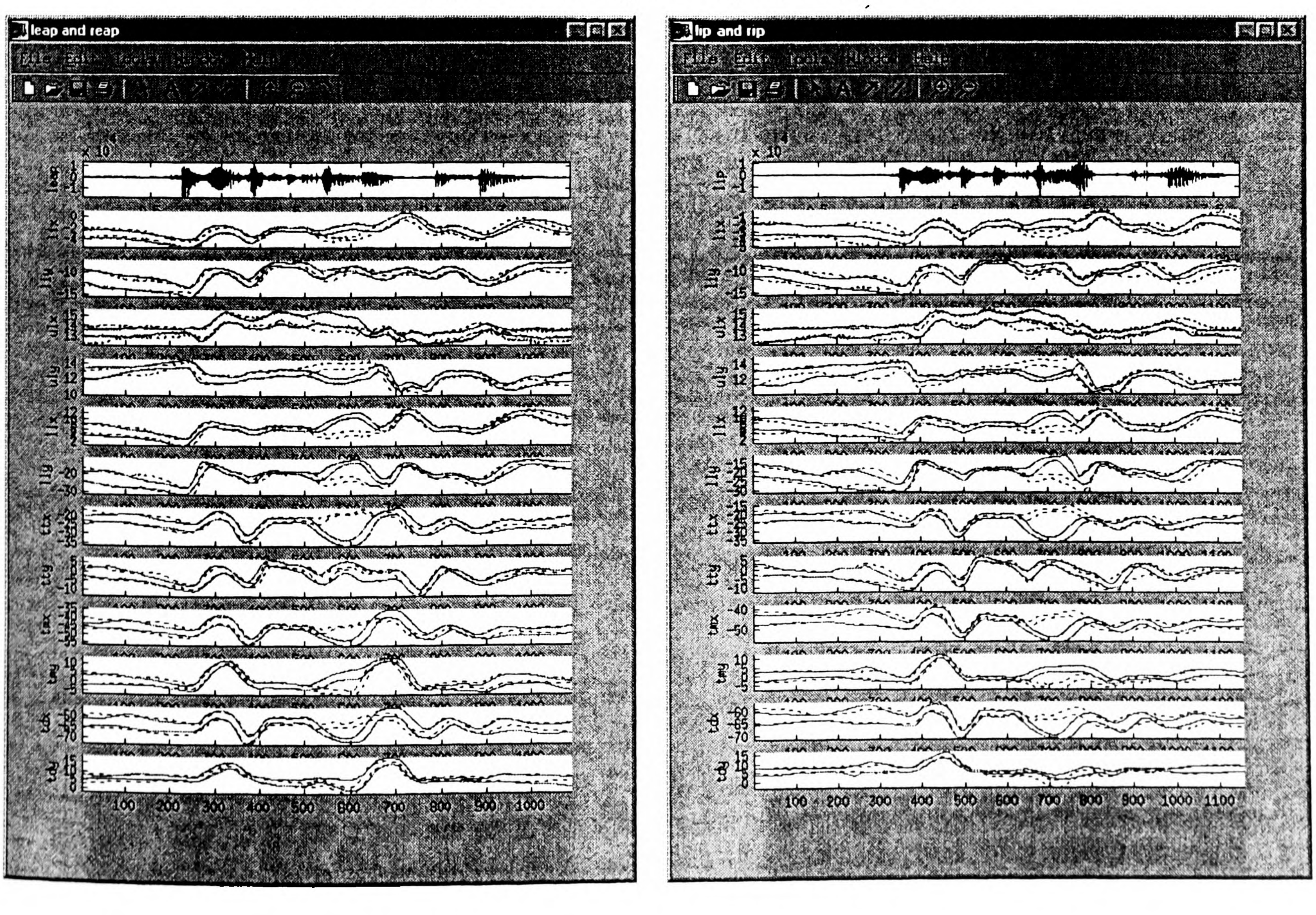
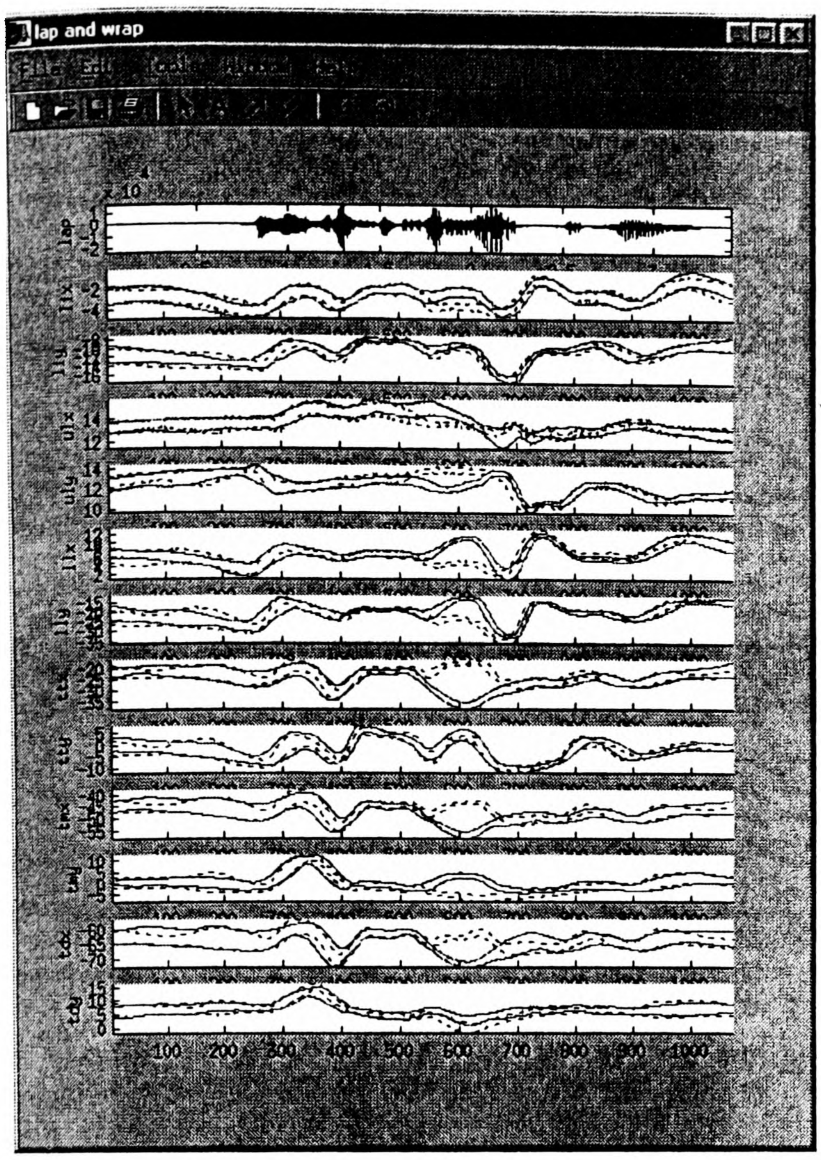
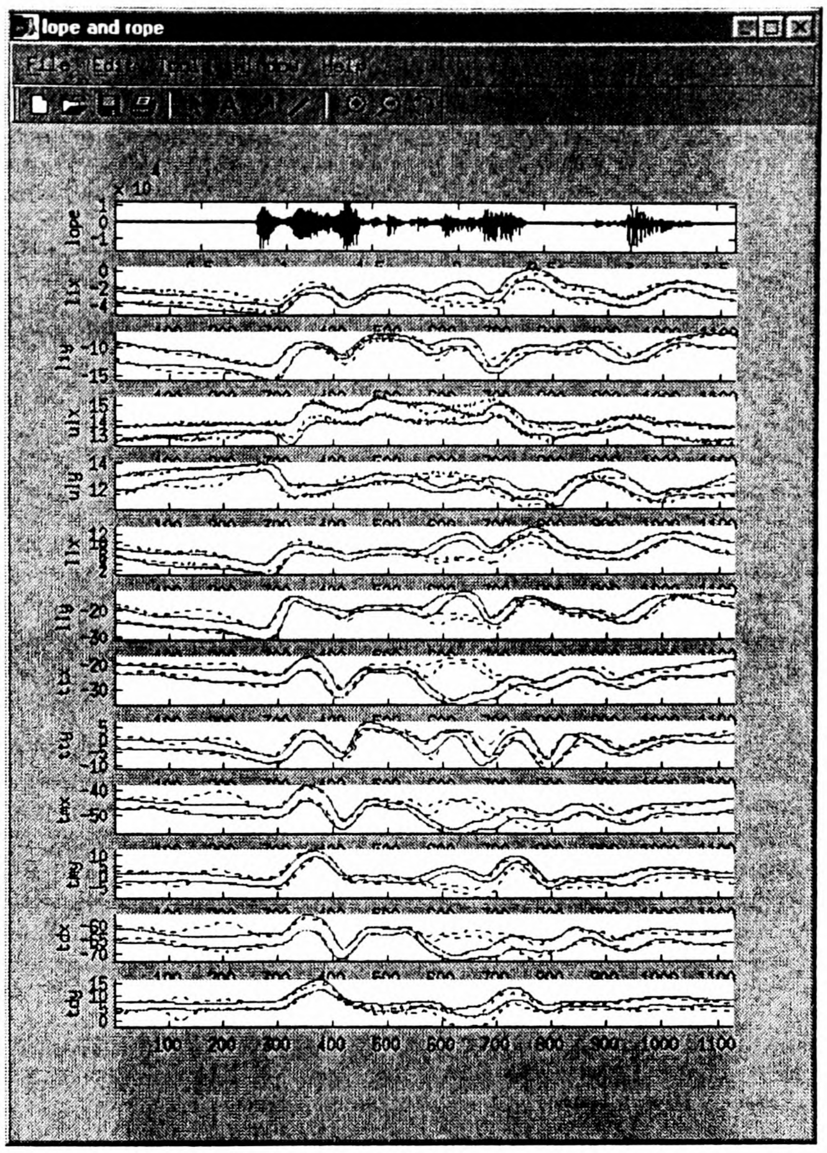
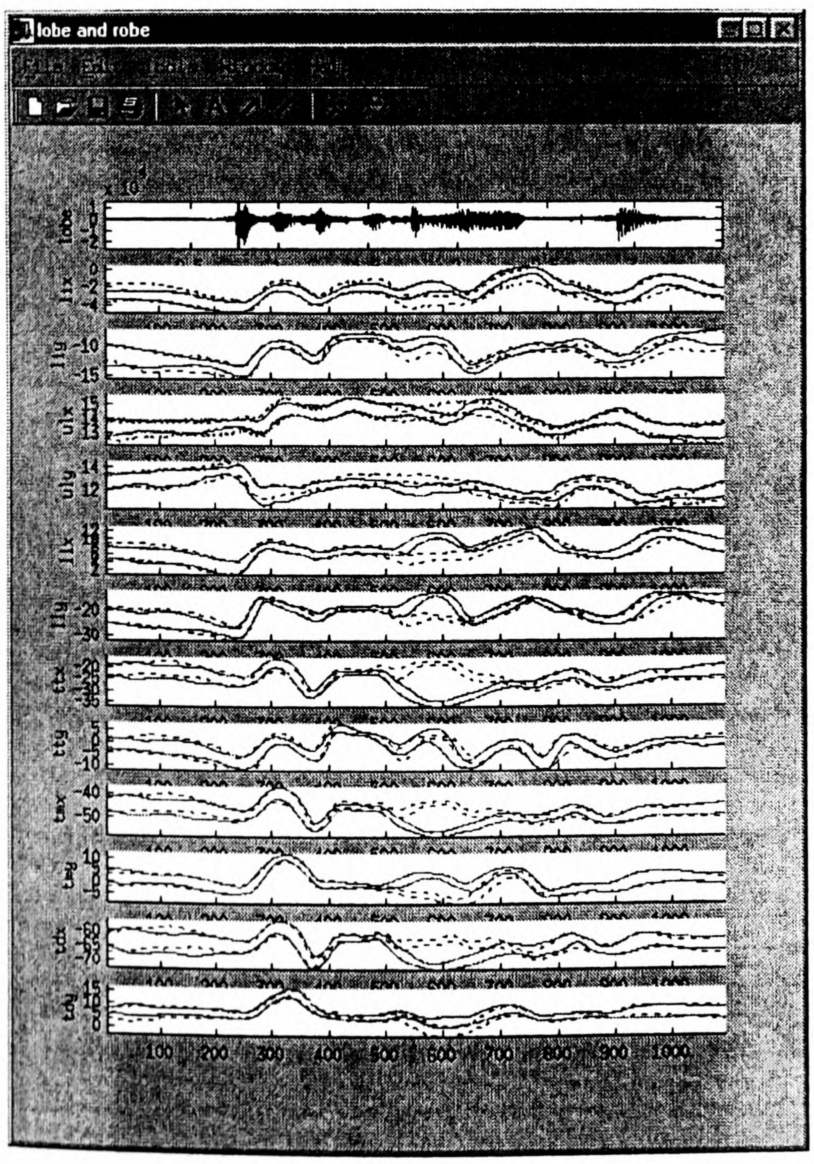


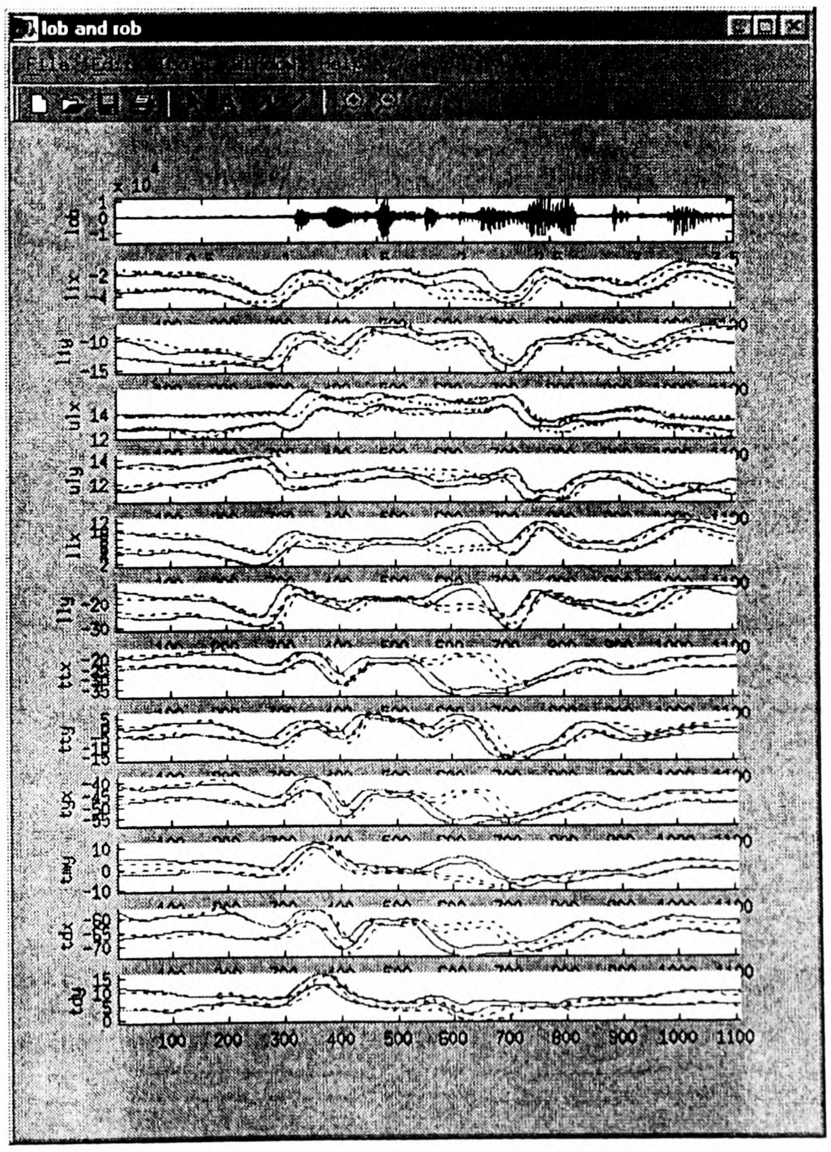
Figure 4.3: 95% confidence intervals for EMA data, subject AS. Solid lines indicate /r/ articulatory traces, dashed lines /l/.




lap and wrap


lope and rope


lobe and robe


lob and rob

The statistical analysis of time-series data can be complicated (Chatfield 1989). Repeated *t*-tests across time-series cannot be done without also employing correction factors for the autocorrelation of the series and the large number of tests. One method of correction for autocorrelation is to calculate the lag one autocorrelation coefficients and to use these to estimate the variance for each sample. The formula for the variance of $\bar{y}$ (the averaged time-series data for one coil of one word) is given by:

$$v(\bar{y}) = \frac{\sigma^2}{n}\left[1 + \frac{2(n-1)}{n}\rho_1\right]$$

where $n = 7$ (the number of repeats for EB) and σ^2 is estimated by s^2 (the known variance). ρ_1 = the lag one autocorrelation coefficient, calculated by putting $k = 1$ in:

$$\rho_k = \frac{\sum_t^{n-k}(y_t - \bar{y})(y_{t+k} - \bar{y})\frac{1}{n-k}}{\sum_t (y_t - \bar{y})^2 \frac{1}{n-1}}$$

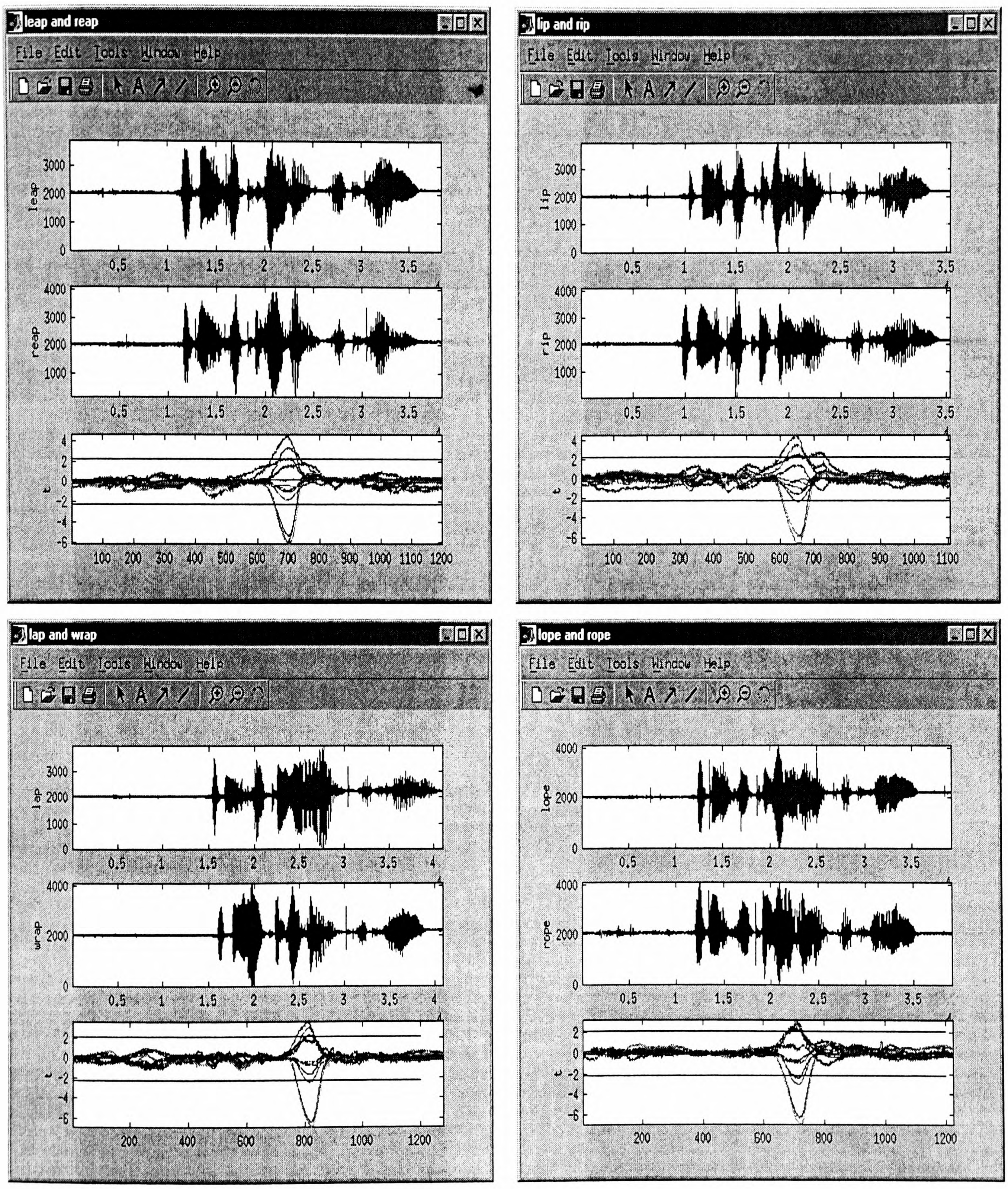
y_t is the value of the time series at time t, and $\bar{y}$ the mean value calculated over the whole of y (Box *et al.* 1978). However this assumes that all the correlations after lag one are zero, which is not the case for this articulatory data. Nevertheless, performing this correction was deemed better than leaving the data entirely uncorrected. For each word pair, the *t*-statistic was thus calculated for the data from the aligned /l/ and aligned /r/ coils at each sampling point (x_1 and x_2) using the square root of the estimated variance as the standard error:

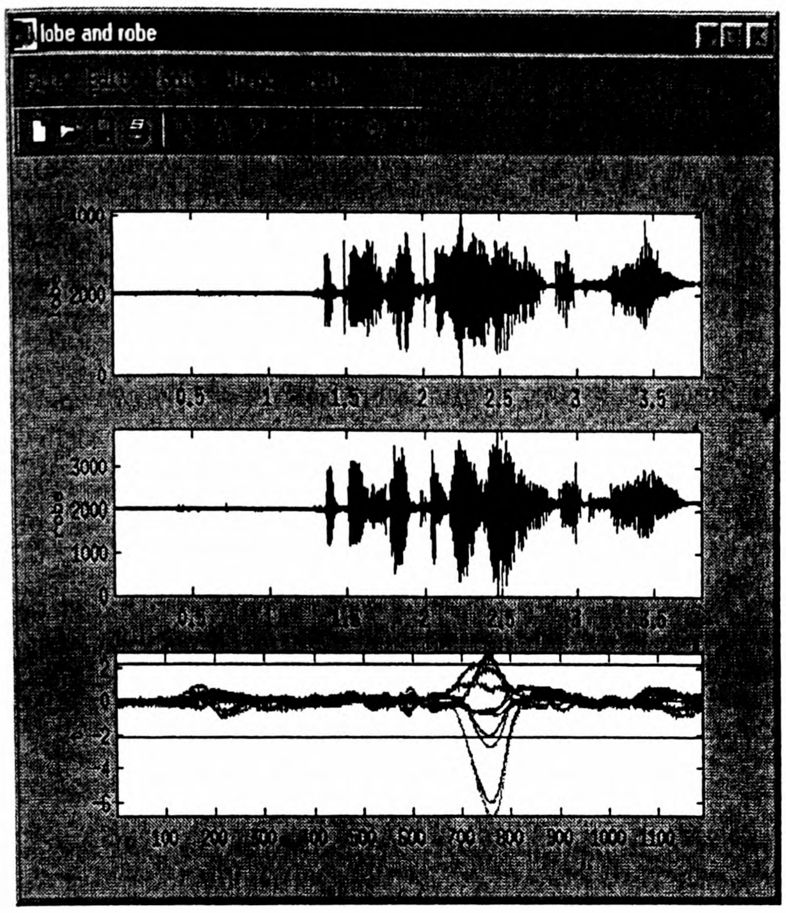
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{v(\bar{x}_1) + v(\bar{x}_2)}}$$

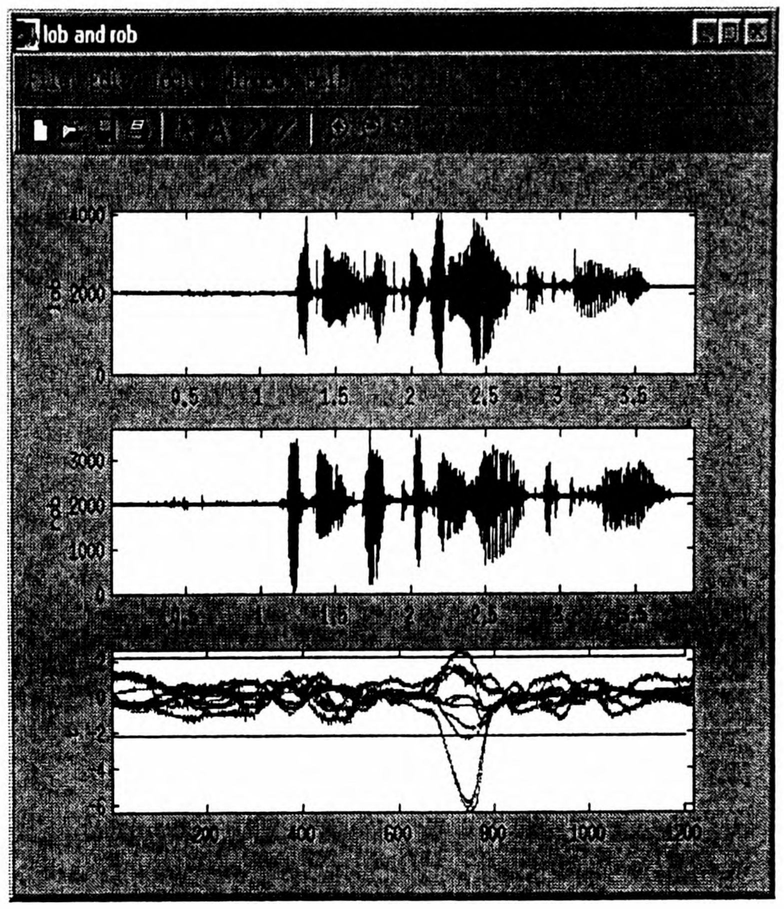
The results of these calculations are displayed for subject EB, together with corresponding acoustic data from one token of each of the utterances in Figure 4.4. 5% significance lines ($t > 2.17$, df = 12, two-tailed test) are displayed with the *t*-test results. These lines suggest that there are significant differences between the /l/ and /r/ data sets local to the liquid and parts of the adjacent vowels (where the t curves lie outside the 5% confidence intervals), with longest stretches ranging from 172 ms for the *lope*/*rope* pair to 262 ms for the *lip*/*rip* pair. All pairs have differences in lower lip and tongue x data, all except *lob*/*rob* in upper lip x data, and all except *leap*/*reap* and *lip*/*rip* in lower lip y data. Additional differences are found in lower incisor x (*lip*/*rip*, *lap/wrap*, *lob/rob*) and lower incisor x and y (*lope/rope*). The longest extent of differences at this level of significance are, with the exception of lower incisor for *lip/rip*, the tongue x data, with differences extending over about 100 ms.

However, this level of significance does not include correction for the large number of tests undertaken. The Bonferroni correction requires a significance level of roughly $p < 0.00004$ (or $\frac{0.05}{1200}$), as roughly 1200 time points were tested, or t considerably greater than 4. As inspection of the figure shows, this would certainly rule out all long-domain effects. To a certain extent, using repeated tests across the whole utterance is unnecessary however, as differences are only expected within a limited domain. Thus it is possible that a lower significance threshold could be used, as has been shown for event-related brain potential data (Guthrie and Buchwald 1991). Nevertheless, inspection of the waveforms for the t-test results highlights a problem which arises from viewing the data in this way: although the articulatory signals are well-aligned, the waveforms are not perfectly matched. This is unsurprising, as the data were naturally produced sentences all with slightly different timing and duration. Differences in the confidence intervals or t-test results thus plotted cannot always be given a straightforward acoustic interpretation. For this reason, these methods of investigation seemed unpromising for the identification of long-domain differences, and were not pursued further.

Figure 4.4: *T*-test results, plotted with 5% significance limits (horizontal lines) for all time-aligned EMA data (all *x* and *y* coil traces) from subject EB, by word pair. Waveforms for one /l/ and one /r/ token are shown above the *t* curves (/l/ token above the /r/ token).







4.2.2. Dynamic time warping

Whereas the alignment method used above simply shifted the traces earlier or later in order to minimise the difference between them, dynamic time warping (DTW) was used to align the traces *non-linearly*. DTW is a method of non-linearly aligning signals that has previously been used in the analysis of speech data by Strik and Boves (1990), Slater and Coleman (1996) and van Santen (1997). The technique was used here to align the multiple repetitions of the EMA data more precisely than the shifting described above, removing small variations in the timing of articulatory events by non-linearly aligning the signals to each other. This corrects for timing discrepancies between tokens, reducing noise in the data. The /l/ and /r/ pairs were aligned by using DTW to remove timing differences between them, so that the magnitude and direction of contrasting articulatory events could be compared. This procedure removes any temporal differences between /l/ and /r/ utterances and thus may in fact be removing some long-domain differences. Nevertheless it allows comparison across signals in the certainty that corresponding articulatory events are being compared.

For each word, the repetitions were aligned using a standard DTW program developed in-house (Slater and Coleman 1996, following Macchi *et al.* 1990, *cf.* van Santen 1997). A centroid or prototype token was chosen, so that all the other tokens could be warped to one time path. This method was adopted, rather than computing a median for warping, so that the time course of the warped utterances was taken from a naturally produced one. The prototype was chosen as the token closest to all of the other tokens, in terms of overall Euclidean distance from the others. In order to compute the distance the alignments of repeats described in the previous section were used so that distances would be at a minimum and errors less likely to occur. The distance was then calculated as the sum of the Euclidean distances between each coil position at each time point. (x, y) pairs were used to calculate the distances, although warping was done on x and y traces separately. For each token the distances to the other repeats were summed, and the token with the smallest overall distance to the other tokens was used as the reference for warping.

All of the EMA coil traces of a token were warped simultaneously and in the same way, based on the best overall fit to the reference token. The DTW algorithm calculated a warp function based on a distance matrix calculated from all of the EMA coils together (sum of the Euclidean distances from the token to the reference). The warp function is the sequence of corresponding time points from the reference signal and test signal (token to be warped) with least overall distance between the two tokens. In this way DTW eliminates differences between tokens. This procedure results in 12 sets of time-aligned tokens, one for each word. In order to make comparisons between the members of a word pair, the /r/ tokens were warped to the corresponding /l/ tokens. This was done by calculating the warp path required to map the prototype /r/ token to the prototype /l/ token of the pair, and then using this path to warp all of the /r/ tokens.

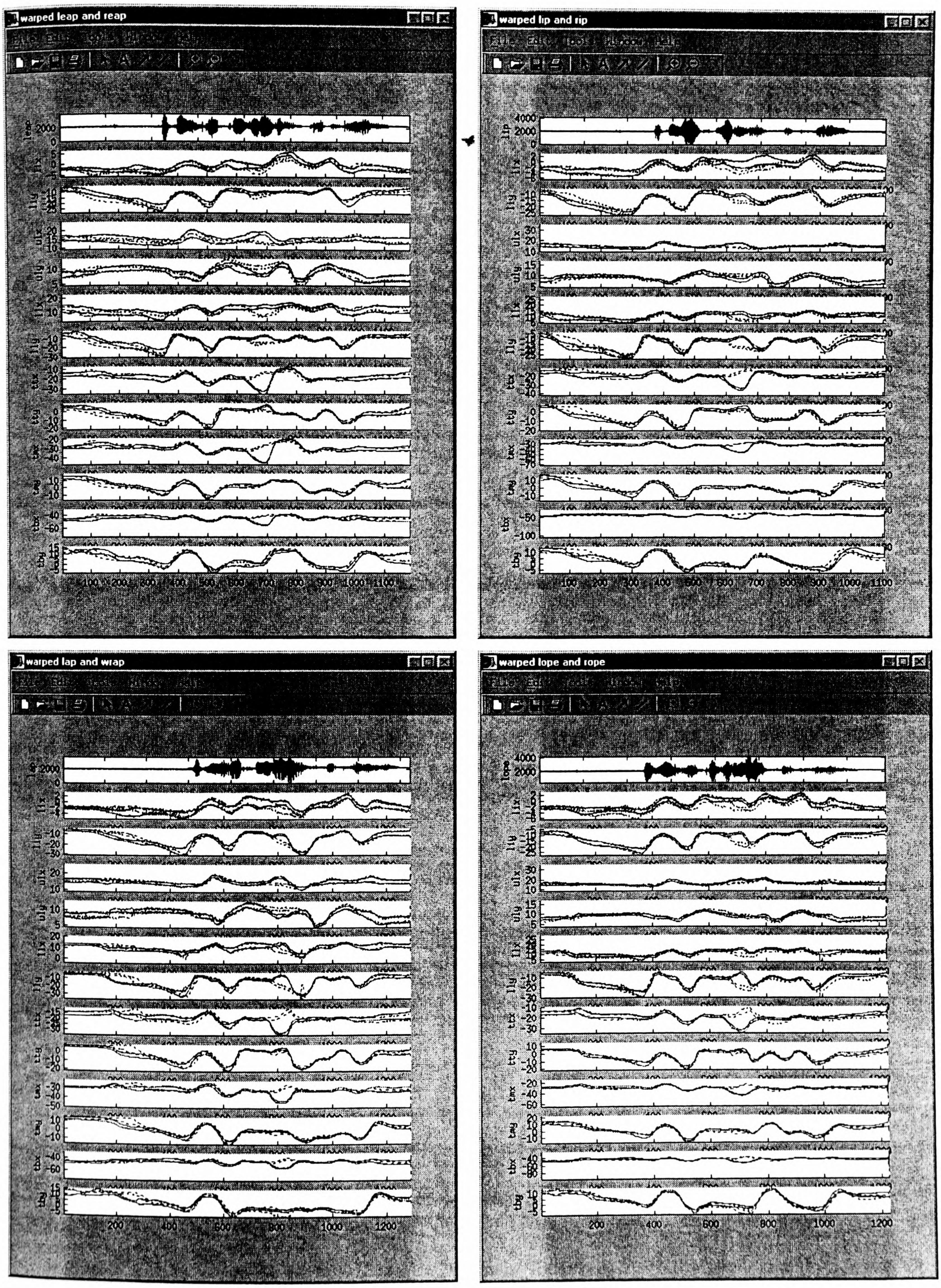
As discussed in 4.2.1, calculating statistical tests on time series is complicated and may hinder the identification of long-domain differences. 95% confidence intervals were calculated instead, based on the warped and aligned repeats for each word pair separately. These are displayed for all EMA coil traces for subject EB in Figure 4.5, MB in Figure 4.6 and AS, Figure 4.7. A waveform is provided as a guide to the warped data. However the acoustic signals could not be warped, and thus the signals corresponding to the reference /l/ tokens are displayed. These tokens were never warped, but had all other tokens warped to fit their time paths.

For EB differences can be seen in the figures for most of the coils, although the extent of differences is not great for most coils. This may be partly due to the effect of time warping, which reduces temporal differences. However upper lip y, lower incisor *x* and tongue back *y* show fairly long extents of difference for some word pairs. For example, lower incisor *x* differs over the two syllables before the liquid (/tədə/ from *uttered a*) and the syllable containing the liquid in *lip/rip* and *lope/rope* utterances. The lower incisor is lower in the /r/ context, showing protrusion in this context. Upper lip *y* shows a similar extent of differences for the *leap/reap* and *lope/rope* pairs, with lowering in the /r/ context. Tongue back *y* has relatively long perseverative differences in the *lob/rob* pair, with raising in the /r/ context over the word pair and following word *at* (/ət/). The tongue *x* differences are particularly large and show marked retraction of the tongue around an /r/ spreading to adjacent vowels (lower values in the /r/ context).

The warped data for MB shows some effects in the tongue data which were not immediately apparent from the aligned data: tongue mid is consistently raised in the /r/ context (higher tmy), and tongue tip is lowered (lower tty). Lower lip raising and protrusion is also apparent for the /r/ context (higher lly, liy, llx and lix). Relatively long-distance differences can be seen in lower incisor *x* for the *leap/reap* and *lip/rip* pairs, with protrusion over the sequences /ɹipət/ and /dəɹɪp/, respectively.

The time warping technique proved slightly problematic for AS's data. The word pairs *lope/rope* and *lobe/robe*, when warped based on the distance matrix calculated from all the coil data, warped the token incorrectly. This can be seen from a comparison of the warped data in Figure 4.7 with the unwarped data in Figure 4.3: at around the 800th sample there is a discrepancy. The /r/ data is stretched too far rightwards. These utterances are therefore excluded from the analysis. However, the overall results from the warped data are the same as for the aligned data. Both tongue raising and retraction (higher *y* and lower *x*) and lip rounding (higher *x* and *y* for lower lip and lower incisor, lower *y* and higher x for upper lip) are found in the /r/ context. Differences extend over the syllable before and after the liquid, in most cases. The *lob/rob* pair has differences extending into the second syllable before the liquid (*i.e.* over /tədə/ from *uttered a*) in tongue back *y* (raised in the /r/ context), tongue tip *x* (retracted in the /r/ context) and lower incisor *x* (protruded in the /r/ context).

Figure 4.5: 95% confidence intervals for thę warped EMA data, subject EB. Solid lines indicate /r/ articulatory traces, dashed lines /l/. The audio signal for the reference /l/ token is shown above.



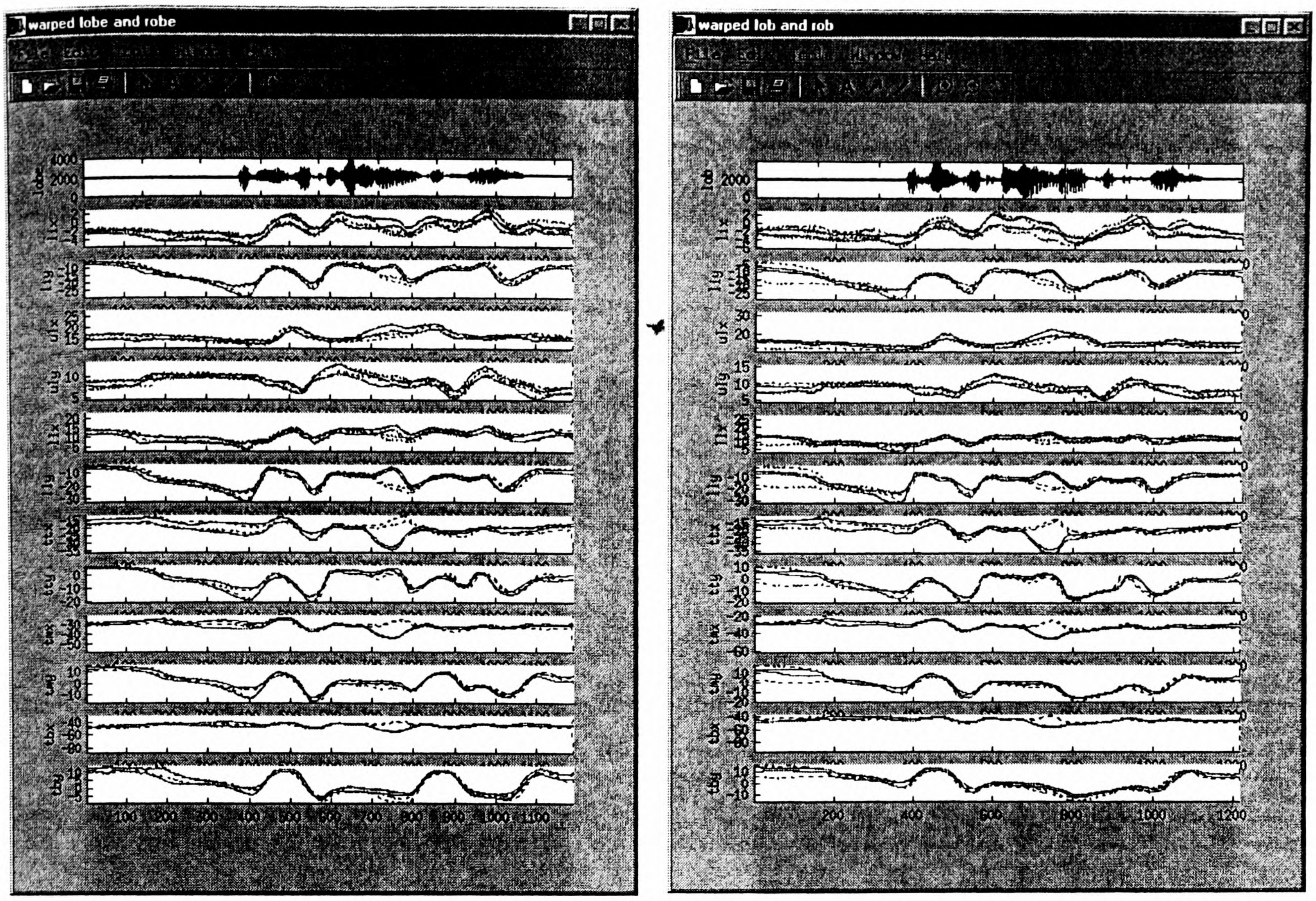
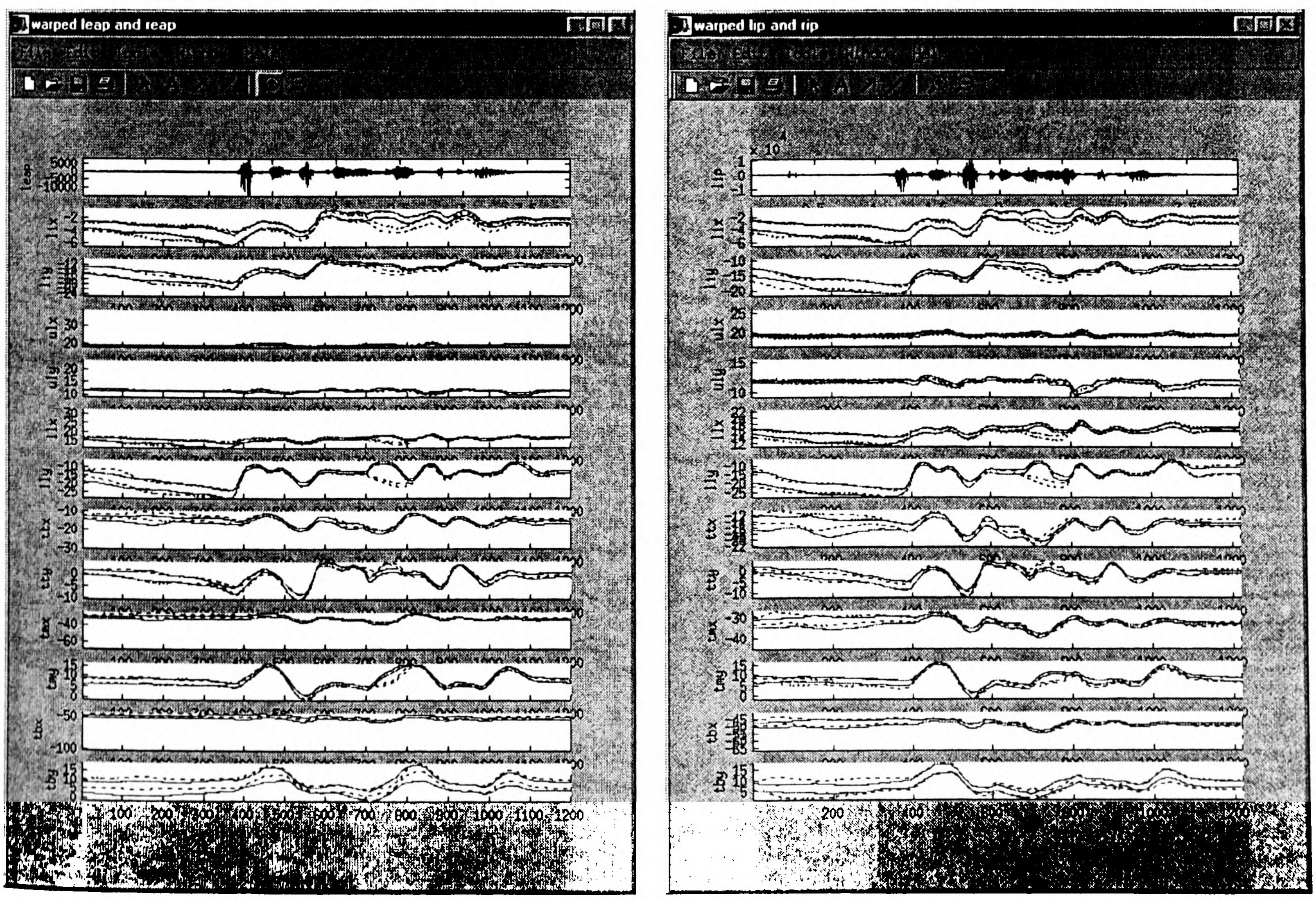
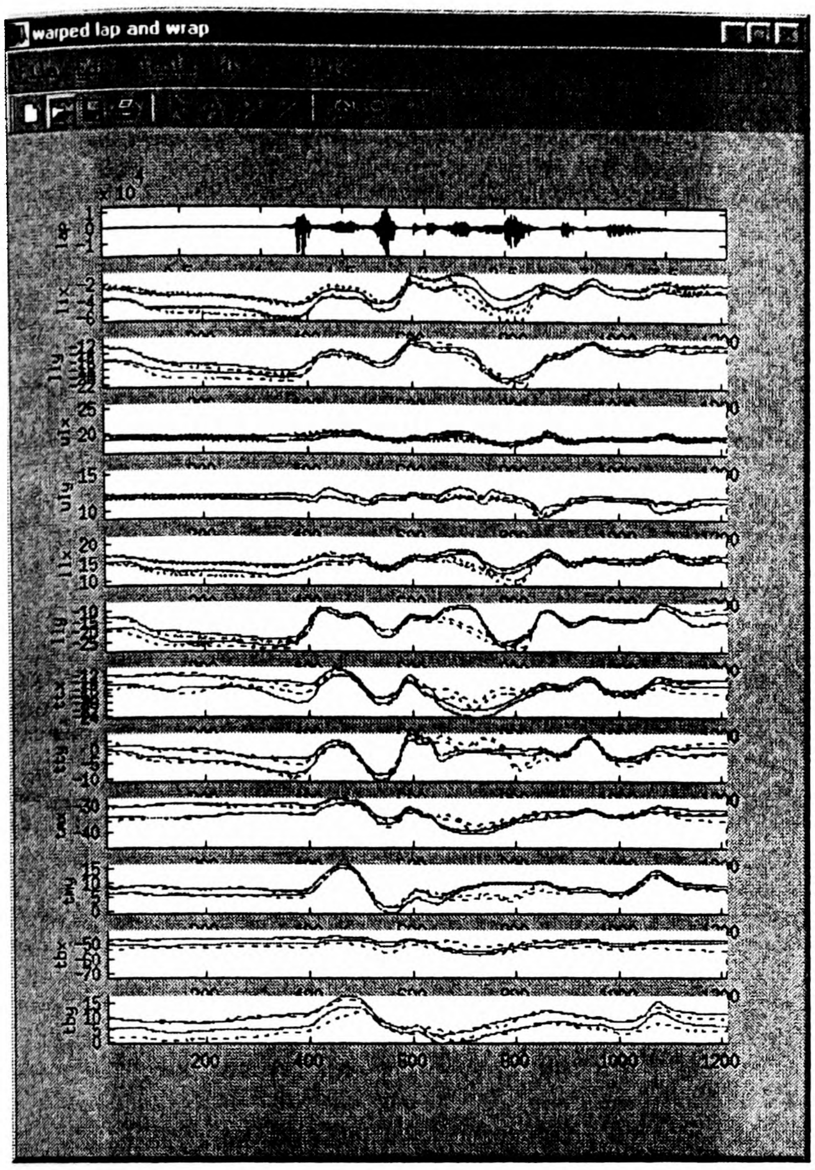
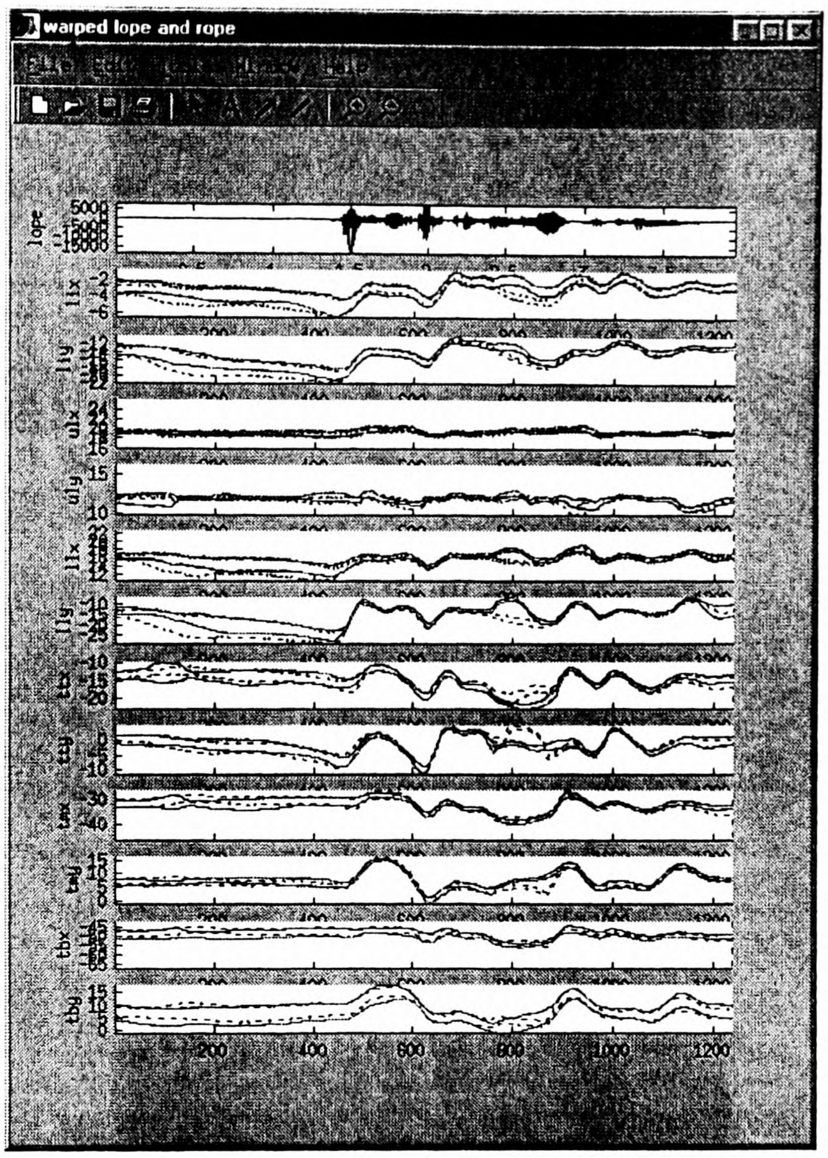
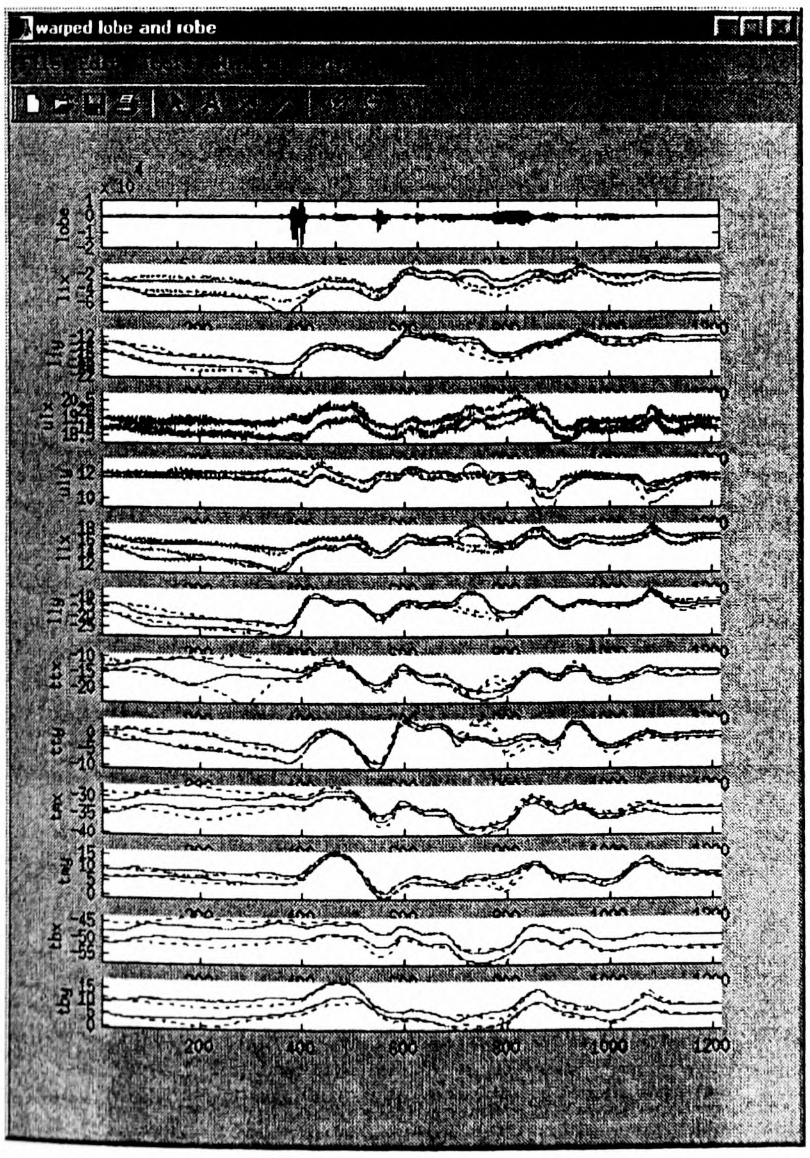


Figure 4.6: 95% confidence intervals for the warped EMA data, subject MB. Solid lines indicate /r/ articulatory traces, dashed lines /l/. The audio signal for the reference /l/ token is shown above.




warped lap and wrap


warped lope and rope


warped lobe and robe

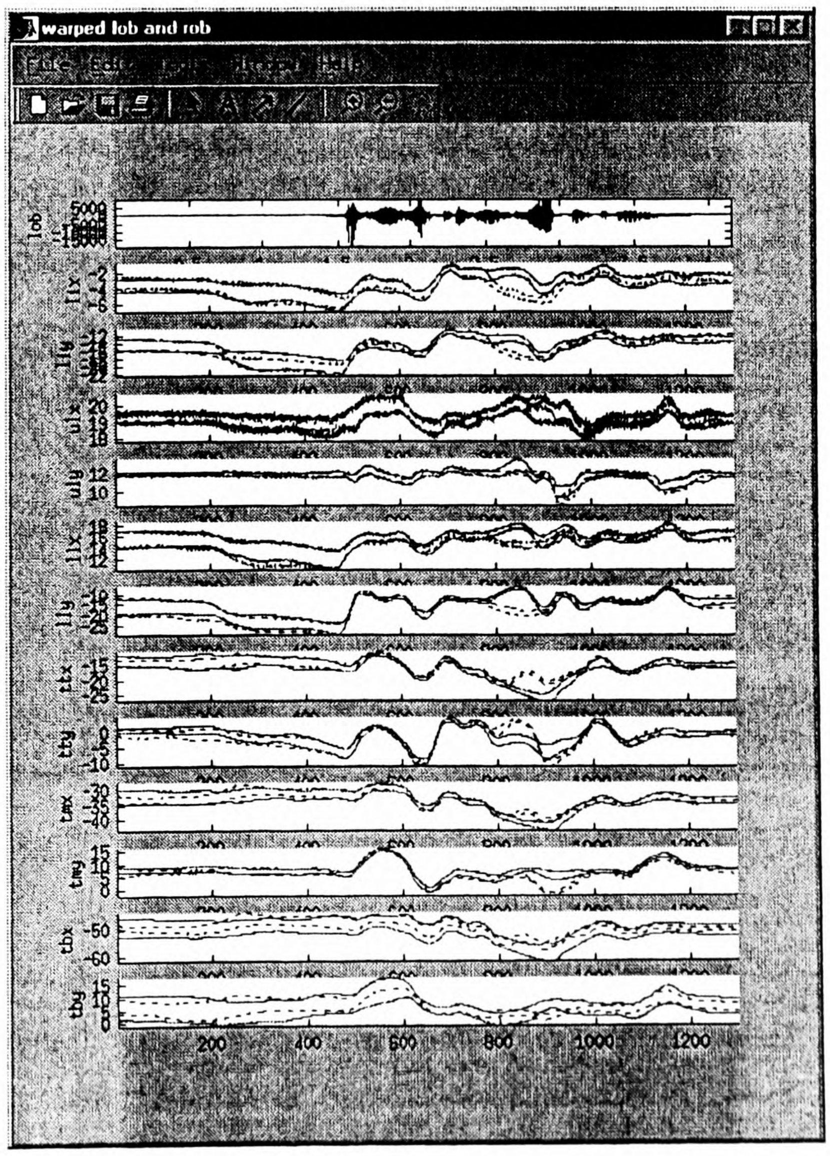
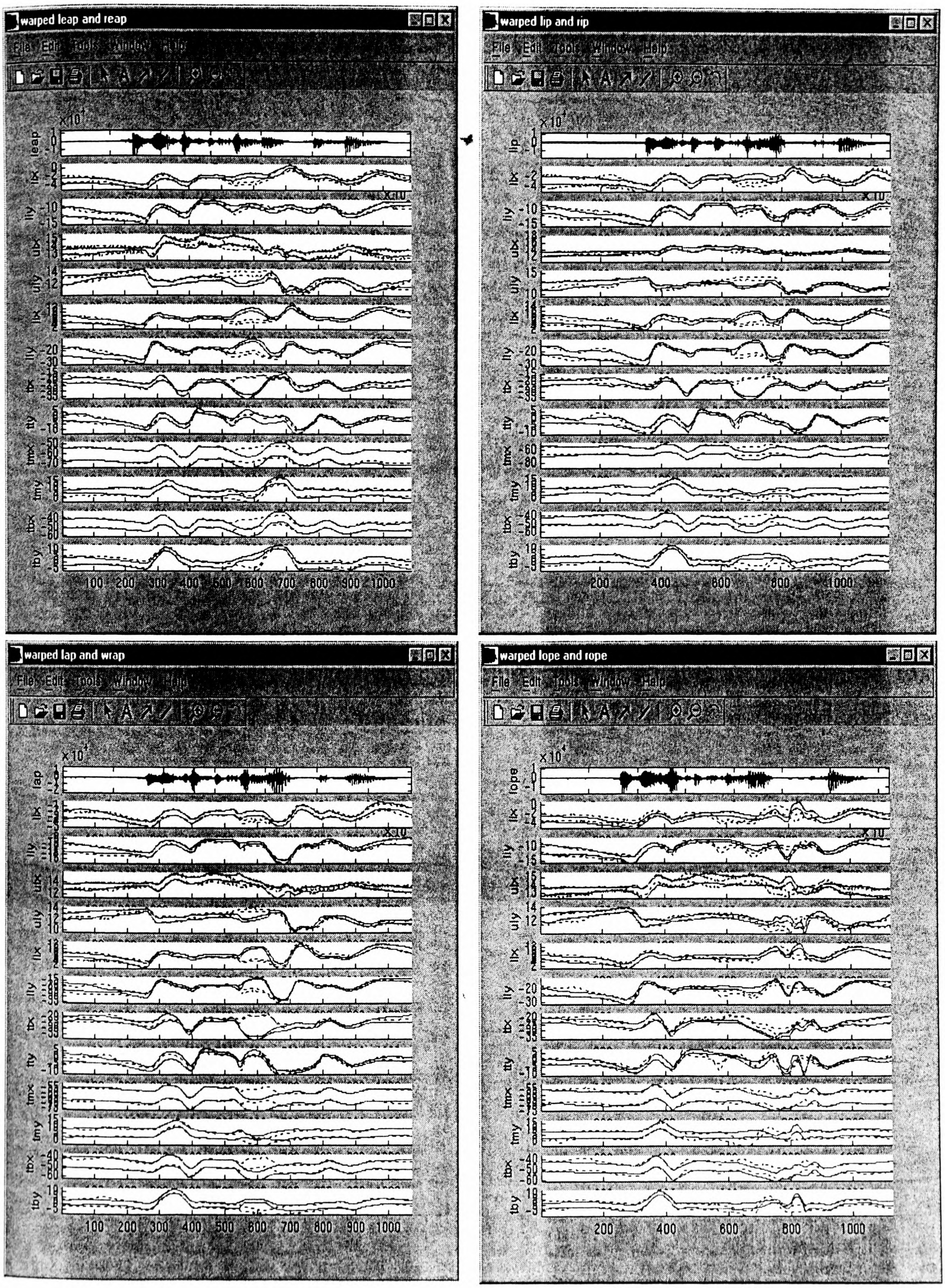
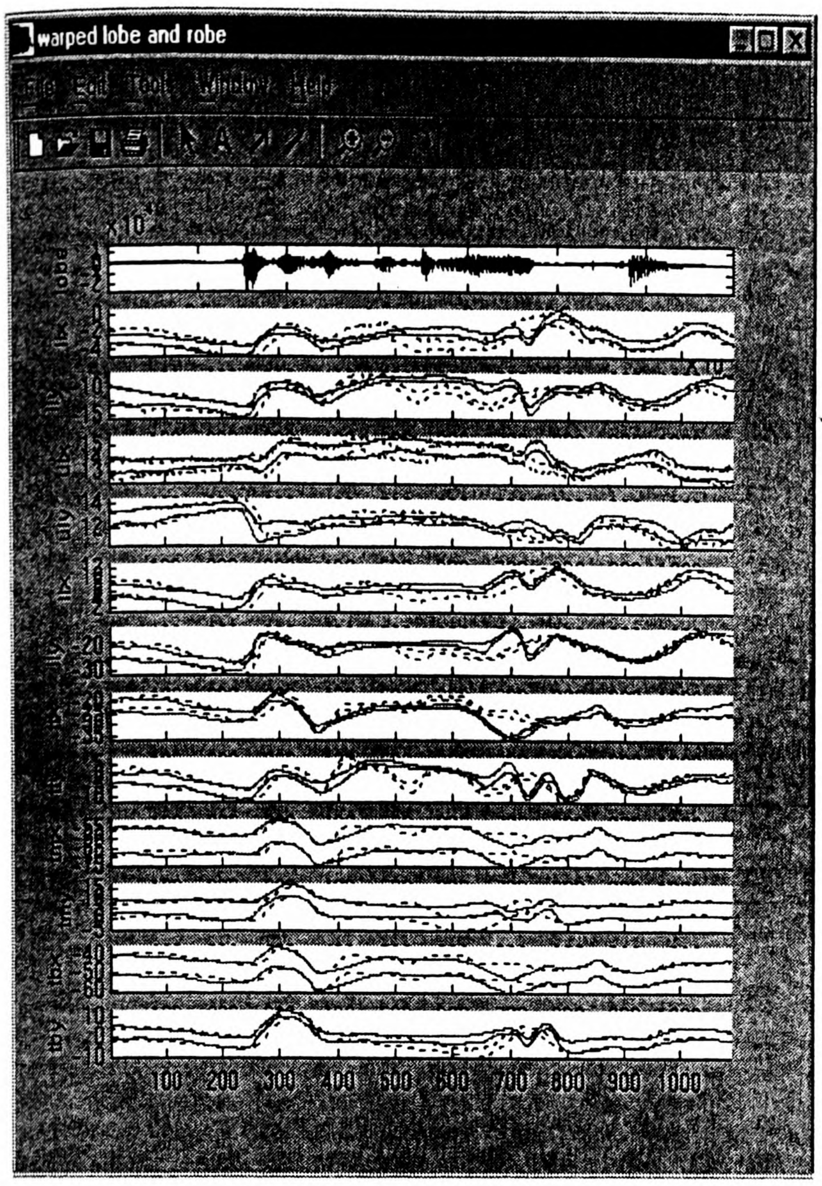
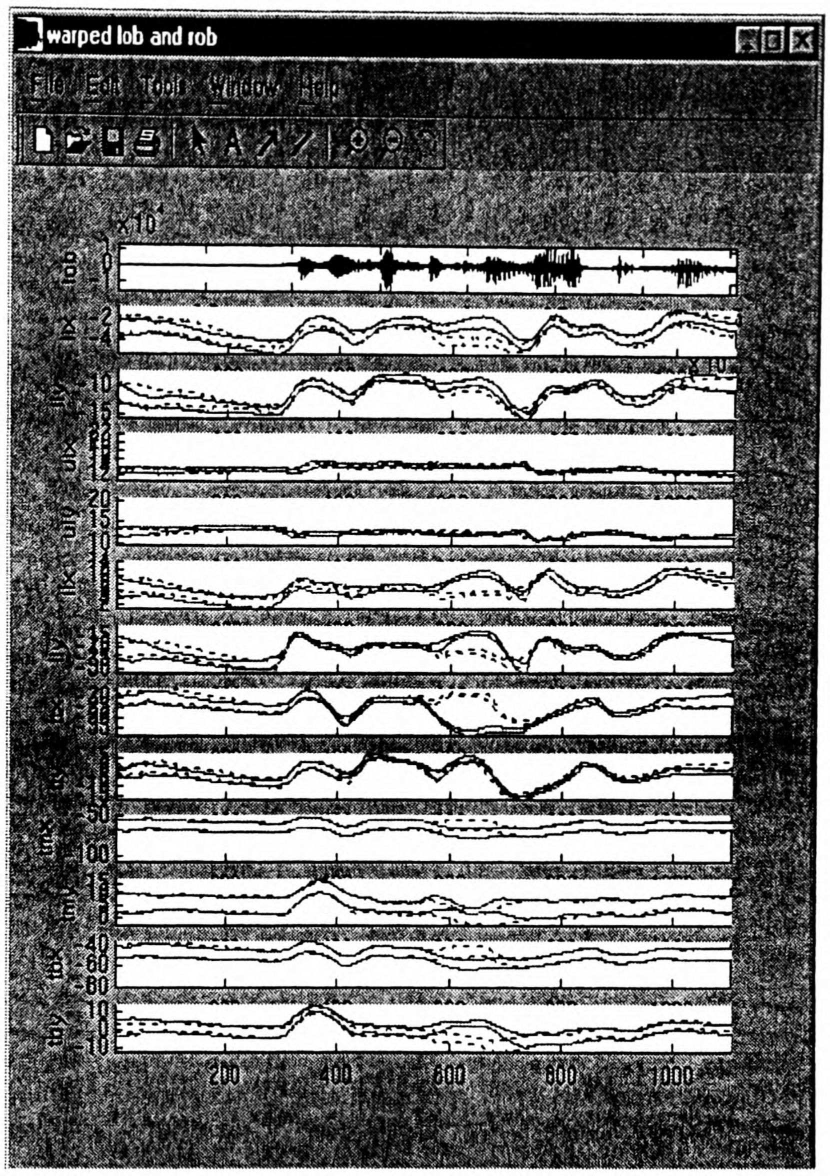

warped lob and rob

Figure 4.7: 95% confidence intervals for the warped EMA data, subject AS. Solid lines indicate /r/ articulatory traces, dashed lines /l/. The audio signal for the reference /l/ token is shown above.







In short, the time-series manipulation and statistical analysis of the data revealed rather little in the way of long-domain effects. However, local effects can be seen extending over the vowels adjacent to the liquid. Some long-distance effects are apparent from this analysis, with differences occasionally extending up to two syllables before and/or after the liquid.

4.3. Long-distance coarticulatory effects in vowels

The small extent of significant differences found in the previous section might be the result of too much variability, due to the lack of precision in determining measurement points. Sections 4.3.1 and 4.3.2 below describe the results of measurements made on the unaligned data, with measurement points determined by hand for each token separately. The articulatory data are described and correlations with the acoustic data examined.

4.3.1. Anticipatory effects

Long-distance differences were sought using the method described for local coarticulatory effects, as one of the reasons for the lack of effect found using the *t*-

tests may have been the use of aligned tokens. For anticipatory differences, measurements were made in the schwa of *uttered* for all tokens as described in previous sections. In this case segmentation was relatively straightforward as the vowel was surrounded by stops. Periodicity (voicing) was used as the marker of vowel presence and aspirated or voiceless sections were discarded. The EMA data are presented in Figure 4.8 (EB), Figure 4.9 (MB) and Figure 4.10 (AS) and the acoustic data in Figure 4.11 (EB), Figure 4.12 (MB) and Figure 4.13 (AS), data divided by liquid context.

Figure 4.8: Articulatory data, schwa in *uttered*: subject EB.

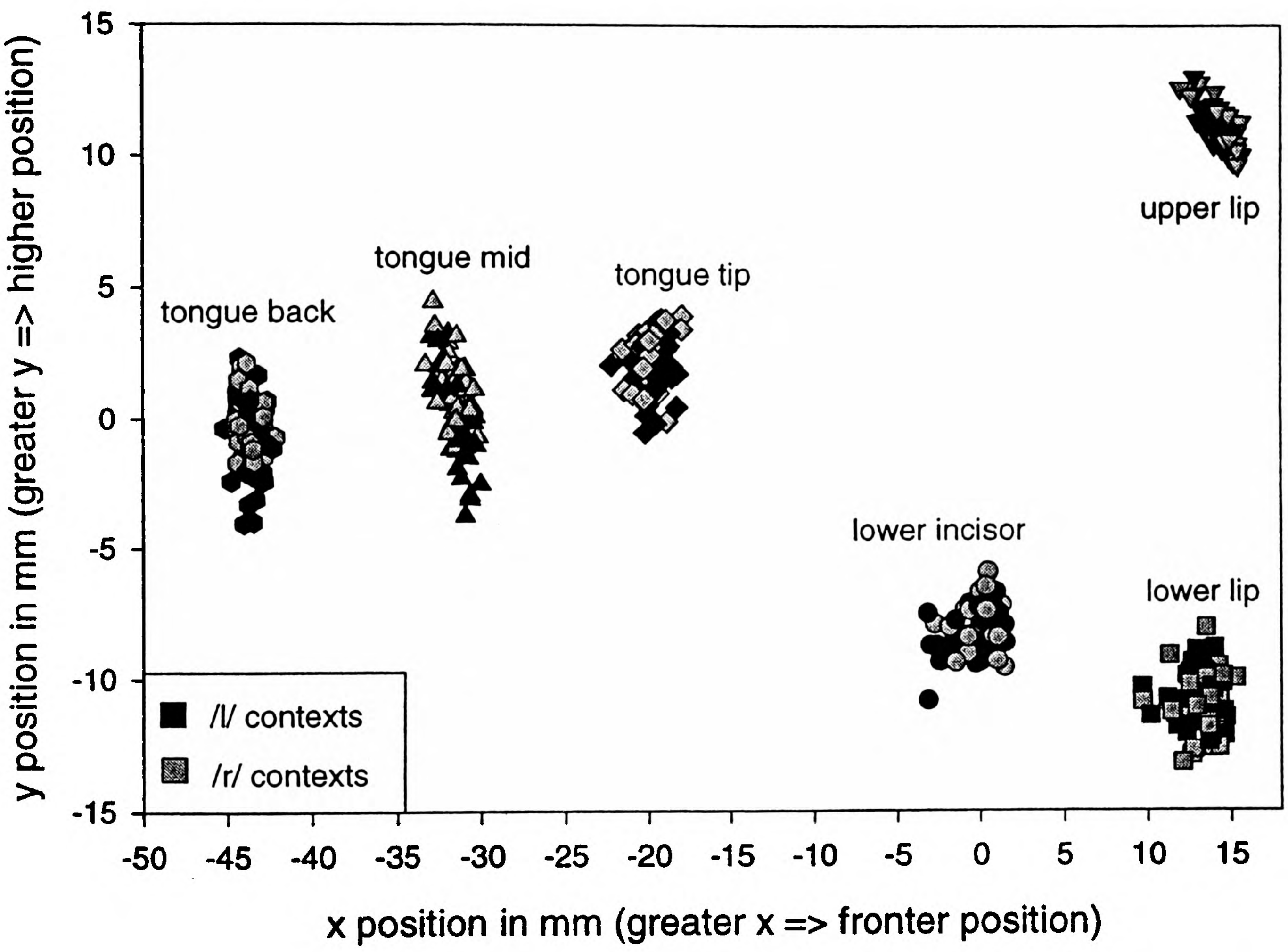


Figure 4.9: Articulatory data, schwa in *uttered*: subject MB.

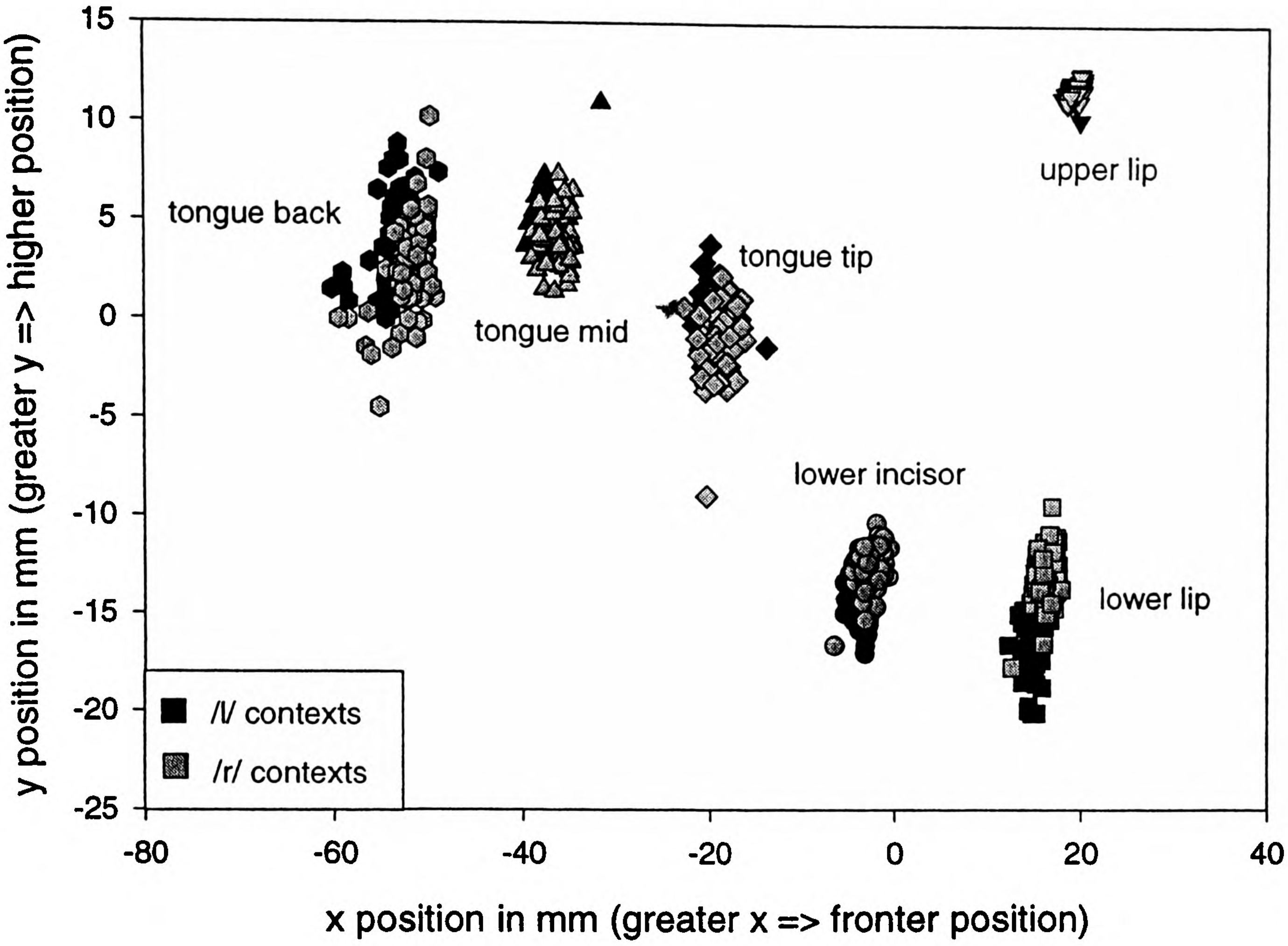


Figure 4.10: Articulatory data, schwa in *uttered*: subject AS.

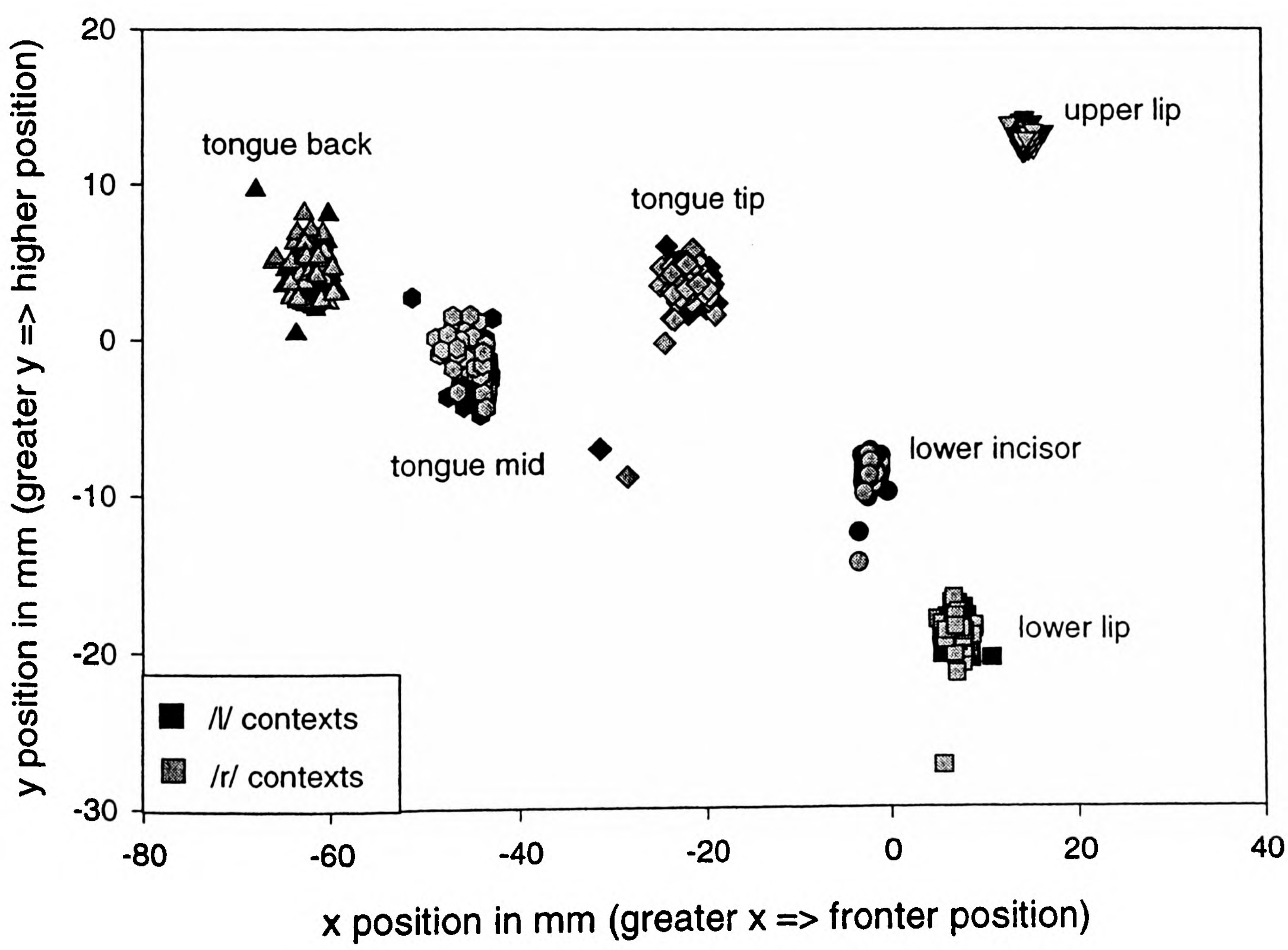


Figure 4.11: Acoustic data, schwa in *uttered*: subject EB.

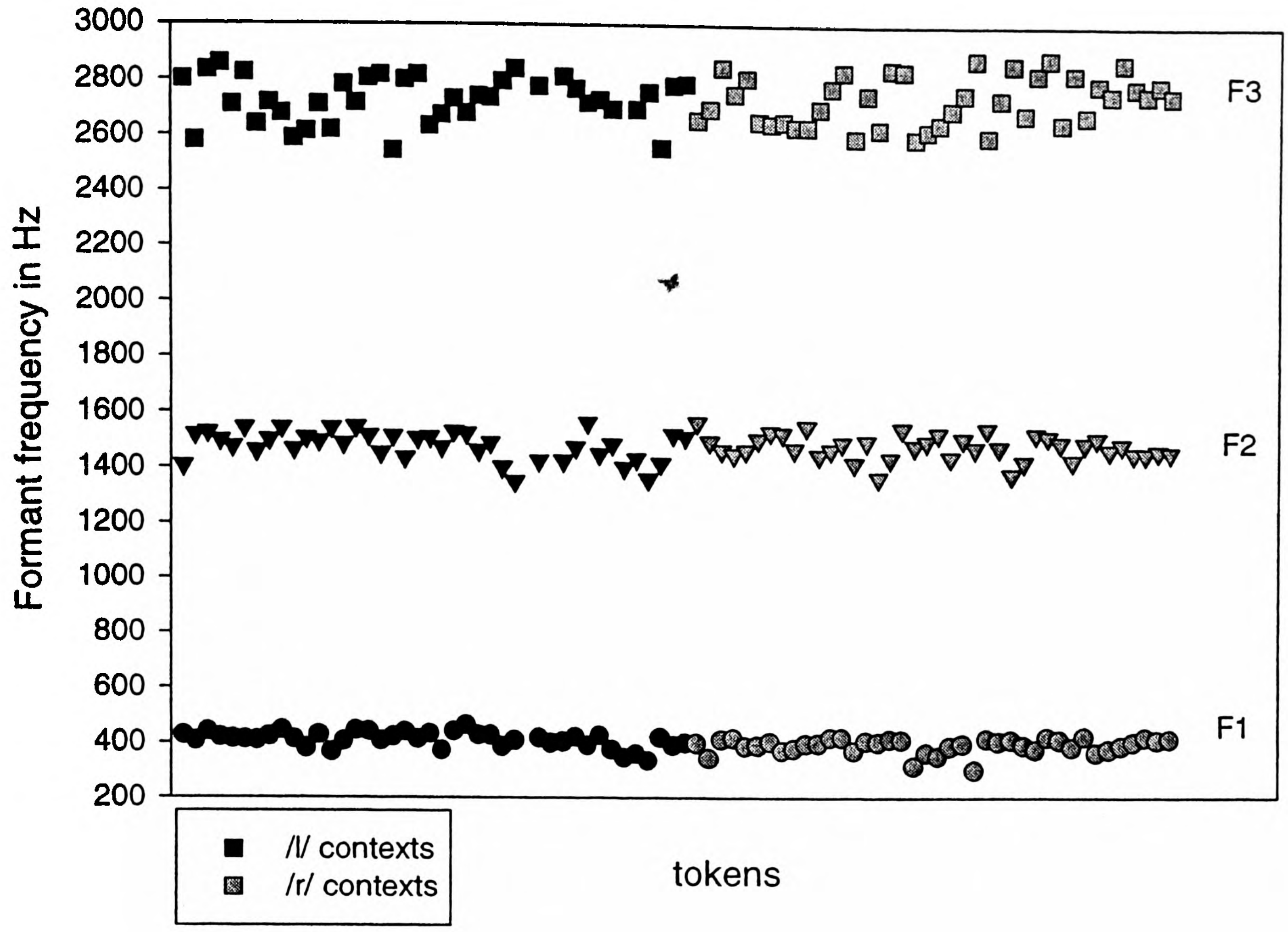


Figure 4.12: Acoustic data, schwa in *uttered*: subject MB.

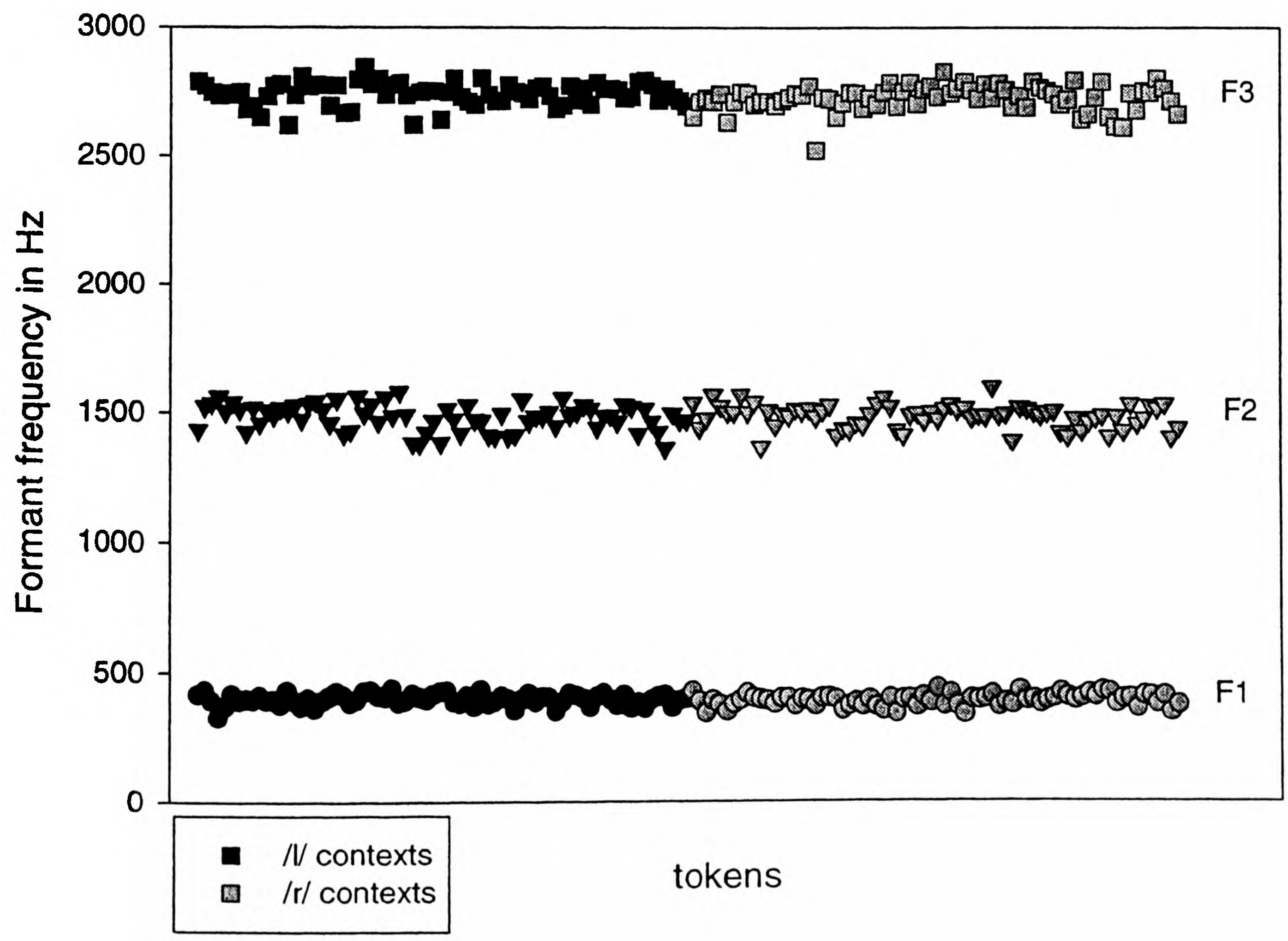
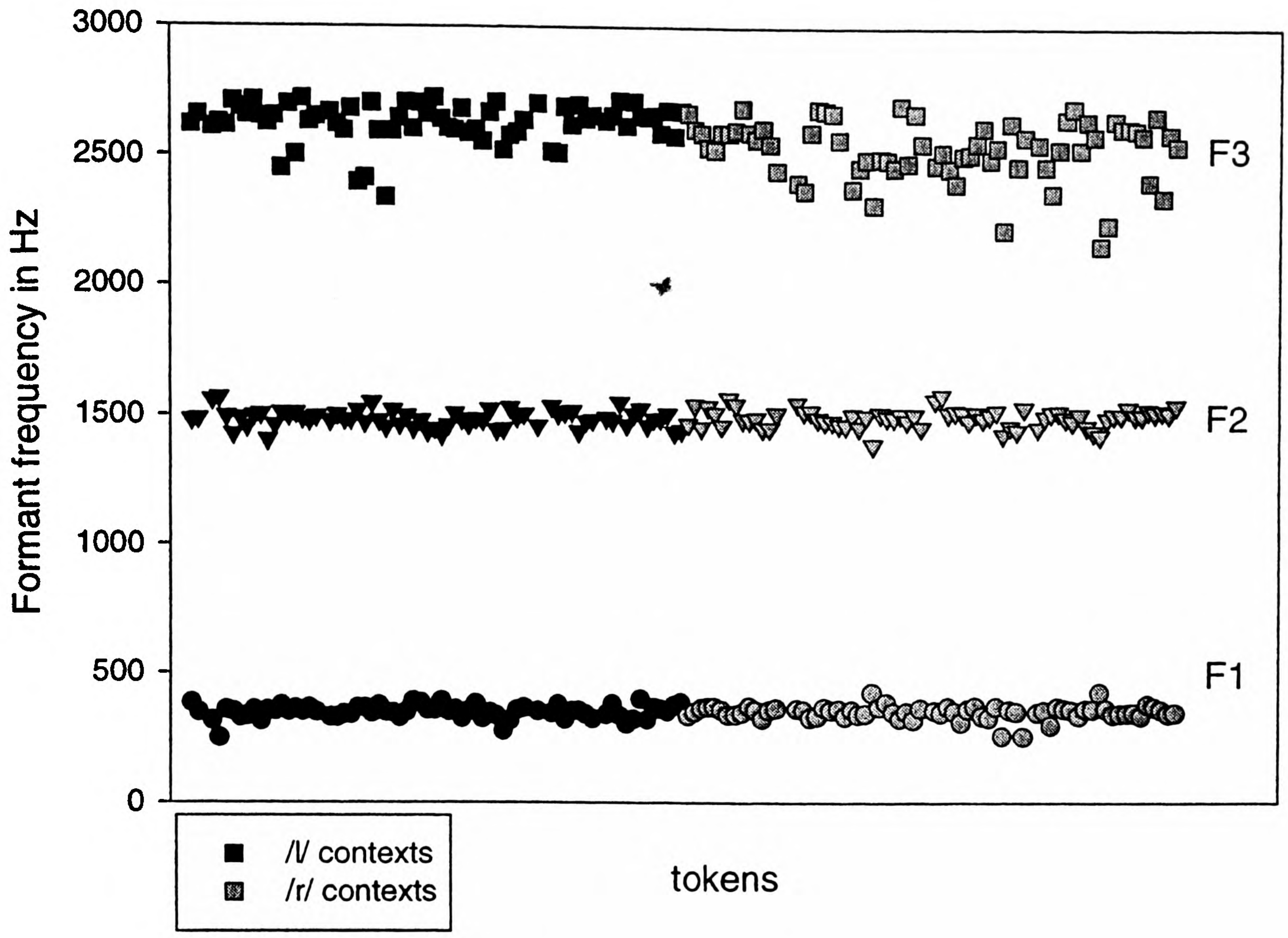


Figure 4.13: Acoustic data, schwą in *uttered*: subject AS.



For each subject, a multivariate general linear model was constructed for the articulatory and acoustic data, with word pair and liquid as factors.

For EB, Levene's test showed that lower incisor *x* and F_3 did not have equal variance of errors ($F(11,67) = 3.086$, $p < 0.002$ and $F(11,67) = 2.963$, $p < 0.003$). One-sample Kolmogorov-Smirnov tests showed no departures from the normal distribution. Both factors liquid and word pair were highly significant (Pillai's test: $F(15,53) = 5.638$, $p < 0.001$ and $F(75,285) = 1.742$, $p < 0.001$), and there was no interaction between them. The significance of the factor liquid is the result of differences in F_3 ($p < 0.004$), tongue mid *y* ($p<0.011$), upper lip *x* ($p< 0.001$) and upper lip *y* ($p < 0.046$).

For MB, tongue back *x* is not normally distributed (a one sample Kolmogorov-Smirnov test gives $Z = 2.539$, $p < 0.001$) and lower incisor *y* does not have equal variance ($F(11, 132) = 3.031$, $p < 0.001$). The factor liquid is significant ($F(15, 118) = 3.131$, $p < 0.001$), word pair and the interaction between liquid and word pair are not. Significance of the factor liquid is due to lower lip *y* ($p < 0.004$), upper lip *x* ($p < 0.044$), F_1 ($p < 0.045$) and F_3 ($p < 0.027$).

AS's data has tongue tip *y*, lower incisor *y* and F_3 not normally distributed ($Z = 1.625$, $p < 0.01$, $Z = 1.875$, $p < 0.002$ and $Z = 1.702$, $p < 0.006$). F_3 and tongue tip *y* also have significantly unequal variance of errors by Levene's test ($F(11, 122) = 2.579$, $p < 0.006$ and $F(11, 122) = 4.069$, $p < 0.001$ respectively). The multivariate GLM shows liquid alone as a significant factor ($F(15, 108) = 3.500$, $p < 0.001$). Significance of the factor liquid is due to tongue back y ($p < 0.022$), tongue middle x and y ($p < 0.014$ and 0.018), tongue tip x ($p < 0.030$) and F_3 ($p < 0.001$).

Additional statistical tests were conducted to check these results. Non-parametric 2 independent sample Kolmogorov-Smirnov tests, with liquid as a factor, were conducted for the variables with unequal variance or non-normal distribution (F_3 for EB and AS). These confirmed the GLM results, although the test is less sensitive.

The differences in means for significant variables are given below in Table 4.1 for EB, Table 4.2 for MB and Table 4.3 for AS. There are some clear similarities between the three subjects: each subject has a lower F_3 in the /r/ context (a negative difference in means). This replicates the results found in the acoustic analyses (chapter 3, section 3.4). MB also has a lower F_1 value in the /r/ context, although this is a very small difference in means and may well be a spurious result. EB and MB have significant differences in lip placement: both have a fronter or more protruded upper lip (greater x value) in the /r/ context, EB has a lower upper lip (lower y value) and MB a higher lower lip (greater y value) in the /r/ context. This is all suggestive of anticipatory lip rounding in the /r/ context. Differences in tongue placement are found for EB and AS. For EB and AS, tongue mid is higher (greater y value) in /r/ than in /l/ contexts. AS also shows a higher tongue back (greater y value) and retracted tongue middle and tip (lower x values) in the /r/ context. These differences anticipate the articulatory configuration that these two subjects show for /r/: a retracted and raised tongue body. Note that the articulatory differences are all less than 1mm in magnitude although the acoustic differences range up to 100 Hz for AS.

Table 4.1: Means, standard deviations and differences between the means, EMA data (in mm) and acoustic data (in Hz), schwa in *uttered*: subject EB.

	LIQUID	N	Mean	Std. Deviation	Difference between means
TMY	r	41	1.0412	1.3583	.92
	l	41	.1251	1.7184	
ULX	r	41	14.7483	.7580	.53
	l	41	14.2198	.6233	
ULY	r	41	10.9212	.8130	-.31
	l	41	11.2298	.5858	
F3	r	40	2706.8250	73.3275	-51.23
	l	39	2758.0513	91.7330	

Table 4.2: Means, standard deviations and differences between the means, EMA data (in mm) and acoustic data (in Hz), schwa in *uttered*: subject MB.

	LIQUID	N	Mean	Std. Deviation	Difference between means
LLY	r	72	-13.4108	.9581	.52
	l	72	-13.9309	1.1366	
ULX	r	72	19.0644	.4090	.15
	l	72	18.9115	.4758	
F1	r	72	388.2500	22.3309	-7.31
	l	72	395.5556	22.2254	
F3	r	72	2721.4722	50.6554	-17.65
	l	72	2739.1250	44.3577	

Table 4.3: Means, standard deviations and differences between the means, EMA data (in mm) and acoustic data (in Hz), schwa in *uttered*: subject AS.

	LIQUID	N	Mean	Std. Deviation	Difference between means
TBY	r	68	4.8976	1.2990	.58
	l	67	4.3166	1.4430	
TMX	r	68	-45.5735	1.3435	-.57
	l	67	-45.0042	1.3240	
TMY	r	68	-1.0862	1.2932	.58
	l	67	-1.6659	1.4470	
TTX	r	68	-22.1668	1.5479	-.63
	l	67	-21.5417	1.7260	
F3	r	67	2530.3731	116.8739	-96.01
	l	67	2626.3881	77.8692	

The results of the GLMs and nonparametric tests demonstrate that there are small, significant differences in articulation between liquid context for all subjects. These differences differ between subjects, but show the same overall pattern of anticipatory coarticulation involving either lip rounding or tongue retraction and raising (or both) in the /r/ context relative to the /l/ context.

Spearman correlations were calculated for the acoustic and articulatory data for each subject. EB's data showed only 1 significant acoustic-articulatory correlation ($\alpha = 0.01$), whereas MB and AS had 5 and 19 respectively. EB has a positive correlation between tongue mid *y* and F_2. MB has negative correlations between F_1 and tongue tip, mid and back *x*, lower lip *y* and upper lip *x*. AS has negative correlations between F_1 and tongue tip *x*, tongue tip *y*, tongue middle *x* and lower lip *y*; positive correlations between F_2 and lower incisor *x*, lower lip *x*, upper lip *x*, tongue back *y* and tongue mid *y*. For AS, F_3 is positively correlated with lower lip and tongue tip *y* and tongue tip mid and back *x* and negatively correlated with lower incisor, lower lip and upper lip *x*, tongue mid and back *y*.

These correlations are confounded with word pair and liquid, and thus partial correlations controlling for liquid and word pair were conducted for all the articulatory and acoustic data. For EB the correlation between F_1 and tongue tip *y* was significant ($r = -0.3613$, $p < 0.001$), and for MB the correlation between F_1 and tongue back *x* ($r = -0.3805$, $p < 0.001$). AS has more significant correlations for this vowel: F_1 and lower incisor *x* ($r = 0.2795$, $p < 0.001$), F_1 and lower lip *x* ($r = 0.3115$, $p < 0.001$), F_3 and lower lip *y* ($r = 0.3540$, $p < 0.001$) and F_3 and tongue mid *x* ($r = 0.4711$, $p < 0.001$).

4.3.2. Perseverative effects

Perseverative differences were sought in the vowel of *at* from the frame sentence 'Have you uttered a — at home?'. The earlier acoustic analyses (chapter 3, section 3.4) showed that there were reliable acoustic differences at this point for MB and AS, but not for EB. Nevertheless, both pre-tests of EB's speech (with and without the EPG palate) showed that he sometimes made an acoustic distinction at this point. This subtle distinction was lost under experimental conditions, probably due to the invasive nature of the technique, subject discomfort and the small number of repeats captured. It was decided to check whether an articulatory difference could be found for all three subjects. Measurements were taken as for the previous section: the articulatory data is displayed in Figure 4.14, Figure 4.15 and Figure 4.16. Figure 4.17, Figure 4.18 and Figure 4.19 contain the acoustic data for EB, MB and AS respectively.

Figure 4.14: Articulatory data, vowel in *at*: subject EB.

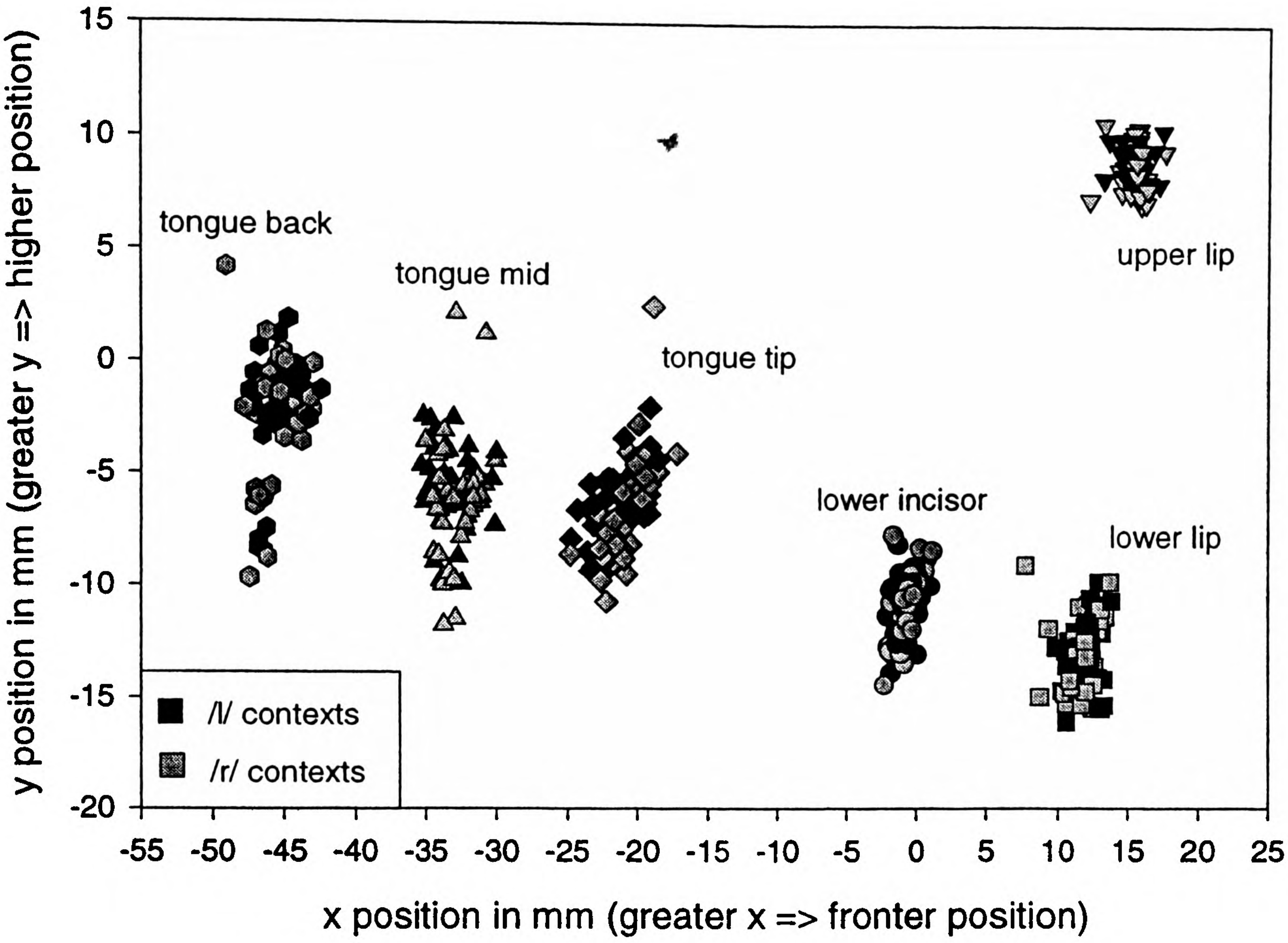


Figure 4.15: Articulatory data, vowel in *at*: subject MB.

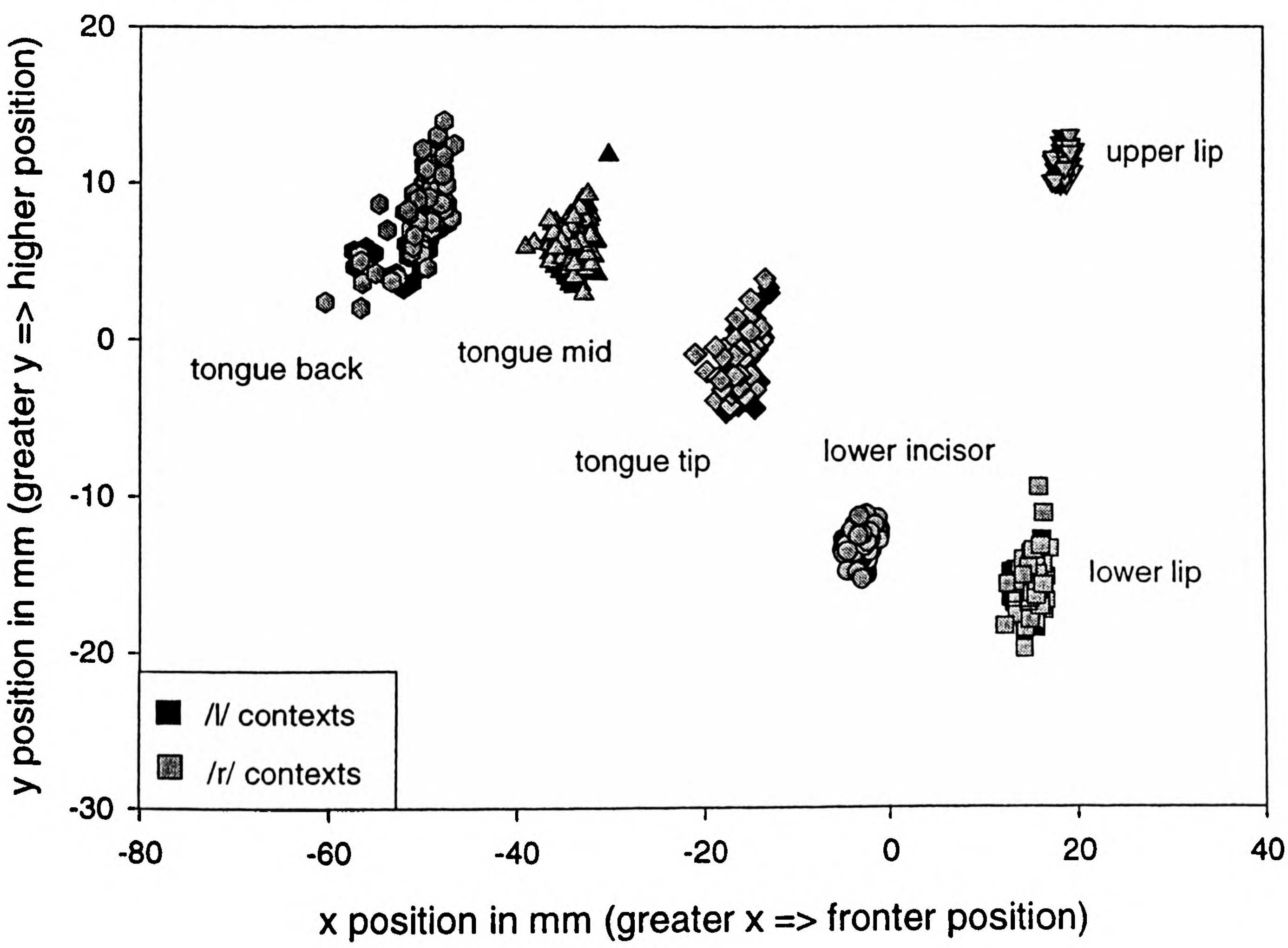


Figure 4.16: Articulatory data, vowel in *at*: subject AS.

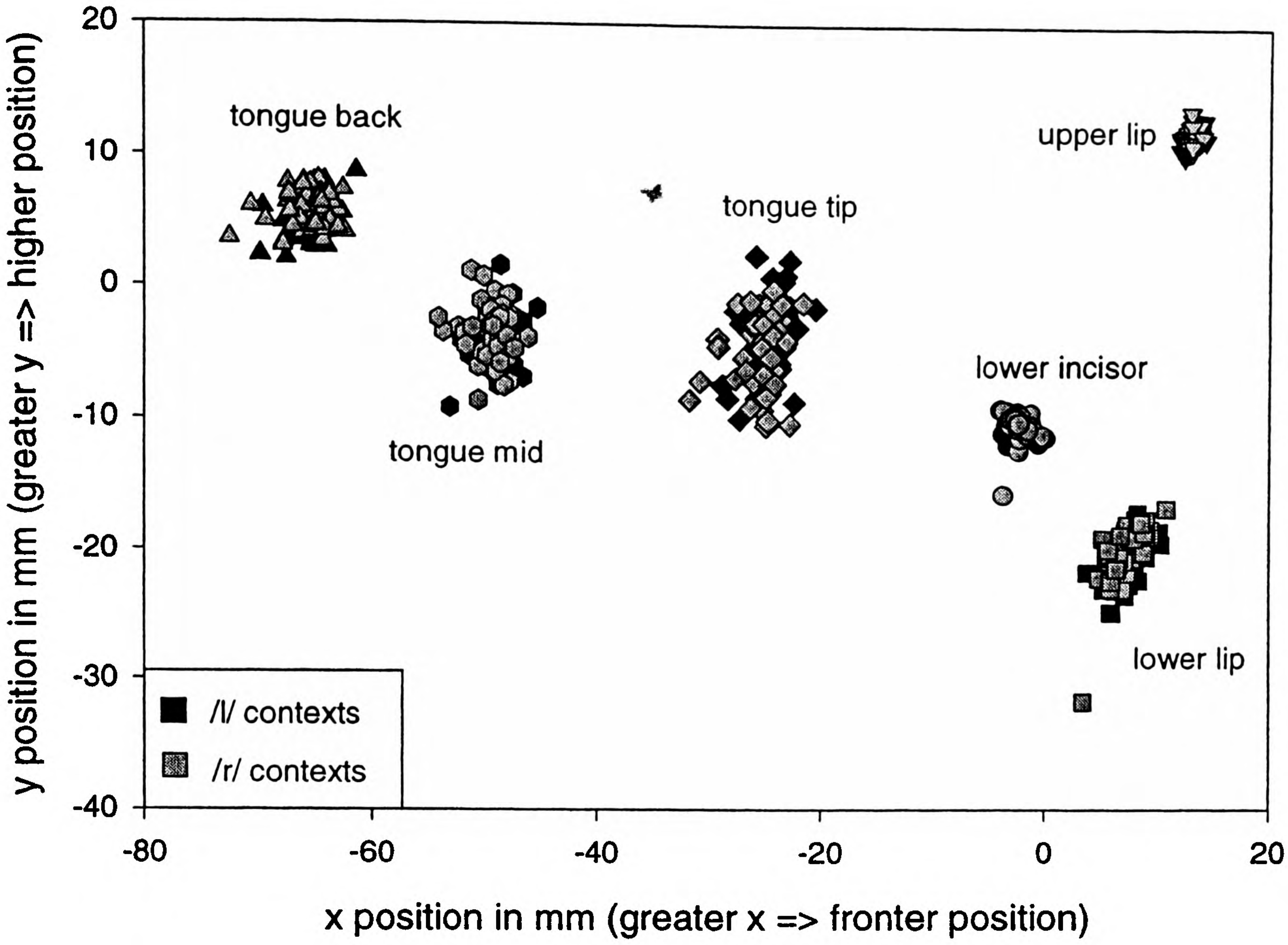


Figure 4.17: Acoustic data, vowel in *at*: subject EB.

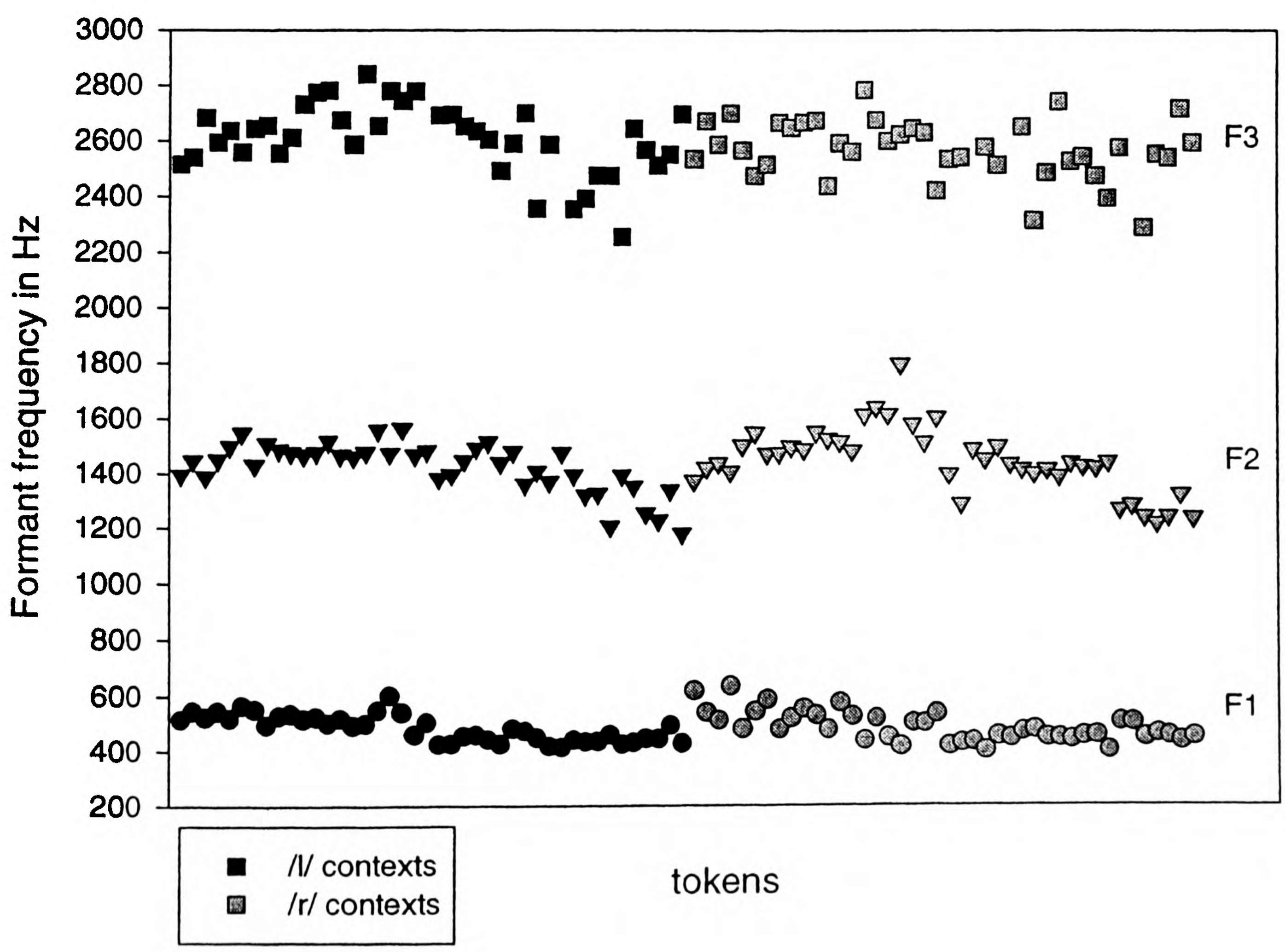


Figure 4.18: Acoustic data, vowel in *at*: subject MB.

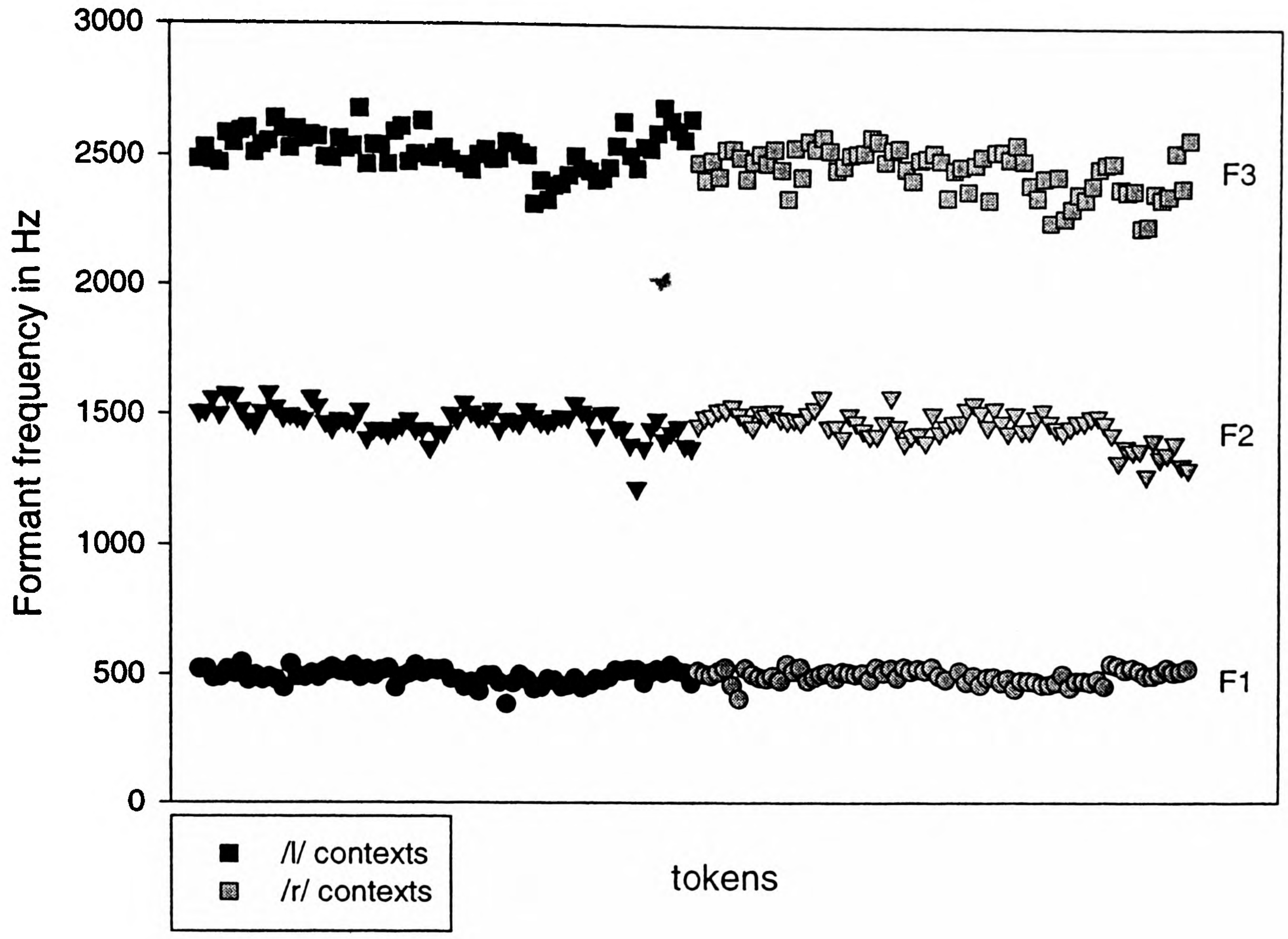
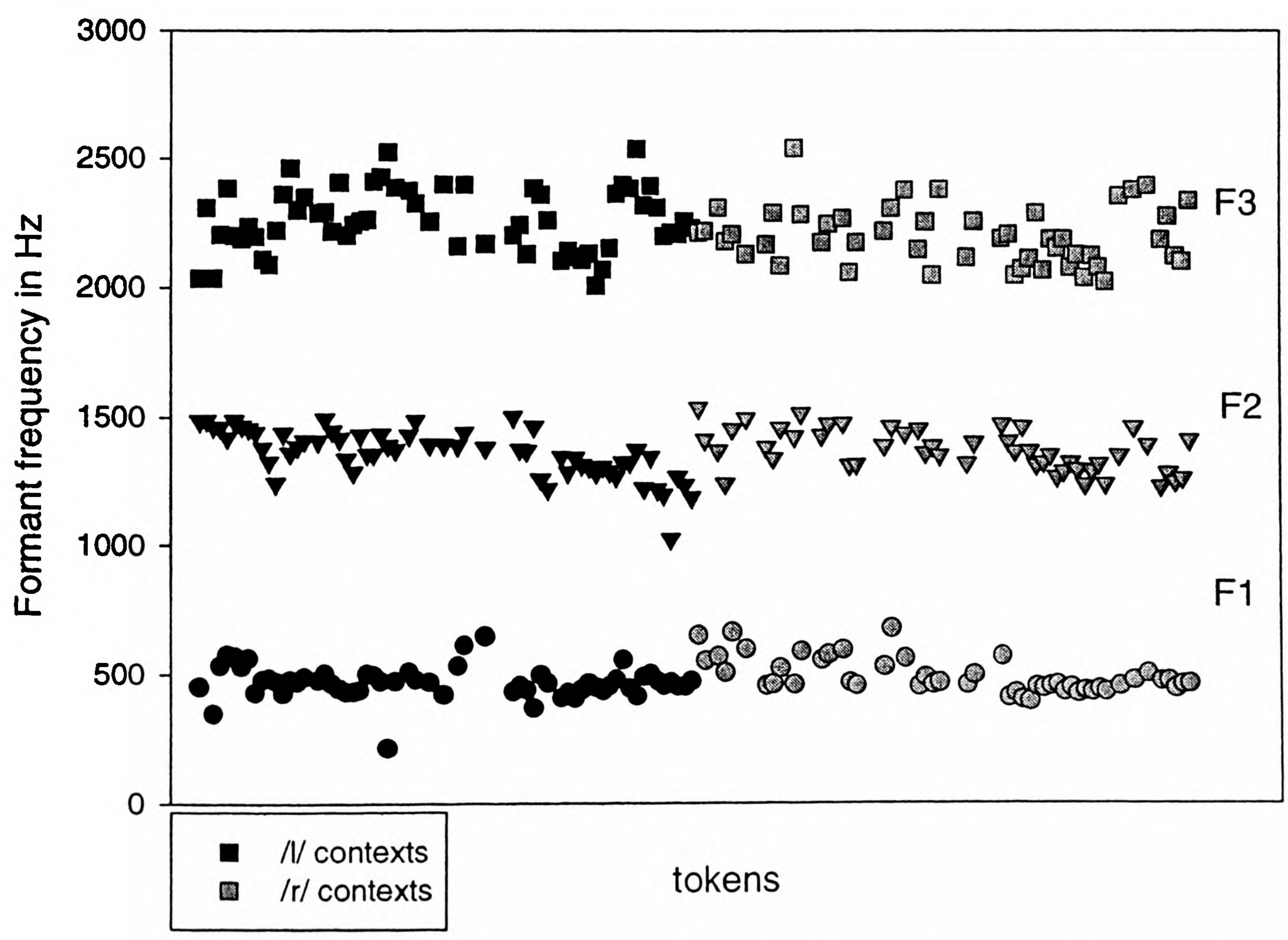


Figure 4.19: Acoustic data, vowel in *at*: subject AS.



A multivariate general linear model was constructed for each data set in *SPSS* as for previous sets.

For EB, Levene's test for equality of error variances produced six significant results ($\alpha = 0.01$): F_1, upper lip *x*, lower lip *x*, tongue tip *y*, tongue middle *y*, tongue back *x* and F_1. One-sample Kolmogorov-Smirnov tests for the normal distribution produced no significant results. The factors liquid and word pair are highly significant (Pillai's test: $F(15,53) = 5.330$, $p < 0.001$ and $F(75,285) = 5.131$, $p < 0.001$) and there is an interaction between word pair and liquid ($F(75,285) = 1.686$, $p < 0.001$). The interaction is significant for the variables lower incisor *y* ($p < 0.004$), tongue mid *x* ($p < 0.028$) and F_2 ($p < 0.005$). The factor liquid is significant for three articulatory variables: tongue tip and mid *x* ($p < 0.001$ for both) and tongue back *y* ($p < 0.008$), and the factor word pair for all variables except lower incisor *x*. To explore the interaction between liquid and word pair, independent samples *t*-tests were conducted for the variable which showed an interaction: tongue mid *x*, for each word pair separately. Tongue mid *x* was significant for three word pairs: *leap/reap*, *lip/rip* and *lobe/robe*, and the difference in means was in the same direction for all significant pairs: fronter (greater *x*) in the /r/ context. Because of the significant results of Levene's test, an independent samples *t*-test was conducted with liquid as factor (the data is normally distributed). The homogeneity of variance test showed no significant results, and only one variable was significant: tongue tip *x* ($t(82) = 2.559$, $p < 0.012$). The lack of other significant variables probably results from ignoring word pair as a factor.

For MB, tongue back *x* is significantly non-normally distributed ($Z = 1.931$, $p < 0.001$) and Levene's test gives a significant result for this variable ($F(11,130) = 2.802$, $p < 0.003$). The GLM has liquid and word pair as significant factors (Pillai's test: ($F(15,116) = 6.165$, $p < 0.001$ and $F(75,600) = 6.709$, $p < 0.001$), with a significant interaction between them ($F(75,600) = 1.941$, $p < 0.001$). The interaction term is due to tongue mid *x* and *y* ($p < 0.015$ and 0.001) and F_3 ($p < 0.001$). Word pair is significant for all variables except lower incisor, lower lip and upper lip *x*. Liquid is significant for lower incisor *x* ($p < 0.034$) and tongue back *y* ($p < 0.015$), and highly significant for tongue mid *x* ($p < 0.006$), tongue mid *y* ($p < 0.001$) and F_3 ($p < 0.001$). To explore the interaction between liquid and word pair, an independent samples *t*-test was conducted for tongue mid *x* and *y* and F_3 for each word pair separately with liquid as factor. Tongue mid *x* was significant for *leap/reap* and *lob/rob*, tongue mid *y* for *lip/rip*, *lobe/robe* and *lob/rob* and F_3 for *leap/reap*, *lip/rip* and *lob/rob*. As was found for EB, the differences in means for the significant word pairs are all in the same direction (lower F_3, higher and backer tongue mid in the /r/ context).

The data from AS has one significantly non-normally distributed variable, F_1, for which $Z = 1.643$, $p < 0.009$. Levene's test produces significant results for lower lip *y* ($F(11, 97) = 2.842$, $p < 0.003$), upper lip *x* and *y* ($F(11, 97) = 2.891$, $p < 0.002$ and $F(11, 97) = 2.538$, $p < 0.007$) and F_1 ($F(11, 97) = 3.726$, $p < 0.001$). The GLM has both liquid and word pair significant (Pillai's test: ($F(15,83) = 3.995$, $p < 0.001$ and $F(75,435) = 3.710$, $p < 0.001$), but the interaction between them is not. Word pair is significant for all variables except tongue mid and back *x* and lower incisor *y*. Differences due to liquid are found in tongue back *y* ($p < 0.040$), tongue mid *y* ($p < 0.028$), tongue tip *x* ($p < 0.012$), and F_3 ($p < 0.001$). As for EB, an independent samples *t*-test with liquid as factor was calculated. This gave only tongue tip *x* ($t(108) = -2.324$, $p < 0.022$) and F_3 ($t(108) = -2.665$, $p < 0.009$) as significant.

Means for all significant variables are given in Table 4.4, Table 4.5 and Table 4.6. For EB the tongue back is higher (by 0.66 mm) for /r/ contexts than /l/ contexts and the tip and mid are significantly fronter (by 0.84 and 0.56 mm respectively). MB has lower F_3 in /r/ contexts (by about 70 Hz), higher tongue back and mid (about 0.80 mm), backer tongue mid and fronter lower incisor (by about 0.30 mm). AS shows higher tongue back and mid (0.27 and 0.50 mm), backer tongue tip (0.76 mm) and lower F_3 (about 60 Hz) in the /r/ context.

Table 4.4: Means, standard deviations and differences between the means, EMA data (in mm), schwa in *uttered*: subject EB.

	LIQUID	N	Mean	Std. Deviation	Difference between means
TBY	r	42	-1.9048	2.2619	.66
	l	42	-2.5598	2.8018	
TMX	r	42	-32.7598	1.3511	.56
	l	42	-33.3193	1.3203	
TTX	r	42	-20.8376	1.4568	.84
	l	42	-21.6733	1.5349	

Table 4.5: Means, standard deviations and differences between the means, EMA data (in mm) and acoustic data (in Hz), schwa in *uttered*: subject MB.

	LIQUID	N	Mean	Std. Deviation	Difference between means
LIX	r	72	-2.8352	.8635	.30
	l	72	-3.1327	.8122	
TBY	r	72	8.3855	2.5434	.82
	l	72	7.5640	2.0274	
TMX	r	72	-34.0522	1.4691	-.27
	l	72	-33.7777	1.1449	
TMY	r	72	6.3807	1.2438	.80
	l	72	5.5783	1.4259	
F3	r	72	2453.4167	85.4275	-68.60
	l	71	2522.0141	75.3380	

Table 4.6: Means, standard deviations and differences between the means, EMA data (in mm) and acoustic data (in Hz), schwa in *uttered*: subject AS.

	LIQUID	N	Mean	Std. Deviation	Difference in means
TBY	r	49	5.6012	1.3043	.27
	l	61	5.3311	1.4768	
TMY	r	49	-3.4607	1.9694	.50
	l	61	-3.9580	1.7821	
TTX	r	49	-25.3463	1.8910	-.76
	l	61	-24.5899	1.5239	
F3	r	49	2199.1837	112.9177	-60.54
	l	61	2259.7213	122.6438	

Results common to all three subjects are a lower F_3 in the /r/ context (MB and AS) and raised tongue position for all three subjects. AS and MB have retracted tongue position (tip and mid respectively) in the /r/ context, whereas EB has a fronter tongue (tip) position. The result for EB is not entirely in accordance with our expectations: we might have expected a small amount of perseverative tongue tip retraction after an /r/, as he produces strongly retracted /r/s.

Note that for two of the three vowels measured directly following the liquid, significantly backer tongue tip positions were found in the /r/ context for EB. The differences in tongue tip position between /l/ and /r/ contexts were large: between 2 and 3.5 mm. A possible explanation for the long-domain perseverative tongue tip result for EB (if it is not just a spurious difference) is that movement away from the back position for the /r/ may be faster and thus achieve a fronter target than that away from an /l/ for this particular subject. This difference is not yet manifested in the vowel directly following the liquid and may be a result of mechanical forces, without having any acoustic result correlated with liquid.

To explore further the relationship between the articulatory and acoustic data, Spearman correlations were calculated. There were several significant correlations with $\alpha = 0.01$. For EB, tongue tip and mid *x* are positively correlated with all three formants, as is lower lip *y*; lower incisor and upper lip *y* are positively correlated with F_1 and F_2, tongue tip *y* and tongue back *x* with F_3. For MB and AS there are many Spearman correlations between articulatory and acoustic variables. Correlations of interest are those between F_3 and the articulatory variables, as F_3 is significantly different in the two liquid contexts. For MB, F_3 is negatively correlated with lower lip *x* and *y*, upper lip *x* and tongue mid *y*, and it is positively correlated with tongue tip *x* and upper lip *y*. Thus F_3 decreases with lip rounding (protrusion of the lips, raising of the lower lip and lowering of the upper lip), raising of the tongue mid and backing of the tongue tip. For AS, F_3 is negatively correlated with lower incisor *x*, lower lip *x* and *y* and positively correlated with tongue tip *x* and *y* and upper lip *x*. Thus F_3 lowers with lip rounding and tongue tip retraction, as for MB. These articulatory positions are all characteristic of the /r/ context in which lower F_3 occurs.

Partial correlations controlling for liquid and word pair were also calculated. For EB F_2 is positively correlated with tongue tip *y* and tongue mid and back *x*. Thus F_2 increases as the tongue body is fronted and the tip raised. The correlations of F_2 with the lip and incisor data have been accounted for by variability owing to vowel and liquid context, as have all correlations with F_1 and all except tongue tip *x* with F_3. This last correlation shows that F_3 increases with fronter tongue tip position. For MB the partial correlations show that F_1 is negatively correlated with lower lip *y* and tongue mid *x*, F_2 is positively correlated with tongue mid *x*, F_3 negatively with upper lip *x*. Thus F_1 lowers with raising of the lower lip, F_2 with tongue lowering and F_3 with upper lip protrusion. For AS, F_1 is negatively correlated with lower lip *y* and tongue back *x*, F_2 is positively correlated with tongue tip and mid *y* and upper lip *x*. F_3 is positively correlated with tongue tip and upper lip *y* and negatively with lower incisor *x*, lower lip *x* and *y*. As for MB, F_1 lowers with raising of the lower lip, but also with fronting of the tongue back. F_2 lowers with tongue tip and mid lowering and upper lip retraction. F_3 lowers with tongue tip and upper lip lowering, lower lip raising, lower incisor and lower lip protrusion.

4.4. Long-distance coarticulatory effects in consonants

The EPG data from all subjects was examined for long-domain coarticulatory effects in consonants as it presents a complementary picture to the EMA data. Tongue palate contact could be examined effectively in the consonants of syllables remote from the liquid, but not in vowels as these show very little tongue-palate contact. One contact on EB's artificial palate was broken: the subject's leftmost contact five rows from the front. As only alveolar stops were examined using the EPG data, an analysis was nevertheless conducted on the remaining data for this subject. For each subject, the corresponding EMA data for the consonants was examined for long-domain differences and compared with the EPG data.

4.4.1. Anticipatory effects

Anticipatory coarticulation was examined at two points: the points of maximal tongue-palate contact in the /t/ and /d/ of *uttered.* These were determined by finding maxima in the number of EPG contacts over the relevant stretches of speech signal. The waveform was primarily used to identify the closure stretch, aided by the tongue tip *y* trace. In most cases the /t/ and /d/ were easy to identify. Occasionally EPG data were not aligned with the EMA and acoustic data as the EPG equipment failed to begin recording on time. When it was clear to which articulatory events the relevant EPG contact peaks corresponded, the data were used in this analysis, although EMA data corresponding to these points could not be captured, as the signals could not be reliably synchronised.

The EPG data were analysed in *SPSS*, using non-parametric tests over various summary measures to determine where differences in contact patterns occurred. Non-parametric tests were used as unequal variances and non-normal distribution made the data unsuitable for the use of parametric tests. Rows are labelled 1 to 8, in order of increasing backness, with 1 the front row and 8 the back row. The summary measures included the number of contacts in each row (rows 1 to 8), the total number of contacts (total) and total number of contacts in rows 1 to 3 (front), and rows 4 to 8 (back). Measures of laterality and centrality were also made: lateral contact was measured as the total number of contacts made in the two contact positions at each edge of each row (four per row), central contact as total contact minus lateral contact.

Each consonant was analysed separately and in some cases rows were further divided in the search for relevant differences in contact patterns.

For subject EB, several differences in contact patterns for the /t/ of *uttered* in different liquid contexts were found. The results were calculated in *SPSS* using the Kolmogorov-Smirnov nonparametric 2-independent samples test. Row 2, row 3, total number of contacts and number of front contacts are significantly different across liquid contexts ($Z = 1.418, p < 0.036, Z = 2.182, p < 0.001, Z = 1.964, p < 0.001, Z = 1.855, p < 0.002$, respectively). As there were 7 repeats of each of the 6 words in each category, there were 42 observations in the summarised data. (MB and AS have 12 repeats and thus 72 observations.) In all cases there was more contact in the /l/ than in the /r/ context. The relevant means are given in Table 4.7, from which it can be seen that the difference in means is small, no more than 2 contacts and in row 3 less than one. Nevertheless the results are significant.

Table 4.7: Mean number of contacts and standard deviations, EPG data for /t/ of *uttered*: subject EB.

	LIQUID	N	Mean	Std. Deviation	Difference in means
ROWTWO	r	42	5.2857	2.1446	-1.55
	l	42	6.8333	1.5604	
ROWTHREE	r	42	2.2619	.7982	-.60
	l	42	2.8571	.7513	
TOTAL	r	42	23.3333	3.2734	-2.02
	l	42	25.3571	2.3040	
FRONT	r	42	13.0714	3.0153	-2.38
	l	42	15.4524	2.1322	

Analysis of the /d/ of *uttered* shows significant differences between the /l/ and /r/ context, but in a more limited area. Number of central contacts and contacts in row two proved significant ($Z = 1.673$, $p < 0.009$ for both); this indicates that the difference lies in the central contacts in row two. The means of these measures (Table 4.8) show that there is more contact in the /r/ context.

Table 4.8: Mean number of contacts and standard deviations, EPG data for /d/ of *uttered*: subject EB.

	LIQUID	N	Mean	Std. Deviation	Difference in means
ROWTWO	r	42	6.5714	1.8365	1.29
	l	42	5.2857	1.7708	
CENTRAL	r	42	4.4048	2.0489	1.17
	l	42	3.2381	1.9977	

Subject MB shows fewer significant results: for the /t/ of *uttered* there are significant difference in lateral and possibly back contact (Z = 1,708, $p < 0.006$, Z = 1.400, $p < 0.040$, respectively). Table 4.9 gives the means which show slightly more contact for /t/ in the /r/ context. The /d/ shows no significant differences at all.

Table 4.9: Mean number of contacts and standard deviations, EPG data for /t/ of *uttered*: subject MB.

	LIQUID	N	Mean	Std. Deviation	Difference in means
BACK	r	71	11.6056	1.1017	.36
	l	72	11.2500	1.5082	
LATERAL	r	71	10.2958	1.6159	.64
	l	72	9.6528	1.4551	

AS shows significant differences in the /t/ of *uttered* as did MB: row 3 and front are significantly different (Z = 1.417, $p < 0.036$ for both), with one more contact on average in the /l/ context (see Table 4.10 for the differences in means). The /d/ of *uttered* shows the same pattern for this subject: row 2 and front are significantly different (Z = 1.667, $p < 0.008$, Z = 1.417, $p < 0.036$, respectively) with one or two more contacts in the /l/ context (means are given in Table 4.11).

Table 4.10: Mean number of contacts and standard deviations, EPG data for /t/ of *uttered*: subject AS.

	LIQUID	N	Mean	Std. Deviation	Difference in means
ROWTHREE	r	72	1.43	1.97	-1.04
	l	72	2.47	2.35	
FRONT	r	72	1.8472	2.6466	-1.76
	l	72	3.6111	4.0195	

Table 4.11: Mean number of contacts and standard deviations, EPG data for /d/ of *uttered*: subject AS.

	LIQUID	N	Mean	Std. Deviation	Difference in means
ROWTWO	r	72	1.85	2.47	-1.44
	l	72	3.29	2.69	
FRONT	r	72	6.7639	5.0084	-2.76
	l	72	9.5278	4.5470	

The EPG results are interpreted as follows: subjects EB and AS show the same pattern for the /t/ in the /l/ context, more contact between the front of the tongue and the front part of the palate (in the central contacts for EB). This is a result of fronter tongue position preceding an /l/. MB shows fewer contacts in the /l/ context in back and lateral contacts, which may be due to the same difference: fronter and more central tongue palate contact in the /l/ context. The data for the /d/ is less straightforward. MB has no significant differences and AS and EB have differences in different directions. AS has more contact in the front part of the palate and row two in the /l/ context, which fits with his other data, but EB has less central contact and contact in row two in the /l/ context. In general EB has considerably more contact in the front part of the palate than AS does, as can be seen from the means for front contact (13 to 15 for EB *vs.* 1 to 4 for AS in the /t/ of *uttered*). Thus greater contact in row two for EB may represent a process of tongue retraction and a backer closure position (coarticulation with a following /r/), and thus a fronter closure position in the /l/ context, which is almost certainly what AS's data points to.

The corresponding EMA data was extracted automatically and examined. Multivariate GLMs with liquid and word pair as factors were constructed to explore differences in each coil corresponding to liquid context. The consonants were analysed separately.

For EB, Levene's test and the Kolmogorov-Smirnov test showed no significant values for either consonant. The model for the /t/ of *uttered* had liquid as the only significant variable (Pillai's test $F(12,59) = 2.866$, $p < 0.004$). This difference is a result of differences in upper lip, lower lip and lower incisor protrusion ($p < 0.007$, $p < 0.019$ and $p < 0.004$) and tongue back height ($p < 0.031$) with more protruded lip and jaw in the /r/ context and a higher tongue back. Means range from 0.5 to 0.8 mm. For /d/, liquid is again the only significant factor (Pillai's test $F(12,59) = 3.168$, $p < 0.002$) with differences attributable to upper lip x ($p < 0.008$) and y ($p <$

0.048), with the lip more protruded and lower in the /r/ context (by roughly 0.5 and 0.4 mm). Note that these differences are as expected, and indicative of lip rounding in the /r/ context. There are no differences in tongue tip placement, which shows that EMA data is complementary to EPG data; they measure different things.

For MB a few variables are non-normally distributed: for /t/ tongue back x, tip y and lower incisor y (Z = 2.417, 2.331 and 1.688, $p < 0.001$, 0.001 and 0.007 respectively), and for /d/ tongue back x and tongue tip y (Z = 2.595 and 1.893, $p <$ 0.001 and 0.002 respectively). Levene's tests produce no significant results. The model for /t/ has neither liquid nor word pair significant. The model for /d/ has liquid significant (Pillai's test: $F(12,120) = 3.363$, $p < 0.001$) as a result of a difference in lower lip y ($p < 0.001$) which is higher in the /r/ context by approximately 0.64 mm. The factor word pair is not significant.

For AS there are no significant results for Levene's test and the Kolmogorov-Smirnov tests. For /t/ the factor liquid is significant. The differences lie in tongue back and mid y ($p < 0.011$ and 0.016) and mid and tip x ($p < 0.028$ and 0.006). The tongue is retracted and raised in the /r/ context by more than 0.5mm in each case. For /d/ liquid is again the only significant factor (Pillai's test: $F(12,121) = 4.306$, $p < 0.001$), with differences in tongue back and mid y, tongue mid and tip x and upper lip y ($p <$ 0.001 for all). The differences in means for the tongue data are all greater than 0.7 mm and show retraction and raising in the /r/ context. The upper lip is lower by about 0.22 mm in the /r/ context. This data confirms the interpretation of the EPG: that tongue tip closure is fronter in the /l/ context for this subject.

4.4.2. Perseverative effects

In the perseverative direction, EPG data were extracted at the point of maximum contact for the /t/ of *at* and 2 sample Kolmogorov-Smirnov tests performed as for the previous section.

For EB there were no significant differences in the EPG data in this position. MB had one significant difference: row eight (Z = 1.620, $p < 0.011$) with slightly more contact in the /r/ context (only 1/2 a contact difference in means), suggesting backer tongue position. AS shows many more differences. Row two, row three, front and central contact are all different (Z = 1.750, 1.667, 1.833 and 1.667, $p < 0.004$, 0.008, 0.002 and 0.008 respectively), with less contact in the /r/ context in all cases by one or two contacts (see Table 4.12). This is strong evidence of tongue retraction in the /r/ context.

Table 4.12: Mean number of contacts and standard deviations, EPG data for /t/ of *at*: subject AS.

	LIQUID	N	Mean	Std. Deviation	Difference in means
ROWTWO	r	72	.35	1.32	-1.18
	l	72	1.53	2.25	
ROWTHREE	r	72	3.61	2.82	-1.68
	l	72	5.29	2.52	
FRONT	r	72	4.0278	3.5681	-2.92
	l	72	6.9444	4.2621	
CENTRAL	r	72	4.8194	2.3094	-1.35
	l	72	6.1667	2.7373	

The EMA data for this closure for the three subjects was analysed by constructing GLMs with word pair and liquid as factors. For EB tongue tip and mid *x* were not normally distributed. Lower lip *y*, tongue tip and mid *x* and back *y* had unequal error variance as shown by Levene's test. The GLM had word pair a significant factor ($F(60,315) = 1.636$, $p < 0.004$), but not liquid.

For MB tongue back *x* had unequal error variance and tongue back and mid *x*, tip *y* and lower incisor *y* are not normally distributed. Both liquid and word pair are significant in the model ($F(12,120) = 3.435$, $p < 0.001$ and $F(60,620) = 3.392$, $p < 0.001$). Differences in liquid are due to tongue tip *x* ($p < 0.004$) which is retracted by over 0.5 mm in the /r/ context.

For AS the data are all normally distributed, but upper lip *x* has unequal error variance. The model has both liquid and word pair and the interaction between them significant ($F(12,121) = 5.680$, $p < 0.001$, $F(60,625) = 3.258$, $p < 0.001$ and $F(60,625) = 1.519$, $p < 0.009$). The interaction term is due to differences in tongue back and mid *y* ($p < 0.036$ and 0.022) and tongue mid *x* ($p < 0.023$). Differences in liquid are due to lower incisor *y* ($p < 0.001$), lower lip *y* ($p < 0.023$), tongue tip *x* ($p < 0.001$) and upper lip *y* ($p < 0.028$). Tongue mid *x* is also significantly retracted in the /r/ context by about 0.5 mm ($p < 0.015$), but this needs to be treated with caution as this variable is significant for the interaction term. Closer inspection shows that the difference in means is in the same direction for all but the *lobe*/*robe* pair. The other differences in means show a raised lower lip, upper lip and incisor in the /r/ context (by 0.13 to 0.34 mm) and a retracted tongue tip (0.88 mm).

4.5. Summary and conclusions

4.5.1. Long-domain differences

Long-domain articulatory differences were found using the techniques and methodology described in this chapter. Both the tongue and the lips play a role in long-domain coarticulatory effects, but individual subjects may differ slightly with respect to the strategies that they use. Table 4.13 summarises the findings of the point-wise analysis described in this chapter, including consistent results from the EMA, acoustic and EPG data. Notice that all subjects show a combination of lip rounding and tongue retraction (together with lower F_3) in the /r/ context, but that precise details differ from subject to subject. EB, for example, shows few perseverative effects. MB shows more consistent lip rounding effects and AS more consistent tongue retraction and raising. Differences in the extent of significant variables might be a result of variations in coil placement and anatomy, as well as, or instead of, differences in speaker strategy. Different lip size and thickness made it difficult to determine where to place lip coils consistently for all subjects, for example. Nevertheless, the overall picture is clear: there are long-distance coarticulatory effects associated with the difference between an /l/ and an /r/ for all three subjects.

The effects found in this experiment are consistent with Tunley's (1999) finding of a lowered F_3 in an /r/ relative to an /h/ context, although the data are not directly comparable as in this experiment /l/ and /r/ are compared. The effects found have greater extent than those reported by Tunley, and confirm the existence of long-domain resonances. These effects may be attributed solely to the /r/, as is suggested by the perceptual studies described in chapter 2, in which /r/ has more salient cues for perception than /l/. However, the contrast with /l/ appears to extend over a longer domain than the contrast with /h/ reported by Tunley. Further studies with different consonant contrasts are required to resolve this question.

Table 4.13: Summary of results for all three subjects, effect in /r/ context described.

	t	ə	d	ə	r/l	V	C	ə	t
EB	tongue retracted lip rounding	tongue raised lip rounding lower F3	tongue retracted lip rounding	tongue retracted, raised lip rounding lower F3, F1		tongue retracted lip rounding lower F3		tongue fronter, raised	
MB	tongue retracted	lip rounding lower F3, F1	lip rounding	tongue fronted, lowered lip rounding lower F3, F1		tongue retracted, raised lip rounding lower F3		tongue retracted, raised lip rounding lower F3	tongue retracted
AS	tongue retracted, raised	tongue retracted, raised lower F3	tongue retracted, raised lip rounding	tongue retracted, raised lip rounding lower F3, F1		tongue retracted, raised (tip lowered) lip rounding lower F3		tongue retracted, raised lower F3	tongue retracted both lips raised

4.5.2. The relationship between EMA and acoustic data

Interpretation of EMA data is notoriously complicated. Most studies to date have been concerned with the articulatory data in relation to predefined acoustic landmarks, with little work done on the relationship between articulation as measured by EMA and the acoustic output (Hoole 1999). Whilst EMA is a relatively non-invasive technique that gives an interesting view of the movement of the articulators, it is limited: measurements are two-dimensional, at a few fixed midsagittal flesh-points. Thus it is necessary to determine empirically whether EMA measurements are subtle enough to show articulatory changes that lead to small acoustic differences, or whether differences in formant frequencies can be systematically related to small differences in EMA measurements. This has been partly attempted above, and the results will now be discussed in more detail.

Correlations between articulatory and acoustic data were conducted for all of the sets of vowels measured. These were the bold vowels in: 'Have you utt**e**red **a** — at home?' for all repeats (7×6 word pairs for EB, 12×6 for AS and MB) and the vowels following the liquid in X = *lip*, *rip*, *lap*, *wrap*, *lob* and *rob* (7×3 word pairs for EB) and *leap*, *reap* (12×4 word pairs for MB and AS). There were many correlations between articulatory and acoustic data, some of which might be attributable to liquid and word pair context. Thus partial correlations controlling for the effect of liquid and word pair were conducted, and the results from these correlations will be discussed.

In the data for the vowel following the liquid, described in chapter 3 (section 3.5.2, perseverative effects), for EB the small variations in formant frequency remaining after liquid and vowel variation has been considered are not correlated with the differences in articulation as measured by EMA. The partial correlations conducted returned no convincingly significant correlations. It is possible that measurement error and inherent variability together with the relatively small number of observations within each vowel (7 tokens) do not allow a pattern to emerge for this subject. MB and AS have many significant correlations between articulatory and acoustic data for this vowel, once liquid and word pair have been accounted for. F_1 increases with retraction and lowering of the articulators, F_2 increases with fronting and raising of the articulators. F_3 increases with tongue raising for AS and decreases with upper lip lowering for MB.

Significant partial correlations are found for the schwa preceding the liquid (chapter 3, section 3.5.1, anticipatory effects). AS and MB have more correlations than EB. Generalising across subjects, we find that F_2 increases with fronting and raising of the articulators. F_3 decreases with tongue retraction for MB and AS, and F_1 increases with tongue retraction and lip rounding for AS.

For the schwa in the anticipatory domain (that of *uttered*, described in 4.3.1) there are also some significant correlations. F_1 increases with tongue tip lowering (EB), tongue back retraction (MB) and fronting of the lower incisor and lower lip (AS). F_3 increases with raising of the lower lip and fronting of the tongue mid (AS).

The vowel in *at* (4.3.2 above) shows fairly strong partial correlations (controlling again for variation introduced by word pair and liquid). F_1 lowers with raising of the lower lip (MB, AS) and also with fronting of the tongue back (AS). F_2 increases as the tongue body is fronted and raised (EB, MB, AS) and the upper lip retracted (AS). F_3 decreases with tongue tip lowering and retraction (EB, AS) and lip rounding (MB, AS).

The data exhibit very few general trends, and the results appear different for each vowel and often different across subjects. The only consistent result is that F_2 is positively correlated with fronting and raising of the articulators. One subject, EB, exhibits far fewer correlations than the other subjects do. This may well be a result of the smaller data set collected for EB. Differences between the different vowels are more interesting, but difficult to account for.

The relationship between EMA and acoustic data appears to be a complicated one. Nevertheless, the EMA data produces interesting results about coarticulation, and when acoustic-articulatory correlations are found, they tend to confirm the predictions of acoustic theory about correlations between vocal tract configurations and formant frequencies (*e.g.* Stevens 1998)

4.5.3. The relationship between EMA and EPG data

EMA and EPG data prove complementary in this study. Where long-domain differences in EPG contact patterns were found, the EMA data sometimes showed no corresponding significant differences in tongue position, but did show long-domain differences at other points on the articulators. This is hypothesised to be a result of the placement of EMA coils and the measurement resolution of EMA *vs.* EPG. EMA reliably measures with an accuracy of up to 0.5mm, whereas positional differences resulting in contact *vs.* lack of contact between the tongue and the palate may be considerably smaller. The EMA tongue tip coil was placed approximately 5mm back from the tip of the tongue and thus measures only at that point, and does not measure activity of the tongue tip. EMA is better suited to giving a general picture of the position of the tongue, lips and jaw, whereas EPG provides a fairly fine-grained view of tongue-palate contact. As complementary tools for investigating sounds with and without palate contact, both techniques add to the picture of long-domain coarticulatory effects.

CHAPTER FIVE

MODELLING ARTICULATORY DATA

5.1. Introduction

Coarticulatory effects of the extent and nature reported in chapter 4 have not previously been demonstrated in an articulatory and acoustic study. As the existence of such long-distance effects is not well known, most current theories of coarticulation were not intended to account for them and need revision in the light of the data presented in chapter 4. In particular, the school of articulatory phonology (Browman and Goldstein 1986, 1989, 1990b, 1992a) models coarticulation as the overlap of articulatory gestures. This chapter describes the application of this method of modelling to the EMA data analysed in chapters 3 and 4, to test whether overlap is able to account for these long-distance coarticulatory effects. This method was chosen as it is well known and explicitly described (Browman and Goldstein 1985, 1990a) and therefore possible to implement. Utterances are modelled as sequences of gestures, described by a set of parameters, including one controlling overlap of gestures. Statistical analysis of the parameter sets shows which are important in describing the difference between /l/ and /r/ utterances. The results and their implications for this theory of coarticulation are discussed. Gestural overlap does not play a large role in describing the differences, and thus the theory of coarticulation as the result of overlap of gestures is shown to be insufficient.

5.2. Dynamic modelling of articulation

The coproduction model of coarticulation (*e.g.* Fowler and Saltzman 1993), outlined in chapter 1 (section 1.4.2), considers coarticulation to be the result of temporal variation and overlap of stable gestures. The modelling of articulation developed by researchers working within or alongside this framework is detailed and explicitly presented (*e.g.* Browman and Goldstein 1985). This makes it possible to implement a model of gestures based on this framework and to test whether liquid resonances can be modelled as coproduction.

Browman and Goldstein (1985) argue that phonetic structures are patterns of articulatory movement, or gestures, not static configurations. They suggest that a phonetic representation should be a characterisation of how a physical system (the vocal tract) changes over time. Specifying time directly in a description is problematic, as it is well known that temporal variation is introduced by change in speaking rate and stress. Thus a description of articulation (or phonetic structures) as a system which produces behaviour that is organised in time but does not require time as a control parameter is preferable (Fowler 1980).

The dynamical approach to action provides just such a framework (Fowler *et al.* 1980; Kelso *et al.* 1983; Kelso *et al.* 1986). Gestures or units of action are hypothesised to be coordinative structures, functional groupings of muscles that are deployed as a unit in a motor task. Coordinative structures are proposed as a solution to the degrees of freedom problem, as they reduce the number of degrees of freedom involved in speech motor control (Fowler *et al.* 1980). Actions (or gestures) are characterised by underlying dynamic systems which can autonomously regulate the activities of sets of muscles over time. Dynamic systems provide structures that can characterise articulatory movement, explicitly generating patterns of articulator movement in time and space (Browman and Goldstein 1985). An example of a dynamic system, which is used in articulatory modelling, is a simple mass-spring system. Browman and Goldstein (1985) model articulatory trajectories as the output of a simple mass-spring system by fitting sinusoids to the observed trajectories, and show that this model is strikingly successful at fitting the articulatory data.

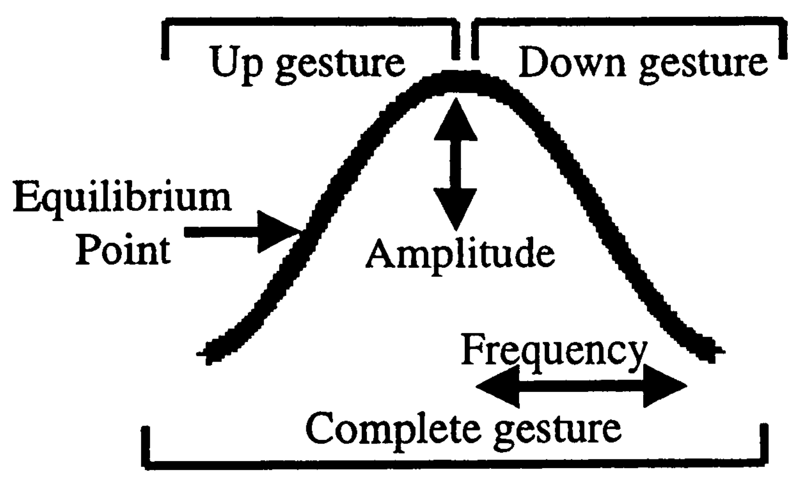
This model of articulation is closely tied to articulatory phonology, which holds that gestures, not features or phonemes, are the units of phonology (Browman and Goldstein 1986, 1990b, 1992a). Phonological processes are modelled as operations on gestures, and coarticulation as the overlap (sometimes total) of adjacent gestures. Testing the applicability of the coarticulation as coproduction model has implications for articulatory phonology: if long-distance coarticulatory effects cannot be modelled as overlap then the theory requires extension or adaptation in order to account for the effects which are demonstrated in this thesis. Articulatory phonology attempts to account for phonological phenomena through articulatory phonetic description. Thus it aims to account for phonetic and phonological phenomena. This chapter shows that the theory's articulatory descriptions fail to account for long-distance coarticulatory effects, be they phonological or phonetic. The phonetic/phonological nature of these coarticulatory effects and possible representations within articulatory phonology are discussed in chapter 6.

5.3. Method

5.3.1. Implementation of a model of gestures

A model of gestures as the result of damped second order differential equations was implemented in *Matlab*, based on Browman and Goldstein (1990a; 1985). *Matlab* modules were created to plot gestures from a set of parameters. Gestures are modelled as two halves (an upward/closing part and a downward/opening part) with parameters controlling the amplitude, frequency and equilibrium points of the half gestures (see Figure 5.1), as well as damping and overlap with following gestures. The modelling of gestures as a series of two half gestures is an adaptation of Browman and Goldstein (1985). Half gestures were used as these allowed much greater accuracy in modelling the data. Browman and Goldstein (1985) contrast two methods of modelling: one in which the model parameters are changed at displacement extrema and one in which they are changed at velocity extrema. Both models fit the data well. The model implemented in this chapter corresponds to the model in which parameters are changed at displacement extrema, as this generates gestures of the type demonstrated in Browman and Goldstein (1990a; 1990b).

Figure 5.1: Model of a gesture with some parameters labelled, showing division of the gesture into two halves.



The equation used to calculate a half gesture (*G*) as a function of sample number (*t*) is: $G(t) = x_0 + Ae^{tk}\sin(\omega t)$, where x_0 is the equilibrium point of the half gesture, *A* the amplitude, ω the angular frequency and *k* the damping factor.

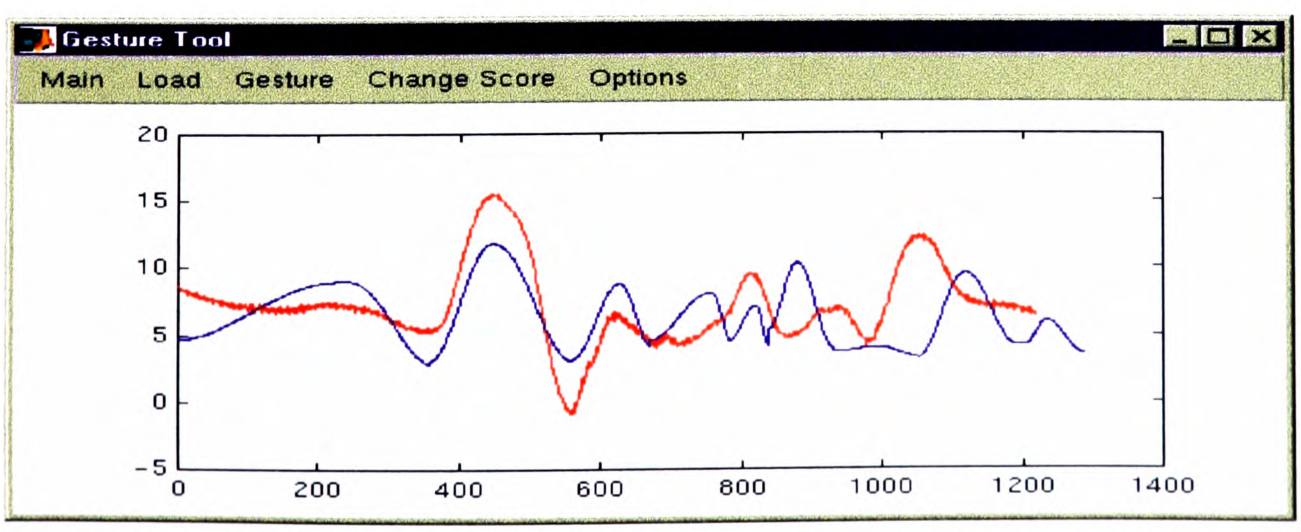
The full list of parameters used in the model for a complete gesture is given in Table 5.1, together with a description of the part of the gesture that the parameter controls. Equilibrium point (x_1, x_2), amplitude (A_1) and frequency (ω_1, ω_2) are as shown in Figure 5.1, damping (k) controls the steepness of the gesture. Amplitude of the down gesture is calculated automatically from the equilibrium point of the down gesture and the amplitude of the up gesture and thus does not need to be extrinsically specified. Overlap angle (ϕ) controls how much of the gesture overlaps with the following gesture.

Table 5.1: Aspects of the gesture that each parameter controls

Parameter label	Function of parameter
x_1	Up equilibrium point
A_1	Up amplitude
ω_1	Up frequency
x_2	Down equilibrium point
ω_2	Down frequency
k	Damping
ϕ	Overlap angle

A tool was created in *Matlab* to draw consecutive gestures, to alter a gesture and its overlap with another gesture by changing parameters, and to compare and calculate the fit of a model and real data. This tool was used to model a subset of the EMA data. A window from the tool is shown in Figure 5.2, it contains a data trace (in red) and a set of gestures created by the model (in blue) which could then be altered using the tool to fit the data trace.

Figure 5.2: Main window from the gesture tool created in *Matlab* and used to fit models to data.



For all three speakers, for all word pairs, two coil traces (tongue mid y and upper lip x) were modelled. These two coils were chosen because they represent both tongue and lip movement in directions of interest: tongue raising and lip protrusion. The upper lip has been of particular interest in studies of coarticulation, as its movement is assumed to be independent of other articulators (Farnetani 1999, 144). The modelling was done by manual adjustment of the parameters controlling successive gestures, using the gesture tool. For the tongue trace models, nine discrete gestures were identified, seven for the upper lip x. These correspond roughly to linguistic segments as shown in Table 5.2 and Table 5.3.

Table 5.2: Tongue mid raising gestures and corresponding segmental labels.

Gesture	1	2	3	4	5	6	7	8	9
Segment	-	u	t	d	LIQV	C	t	ʊ	-
Utterance	hæv	ju ʌ	tə	də	LIQV	C	ət	həʊm	

Table 5.3: Upper lip fronting gestures and corresponding segmental labels.

Gesture	1	2	3	4	5	6	7
Segment	-	u	LIQV	C	əʊ	m	-
Utterance	hæv	ju ʌ	tədəLIQV	C	əthəʊ	m	

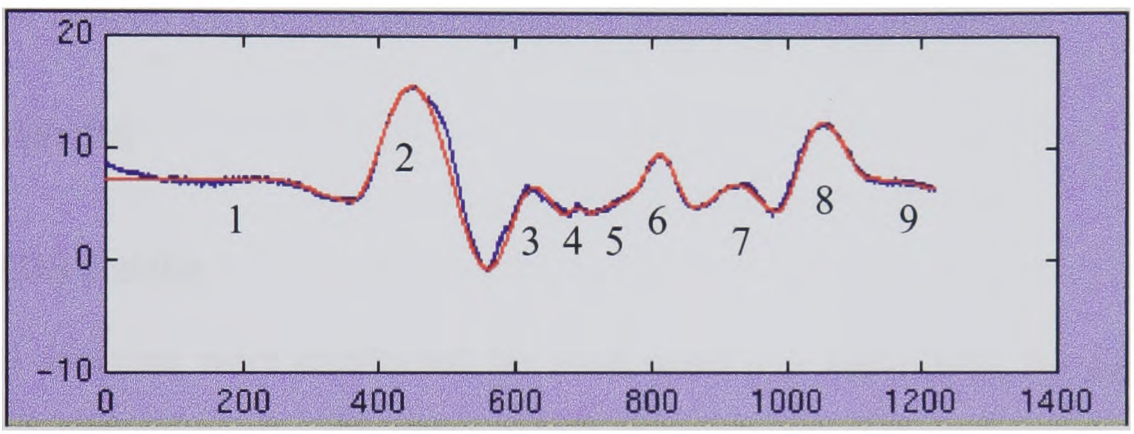
Not all gestures are evident in all tokens for all word pairs. However, in order to maintain consistency, each utterance token was modelled as though it contained the same number of gestures as every other token of that utterance, although some gestures might in consequence have very small, or zero, amplitude. The number of gestures modelled for each trace was decided on the basis of inspection of multiple repetitions of utterances across speakers. When a gesture occurred consistently in more than 1/3 of the traces, it was included. It was fairly straightforward to decide on the number of gestures to model, as the traces were all quite similar.

5.3.2. Accuracy of the model

In general, utterances could be modelled to within the reported measurement error of 0.5 mm of the EMA system. The fit of the model was estimated using a root mean square (rms) measure: $\sqrt{\frac{1}{n}\sum_{1}^{n}(d_j - m_j)^2}$ where n is the number of samples, d_j is the *j*th point of the data trace and m_j the *j*th point of the model. These rms error measurements were all less than 1 mm, with most under 0.5 mm.

Figure 5.3 shows tongue mid *y* data from the utterance 'Have you uttered a lip at home?' (subject MB) in blue and the corresponding model in red. Nine gestures are modelled. The first and last gestures were modelled less carefully than the other gestures, as these are produced respectively before and after the utterance. In some cases, this led to higher rms error values for the whole utterance. The gestures are identified by number in Figure 5.3. These numbers correspond to linguistic segments as discussed in 5.3.1. The rms error of this model is 0.5mm.

Figure 5.3: Tongue mid *y* data (blue) and model (red) for the utterance 'Have you uttered a lip at home?', subject MB. Numbers 1 to 9 identify the nine gestures modelled.



5.3.3. Statistical Analysis

In order to reduce the dimensionality of the data and to look for patterns of interest, multidimensional scaling (MDS) was used. The technique produces a spatial representation of the relationship between different objects (in this case gestures described by parameters), based on a calculation of dissimilarity or similarity (Woods *et al.* 1986). MDS aims to find a configuration in a low dimensional space, usually Euclidean, in which points in the space represent the objects (gestures), and distances between the points in the space match as well as possible the original dissimilarities between objects (Cox and Cox 1994, 1). Thus it aims to find the 'best' fitting set of coordinates to describe the data, and the appropriate dimensionality in which to do this.

The transformation of dissimilarities into distances may be done in many ways. The data in this chapter are analysed using Sammon's method. This is a metric least-squares method, which uses a non-linear mapping from dissimilarities to Euclidean distances, attempting to minimise a weighted least-squares criterion. It begins with a random initial configuration and uses a steepest descent algorithm to search for a minimum value of the fit criterion ('stress'). For further technical details see Everitt and Rabe-Hesketh (1997, 35). Evaluation of the goodness-of-fit in MDS is based on a 'stress' value, with 0% stress interpretable as perfect fit, 5% as good, 10% as fair and 20% as poor (Everitt and Rabe-Hesketh 1997, 39).

The MDS analysis was conducted for each subject and word pair separately. The three-dimensional plots created from the MDS analysis were used to identify gestures which might differ significantly with liquid (/l/ or /r/) context. Non-parametric two-sample Kolmogorov-Smirnov tests were conducted on the parameters of these gestures, for each word pair separately, with liquid as grouping variable. Performing multiple statistical tests increases the chance of finding spurious results. In order to reduce the likelihood of spurious results, MDS was used to identify gestures of interest and reduce the number of statistical tests performed.

5.4. Results

5.4.1. MDS results

The MDS analyses were conducted for each word pair separately, for each subject. The fit is in general good, with stress values always under 10% and usually under 5%. Each plot produced by an MDS analysis was examined for gestures with distinct groupings according to liquid. Non-parametric tests were then conducted for gestures with distinct liquid groupings, and the results of these tests are reported below in section 5.4.2 (tongue mid *y*) and section 5.4.3 (upper lip *x*).

For consistency within data sets (*i.e.* data from an articulator of a subject), the same set of gestures was examined for each word pair, even when differences were not apparent from the plots in all word pairs. The MDS analyses justified the omission of some gestures from the tests. However, the number of gestures excluded was fairly small. Only the first and last gestures could be omitted consistently as most of the other gestures appeared to differ between /l/ and /r/ contexts for at least one word pair. This demonstrates that the modelling method is sensitive enough to reflect long-domain coarticulatory differences and provides an alternative analysis of the extent of differences which is not tied to the notion of the segment, or to pointwise measurements.

As the MDS plots are not easily interpretable other than to show which gestures might differ with liquid context, they are not all displayed here. Only one example will be discussed: the analysis for EB's tongue mid *y* data. The MDS analyses, conducted for each word pair separately, fit well and have stress values all under 5% (0.05). The stress values are displayed in Table 5.4.

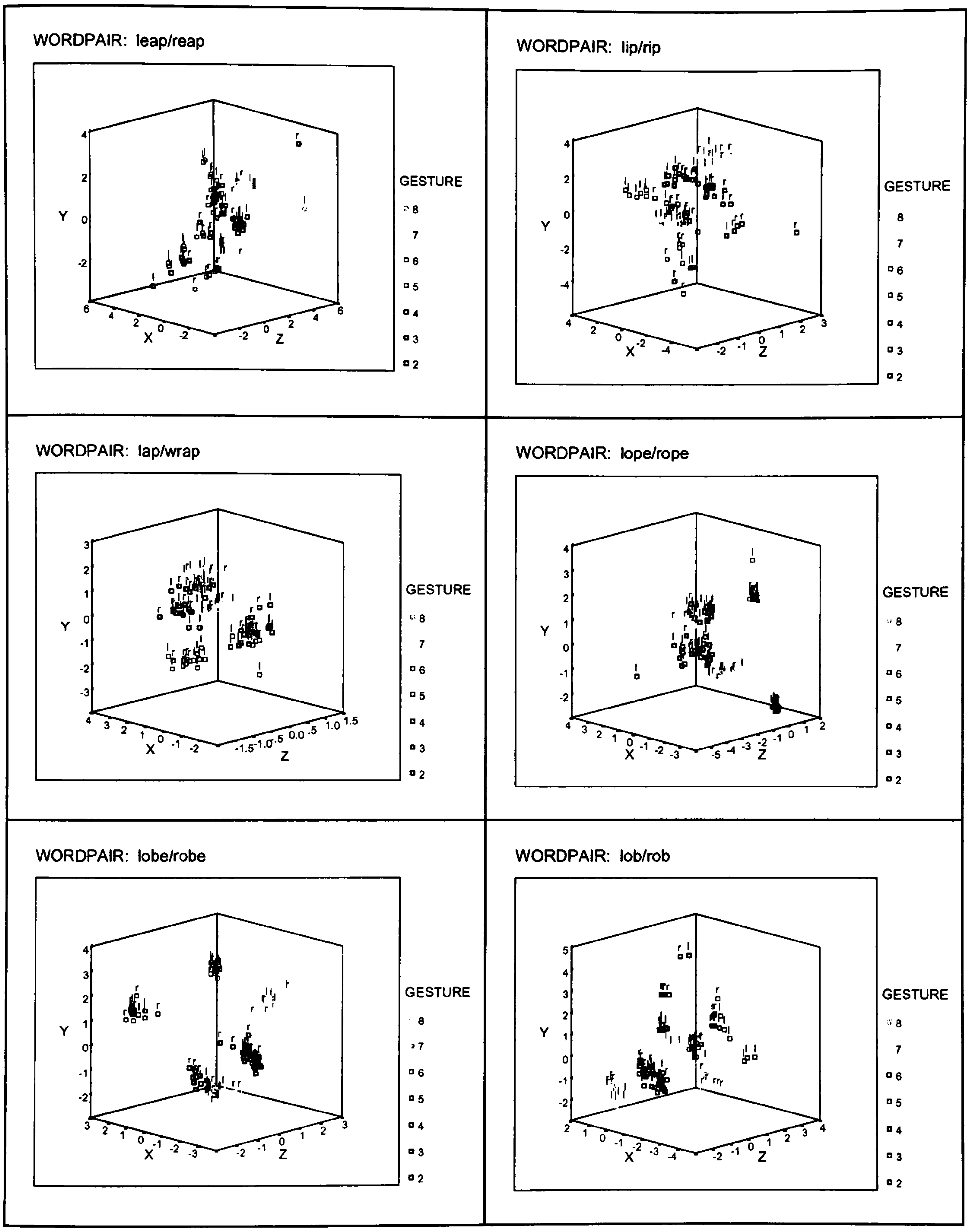
Table 5.4: MDS stress values for tongue mid *y* data, subject EB.

Word pair	leap/reap	lip/rip	lap/wrap	lope/rope	lobe/robe	lob/rob
Stress	0.014	0.016	0.040	0.039	0.009	0.009

The corresponding three dimensional plots are displayed in Figure 5.4. Only gestures 2 to 8 are plotted, as gestures 1 and 9 do not correspond consistently to speech segments, and the graphs become cluttered when all the data points are plotted. Note that the MDS analysis successfully separates gestures (plotted in different colours). Grouping according to liquid (data points labelled 'l' or 'r') is less obvious, although separation of the some /r/ and /l/ groups can be seen for most word pairs. For example, distinct /r/ and /l/ groupings can be seen for gestures 4 (dark blue), 5 (magenta) and 6 (light blue) of the pair *leap/reap*. For the pair *lob/rob*, gestures 4 (dark blue), 5 (magenta) and 7 (yellow) appear to differ between /l/ and /r/ contexts. The *lip/rip* pair shows a possible difference in gestures 3 (green), 4 (dark blue), 5 (magenta), 6 (light blue), and 7 (yellow). Based on inspection of the plots, non-parametric tests were conducted on gestures 3 to 7. The results are reported in 5.4.2 below.

Similar analyses were conducted for the tongue mid *y* and upper lip *x* data sets for each of the three subjects. On the basis of the analyses, tests were conducted on gestures 3 to 8 for the tongue mid *y* data of MB and AS, and gestures 2 to 6 for the upper lip *x* data for all subjects.

Figure 5.4: MDS three-dimensional plots for tongue mid *y* data, subject EB



5.4.2 Tongue movement (tmy)

For EB, non-parametric 2 sample Kolmogorov-Smirnov tests were conducted on the parameters of gestures 3 to 7, for each word pair separately, as these gestures appear from the MDS plots to differ with liquid context for some word pairs. The significant results ($\alpha = 0.05$) are given in Table 5.5. The parameters which differ significantly are those controlling equilibrium points (x_1, x_2), amplitude (A_1), duration of the gestures (ω_1, ω_2) and overlap (ϕ). Damping (k) differs only once.

Table 5.5: Parameters (*e.g.* x_1) which differ significantly according to liquid context, given by word pair and gesture, for tongue mid *y* data, subject EB.

	gest. 3	gest. 4	gest. 5	gest. 6	gest. 7
leap/reap		ω_1	A_1, ω_1, ω_2, ϕ	x_1, x_2, ω_2, k, ϕ	
lip/rip	ω_2	ω_2	x_1, ω_1	x_1	ω_1, ω_2
lap/wrap	ϕ	ω_2, ϕ	x_1, ω_2	A_1	
lope/rope	ω_1, x_2, ω_2	x_1, ω_2	ω_1, x_2	x_2	ϕ
lobe/robe		x_1			
lob/rob		ω_1	x_1, x_2, ω_2		ϕ

Table 5.6 contains the direction of differences for the significant parameters, summarised over word pair, as we are interested in differences which are not related to word pair. A positive difference (+) indicates a larger value for the parameter in the /l/ context, a negative difference (–) indicates a larger value in the /r/ context. The extent of difference is from gesture 3 to gesture 7, which corresponds to the linguistic material 'ttered a LIQVC at'. The most consistent differences are found in gesture 5, which corresponds to the liquid and following vowel. The up gesture's equilibrium point is higher in the /r/ context (larger x_1) and its duration is shorter (larger ω_1); the down gesture's duration is longer (smaller ω_2) in the /r/ context. This suggests a higher tongue position (greater *y* value) for this gesture. The duration differences are more difficult to interpret: there is a shorter upwards gesture and a longer downwards gesture of the tongue mid in the /r/ context than in the /l/.

Table 5.6: Direction of differences for parameters which differ significantly according to liquid context, given by gesture, for tongue mid *y* data, subject EB.

	gest. 3	gest. 4	gest. 5	gest. 6	gest. 7
x_1		– –	– – –	+ +	
A_1			+	–	
ω_1	+	+ –	– – –		+
x_2	–		+ –	+ +	
ω_2	+ –	+ – –	+ + +	–	–
k				–	
ϕ	–	–	–	–	– –

For MB's data, the parameters of gestures 3 to 8, which appeared to differ with liquid context, were tested using non-parametric Kolmogorov-Smirnov tests. The results of the tests are given in Table 5.7, which displays the parameters which differ significantly (α = 0.05) with liquid context. This speaker has a large number of significant differences, ranging from gesture 3 to gesture 8. Note that *k* does not regularly differ with liquid context, but all the other parameters differ significantly in some contexts.

Table 5.7: Parameters which differ significantly according to liquid context, given by word pair and gesture, for tongue mid *y* data, subject MB.

	gest. 3	gest. 4	gest. 5	gest. 6	gest. 7	gest. 8
leap/reap		ω_2, ϕ	x_1, A_1, ω_1, x_2, ω_2	x_1, A_1, ω_1, ω_2	ϕ	
lip/rip		ω_2	x_1, A_1, ω_1, x_2, ω_2	x_1, A_1, ω_1, x_2, ω_2, ϕ	x_1, A_1, ϕ	ω_1, x_2, ω_2, ϕ
lap/wrap	ω_1	A_1, ω_1	x_1, A_1, ω_1, x_2, ω_2	x_1, A_1, ω_1, x_2, ω_2	ω_1, ϕ	ω_2, ϕ
lope/rope	x_2, ω_2	x_1, ω_1, ω_2, ϕ	x_1, A_1, ω_1, x_2, ω_2	x_1, A_1, ω_2, ϕ	x_2, ϕ	A_1, ω_1, x_2, ω_2, ϕ
lobe/robe	ω_1	ω_1, ω_2, ϕ	x_1, A_1, ω_1, x_2, ω_2, k	x_1, A_1, ω_2, ϕ	A_1, ω_1	ω_2, ϕ
lob/rob	ω_1	A_1, ω_1, ϕ	x_1, A_1, ω_1, x_2, ω_2	x_1, A_1, ω_1	A_1, ω_1	x_1, ω_1, ϕ

Table 5.8 shows the direction of differences for significantly different parameters for subject MB, summarised across word pair. The differences are quite consistent, and span gestures 3 to 8 (the linguistic material 'ttered a LIQVC at home'), with particularly evident effects in gestures 4, 5, 6 and 8. Gesture 4 has a longer duration down gesture in the /r/ context (smaller ω_2). Gesture 5 has a higher equilibrium point (larger x_1) and larger amplitude (larger A_1) up gesture and a higher equilibrium point (larger x_2) and shorter duration down gesture (larger ω_2). Gesture 6 has a higher equilibrium point (larger x_1), smaller amplitude (smaller A_1) and shorter duration (larger ω_1) up gesture. Gesture 8 has significantly more overlap with the following gesture in the /r/ context (smaller ϕ). Interpretation of these results is not straightforward, but a higher tongue position in the /r/ context is suggested by the higher equilibrium points.

Table 5.8: Direction of differences for parameters which differ significantly according to liquid context, given by gesture, for tongue mid *y* data, subject MB.

	gest. 3	gest. 4	gest. 5	gest. 6	gest. 7	gest. 8
x_1		–	– – – – – –	– – – – – –	–	–
A_1		+ +	– – – – – –	+ + + + + +	+ + –	–
ω_1	+ – –	+ – – –	+ + + + + +	– – – –	+ + –	+ + +
x_2	–		– – – – – –	– –	–	– –
ω_2	–	+ + + +	– – – – – –	+ + + + –		+ + – –
k			–			
ϕ		+ – – –		+ + –	+ – – –	+ + + + +

For AS, gestures 3 to 8 appeared potentially significantly different across liquid contexts from the MDS plots. Non-parametric two-sample Kolmogorov-Smirnov tests were conducted on the parameters of these gestures for each word pair separately with liquid as factor. The parameters which differ significantly with liquid context ($\alpha = 0.05$) are summarised in Table 5.9. From this table it can be seen that most differences are found in gestures 5 and 6, with some differences found as remote as gestures 3 and 8, the same stretch of linguistic material as for subject MB.

Table 5.9: Parameters which differ significantly according to liquid context, given by word pair and gesture, for tongue mid *y* data, subject AS.

	gest. 3	gest. 4	gest. 5	gest. 6	gest. 7	gest. 8
leap/reap		ω_2	$x_1, A_1, \omega_1, x_2, \omega_2, \phi$	x_1, A_1	A_1, ω_1, ω_2	ϕ
lip/rip		x_2, ϕ	$x_1, A_1, \omega_1, x_2, \phi$	$x_1, A_1, \omega_1, x_2, \omega_2, \phi$		ϕ
lap/wrap			ω_1,4	$x_1, A_1, \omega_1, x_2, \omega_2$	x_1, x_2	A_1, ω_1
lope/rope	ω_2	x_1, ω_1	$x_1, A_1, \omega_1, x_2, \omega_2$	$x_1, A_1, \omega_1, \omega_2, \phi$		
lobe/robe	ω_2	ϕ	$x_1, A_1, \omega_1, x_2, \omega_2$	$x_1, A_1, \omega_1, \omega_2$	ω_1, ϕ	k
lob/rob	x_2, ϕ	x_1, x_2	x_1, A_1, ω_1, x_2	x_1, A_1, ω_1	ϕ	

The parameters which differ most consistently are x_1, A_1, ω_1, x_2, ω_2 and ϕ. The directions of differences are given in Table 5.10.

Table 5.10: Direction of differences for parameters which differ significantly according to liquid context, given by gesture, for tongue mid *y* data, subject AS.

	gest. 3	gest. 4	gest. 5	gest. 6	gest. 7	gest. 8
x_1		– –	– – – – –	– – – – – –	–	
A_1			– – – – –	+ + + + + –	+	+
ω_1		–	+ + + + + –	+ – – – –	– –	–
x_2	–	– –	– – – – – –	– –	–	
ω_2	– –	–	+ – –	+ + + +	+	
k						–
ϕ	+	– –	+ +	+ +	– –	+ –

The strongest and most consistent effects are in gestures 5 and 6 (the liquid, vowel and following consonant). Gesture 5 has higher equilibrium points for the up and down sections, a higher up amplitude and a longer duration up gesture in the /r/ context. Gesture 6 has higher equilibrium points, a lower up amplitude, shorter up duration and longer down duration in the /r/ context. This subject has higher equilibrium points and thus a higher tongue position in the /r/ context, as do the previous two subjects. Relating these results to the segmental material, we see that there are reliable differences in the liquid and following vowel and consonant and smaller differences extending from two syllables before the liquid, to two after it.

5.4.3. Lip movement (ulx)

For EB's upper lip x data, two-sample Kolmogorov-Smirnov tests were conducted on the parameters of gestures 2 to 6, with liquid as factor, as these gestures appeared potentially significantly different across liquid contexts. Significant parameters (α = 0.05) are listed in Table 5.11.

Table 5.11: Parameters which differ significantly according to liquid context, given by word pair and gesture, for upper lip x data, subject EB.

	gest. 2	gest. 3	gest. 4	gest. 5	gest. 6
leap/reap		$x_1, A_1, \omega_1, x_2, \omega_2$		A_1, ω_2	ω_1
lip/rip		$x_1, A_1, \omega_1, x_2, \omega_2$	ϕ		
lap/wrap	ϕ	$x_1, A_1, x_2, \omega_2, \phi$	x_2		
lope/rope	ω_2, ϕ	x_1, A_1, x_2	$x_1, A_1, \omega_1, \omega_2, \phi$	A_1, ω_1	A_1, x_2
lobe/robe		x_1, A_1, x_2, ϕ			ω_1
lob/rob	ϕ			ω_2	

The directions of differences, summarised across word pair, are given in Table 5.12. Most consistent differences are found in gesture 3 (the liquid), where the up gesture has a longer duration (lower ω_1 value), higher equilibrium point (greater x_1) and greater amplitude (greater A_1); the down gesture has a shorter duration (greater ω_2) and higher equilibrium point (greater x_2) in the /r/ context. This is indicative of a stronger lip protrusion gesture in the /r/ context, as the articulator is fronter (greater x values). When the overlap parameter (ϕ) is significantly different, it is greater in the /r/ context, showing less overlap with following gestures in this context.

Table 5.12: Direction of differences for parameters which differ significantly according to liquid context, given by gesture, for upper lip x data, subject EB.

	gest. 2	gest. 3	gest. 4	gest. 5	gest. 6
x_1		– – – – –	–		
A_1		– – – – –	+	+ –	+
ω_1		+ +	–	+	– –
x_2		– – – – –	+		+
ω_2	–	– – –	+	+ +	
ϕ	– – –	– –	– –		

For MB's data, the Kolmogorov-Smirnov test results are shown in Table 5.13, which lists the parameters which differ significantly with liquid context ($\alpha = 0.05$) by gesture and word pair. Parameters from gestures 2 to 6 were tested.

Table 5.13: Parameters which differ significantly according to liquid context, given by word pair and gesture, for upper lip x data, subject MB.

	gest. 2	gest. 3	gest. 4	gest. 5	gest. 6
leap/reap		$x_1, A_1, \omega_1, x_2, \omega_2$	ω_1	x_1	ω_2
lip/rip	ω_2	$x_1, A_1, \omega_1, x_2, \omega_2, \phi$	$\omega_1, x_2, \omega_2, \phi$	ω_1	ω_2
lap/wrap	ϕ	$x_1, A_1, \omega_1, x_2, \omega_2$	ω_1, ϕ	ω_1	ω_1
lope/rope	ϕ	A_1, ϕ	$x_1, \omega_1, \omega_2, \phi$	ω_1, ϕ	
lobe/robe	ϕ	ϕ	x_1, A_1, ω_2, ϕ	x_1	A_1, ω_1, ω_2
lob/rob		ω_1, ϕ	ω_2, ϕ		

The directions of differences, summarised across word pair, are given in Table 5.14. Most consistent differences are again found in gesture 3. As for EB we find evidence of strong lip protrusion: the up gesture has a longer duration (lower ω_1 value), higher equilibrium point (greater x_1) and greater amplitude (greater A_1); the down gesture has a shorter duration (greater ω_2 value) and higher equilibrium point (greater x_2 value) in the /r/ context.

Table 5.14: Direction of differences for parameters which differ significantly according to liquid context, given by gesture, for upper lip *x* data, subject MB.

	gest. 2	gest. 3	gest. 4	gest. 5	gest. 6
x_1		– – –	+ –	– –	
A_1		– – – –	+		–
ω_1		+ + + +	+ – – –	+ – – –	+ –
x_2		– – –	+		
ω_2	–	– – –	+ + –		+ – –
ϕ	+ + –	+ + + –	+ + + –	– –	

For AS, non-parametric tests were also conducted on gestures 2 to 6. The parameters which differed significantly with liquid context ($\alpha = 0.05$) are given in Table 5.15, by gesture and word pair.

Table 5.15: Parameters (*e.g.* x_1) which differ significantly according to liquid context, given by word pair and gesture, for upper lip *x* data, subject AS.

	gest. 2	gest. 3	gest. 4	gest. 5	gest. 6
leap/reap		ω_2	x_1, ω_1, x_2		ω_1
lip/rip		x_2, ω_2	x_1, x_2, ω_2		ω_2, ϕ
lap/wrap		x_2, ω_2, ϕ	x_1, x_2, ω_2	x_1, ω_1, x_2, ω_2, ϕ	x_1, A_1, ω_1, x_2, ω_2, ϕ
lope/rope	ϕ		ϕ	x_1, A_1, ω_2	A_1, ω_1, ϕ
lobe/robe		ϕ	ϕ	ω_2	
lob/rob	A_1		ω_2, ϕ	ω_1, ω_2, ϕ	ω_1

The directions of differences are given in Table 5.16. There are some consistent effects across word pair: gesture 3 (the liquid) has a shorter duration down gesture (greater ω_2 value) with a higher equilibrium point (greater x_2 value) in the /r/ context; gesture 4 (the consonant after the liquid) has higher up and down equilibrium points (x_1, x_2) and a longer duration down gesture (ω_2) in the /r/ context. This suggests more lip protrusion on this consonant in the /r/ context, as the articulator is fronter.

Table 5.16: Direction of differences for parameters which differ significantly according to liquid context, given by gesture, for upper lip x data, subject AS.

	gest. 2	gest. 3	gest. 4	gest. 5	gest. 6
x_1			– – –	+ +	+
A_1	–			–	+
ω_1			+	+ –	+ – –
x_2		– –	– – –	+	+
ω_2		– – –	+ + +	– – –	+ – –
ϕ	+	+ –	+ – –	+ +	– –

5.5. Summary and conclusions

5.5.1. Summary of modelling results

The results of the modelling show that the parameters which differ most with liquid context are those controlling position of the articulators: equilibrium points (x_1, x_2) and amplitude (A_1); duration of the gesture: frequency (ω_1, ω_2) and overlap (ϕ). Only damping (k) is not usually significantly different across liquid contexts. This is not surprising, as this parameter was the one that was least manipulated during the modelling process. Damping might in fact be dependent on the nature of the articulator modelled, which could explain why it did not vary greatly with gesture. Note that the overlap parameter (ϕ) appears important in categorising the differences between /l/ and /r/ contexts, although the direction of the effect varies with word pair, and with subject.

The extent of effects consistent across word pairs is show in Table 5.17 and Table 5.18 for all subjects. All of these effects are in both duration of the gesture and position of the articulators. The positional findings are consistent with the pointwise analysis of the EMA data. The extent of effects which are consistent across word pairs is smaller, possibly because fewer variables were modelled.

Table 5.17: Consistent effects for all subjects for tongue mid *y*.

Gesture	1	2	3	4	5	6	7	8	9
Segment	-	u	t	d	LIQV	C	t	ʊ	-
Utterance	hæv	ju ʌ	tə	də	LIQV	C	ət	həʊm	
EB					*				
MB				*	*	*		*	
AS					*	*			

Table 5.18: Consistent effects for all subjects for upper lip *x*.

Gesture	1	2	3	4	5	6	7
Segment	-	u	LIQV	C	əʊ	m	-
Utterance	hæv	ju ʌ	tədəLIQV	C	əthəʊ	m	
EB			*				
MB			*				
AS			*	*			

When individual word pairs are considered, the extent is longer for some word pairs. The tongue mid *y* data differ from gestures 3 to 7 for the word pairs *lope/rope* and *lip/rip* for EB. Subject MB has differences in tongue mid *y* extending from gestures 3 to 8 for the pairs *lope/rope*, *lap/wrap*, *lobe/robe* and *lob/rob*. AS has differences in tongue mid *y* from gestures 3 to 7 for *lob/rob* and 3 to 8 for *lobe/robe*. The upper lip *x* data shows fewer differences of such long extent. EB has differences in *lope/rope* extending from gestures 2 to 6 and MB has differences in *lip/rip*, *lap/wrap* and *lobe/robe* over the same extent. The word pair *leap/reap* is the only one not to show such long-distance effects. This is consistent with work by Recasens (1985, 1989) which shows that /i/ has a higher degree of coarticulatory resistance than other vowels in Catalan and American English. /l/ and /r/ resonances may not spread through a vowel that is highly resistant to coarticulatory influence. However, Tunley (1999) finds that /i/ in standard Southern British English is more susceptible to rhotic colouring effects (formant lowering) than /æ/. She suggests that /i/ in standard Southern British English is particularly susceptible to the coarticulatory influence of /r/. Tunley finds no coarticulatory effects of /r/ on schwa in the sequence /riCə/, but does find significant F_2 lowering in the sequence /rɪCə/. This suggests that there is a complicated relationship between the susceptibility of a particular vowel to coarticulatory effects, and the existence of coarticulatory effects in adjacent vowels.

The modelling results confirm the pointwise analysis of the EMA data: there is evidence of tongue raising and lip protrusion in the /r/ context relative to the /l/ context. The results are less striking, probably because only two variables were examined out of the set of twelve. Nevertheless the modelling confirms earlier results and introduces an additional result not previously seen: duration of gestures appears to differ with liquid context. This is not surprising for gestures involving the liquid, however duration effects are found in gestures which correspond to segments adjacent to the liquid. The pointwise and dynamic time-warping analyses are unable to capture duration effects: one captures static data, the other analysis compensates or corrects for differences in duration.

5.5.2. Implications for the coproduction theory of coarticulation

The coproduction account of coarticulation claims that coarticulation is the result of variability in timing, and that gestures themselves are relatively stable and do not vary. The results of this chapter show that, while variability in timing is fairly extensive, there is also variability in gestural shape which cannot easily be attributed to timing differences. This suggests that gestural overlap is not sufficient to describe long-distance coarticulatory differences; differences are found in duration and position of the gesture as well as gestural overlap.

Overlap may account for some coarticulatory effects, but it cannot account for the coarticulatory effects described here without some additional mechanism. Of course, the data modelled do not provide a complete account of the articulation involved, and there may be many other factors influencing the shape and timing of articulations. Nevertheless, a straightforward account in terms of overlap of gestures appears insufficient to model this data. Given that long-distance acoustic distinctions are consistently produced and that these appear to be perceptually salient, a model which allows for an interaction between acoustic and kinematic goals (*e.g.* Perkell *et al.* 1997), or even one based solely on auditory perceptual targets (*e.g.* Guenther *et al.* 1998), seems more satisfactory.

Some of the criticisms that have been made of Browman and Goldstein's theory of articulatory phonology are that this approach fails to recognise the cognitive aspects of linguistic organisation (Clements 1992; Coleman 1992a), that it cannot capture certain phonological phenomena such as total place assimilations which involve restructuring of gestures (Kohler 1992), and that it cannot account for the patterns of sounds that occur in a language (Ladefoged 1990). In other words, the theory, whilst trying to account for phonology via phonetics, fails to distinguish between the arbitrary characteristics of a language's phonology and the details of its phonetic implementation. This chapter shows that the theory as it stands is also unable to account for some aspects of fine-grained phonetic detail. In particular, it cannot account for long-distance coarticulatory effects of tongue body height and upper lip protrusion without modification. Possible modifications to the theory are discussed in chapter 6 (section 6.4.4), which contains a discussion of how several current theories, none of which are able to account for long-distance coarticulation, might be adapted to account for the findings of this thesis.

CHAPTER SIX

CONCLUSION

6.1. Introduction

In this chapter the main findings of the thesis are summarised and their theoretical implications discussed. Coarticulatory effects of the extent and nature demonstrated in this thesis have not previously been reported. They thus pose a challenge to current theories of coarticulation, which were not intended to account for such effects. Theoretical issues raised are the status of resonance effects as phonetic or phonological, their implications for the segmental view of speech and for lexical representations. Representing the effects within standard generative phonology and articulatory phonology is shown to be problematic. The gestural overlap and window models of coarticulation are discussed and shown to require modification if they are to account for long-domain effects. It is argued that a radically different approach, the view that fine-grained phonetic detail (such as coarticulatory effects) is stored in the lexicon, is more consistent with the experimental findings. Finally, future directions for research arising from the experimental findings are outlined.

6.2. Summary of results

6.2.1. The perceptual relevance of coarticulatory information

The two perceptual experiments described in chapter 2 show that coarticulatory cues do play a role in perception. Under some circumstances, subjects can identify a word containing a liquid even when the liquid and some adjacent material is deleted. Identification is typically better for /r/s than for /l/s, and cues to an /r/ appear to be more salient and distributed over a greater stretch of the speech signal than cues to an /l/. Perseverative cues to an /l/ appear to be more perceptually salient than anticipatory cues. Listeners can use local coarticulatory (or resonance) information to identify a liquid. Long-distance coarticulatory information is perceptually relevant, as identification of a liquid deteriorates when an /l/ is cross-spliced into the acoustic context of an /r/. Long-domain cues to an /r/ are not salient enough to allow correct identification, but they do conflict with local coarticulatory cues, disrupting recognition.

6.2.2. Acoustic manifestations of long-distance coarticulation

The acoustic results, which are found again and again in several experiments, show that coarticulatory effects may be found in F_1, F_2 and F_3. The most consistent effect is lowering of F_3 in an /r/ context relative to an /l/ context. This effect is found up to two syllables before and after the liquid. With the exception of one subject (EB), anticipatory and perseverative effects appear to be equally likely.

6.2.3. Articulatory evidence

Articulatory evidence for long-distance coarticulatory effects of liquids has been demonstrated. As with the acoustic effects, articulatory effects spread over two syllables in the anticipatory and perseverative direction. EMA and EPG evidence converge to show tongue raising and/or retraction in the /r/ relative to the /l/ context. EMA data show lip rounding in the /r/ relative to the /l/ context. This lip rounding is likely to be responsible for the consistently lowered F_3 found in the /r/ context. Articulatory differences between /l/ and /r/ contexts are surprisingly consistent, and although subtle were successfully discovered using EMA and EPG.

6.2.4. Modelling the articulatory data

A sample of the EMA data was closely modelled using a model based on the dynamic modelling of the school of articulatory phonology. The results of the modelling showed differences in the positioning of the articulators and the duration of gestures associated with the liquid contrast. Gestural overlap did not appear sufficient to model the coarticulatory differences between /l/ and /r/ contexts, which suggests that coarticulatory effects may be attributable to something other than temporal variation of gestures. Acoustic and perceptual goals may play a role, in addition to kinematics.

6.3. Theoretical issues

6.3.1. The extent of coarticulatory effects

The acoustic and articulatory data show that coarticulatory effects may spread over two syllables in the anticipatory and perseverative direction. Effects were consistently found in the l/r word and the bold part of the frame sentence 'Have you **uttered a — at** home?'. The effects found in this context coincide with Kelly and Local's (1986) description of the domain of resonance as metrically bounded, comprising the foot containing the liquid and all unaccented syllables before it. However, the modelling results suggest that effects may exist over an even longer stretch, including the stressed vowel in *home*. As only one context was examined in this thesis, it is not possible to make conclusive statements about the domain. However, it appears that the effects may not be strongly bounded by a phonological domain, but extend over a large extent of speech, decreasing in magnitude with distance from the liquid.

The extent may be influenced by metrical factors such as stress, and dialect (Kelly and Local 1986, 1989). Tunley (1999) discovered long-domain rhotic resonance effects (F_2 and F_3 lowering) in unstressed and not stressed vowels. Heid and Hawkins (2000) have found that F_3 and F_4 differences due to /l/ and /r/ may spread across stressed vowels, although the effects are not found in the stressed vowels themselves. These coarticulatory effects are extremely subtle, which suggests why they had not previously been reported or experimentally verified and were found only to a limited extent by Tunley (1999).

The effects reported by Heid and Hawkins extend over a larger phonological domain than that described by Kelly and Local (1986). Heid and Hawkins found effects which spread over two feet and an unaccented syllable before the liquid. The domains described by both sets of researchers include whole feet and partial feet (the unaccented syllables of one foot). The effects span feet and are not bounded by a higher order constituent in standard metrical theory. The domain is therefore an unusual one for metrical theory. The fact that effects appear to skip over stressed vowels to preceding unstressed vowels (Heid and Hawkins 2000) makes the representation even more complicated. These issues will be discussed further in section 6.4.2 on autosegmental-metrical phonology.

The main coarticulatory effects of /l/ and /r/ reported in this thesis were found in the vowel /ə/. Experimental studies of English coarticulation often employ the vowel /ə/ as a neutral vowel through which coarticulatory effects carry. Catford (1988, 158) writes that the label schwa is often used for any obscure, unstressed mid-central vowel, even though these vary in quality. This variation in quality reflects the fact that schwa often shows strong coarticulatory effects. It is claimed by some, on the basis of detailed acoustic work, that schwa is a vowel without an articulatory target of its own, unspecified for tongue position and completely assimilated with its phonemic context (Bates 1995; van Bergem 1995). The proponents of a target for schwa, Browman and Goldstein, found, using articulatory data, that a V to V trajectory across a schwa (/pipəˈpipə/) was warped slightly towards the neutral or relaxation position. They suggest that the best target for articulatory synthesis is the mean of all other vowels, and that this target should be completely overlapped by a following vowel (Browman and Goldstein 1992). It is apparent from these accounts that schwa is highly susceptible to coarticulatory effects.

The difference in coarticulatory effects found in stressed and unstressed vowels, such as schwa, may be accounted for theoretically in terms of degrees of specification of the vowel. In a study concerned with vowel reduction in southern British English, Bates (1995) found that long vowels and stressed vowels were least susceptible to reduction, although full vowels varied with respect to context-sensitivity. Bates concluded that different vowels may have different degrees of specification (more or less precisely defined targets) for any feature, and that separate degrees of specification must exist for stressed and unstressed vowels. An alternative explanation for the reduction of coarticulatory effects in stressed syllables is given by de Jong *et al.* (1993). They adopt the model of coarticulation as overlap of articulatory gestures and suggest that reduction of articulatory overlap (coarticulation) in stressed syllables is due to hyperarticulation of these syllables (following Lindblom 1990).

The extent of coarticulatory effects reported in this thesis might well differ if vowels other than schwa were investigated, and it would be premature to draw conclusions about the extent of resonance effects at this stage. However, the effects have been shown to be wide-ranging and not restricted to the phonological foot. The extent of these effects is far greater than usually reported. Recasens (1999b), for example, describes coarticulatory effects in VCV and VCCV sequences as long-range coarticulatory effects. The effects reported in this thesis spread over the stretch CVCVCVCVC. Further work is required to determine the maximum extent of the effects and whether they are bounded by a prosodic category or extend for a maximum segmental or temporal stretch.

A further issue associated with the extent of coarticulatory effects is whether there are differences between anticipatory and perseverative effects. Recasens (1999a), surveying results on lingual coarticulation, reports that the onset of anticipatory effects is more consistent than the offset of perseverative effects, suggesting that anticipatory effects reflect gestural formation (planning) and perseverative effects reflect mechano-inertial context factors. The results of the articulatory experiment tend to support the observation about consistency of effects, as anticipatory effects were more consistent across speakers. Farnetani (1999) points out that most studies of labial coarticulation have focussed on anticipatory coarticulation, and reports only one study (of French) which directly examined carryover lip movement. The anticipatory effects examined most often are those of the following vowel on the first vowel in VC_nV sequences, and the results reported are sometimes contradictory (Farnetani 1999). Results such as those reported in this thesis, which shows anticipatory and perseverative labial and lingual coarticulatory effects extending over two syllables in either direction from the influencing consonant, appear not to have been previously demonstrated.

6.3.2. Phonetic or phonological?

The analysis of resonance which Kelly and Local (1986, 1989) present is informed by Firthian prosodic analysis (Firth 1948). Resonance is described as a prosody associated with a certain linguistic domain. Lass (1984, 243) writes that a phonetic property or bundle of properties qualifies as a prosody if it extends over more than one segment or has implication for more than one place in a structure, where place refers to sequential position and grammatical function. Resonance fits into this definition, and the questions of domain and function which Kelly and Local raise can be rephrased in terms of prosodies. Kelly and Local argue that resonance is part of the structure of the language and hence part of its phonology.

Throughout this thesis the terms 'resonance' and 'coarticulatory effects' have been used synonymously. However, they have quite different theoretical flavours: coarticulation is usually described as the influence of one segment on another, whereas resonance is described as a feature associated with a domain. Accounts of coarticulation often view it as phonetic, and not phonological. Ohala (1993) discusses the relationship between coarticulation and phonology. He writes that "phonology is a source of information on coarticulation in speech" (1993, 155), implying that coarticulation is a phonetic phenomenon which, if fossilised (phonologised), results in phonological patterns and sound change. Beckman (1999) discusses the implications of coarticulation for phonological theory and the relationship between phonetics and phonology. She argues that coarticulation is phonetic in nature, but can lead to useful insights about phonological organisation, particularly as it "provides another kind of controllable variation for the phonologist to use in uncovering the speaker's intent" (Beckman 1999, 205). Kelly (1995, 348) describes resonance as "a category of co-articulation in the broad sense of the word", but points out that his observations are not easily formulated in a segmental model. The issue of segmental *vs.* non-segmental models will be discussed further in section 6.3.3.

The relationship between phonetics and phonology and the question of how resonance effects should be classified depends upon one's theoretical perspective. Probably the most widely held, text-book view is that phonetics deals with concrete, physical manifestations of speech sounds; phonology with abstract, psychological manifestations, the nature of human language and the genetic endowment which makes it possible (*e.g.* Spencer 1997, Ohala 1997). Keating (1990b, 1996) writes that phonology deals in discrete symbolic elements, whereas phonetics deals in continuous dimensions, or gradient effects, and relates the phonological idealisations to speech. Keating argues that phonology and phonetics account for different phenomena, at different levels in the grammar. Language specific differences are both phonological (categorical) and phonetic (gradient). Both types of difference should be modelled in the grammar as they are part of speakers' knowledge of their language (Keating 1990b). Ladefoged (1990) also maintains that a distinction should be kept between the physiological notions of phonetics and the mentalistic notions of phonetics. In many views, small coarticulatory effects which spread across word boundaries might be viewed as phonetic detail and not part of the structure of the language or lexical items.

Claiming that the resonance distinction is phonological, as Kelly and Local do, amounts to claiming that clear and dark are two abstract properties (of words or segments or larger stretches of speech). Resonance differences between dialects suggest that resonance is part of the phonology of different dialects. A Firthian prosodic analysis of these effects is, however, not wholly consistent with the data. Resonance effects are relatively large in the central portion of the domain and taper off, decreasing in size, towards the edges. Thus, although the coarticulatory effects of a liquid can be described as distributed over a domain, they are not evenly, but gradiently, distributed over that domain. Detailed phonetic exponency rules, which interact with position in the domain, would be required to explain the gradient nature of the effects. The strength of the prosodic approach is that the effects can be represented phonologically using a single label. The weakness, however, is that the phonological representation does not give any indication of the nature of the effects, and all of the phonetic detail would have to be captured in detailed exponency rules.

The coarticulatory effects of a liquid, however, can be modelled as phonetic even if the existence of the clear/dark distinction falls within the realm of phonology. The experimental results (subtle and variable coarticulatory effects, sometimes perceptually relevant and sometimes not) place the coarticulatory effects of liquids firmly within Keating's definition of "phonetic". Thus Keating's (1990a) "window" model of coarticulation, discussed in section 6.4.3 below, is of interest as a possible way of accounting for the effects.

Following Keating, Cohn (1993) discusses the relationship between phonetic and phonological rules. In her view, language-specific rules can be divided into two categories: phonological and phonetic rules. Gradient behaviour is a result of phonetic implementation and categorical behaviour a result of phonological implementation. Discrete phonological representations are mapped to quantitative phonetic ones, realised in time and space, by interpolation of phonetic targets. In this view, the clearness or darkness of a liquid might be phonological, if it is categorical, whereas its coarticulatory effects on other segments are phonetic. Cohn's work on nasalisation examines nasal airflow patterns for gradient (cline-like, phonetic) or categorical (plateau-like, phonological) behaviour. The articulatory data collected in this thesis are not as amenable to a binary distinction between plateaux and clines. The main analysis of the data is done on discrete points; changes during a segment were not examined in detail. However, the confidence intervals calculated for differences between /l/ and /r/ utterances (chapter 4, section 4.2), show that effect sizes decrease with distance from the liquid. This is suggestive of gradience or interpolation of some kind, and in Cohn's view, the distant coarticulatory effects would be a result of phonetic rules.

Fowler (1990) argues against Keating's (and hence Cohn's) model of the relationship between phonetics and phonology, and the formulation of phonetic rules which are incorporated into the grammar of a language. Fowler (1990, 477) writes that "we must be very careful to distinguish what talkers do on purpose from what just happens anyway". In her view, coarticulation (as distinct from assimilation, which is grammatical) "just happens when phonological segments are serially ordered in speech" (1990, 484). It is not a grammatical process and does not need representation. Fowler suggests that cross-language differences in coarticulation may be a result of inhibition of coarticulatory behaviour. She claims that it is inhibition that is grammatical, rather than coarticulation, although she presents no evidence for inhibitory effects. Following Fowler's discussion, long-distance coarticulatory effects can only be phonetic and should not be modelled as part of the grammar. They are not phonological, as they do not cause category changes of phonemes (*i.e.* they are coarticulatory and not assimilatory effects). Only the absence of coarticulatory effects should be represented in the grammar. Fowler claims that coarticulatory patterns arise from overlapping of articulatory gestures, caused by simultaneous activation of the same articulators. The modelling study in chapter 5 attempted to model coarticulation as overlap of articulatory gestures, but was only partially successful. This suggests that if overlap is to account for coarticulation, a greater level of abstraction is required, perhaps with hidden or underlying gestures. Accounting for coarticulatory effects by extending the articulatory overlap model will be discussed further in section 6.4.3 below.

Current phonetic and phonological theories were not intended to account for coarticulatory effects of the extent described in this thesis. In section 6.4, the representation of resonance within current phonetic and phonological theory will be discussed and models best able to deal with the phenomenon identified. If long-distance coarticulatory effects are 'automatic' and the result of mechanics, then they need to be accounted for by phonetic theory. If they are not automatic, they must be accounted for by either phonetic or phonological theory. In particular, the extent of effects and their nature must be captured in an adequate theory. Should Kelly and Local prove correct in their suggestion that the effects are bounded by a phonological domain, then phonological theory must describe this domain, with the details of the coarticulatory effects worked out in rules of phonetic implementation. The data presented in this thesis do not allow conclusive argument for any one position, so the ability of both phonological and phonetic theories to account for the data will be explored.

6.3.3. Segmental theory

As mentioned above, Kelly (1995) points out that resonance and coarticulation may refer to the same phenomenon, but are not directly compatible because of different assumptions about the segmental/non-segmental nature of speech. Abercrombie (1991, 31) states that "the segment is an invaluable methodological fiction, but it is not an indispensable one", and debate about the status of the segment is found in all fields of phonetics, despite the overwhelming use of segmental notation. Note however, that segmental notation is compatible with non-segmental interpretation or analysis (Local 1995, Coleman 1998b).

In work on perception, views range from Nearey (1990, 1997), who argues that the segments are the units of perception, to Hawkins (1995), who argues for a non-segmental theory of speech perception. Acoustic analysis of speech shows no one-to-one correspondence of the acoustic output and linguistic segments (Repp 1981). Researchers working on speech synthesis have claimed that segmental units are not the best way to produce synthetic speech (Browman & Goldstein 1990a, Hawkins *et al.* 1998, Coleman 1992b, Dirksen and Coleman 1997). Radically different approaches to phonological analysis include non-segmental analyses in the Firthian tradition (*e.g.* Firth 1948, Palmer 1970, Kelly and Local 1989, Local 1995) and segmental, or phonemic, approaches (*e.g.* Clements 1992, Kenstowicz 1994). Theories of lexical representation include the view that words are stored as sequences of discrete segments (Stevens 1998) and the view that words are stored as auditory memories (Coleman 1998a). Most modern phonological theories are phonemic/segmental theories, whereas research in phonetics often recognises non-segmental phenomena (*e.g.* Local 1995, Keating 1995). In short, the status of segments is an unresolved issue in most areas of speech analysis. Several compromises have been attempted. Keating (1995) attempts to resolve the issue by demonstrating that segmental phonology can map onto non-segmental phonetics. Ohala (1992) claims that the segment, which is now an important unit in speech, is not a primitive category but a derived category, which arose from the correlation of transition pairs with temporal coordination of the articulators (*e.g.* [b] is a generalisation from [bi], [bɛ], [bu] *etc.*).

As Kühnert and Nolan (1999) point .out, the traditional conception of coarticulation rests crucially on a phonological model which has underlying discrete segments, which are integrated in regular ways in speech production. The 'problematic' nature of coarticulation results from the view that invariant underlying segments are altered in speech. Beckman (1999) questions the assumption of necessity of sequentially ordered segments, or the 'alphabetic model':

> Is there any phonological evidence for alphabetic segments as discrete units of sequential organisation, i.e. as countable prosodic constituents below the syllable? The best answer at this time is a tentative 'no' ... One of the most important implications of coarticulation for phonological theory is that phonetic evidence for the segment as the basic unit of prosodic information is even less likely to be forthcoming. (Beckman 1999, 221).

Beckman (1999) argues that the role of segments within phonology is to specify groupings (or 'bundles') of features, and that the phenomenon of coarticulation demonstrates that some gestures (described by features) must be coordinated with each other in precise ways. Coarticulation is not a problem, but a natural result of the structure of spoken language.

The results of this thesis add support to the position that a discrete segmental analysis may not be the most appropriate way to analyse speech. The results show that non-segmental phenomena play a role in perception, and that distinctions traditionally associated with segments (/l/ *vs.* /r/) are in fact produced over long sections of speech. The contrast between an /l/ and an /r/ is not merely a contrast between two segments: it is not produced at only one place in structure, but is associated with larger sections of speech. The size of the section with which the contrast is associated may be larger than the word, as discussed in 6.3.1 above.

6.3.4. Lexical representations

The question of segments extends to the question of lexical representation and whether representations are segmental in nature or not. The mainstream view is that lexical items are stored as sequences of segments, defined by features (*e.g.* Stevens 1998, Kenstowicz 1994). The debate concerns whether segments are fully specified for all features in lexical representations, or underspecified for some features (*e.g.* Archangeli 1988). Keating (1988) suggests that underspecification may be found in both phonetic and phonological representations, with unspecified features possibly filled in by interpolation from surrounding values for that feature. Her window model of coarticulation, and other feature-spreading models, rely crucially on the notion that features can be underspecified (Beckman 1999).

It has been suggested that redundant features (those predictable from other features) are underspecified in lexical representations and introduced by phonological linking rules (Stevens *et al.* 1986). However, there is evidence that 'redundant' features perform a useful function in speech perception. Stevens *et al.* (1986) claim that some redundant features enhance distinctive features by strengthening the acoustic representation of the distinctive features. Following on from this work, de Jong (1995) argues that redundant features (in this case roundness of back vowels in American English) add robustness to speech, increasing the acoustic salience of a contrast and allowing a speaker more ways of maintaining a contrast under conflicting motor commands from other segments. Perkell *et al.* (1993) provide evidence of trading relations between articulators (lip rounding and tongue back raising in the production of /u/) to preserve acoustic goals. In this case, the redundant feature of lip rounding helps in the production of comprehensible speech. Realisation of this feature interacts with realisation of the features [back] and [high]. This evidence suggests that 'redundant' features should be specified in phonological representations, as they perform a useful function in speech perception and interact with other 'non-redundant' features.

As the coarticulatory effects of /l/ and /r/ are found across word boundaries, they would be described by post-lexical rules in standard generative phonology, whether phonetic or phonological (*cf.* Cohn 1993, Kenstowicz 1994). Clearness and darkness could be part of the lexical representation of words, but the effects which spread across word boundaries cannot be lexically represented, as the words differ depending on the environment in which they occur. Defining the domain over which spreading occurs remains a problem, regardless of what type of rule is proposed.

6.4. Representations in phonological and phonetic theory

An adequate representation of long-distance coarticulatory effects should consist of two things: a feature/features representing the effects and a means of describing the extent of these effects, taking into account the role of stress and prosodic structure. Although no one phonological theory is able to account for the effects, some theories contain aspects which might allow an account of these effects, *i.e.* features may be sought within standard feature theory and accounts of extent within a combination of autosegmental and metrical phonology. Solutions based on these theories are argued below to be problematic and inadequate. The phonetic implementation of coarticulatory effects is discussed within two competing frameworks: the overlap model (of articulatory phonology) and a feature spreading model (the window model). Both of these models are shown to need modification if they are to account for long-distance coarticulatory effects and are argued to be unsatisfactory.

6.4.1. Feature theory

The coarticulatory differences described in this chapter consist of differences in tongue position (backer or fronter, higher or lower) and lip position (rounded or less rounded). /r/ contexts induce tongue retraction and raising and lip rounding, /l/ contexts have relatively fronter and lower tongue position and unrounded lips. An important question which has not thus far been addressed is whether the effects represent a contrast between /l/ and /r/, or a contrast between /r/ and neutral (*cf.* Heid and Hawkins 2000). Local coarticulatory effects can be attributed to the influence of both /l/ and /r/, as is shown by Tunley's (1999) experiment, which contrasts /l/, /r/ and /h/ environments and finds /l/ and /r/ related effects. However, long-distance effects may be attributable to the influence of /r/ alone, as is suggested by the perceptual evidence that /r/s are identifiable from remote cues whereas /l/s are not (experiment 2, chapter 2). In this case, features for tongue retraction and raising and lip rounding would be required.

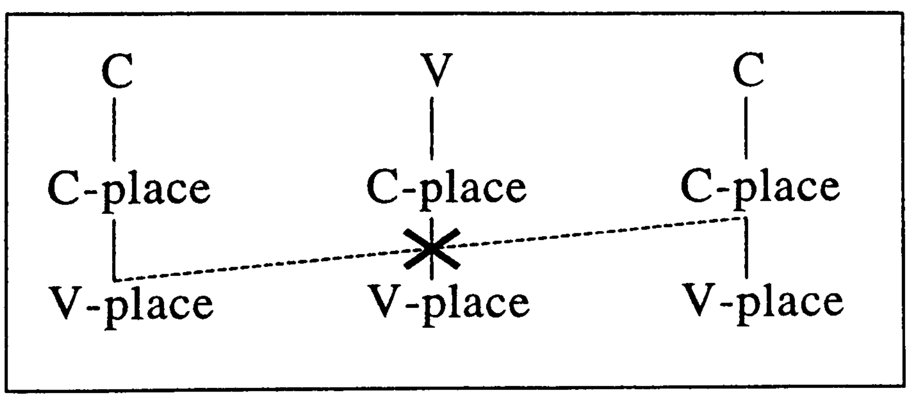
Possible candidates within standard distinctive features theory are [back], [round] and [high]. ([RTR], or [–ATR], does not produce the right effect: it is lowering and retraction of the tongue rather than raising and retraction.) However, use of any or all of these features is problematic, as the coarticulatory effects do not cause category changes. So if, for example, [back] is used to represent /r/ effects we are faced with the problem that front vowels produced with a backer place of articulation in /r/ contexts do not become categorically back. A similar problem occurs if either [round] or [high] is used. For example, in the feature system proposed for British English phonemes in a recent textbook (Davenport and Hannahs 1998, based on the SPE feature system of Chomsky and Halle 1968), spreading the features [+back] and [+round] to /ə/ results in a change of category to the vowel /ɔ/. Thus, spreading the features [+back] and [+round] to the schwa in the word *a* of the frame sentence 'Have you uttered a — at home?' would result in a lexical change of *a* to *awe*. This change is not found in the experimental data: l/r coarticulatory effects do not cause lexical changes. Similar problems obtain for the feature [+high]: spreading [+high] to /ɔ/ results in a category change to /ʊ/. [back], [round] and [high] cannot be used to express coarticulatory differences in this system, and similar problems exist for any other standard system.

The problem is a general one, founded in the nature of distinctive feature systems: they are intended to be efficient systems for expressing category differences, not small within-category differences. Changing the feature specification for a phoneme usually results in a category change. Thus features chosen for these coarticulatory effects would have to be distinct in some way from the standard feature inventory (whether for example on a different tier, attached to different levels of a tree, or specified later in a derivation), as they are not primary place of articulation features. They should apply to both vowels and consonants, with phonetic interpretation adjusting articulation without producing category changes.

In the theory of feature geometry, Clements (1991) and Clements and Hume (1995) suggest a unified set of place features for consonants and vowels which differ in their realisation depending on whether they are under a consonant-place or vowel-place node. This theory allows a distinction to be drawn between primary and secondary place of articulation features for consonants. They suggest that secondary articulation for a consonant is a result of specification of vowel-place features for that consonant. The secondary articulations most relevant for l/r coarticulatory effects, labialization, velarization and palatalization, are represented by the features [labial], [coronal] and [dorsal] under vowel-place nodes. Vowels also have consonant-place features, and as the theory attempts to treat consonants and vowels alike, the vowel-place features are dependent of the consonant-place node for both consonants and vowels. The specification of separate features for secondary articulation means that the problems induced by spreading primary place of articulation features are avoided. As with other feature systems, spreading primary place of articulation features is not possible, as this would cause category changes of the type described above: *e.g.* [dorsal] spread to a /t/ results in /k/. The theory provides an alternative. Coarticulatory effects might be induced by spreading the secondary articulation (vowel-place features) of the liquid consonant to other segments.

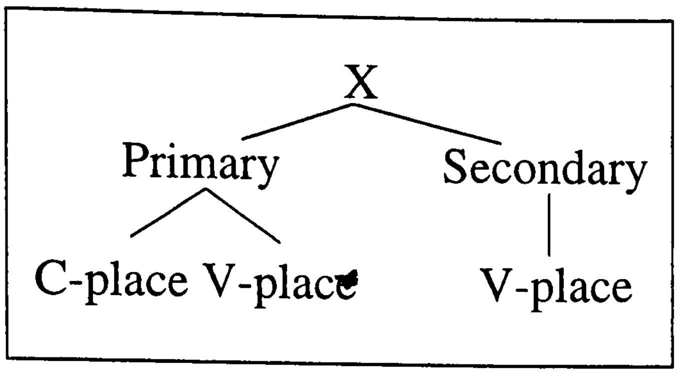
However, spreading of vowel-place features is either blocked by vowels (Figure 6.1) or results in categorical changes in their feature structure. Spreading [dorsal] to a front ([coronal]) vowel changes that vowel to a back ([dorsal]) vowel. Spreading [labial] to the unrounded vowel /ʌ/, would result in the rounded vowel /ɒ/, thus *rub* might change to *rob*. Reversing the argument, spreading [labial] to a vowel which was already specified as [labial] would have no impact in the theory as only one degree of labiality is specified. Yet the experimental data shows that /ɒ/ is produced with more rounding in an /r/ context than in an /l/ context. Spreading of [–labial] or [–round] from an /l/ would not work either as the rounded vowels are still produced with rounding in the /l/ context. The difference is in degree of rounding, with more rounding in an /r/ context. The theory is not equipped to handle these subtle coarticulatory effects which spread across vowels and consonants alike without inducing category changes. This may be a strong reason for the standard assignment of coarticulatory effects to the realm of phonetics, not phonology.

Figure 6.1: Feature geometry: spreading of secondary articulations across vowels is blocked.



One way in which feature geometry might be modified in order to represent these articulatory effects is to separate primary and secondary articulation, and to represent these on separate tiers. This proposal is illustrated in Figure 6.2. In this system secondary articulation features can spread without affecting primary place features, as the spreading takes place on an independent tier.

Figure 6.2: Separating primary and secondary place of articulation in feature geometry.



An alternative to articulatory features, with a long history (Jakobson *et al.* 1952), is a system of features based on acoustic and auditory properties. A feature which might be useful in capturing the different coarticulatory effects of /l/ and /r/ sounds is [flat]. [+flat] sounds have lower formant frequencies than corresponding [–flat] (or plain) sounds, with the contrast articulated by a narrowed aperture e.g. by lip-rounding or pharyngealisation. Thus [+flat] could be used for sounds in the /r/ context, with those in the /l/ context [–flat], or possibly [+sharp] (which indicates raised upper formants), for clear /l/. The feature system of Jakobson *et al.* (1952) has the same inability to capture sub-categorical variation as the other theories discussed. Thus if the coarticulatory effects are to be captured using features, a new feature would need to be introduced, or an old one used in a different way. [flat] might be added to the articulatory features to represent the coarticulatory effects of /l/ and /r/. This proposal is unsatisfactory for feature geometry, as it runs counter to the aim of representing articulatory differences with articulatory features organised in a manner which reflects the vocal tract. It is also problematic for standard distinctive feature theory, as features are motivated by category distinctions and the description of natural classes, not sub-categorical variation.

6.4.2. Autosegmental-metrical phonology

Should a feature be found, or a feature system adapted, to represent the coarticulatory effects of English liquids, then a method of spreading that feature throughout the domain of the effects would be required. Autosegmental-metrical phonology presents a possible mechanism and can be used to represent several aspects of l/r coarticulatory effects. Metrical structure is necessary for specification of prosodic domains and boundaries and a description of the effects of stress. Autosegmental theory allows the spreading of features independent of segmental material and prosodic structure (Goldsmith 1990, Kenstowicz 1994, Kager 1995). However, problems with the representation of long-distance coarticulatory effects using this approach exist. These will be discussed in this section.

An autosegmental approach to long-distance coarticulation would consist of spreading of the relevant features within a certain domain. The standard spreading mechanism predicts spreading of a feature until it is blocked by the appearance of segmental material (an anchor) with a(nother) value of that feature associated with it. This is an instance of the Universal Association Convention which matches features with anchors, spreading rightwards. Long-distance coarticulatory effects pose two problems for an autosegmental spreading analysis: the subcategorical nature of effects and the extent over which they are spread, *i.e.* how they are blocked.

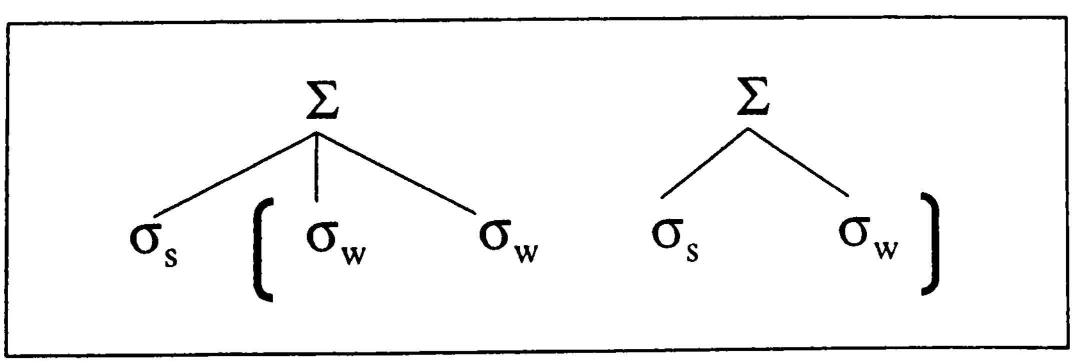
The first problem, that subcategorical effects are not captured well by a phonological feature-spreading mechanism, has been mentioned in the preceding section. Although the autosegmental mechanism allows for independent spreading of features, this spreading alters the feature structure and hence category of segments. One solution to the problem is to assign this feature spreading to a different stage in the phonological derivation, and claim that features spread at a later stage have different phonetic interpretation. This complex solution requires detailed phonetic implementation rules, with different interpretations for features acquired at different stages in a derivation. A more attractive alternative is to associate features with different structural positions (different anchors) and to assign phonetic interpretations to features dependent on the structure (anchor) with which they are associated. For example, features could spread from syllable to syllable, rather than segment to segment (or X slot to X slot). This solution is preferable, and compatible with autosegmental spreading, which is defined as spreading from anchor to anchor, where anchors need not be segments but can be other structural units. Metrical structure provides an appropriate framework over which to spread the effects; this will be discussed further below.

Related to the problem of category changes caused by feature spreading is the fact that coarticulatory effects reduce in size towards the edges of the domain. Autosegmental spreading has no means of representing this, as features either spread or do not, they can not spread partially or in a gradient manner. This is a more serious problem for the approach. The gradient coarticulatory effects would have to be captured by detailed phonetic interpretation rules. As these rules would need to refer to structure as well as segmental material, they might well render the autosegmental spreading unnecessary.

The second problem for autosegmental spreading, that of constraining or blocking spreading, confirms the need for metrical structure in representations. As the extent of coarticulatory effects is related to metrical structure and the effects spread leftwards and rightwards from the liquid, metrical representations are needed to delimit the extent of coarticulation. Autosegmental spreading as described by the Universal Association Convention (Kenstowicz 1994, 317) is insufficient: it matches features to anchors from left to right, and not right to left. Thus, an additional mechanism is required to describe the right to left, or anticipatory, coarticulatory effects. In autosegmental theory, spreading is blocked by the existence of feature specifications for the feature being spread. However, /r/ related effects are not blocked by back rounded vowels (they spread through the vowel /ɒ/ in the word *rob*), nor /l/ effects by front unrounded vowels (they spread through /i/ in *leap*). Thus, it appears highly unlikely that blocking by features can account for the extent of effects. Defining a domain within which spreading occurs is necessary if autosegmental spreading is to work, and metrical phonology provides a suitable framework.

Metrical representations consist of projections from the rhyme of each syllable (σ), these are grouped into feet (Σ) each containing one stressed syllable (marked σ_s) and possibly several weak syllables (σ_w). Feet in turn can be grouped into phonological words or higher constituents. The domain over which the l/r coarticulatory effects spread is larger than the foot, but does not correspond to any existing higher prosodic unit as it does not consist of complete feet, as illustrated in Figure 6.3. The domain is not naturally expressed in the existing framework, although it can be described.

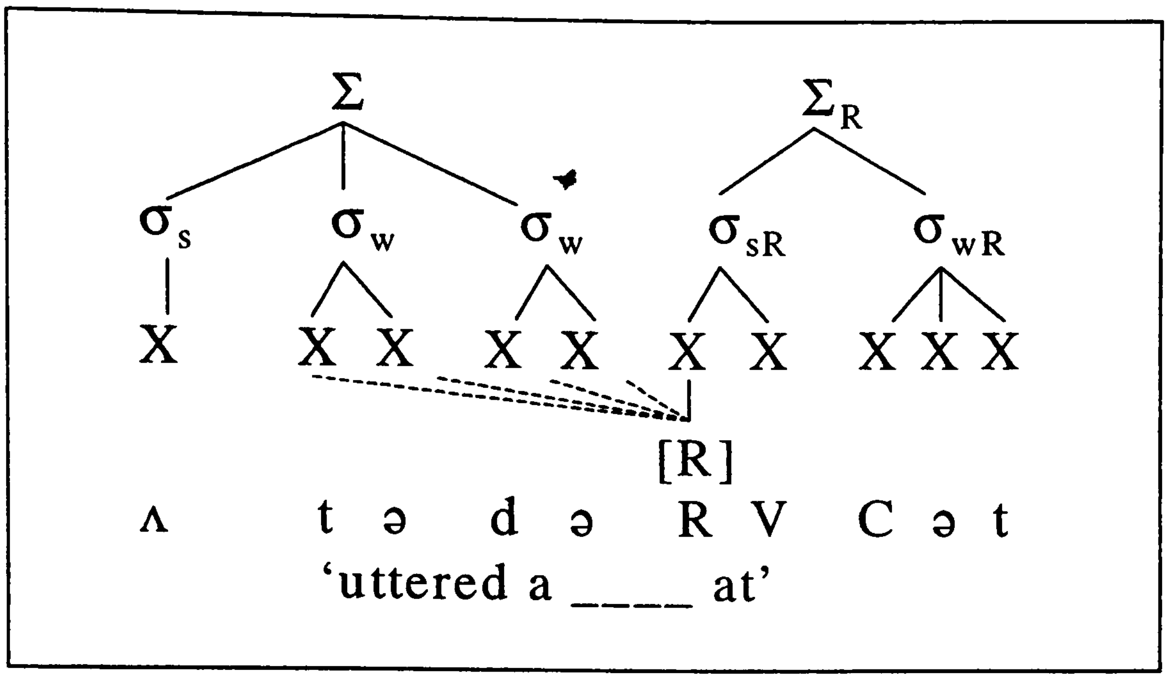
Figure 6.3: The extent of coarticulatory effects, indicated by parentheses. Note that this extent spans feet and does not constitute a well-formed prosodic domain.



A plausible proposal is that resonance (or coarticulatory) effects are constrained by metrical structure. Resonance features are associated to a higher node in the metrical structure, such as the foot or syllable. Features attached to non-segmental nodes (such as the foot or syllable) are given a different phonetic interpretation to those attached to terminal nodes (segments). In this way association of features to metrical structure allows a distinction to be drawn between use of a feature to denote sub-categorical coarticulatory effects (resonance) and use of that feature to describe primary place of articulation. Thus, the feature [round] attached to a syllable node might be used to describe the small coarticulatory rounding effects of an /r/ on non-adjacent syllables, increasing rounding on all segments without making unrounded segments in that syllable categorically round. Differences between the strength of effects in stressed and unstressed syllables might be captured in this proposal if different phonetic interpretations are given for vowels in stressed and unstressed syllables.

A question that remains is how to account for effects that spread beyond the foot in which the liquid occurs. Coarticulation within a foot is easily modelled as the foot node exists in the standard theory. Capturing the extended domain is problematic, as only the weak members of a preceding foot show coarticulatory effects and there is no standard higher node dominating a foot and part of another foot. The best solution is probably to model coarticulation within the foot as associated with the foot node, and coarticulatory effects which spread across feet as the result of autosegmental spreading across structures. An illustration of this proposal is given in Figure 6.4. A resonance feature [R] is shown associated with the foot node (Σ_R) of the foot in which the liquid occurs. This feature is expressed on all the syllables in that foot (σ_{sR} and σ_{wR}), accounting for perseverative coarticulation. The feature also spreads in an anticipatory direction from the liquid to preceding unstressed syllables, until blocked by a stressed vowel. The spreading could be defined on syllable nodes, so that it can be blocked by a stressed syllable. This accounts for the anticipatory domain of coarticulation found in chapter 3.

Figure 6.4: Possible representation of the extent of coarticulatory effects within autosegmental-metrical phonology. A feature [R] is associated with a foot, expressed on all syllables in that foot and spread (indicated by dashed lines) to preceding unaccented syllables.



This solution, although it could be made to work with detailed phonetic implementation rules, is unsatisfactory for several reasons. One of these has been mentioned before: detailed phonetic implementation rules, which are sensitive to position in the resonance domain, are vital for the account to work. These phonetic rules render the autosegmental spreading determined by metrical structure unnecessary. The model described here is incapable of describing the strength of the effects, and the fact that they diminish with distance from the liquid.

Another problem with the account is that the coarticulatory effects do not appear to be a foot-based phenomenon. Their strength when directly adjacent to the liquid strongly suggests that they originate in that position. They are not noticeably greater in the foot within which the liquid occurs than immediately prior to the liquid. In fact, there appears to be little difference between anticipatory and perseverative coarticulation, with anticipatory effects more consistent than perseverative ones. The model illustrated in Figure 6.4 might predict more consistent perseverative effects, as these are structurally determined and not produced by later spreading rules. One piece of evidence does provide some support for this model: the second perceptual experiment (described in chapter 2) suggests that local perseverative cues to an /l/ are stronger than local anticipatory ones.

An alternative, and possibly slightly better, solution might be to originate effects in the syllable containing the liquid and to define autosegmental spreading rules to weak syllables in either direction from the liquid. The coarticulatory effects resulting from autosegmental spreading of features would be smaller than those resulting from association of the features with syllable structure. However, stronger anticipatory effects are found in the vowel before the liquid than other vowels in the anticipatory domain, which belies this simple solution. In addition, the association of features which primarily reflect secondary articulation with a syllable node is not standard within the theory. In short, although a combination of metrical structure and autosegmental spreading rules can be used to construct a model of long-distance coarticulatory effects, the model is unconvincing and complex and fails to reflect the nature of these effects.

6.4.3. Windows

Several theories propose windows for segments to explain phonetic variation. Keating's window model was briefly outlined in chapter 1, section 1.4.2 (Keating 1990a). In the model segments are described by windows, which reflect the variation a segment exhibits. Each relevant articulatory dimension or feature of a segment has a window associated with it. The window is a range of possible spatial values described by a maximum and minimum into which the value for the segment must fall. Windows are independent of context, so the size of the window reflects the amount of contextual variability a segment exhibits. Paths through sequences of windows are plotted by interpolation, resulting in smooth paths. Keating does not specify a precise method of path plotting, but paths are assumed to be smooth and continuous. Paths through VCV sequences could vary considerably, depending on the specifications of the vowels and consonants.

An account of long-distance coarticulatory phenomena using this model is possible and could be checked experimentally by exploring the range of values which segments can assume for the relevant parameters (i.e. their windows). The liquids /l/ and /r/ must be modelled as having different windows for the relevant articulatory and acoustic parameters. Paths from surrounding speech would then be different, because of the difference in the /l/ or /r/ windows through which they must pass. The experimental data suggests a lower F_3 window, and windows representing backer/higher tongue position, smaller lip aperture and more protruded lips for /r/ than for /l/. The windows would have to be quite distinct to exert a long-distance effect on other segments. For paths to remain distinct at a distance, the difference between the /l/ and /r/ should be quite large, unless the windows for the segments affected are very wide and thus allow the paths to remain separate. The window model was designed to account for coarticulatory effects on adjacent segments, not long-distance effects. In particular, it cannot account for the spread of coarticulatory effects through segments which do not show effects. If a segment does not show any difference between two contexts, then the paths through it must be identical. These identical paths cannot subsequently diverge as a result of previous context. Once a segment has ceased to influence the path through adjacent segments, its influence is expended. The only situation in which this might not hold is if the segments are very short. For stressed vowels, which are the segments through which effects have been claimed to spread (Heid and Hawkins 2000), this is unlikely to be the case.

However, the window model may be able to account for some of the phonetic detail of long-distance coarticulatory effects. In particular, reduced effects in stressed vowels are predicted if stressed vowels are represented by very narrow windows. The model is nevertheless not sufficiently developed to account for any hierarchical restrictions on the extent of resonance. The extent of coarticulatory effects would follow automatically from the width and position of windows. In other words, the model is probably capable of handling long-distance coarticulation to the extent that it is a phonetic phenomenon, but requires a phonological model to describe the extent. It is not obvious how a phonological model could influence the extent of coarticulation, as windows are fixed and independent of context. The model does not deal with phonetic variation due to factors other than segmental phonetic context (Farnetani 1997, 393). A possible mechanism would be to allow the phonology to alter the range of values possible for a window, or to shift the window, within the domain of coarticulation. Presumably, the effects of stress would have to be modelled by window changes (narrower windows for stressed vowels, wider windows for unstressed vowels), and the effects of coarticulation might be modelled similarly. Specification at the phonological level with implementation at the phonetic level is entirely in keeping with the Keating's view of the relationship between phonetics and phonology. However, the window model as initially proposed (Keating 1990) is too simple to model long-distance coarticulatory effects adequately.

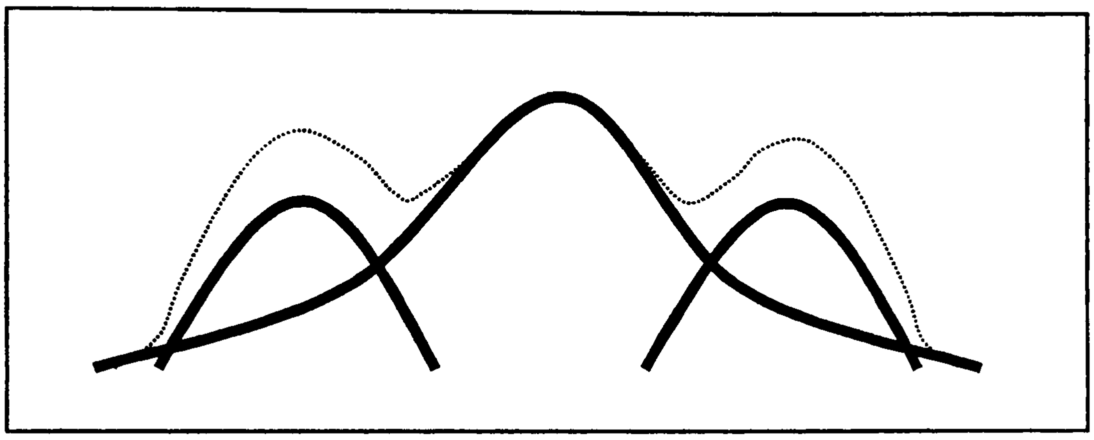
Guenther (1994, 1995a, 1995b) proposes a neural network model which overcomes one of the shortcomings of the window model. He models speech targets as multidimensional convex regions in orosensory coordinates, suggesting that infants learn targets as orosensory coordinates. Targets can increase or reduce in size as a function of speech rate and accuracy of pronunciation, and speech rate effects can therefore be captured in the model. Coarticulation is explained as the reduction of targets based on context, to provide more efficient sequences of articulator movements. The description of target regions overcomes the window model's problem of invariant windows, and is therefore a more promising way of describing coarticulation. Guenther (1995a) demonstrates that the model can account for carryover and anticipatory coarticulation. In another variation on the window model, Blackburn and Young (2000) suggest that probabilities be used to model coarticulation. They describe the midpoint of each phoneme by a smooth probability function (a Gaussian distribution), based on knowledge of variation in articulator positions due to context and relative duration of segments. They contrast their model of coarticulation with Keating's model, showing that it does not specify time intervals during which articulators must satisfy spatial constraints, but only specifies articulator positions at midpoints of phonemes. Thus speech rate effects can be modelled, and it is possible that prosodic effects might also be captured, if the effects of prosody are included in the probability functions. Both of these models deal only with adjacent coarticulatory effects, however, and are not presently sufficiently developed to be extended to account for long-distance effects.

6.4.4. Articulatory overlap

Attempting to expand the model of articulatory overlap to frame an account of long-distance coarticulatory effects of English liquids is problematic, as has been shown in chapter 5. The attempts to model articulatory data suggest that overlap of gestures is insufficient to model these coarticulatory effects. Certainly, overlap of superficially observable gestures as modelled in chapter 5 is insufficient. However, it may be possible to adapt the model in order to account for long-distance coarticulatory effects.

Gestural overlap was implemented in the modelling study as averaging of gestures (following Browman and Goldstein 1992b). However, an alternative proposal about the nature of articulatory overlap exists: overlap may also be interpreted as addition of gestures without averaging (Fowler and Saltzman 1993). With such an additive model of overlap, a lip protrusion gesture which overlaps another lip protrusion gesture will result in additional lip protrusion, not the average of the protrusion described by the two gestures. This model could be used to capture the observation that a rounded vowel in an /r/ context is produced with more lip rounding than the same vowel in an /l/ context. There are two ways in which this might be implemented: /r/ and /l/ gestures may overlap with many adjacent gestures. Alternatively, there may be underlying lip rounding and tongue retraction/raising gestures which influence long stretches of speech. These proposals amount to the same thing, gestures associated with a liquid overlap with several other gestures to cause coarticulation. Figure 6.5 is an attempt to illustrate additive overlap: two gestures are influenced by a longer duration gesture, which affects their amplitude.

Figure 6.5: Use of additive articulatory overlap to represent coarticulatory effects: two shorter duration articulatory gestures are overlapped by a longer gesture in an additive way to produce the outcome (dotted line).

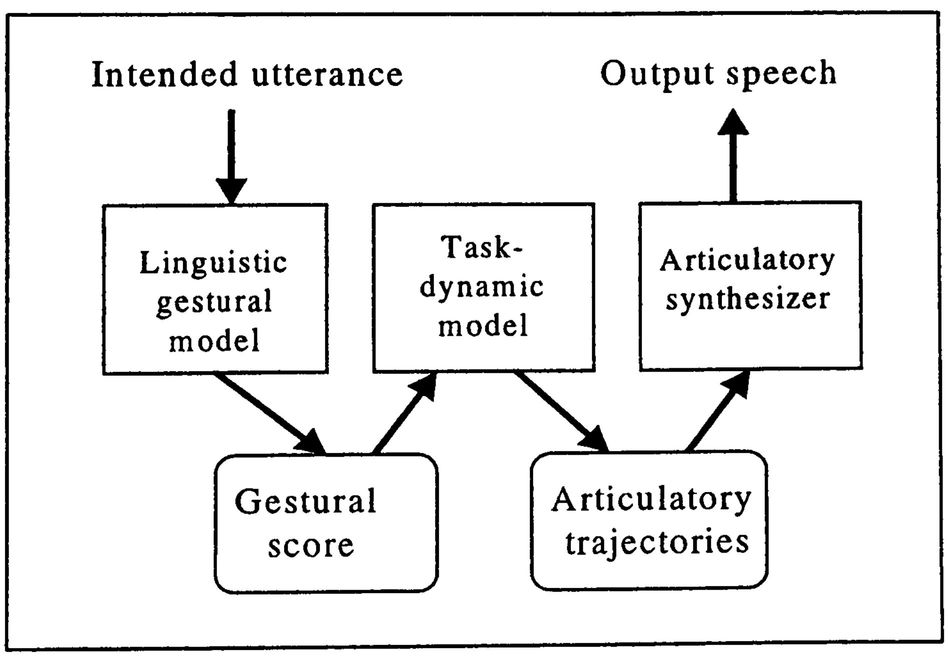


The sub-categorical nature of the l/r coarticulatory effects can be described with the additive version of articulatory overlap, as it captures the fact that the effects add rounding, or increase backness, in a gradient manner. The additive analysis also explains the decrease in magnitude of effects towards the edges of the domain. The gesture has most influence when it is at its maximum, and its influence decreases towards the edges. However, a problem for this analysis is that the articulatory data suggest that effects which span large sections of speech are not present in every segmental position. For example, subject MB shows tongue retraction two syllables before an /r/, but tongue fronting in the vowel immediately preceding the /r/. Acoustic data also suggests that effects may skip stressed vowels (Heid and Hawkins 2000). These results are incompatible with the proposal that an underlying gesture influences all segments with which it overlaps. Specifying different gestural "blending strengths" as in Fowler and Saltzman (1993) might be a means of accounting for absence and presence of effects over a domain. Segments showing no coarticulatory effects would be specified as having much greater blending strength (much greater coarticulatory resistance) than those which show effects.

In the modelling study in chapter 5, an attempt was made to implement blending strengths: a 'blending strength' parameter was introduced to account for different weightings of gestures when they overlapped. However, this parameter was not used for the modelling. The technique adopted of modelling gestures that were superficially obvious and using averaging for overlap meant that overlap weights were not necessary for accurate modelling and merely added to the choice of options for modelling an articulatory trace. The parameter was therefore discarded. However, in a different implementation the concept might be useful. In fact, it is probably necessary for accurate modelling if underlying gestures are proposed which extend over the domain of coarticulatory effects. Determining the values of blending strength is likely to be difficult, however, as the inclusion of this parameter increases the choice of possible models.

Positing these gestures entails introducing a level of abstraction in the analysis, as such gestures are not obvious in the articulatory data. Thus some level of abstraction or phonological specification is required. The description of gestures used in this thesis originated within the computational model of speech production developed at Haskins laboratories. This model (as given by Browman and Goldstein 1990a, 1992b), is illustrated in Figure 6.6.

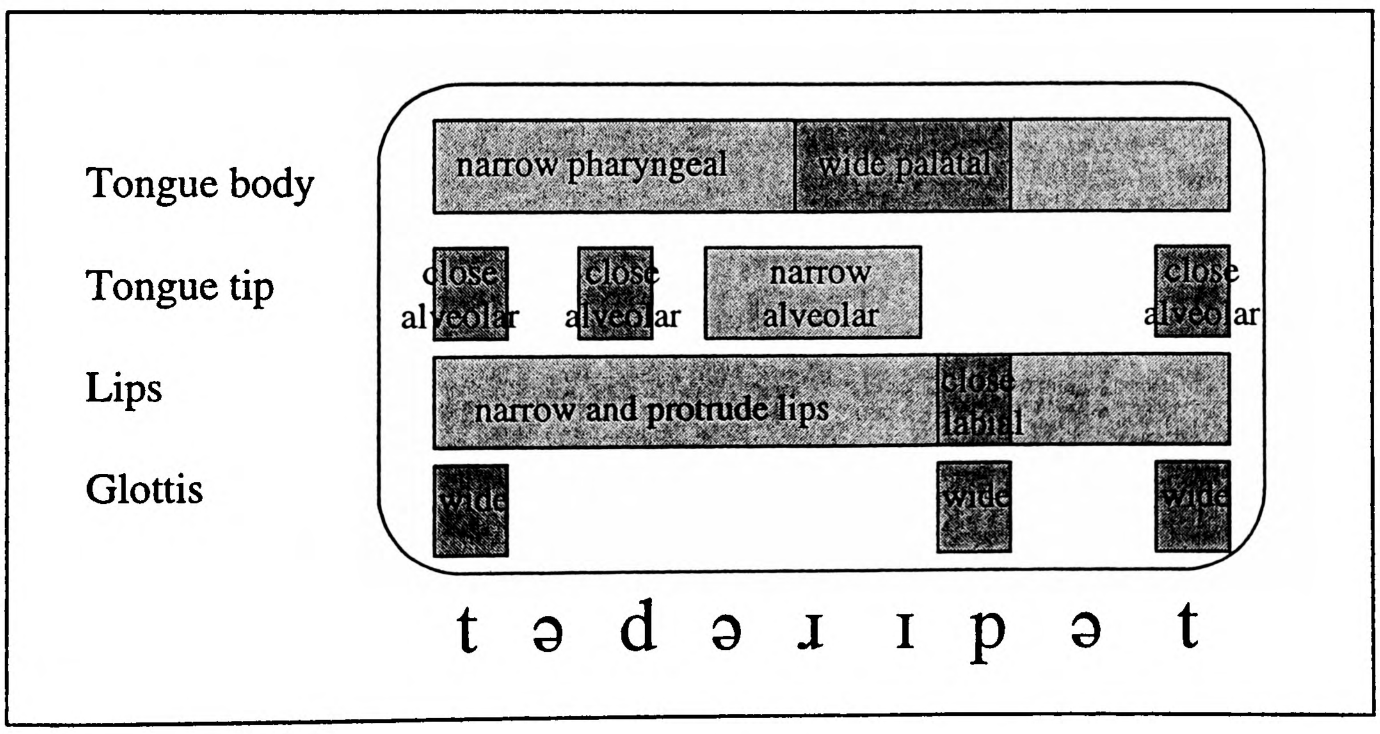
Figure 6.6: Gestural computational model developed at Haskins laboratories (Browman and Goldstein 1990a, 1992b).



Within the model, abstraction takes place at the level of the gestural score, which is created by a 'linguistic gestural model'. Task dynamics is assumed to be universal, so all language specific information resides with the gestural score (Browman and Goldstein 1992b). The gestural score specifies the temporal relationship between gestures, and utterances are modelled as ensembles of (potentially) overlapping gestural units. Phonology is thus based on articulatory units, and the approach is known as articulatory phonology (*e.g.* Browman and Goldstein 1990b).

Articulatory phonology does not provide an obvious way of describing the extent of influence of the gestures, other than by arbitrary specification of gestural activation in the gestural score. Figure 6.7 contains a hypothetical gestural score for /tədəɹɪpət/ ('ttered a rip at' from 'Have you uttered a rip at home?'). Pale boxes represent activation of gestures associated with the /r/, dark boxes represent other gestures ('close alveolar' for the alveolar stops, for example). Gestures attributed to /r/ are pharyngeal narrowing at the tongue body, which is intended to capture tongue retraction and raising, lip narrowing and protrusion for lip effects, and a shorter duration alveolar narrowing associated more directly with the /r/.

Figure 6.7: Hypothetical gestural score for 'ttered a rip at'. Pale blocks represent activation of gestures associated with /r/, dark blocks represent gestural activation for other gestures.



A similar score might be constructed for /l/ utterances, most probably without the long-distance gestures and only a lateral alveolar constriction which influences surrounding vowels. (It is possible that the long-distance effects are rhotic coarticulatory effects, rather than an l/r contrast.) Note that several gestures are completely overlapped by others: 'wide palatal' (for the front vowel /ɪ/ is completely overlapped by the /r/ gesture, as is the 'close labial' (for /p/) by the /r/-related lip gesture. If such long gestures exist, then they must be of lower amplitude than the other gestures, as they were not obvious from inspection of the articulatory data. This score is entirely hypothetical, but if a task dynamic model could derive accurate articulatory trajectories from this score, using an additive approach to overlap, then the theory might represent the effects quite well.

The overlap theory of coarticulation predicts that the extent of gestural overlap is fixed or time-locked (dependent on speech rate) and only dependent on the nature of the gestures involved. As only one frame sentence was examined in the articulatory experiment, this claim cannot be properly evaluated using the data collected. The coarticulatory differences between /l/ and /r/ utterances cannot, however, be a result of differences in the timing of gestures for /l/ and /r/. When dynamic time warping was used to eliminate timing-based differences between /l/ and /r/ utterances, coarticulatory effects were still found (chapter 4, section 4.2.2).

Arbitrary specification of the extent of coarticulatory influence is an unsatisfactory solution as metrical structure influences the extent of coarticulatory effects. Thus a means of specifying the extent needs to be sought. This would presumably be part of the 'linguistic gestural model' (see Figure 6.6) that Browman and Goldstein mention. Articulatory phonology makes little mention of metrical structure. Browman and Goldstein (1988, 1989), followed by Sproat and Fujimura (1993) and Gick (1999a), claim that syllable structure can be seen in gestural organisation. This claim refers to the organisation of gestures within a syllable, allophonic variation conditioned by syllable structure, and the affiliation of consonants with following vowels. It is unclear how this "bottom-up" perspective on syllable structure could be extended to capture coarticulatory effects which are influenced by metrical structure and extend beyond the syllable. However, Browman and Goldstein (1990a) describe a system of tiers within articulatory phonology that are similar in concept to autosegmental tiers. They include an independent rhythmic tier, which contains information about levels of stress. Although structures such as feet and syllables are not directly represented, this is the only place in the theory where metrical structure might be represented.

Byrd (1996) discusses the problem of representing the effects of syllable and phrase structure on articulation within the theory of articulatory phonology.

> As research in speech production becomes more and more integrated with linguistic theory, it has become apparent that articulatory detail cannot be understood except in the context of linguistic structure; such structure includes, but is surely not limited to, the domains of gestures (or

> segments), syllables and phrases. Effects of such phonological structure pervade low-level articulatory behaviour (Byrd 1996, 143).

Byrd is concerned with the effect of linguistic structure on the temporal organisation of speech, and introduces a theoretical framework for describing variability in timing: the phase window framework. Some phasing of gestures is lexically specified: gestures that are relatively stable and 'traditionally considered to constitute a segment'. All other phasing relationships are constrained by a phase window which specifies the degree of variability allowed. Non-lexically specified phase windows are argued to operate post-lexically, on the output of the linguistic gestural model. Complete gestural overlap of the type required for long-distance coarticulatory effects might also be described by rules at this level, but articulatory phonology gives no insight as to how these rules should be formulated or what constrains them. Metrical structure could be added to the rhythmic tier, and gestural activation defined relative to this tier. However, rules of this type could be constructed in any framework and are not specific to articulatory phonology. The articulatory descriptions derived from the gestural score might be an excellent description of the articulatory behaviour. The phonological framework, however, adds nothing and requires considerable elaboration to describe the effects. Furthermore, this addition runs contrary to the spirit of articulatory phonology, which hopes to describe articulatory behaviour in terms of automatic phasing of gestures. Byrd and Saltzman (1998) provide an example of this research goal: they show that temporal lengthening of articulatory gestures at prosodic boundaries can be related to decrease of gestural stiffness. Unless long-distance coarticulatory effects can be shown to be an automatic outcome of some (prosodically related) articulatory parameter, the model of articulatory phonology is challenged by these results.

To conclude, the articulatory modelling developed within this school can be used to describe articulatory data and coarticulatory effects extremely accurately, as was shown in chapter 5. Interpreting overlap as addition of gestures rather than averaging might provide a unified description of the articulatory data. Articulatory phonology, however, seems no better than autosegmental-metrical models in terms of capturing the extent of the effects.

6.5. An alternative approach: lexical storage

The above discussion has been largely concerned with mainstream approaches to phonetics and phonology. There is however, an alternative view. Fine-grained phonetic detail, such as coarticulatory effects, may be lexically stored (Johnson 1997, Goldinger 1997). In such a model the need for phonological representation of coarticulatory effects within a word vanishes as they are directly represented in the lexicon. An articulatory phonology implementation of effects within a word becomes quite plausible if they are stored in the lexicon. Cross-word effects require an explanation, and this issue will be discussed once the theory has been outlined.

The theory of lexical storage of phonetic detail is based on an episodic theory of memory. In episodic theory, each occurrence (or episode) of a stimulus leaves a unique trace in memory. When a stimulus is heard, traces are activated in proportion to their similarity to the stimulus and the most activated trace comes to consciousness and is "recognised" (Goldinger 1997). Goldinger (1997) cites experimental evidence showing that episodic traces of words persist in memory and influence perception and lexical access. He suggests that these episodic memories might constitute the mental lexicon, a lexicon in which perceptual details typically thought of as noise in abstractionist theories are stored:

> The data indicate that detailed episodes are automatically created in perception, and they later effect perception. Thus, it is parsimonious to assume they constitute the basic substrate of the lexicon. (Goldinger 1997, 57)

Additional motivation for detailed lexical representations comes from connectionist models used in automatic speech recognition (Morgan and Scofield 1991) and other psycholinguistic models of speech perception (Gaskell *et al.* 1995, Gaskell and Marslen-Wilson 1997).

Detailed lexical storage might account for listeners' knowledge of variation (*cf.* Pierrehumbert 1994), knowledge of individual talker characteristics (Johnson 1997) and word frequency effects (Goldinger 1997). Coleman (forthcoming) argues that knowledge of phonetic details, phonological statistics and subtle phonetic variation is better explained by a model of lexical representation in which word forms are stored as psychophysical memories. As Coleman (forthcoming) writes, in mainstream phonological theory "knowledge of phonetic details is deputed to a vaguely delineated 'postlexical component', or hoped to follow from the biomechanics of the articulatory system". The hope that phonetic detail might follow from the biomechanics of the articulatory system is likely to be ill-founded, as increasing evidence of language-specific phonetic detail suggests (Nolan 1999). The formulation of post-lexical rules describing long-distance coarticulatory effects has been shown above to be problematic in several mainstream theories. If phonetic detail is stored in the lexicon, then construction of this detail from postlexical rules or properties of the articulatory system is unnecessary, as the detail is already represented.

The results of the perception experiments described in chapter 2 show that an adequate theory of word recognition must allow for the effects of non-adjacent phonetic segments and fine coarticulatory detail. Lexical representations used for speech perception must include fine-grained phonetic detail, as this detail is used in speech perception (Nguyen and Hawkins 1999). A phonemic representation is insufficient, and an underspecified representation even less so (*cf.* Ohala and Ohala 1995). It is possible that there are separate lexicons for perception and production, in which case fine coarticulatory detail might not be specified in the production lexicon. However, it appears unlikely from neurological evidence that there are two distinct lexicons, and the evidence points toward an auditory lexicon (Coleman 1998a).

In this theory, coarticulatory effects within a word are lexically represented. If lexical representations contain such fine-grained phonetic detail, then the search for formal representation of coarticulatory effects within words becomes unnecessary. Coarticulatory detail is lexically stored and speech perception involves matching auditory memories with input. The biggest problem for this approach is the existence of cross-word effects. However, the theory has a response to this. If words are stored as auditory memories, then many different versions of words may be stored, including a continuum of clearer and darker versions which might occur in /l/ or /r/ contexts. If episodic storage is indeed the key to lexical representations, then all (or most) forms heard must be stored. Therefore, coarticulated forms of all words are represented and the only question is how these forms are linked so that appropriate forms are produced in appropriate contexts. If words are represented in psychophysical space, then it is necessary that words are linked in this space, and paths between contextually appropriate word forms exist. Thus the version of *a* most appropriate for an /r/ context might be stored adjacent or connected to /r/ words. Versions appropriate to /l/ contexts are stored with links to /l/ words, etc.

Another question of interest in this approach is the link between the lexicon and production. If the lexicon is auditory in nature, then a mechanism is required by which the fine-grained coarticulatory information stored in the lexicon is recoded into an articulatory plan, for production. This is a general issue for the approach, and not one specific to the production of coarticulation. Models of this sort exist: Guenther (1995a) and Guenther *et al.* (1998) propose a model of speech production in which movement is planned based on acoustic/auditory specifications of speech goals. Positing an articulatory lexicon, to which auditory information must be matched in perception, requires articulatory recoding of the input speech before lexical access. An acoustic lexicon requires recoding for production. The only difference between the two models is the stage at which recoding occurs, whether before lexical access to enable perception or after lexical access to enable production (Coleman 1998a). This particular issue is therefore not a weakness of this model, but a question of broader significance. The model is compatible with an articulatory phonology approach to speech production; lexical storage of detail removes the need for arbitrary rules specifying the extent of coarticulation. Of course, the view that the lexicon is auditory in nature in entirely incompatible with the research programme behind articulatory phonology, but this does not invalidate the usefulness of the articulatory descriptions developed within this approach. A different mapping from the lexicon to articulation is required, as discussed above.

The theory that coarticulatory detail is lexically stored provides a more convincing account of production of the coarticulatory effects described in this thesis, as arbitrarily specified rules are not required. It is able to account for the results of the perception experiments and obviates the need for complex and unsatisfying phonological models. The details of production are probably best accounted for using the articulatory descriptions of articulatory phonology. A model of lexical representations as detailed memories accounts more convincingly for the fine coarticulatory details described in this thesis, and their effect on perception.

6.6. Directions for further research

The experiments described in this thesis demonstrate the existence of long-domain resonance. However, the data is limited to the contrast between /r/ and /l/ and effects on unstressed nonadjacent vowels. Work by Tunley (1999) and Heid and Hawkins (2000) suggests that (non-adjacent) stressed vowels might not show resonance effects, and this requires exploration. Heid and Hawkins' (2000) work demonstrates that metrical structure plays a role in resonance effects, as they find an interaction between stress and resonance effects. Due to the nature of the EMA data collection, and the need to collect many repetitions, metrical structure was kept constant in the articulatory experiment reported in chapters 3 and 4. Future work should control for and vary metrical structure, examining the role of prosody on resonance effects.

Tunley's (1999) work on coarticulation and Bates' work on vowel reduction (Bates 1995) suggests that the strength of coarticulatory effects on a non-adjacent vowel may differ with vowel quality, as discussed in section 6.3.1. This thesis explored long-distance resonance effects only on schwa, as this seemed the vowel most likely to show long-distance coarticulatory effects. Now that these have been established, the effects on other vowels should be explored. This might also allow for the use of less artificial sentences and a wider range of prosodic and phonetic contexts.

In the acoustic studies reported in this thesis, only the first three formant frequencies were considered. Work by others (Hawkins and Heid 2000) suggests that F_4 may be a better indicator of coarticulatory effects of liquids than the other formants. Preliminary findings of theirs suggest that F_4 effects may be larger and have a longer domain, with lowering in the /r/ context and raising in the /l/ context, relative to /h/ contexts. They also suggest that resonance effects of this type may spread through stressed vowels to unstressed vowels. This area of work is very promising, but will require careful acoustic analysis. The domain of resonance effects may yet prove to be even more extensive than that demonstrated in this thesis.

Further work on the perceptual relevance of resonance effects will probably need to be subtler than the gating-type experiments reported here. Possible avenues of exploration include reaction time experiments measuring speed of lexical access, and manipulation of synthetic speech where cues and context are carefully controlled. The effects of l/r coarticulatory cues may not be strong enough to facilitate identification of liquids, but may still affect lexical access. Reaction time tasks may display subtle effects more clearly. Use of synthetic speech in which cues are carefully controlled should allow evaluation of the perceptual relevance of the various acoustic cues (*e.g.* F_3 *vs.* F_4) to the l/r contrast.

One of the main findings of this thesis is that the coarticulatory effects of /r/ are more perceptually salient than those of /l/. This thesis does not explore whether other consonants, or other contrasts between consonants, show similar long-distance effects. It is not clear from the experiments reported here whether /l/ actually has long-distance coarticulatory effects, or whether the long-distance effects found should be solely attributed to /r/. (/l/ clearly has local coarticulatory effects, as does /r/.) As one of the consistent effects found was lip rounding in /r/ contexts, it is possible that other English consonants produced with lip rounding may exhibit similar effects. For example, /ʃ/ is produced with protruded lips relative to /s/, as /r/ is to /l/. Contrasts such as /s/ *vs.* /ʃ/ might show similar effects to the /l/ *vs.* /r/ contrast. An exploration of the role of lip rounding in the English consonant system (*e.g.* Brown 1981) might show that coarticulatory lip rounding is more pervasive in English than commonly thought.

The experimental data point to a need for a better model of coarticulation, one which can account for long-distance effects. Related to this is the issue of whether these coarticulatory effects are planned or automatic, their status in the phonological system of the language and their representation. It is suggested here that coarticulatory effects are lexically stored, but this remains to be demonstrated. Determining the full extent of coarticulatory effects, their interaction with segmental material, and whether they are prosodically determined, is necessary for these questions to be properly addressed.

6.7. Conclusion

Acoustic and articulatory evidence of the existence of long-domain resonances, or long-distance coarticulatory effects associated with the difference between /l/ and /r/ is presented in this thesis. Coarticulatory information associated with the /l/ *vs.* /r/ distinction is shown to be perceptually relevant, and cues to the presence of an /r/ to be wider ranging and more salient than those to an /l/. Some current theories of coarticulation are shown to be in need of revision or expansion given the experimental results presented in this thesis, neither articulatory overlap nor feature-spreading window models account for the extent and nature of effects without modification. A model in which coarticulatory detail is lexically represented is consistent with the results of the experiments. The confirmation of the existence of long-distance coarticulatory effects of English liquids points to the need for further research on the extent and function of coarticulation, and the nature of its representation.

REFERENCES

Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Abercrombie, D. 1991. *Fifty years in phonetics*. Edinburgh: Edinburgh University Press.

Alfonso, P. J., J. R. Neely, P. H. M. M. van Lieshout, W. Hulstijn and H. F. M. Peters. 1993. Calibration, validation and hardware-software modifications to the Carstens EMMA system. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 31: 105-120.

Alwan, A., S. Narayanan and K Haker. 1997. Towards articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *Journal of the Acoustical Society of America* 101: 1078-1089.

Archangeli, D. 1988. Aspects of underspecification theory. *Phonology* 5 (2): 183-207.

Bates, S. R. 1995. *Towards a definition of schwa: an acoustic investigation of vowel reduction in English.* Ph.D. thesis, University of Edinburgh, Edinburgh.

Beckman, M. 1999. Implications for phonological theory. In *Coarticulation: theory, data and techniques*, edited by W. Hardcastle and N. Hewlett. Cambridge: Cambridge University Press. 199-225.

Bell-Berti, F. and K. S. Harris. 1979. Anticipatory coarticulation: some implications from a study of lip rounding. *Journal of the Acoustical Society of America* 65 (5): 1268-1270.

Blackburn, C. S. and S. Young. 2000. A self-learning predictive model of articulator movements during speech production. *Journal of the Acoustical Society of America* 107 (3): 1659-1670.

Bladon, R. A. W. 1979. The production of laterals: Some acoustic properties and their physiological implications. In *Current Issues in the Phonetic Sciences*, edited by H. Hollien and P. Hollien. Amsterdam: John Benjamins. 501-508.

Bladon, R. A. W. and A. Al-Bamerni. 1976. Coarticulation resistance in English /l/. *Journal of Phonetics* 4: 137-150.

Box, G., W. Hunter and J. S. Hunter. 1978. *Statistics for experimenters.* New York: Wiley.

Boyce, S., R. Krakow, F. Bell-Berti and C. Gelfer. 1990. Converging sources of evidence for dissecting articulatory movements into core gestures. *Journal of Phonetics* 18: 173-188.

Branderud, P., R. McAllister and B. Kassling. 1993. Methodological studies of movetrack. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 31: 65-82.

Browman, C. P. and L. Goldstein. 1985. Dynamic modeling of phonetic structure. In *Phonetic Linguistics: Essays in honour of Peter Ladefoged*, edited by V. A. Fromkin. New York: Academic Press. 35-53.

Browman, C. P. and L. Goldstein. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3: 219-252.

Browman, C. P. and L. Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45: 140-155.

Browman, C. P. and L. Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6 (2): 201-251.

Browman, C. P. and L. Goldstein. 1990a. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18: 299-320.

Browman, C. P. and L. Goldstein. 1990b. Tiers in articulatory phonology, with some implications for casual speech. In *Papers in Laboratory Phonology I. Between the grammar and physics of speech*, edited by J. Kingston and M. Beckman. Cambridge: Cambridge University Press. 341-376.

Browman, C. P. and L. Goldstein. 1992a. Articulatory phonology: an overview. *Phonetica* 49: 155-180.

Browman, C. P. and L. Goldstein. 1992b. "Targetless" schwa: an articulatory analysis. In *Papers in Laboratory Phonology II, Gesture, Segment, Prosody*, edited by G. J. Docherty and D. R. Ladd. Cambridge: Cambridge University Press. 26-56.

Brown, G. 1981. Consonant rounding in British English: the status of phonetics descriptions as historical data. In *Towards a history of phonetics*, edited by R. E. Asher and E. Henderson. Edinburgh: Edinburgh University Press. 67-76.

Byrd, D. 1996. A phase window framework for articulatory timing. *Phonology* 13: 139-169.

Byrd, D. and E. Saltzman. 1998. Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26: 173-199.

Carter, P. 1999. Abstractness in phonology and extrinsic phonetic interpretation: the case of liquids in English. *Proceedings of the 14th International Congress of Phonetic Sciences* 1: 105-108.

Catford, J. C. 1988. *A practical introduction to phonetics*. Oxford: Clarendon Press.

Chatfield, C. 1989. *The analysis of time series: an introduction*. London: Chapman and Hall.

Chatfield, C. and A.J. Collins. 1980. *Introduction to Multivariate Analysis*. London: Chapman and Hall.

Chomsky, N. and M. Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Clements, G. N. 1991. Place of articulation in consonants and vowels: a unified theory. *Working Papers of the Cornell Phonetics Laboratory* 5: 77-123.

Clements, G. N. 1992. Phonological primes: features or gestures? *Phonetica* 49: 181-193.

Clements, G., and E. Hume. 1995. The internal organization of speech sounds. In *The handbook of phonological theory*, edited by J. Goldsmith. Oxford: Blackwell. 245-306.

Cohn, A. 1993. Nasalisation in English: phonology or phonetics. *Phonology* 10: 43-81.

Coleman, J. S. 1992a. The phonetic interpretation of headed phonological structures containing overlapping constituents. *Phonology* 9: 1-44.

Coleman, J. S. 1992b. "Synthesis-by-rule" without segments or rewrite rules. In *Talking Machines: Theories, Models, and Designs*, edited by G. Bailly, C. Benoit and T. R. Sawallis. Amsterdam: Elsevier. 43-60.

Coleman, J. 1998a. Cognitive reality and the phonological lexicon: a review. *Journal of Neurolinguistics* 11 (3): 295-320.

Coleman, J. 1998b. *Phonological representations: their names, forms and powers.* Cambridge: Cambridge University Press.

Coleman, J. Forthcoming. Phonetic representations in the mental lexicon. In *Phonology from Phonetics to Cognition*, edited by J. Durand and B. Laks. Oxford: Oxford University Press.

Cox, T. and M. Cox. 1994. *Multidimensional Scaling*. London: Chapman & Hall.

Cruttenden, A. 1994. *Gimson's pronunciation of English.* London: Edward Arnold.

Davenport, M. and D. Hannahs. 1998. *Introducing Phonetics and Phonology*. London: Arnold.

de Jong, K., M. Beckman and J. Edwards. 1993. The interplay between prosodic structure and coarticulation. *Language and Speech* 36 (2-3): 197-212.

de Jong, K. 1995. On the status of redundant features: the case of backing and rounding in American English. In *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Connell and A. Arvaniti. Cambridge: Cambridge University Press. 68-86.

Delattre, P. 1965. *Comparing the phonetic features of English, French, German and Spanish*. London: Harrap.

Delattre, P. C., A. M. Liberman, and F. S. Cooper. 1955. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America* 27: 769-773.

Dirksen, A. and J. S. Coleman. 1997. All-Prosodic Synthesis Architecture. In *Progress in Speech Synthesis*, edited by J. P. H. v. Santen, R. W. Sproat, J. P. Olive and J. Hirschberg. New York: Springer-Verlag. 91-108.

Edwards, V. 1993. A directory of English dialect resources: the English counties. In *Real English*, edited by J. Milroy and L. Milroy. London: Longman. 245-339.

Ellis, L. and W. J. Hardcastle. 2000. Assimilation strategies in the production of alveolar to velar sequences: EPG and EMA data. *Proceedings of the 5th seminar on speech production: models and data*. 117-120.

Everitt, B. S. and S. Rabe-Hesketh. 1997. *The analysis of proximity data*. London: Arnold.

Farnetani, E. 1990. V-C-V lingual coarticulation and its spatiotemporal domain. In *Speech Production and Speech Modelling.*, edited by W. J. Hardcastle and A. Marchal. Dordrecht: Kluwer. 93-130.

Farnetani, E. 1997. Coarticulation and connected speech processes. In *Handbook of the Phonetic Sciences*, edited by W. Hardcastle and J. Laver. Oxford: Basil Blackwell. 371-404.

Farnetani, E. 1999. Labial coarticulation. In *Coarticulation: theory, data and techniques*, edited by W. Hardcastle and N. Hewlett. Cambridge: Cambridge University Press. 144-163.

Farnetani, E. and D. Recasens. 1999. Coarticulation models in recent speech production theories. In *Coarticulation: theory, data and techniques*, edited by W. J. Hardcastle and N. Hewlett. Cambridge: Cambridge University Press. 31-65.

Firth, J. 1948. Sounds and Prosodies. *Transactions of the Philological Society 1948*. 127-152.

Fitzpatrick, L. and A. Ní Chasaide. 1993. Using EMA, EPG and acoustic data to look at quantal theory. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 31: 229-237.

Fitzpatrick, L. and A. Ní Chasaide. 2000. Inferring tongue articulation from simultaneous EMA & EPG: complementary and conflicting data. *Proceedings of the 5th seminar on speech production: models and data.* 201-204.

Fowler, C. A. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8: 113-133.

Fowler, C. A. 1981. Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research* 46: 127-139.

Fowler, C. 1990. Some regularities in speech are not consequences of formal rules: comments on Keating's paper. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, edited by J. Kingston and M. Beckman. Cambridge: Cambridge University Press. 476-489.

Fowler, C. A., P. Rubin, R. E. Remez and M. T. Turvey. 1980. Implications for speech production of a general theory of action. In *Language Production: Speech and Talk*, edited by B. Butterworth. London: Academic Press. 373-420.

Fowler, C. A. and E. Saltzman. 1993. Coordination and coarticulation in speech production. *Language and Speech* 36 (2-3): 171-195.

Fry, D. B. 1979. *The physics of speech*. Cambridge: Cambridge University Press.

Gartenberg, R. 1984. An electropalatographic investigation of allophonic variation in English /l/ articulations. *Work in progress: Phonetics Laboratory, University of Reading* 5: 135-157.

Gaskell, M. G., M. Hare and W. D. Marlesn-Wilson. 1995. A connectionist model of phonological representation in speech perception. *Cognitive Science* 19: 407-439.

Gaskell, M. G. and W. D. Marslen-Wilson. 1997. Integrating form and meaning: a distributed model of speech perception. *Language and cognitive processes* 12 (5/6): 613-656.

Gick, B. 1999a. *The articulatory basis of syllable structure: a study of English glides and liquids*. PhD, Yale University.

Gick, B. 1999b. The organisation of segment-internal gestures. *Proceedings of the 14th International Congress of Phonetic Sciences* 3: 1789-1792.

Giles, S. B., and K. L. Moll. 1975. Cinefluorographic study of selected allophones of English l. *Phonetica* 31: 206-227.

Goldinger, S. 1997. Words and voices: perception and production in an episodic lexicon. In *Talker variability in Speech Processing*, edited by K. Johnson and J. W. Mullinnex. San Diego, CA: Academic Press. 33-66.

Goldinger, S. D., D. B. Pisoni and P. A. Luce. 1996. Speech perception and spoken word recognition: research and theory. In *Principles of Experimental Phonetics*, edited by N. J. Lass. St Louis: Mosby. 277-327.

Goldsmith, J. A. 1990. *Autosegmental and Metrical Phonology*. Oxford: Blackwell.

Goldsmith, J. A. ed. 1995. *Handbook of Phonological Theory*. Oxford: Blackwell.

Gracco, V. L. 1995. Electromagnetic articulography: a brief overview. *Proceedings of the 13th International Congress of Phonetic Sciences* 4: 59.

Grosjean, F. 1996. Gating. *Language and Cognitive Processes* 11 (6): 597-604.

Guenther, F. 1994. Skill acquisition, coarticulation, and rate effects in a neural network model of speech production. *Journal of the Acoustical Society of America* 95 (5): 2924.

Guenther, F. 1995a. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review* 102: 594-621.

Guenther, F. 1995b. A model framework for speech motor development and kinematic articulator control. *Proceedings of the 13th International Congress of Phonetic Sciences* 2: 92-99.

Guenther, F. H., M. Hampson and D. Johnson. 1998. A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review* 105: 611-633.

Guenther, F. H., C. Y. Espy-Wilson, S. E. Boyce, M. L. Matthies, M. Zandipour and J. S. Perkell. 1999. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America* 105 (5): 2854-2865.

Guthrie, D. and J. S. Buchwald. 1991. Significance testing of difference potentials. *Psychophysiology* 28 (2): 240-244.

Hardcastle, W. J. and W. Barry. 1985. Articulatory and perceptual factors in /l/ vocalisation in English. *Work in progress: Phonetics Laboratory, University of Reading* 5: 31-44.

Hardcastle, W. J. and F. Gibbon. 1997. Electropalatography and its clinical applications. In *Instrumental Clinical Phonetics*, edited by M. J. Ball and C. Code. London: Whurr. 149-193.

Hardcastle, W. J. and N. Hewlett, eds. 1999. *Coarticulation: theory, data and techniques.* Cambridge: Cambridge University Press.

Hardcastle, W. J. and J. Laver, eds. 1997. *The Handbook of Phonetic Sciences.* Oxford: Blackwell.

Hardcastle, W. J., F. Gibbon and K. Nicolaidis. 1991. EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics* 19 (3-4): 251-266.

Hardcastle, W. J., W. Jones, C. Knight, A. Trudgeon and G. Calden. 1989. New developments in Electropalatography: A state-of-the-art report. *Clinical Linguistics and Phonetics* 3 (1): 1-38.

Hardcastle, W. J., B. Vaxelaire, F. Gibbon, P. Hoole and N. Nguyen. 1996. EMA/EPG study of lingual coarticulation in /kl/ clusters. Paper read at 1st ESCA Tutorial and Research Workshop on Speech Production Modelling, Autrans, France.

Hawkins, S. 1995. Arguments for a non-segmental view of speech perception. *Proceedings of the 13th International Congress of Phonetic Sciences* 3: 18-25.

Hawkins, S., J. House, M. Huckvale, J. Local and R. Ogden. 1998. ProSynth: An integrated prosodic approach to device-independent, natural-sounding speech synthesis. *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney.

Hawkins, S. and N. Nguyen. Forthcoming. Effects on word recognition of syllable-onset cues to syllable-coda voicing. In *Papers in Laboratory Phonology VI.* Cambridge: Cambridge University Press.

Hawkins, S. and A. Slater. 1994. Spread of CV and V-to-V coarticulation in British English: Implications for the intelligibility of synthetic speech. *Proceedings of the 3rd International Conference on Spoken Language Processing*, Yokohama.

Heid, S. and S. Hawkins. 1999. Synthesizing systematic variation at boundaries between vowels and obstruents. *Proceedings of the 14th International Congress of Phonetic Sciences* 1: 511-514.

Heid, S. and S. Hawkins. 2000. An acoustical study of long domain /r/ and /l/ coarticulation. *Proceedings of the 5th seminar on speech production: models and data.* 77-80.

Henke, W. 1966. *Dynamic articulatory model of speech production using computer simulation.* Ph.D. thesis, MIT.

Hewlett, N. and L. Shockey. 1992. On types of coarticulation. In *Papers in Laboratory Phonology II. Gesture, Segment, Prosody*, edited by D. G. Docherty and D. R. Ladd. Cambridge: Cambridge University Press. 166-183.

Hieronymus. J., M. Alexander, C. Bennett, I. Cohen, D. Davies, J. Dalby, J. Laver, W. Barry and J. Wells. 1990. *Speech Segmentation Criteria for the SCRIBE project.* Unpublished project report.

Hogden, J., A. Lofqvist, V. Gracco, I. Zlokarnik, P. Rubin, and E. Saltzman. 1996. Accurate recovery of articulator positions from acoustics: New conclusions based on human data. *Journal of the Acoustical Society of America* 100 (3): 1819-1834.

Honda, M. and T. Kaburagi. 1993. Comparison of electromagnetic and ultrasonic techniques for monitoring tongue movement. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 31: 121-136.

Hoole, P. 1993. Methodological considerations in the use of Electromagnetic Articulography in Phonetic Research. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 31: 43- 64.

Hoole, P. 1996. Issues in the acquisition, processing, reduction and parameterization of articulograph data. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 34: 158-173.

Hoole, P. 1999. Articulatory-acoustic relations in German vowels. *Proceedings of the 14th International Congress of Phonetic Sciences* 4: 2153-2156.

Hoole, P. and B. Kühnert. 1995. Patterns of lingual variability in German vowel production. *Proceedings of the 13th International Conference on the Phonetic Sciences* 2: 442-445.

Hoole, P. and N. Nguyen. 1999. Electromagnetic articulography. In *Coarticulation: theory, data and techniques*, edited by W. J. Hardcastle and N. Hewlett. Cambridge: Cambridge University Press. 260-269.

Hoole, P., N. Nguyen-Trong and W. Hardcastle. 1993. A comparative investigation of coarticulation in fricatives: electropalatographic, electromagnetic and acoustic data. *Language and Speech* 36 (2-3): 235-260.

Horn, H., G. Göz, M. Bacher, M. Müllauer, I. Kretschmer and D. Axmann-Krcmar. 1997. Reliability of electromagnetic articulography recording during speaking sequences. *European Journal of Orthodontics* 19: 647-655.

Huffman, M. K. 1997. Phonetic variation in intervocalic onset /l/'s in English. *Journal of Phonetics* 25: 115-141.

Hughes, A. and P. Trudgill. 1979. *English accents and dialects*. London: Edward Arnold.

Jakobson, R., G. Fant and M. Halle. 1952. *Preliminaries to speech analysis: the distinctive features and their correlates*. Cambridge, MA: MIT Press.

Johnson, K. 1997. Speech perception without speaker normalization. In *Talker variability in Speech Processing*, edited by K. Johnson and J. W. Mullinnex. San Diego, CA: Academic Press. 145-165.

Johnson, K., E. F. Lemming and R. Wright. 1993. The hyperspace effect: phonetic targets are hyperarticulated. *Language* 69 (3): 505-528.

Jones, D. 1972. *An outline of English phonetics*. Cambridge: Cambridge University Press.

Kager, R. 1995. The metrical theory of word stress. In *The handbook of phonological theory*, edited by J. Goldsmith. Oxford: Blackwell. 367-402.

Kaisse, E. and P. Shaw. 1985. On the theory of Lexical Phonology. *Phonology Yearbook* 2: 1-30.

Katz, W., J. Machetanz, U. Orth and P. Schönle. 1990. A kinematic analysis of anticipatory coarticulation in the speech of anterior aphasic subjects using electromagnetic articulography. *Brain and Language* 38 (4): 555-575.

Keating, P. 1988. Underspecification in phonetics. *Phonology* 5 (2): 275-292.

Keating, P. 1990a. The window model of coarticulation: articulatory evidence. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* edited by J. Kingston and M. E. Beckman. Cambridge: Cambridge University Press. 451-470.

Keating, P. 1990b. Phonetic representations in a generative grammar. *Journal of Phonetics* 18: 321-334.

Keating, P. 1995. Segmental phonology and non-segmental phonetics. *Proceedings of the 13th International Congress of Phonetic Sciences* 3: 26-32.

Keating, P. 1996. The phonology-phonetics interface. *UCLA Working Papers in Phonetics* 92: 45-60.

Kelly, J. 1995. Consonant-associated resonance in three varieties of English. In *Studies in general and English phonetics. Essays in honour of J. D. O'Connor*, edited by J. Windsor Lewis. London: Routledge. 335-349.

Kelly, J. 1989. On the phonological relevance of some non-phonological elements. *Magyar fonetikai füzetak* 21: 56-9.

Kelly, J. and J. Local. 1986. Long-domain resonance patterns in English. *International Conference on Speech Input/Output; Techniques and Applications*. Conference Publication No. 258: 56-59.

Kelly, J. and J. Local. 1989. *Doing Phonology*. Manchester: Manchester University Press.

Kelso, J., E. Saltzman and B. Tuller. 1986. The dynamical perspective on speech production: data and theory. *Journal of Phonetics* 14: 24-59.

Kelso, J., B. Tuller and K. Harris. 1983. A "dynamic pattern" perspective on the control and coordination of movement. In *The production of speech*, edited by P. MacNeilage. New York: Springer-Verlag. 138-173.

Kenstowicz, M. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell.

Kiparsky, P. 1985. Some consequences of lexical phonology. *Phonology Yearbook* 2: 85-138.

Klatt, D. H. 1979. Speech perception: a model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics* 7: 279-312.

Kohler, K. J. 1992. Gestural reorganization in connected speech: a functional viewpoint on articulatory phonology. *Phonetica* 49: 205-211.

Krzanowski, W.J. 1988. *Principles of Multivariate Analysis. A User's Perspective.* Oxford: Oxford University Press.

Kucera, H. and W. N. Francis. 1967. *Computational analysis of present-day American English.* Providence: Brown University Press.

Kühnert, B. 1993. Some kinematic aspects of alveolar-velar assimilations. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 31: 263-272.

Kühnert, B. and F. Nolan. 1999. The origin of coarticulation. In *Coarticulation: theory, data and techniques*, edited by W. J. Hardcastle and N. Hewlett. Cambridge: Cambridge University Press. 7-30.

Kwong, K. and K. Stevens. 1999. On the voiced-voiceless distinction for writer/rider. *Speech Communication Group Working Papers, Research Laboratory of Electronics, Massachusetts Institute of Technology* XI: 1-20.

Ladefoged, P. 1993. *A course in phonetics.* Fort Worth: Harcourt Brace Jovanovich.

Ladefoged, P. 1990. On dividing phonetics and phonology: comments on the papers by Clements and by Browman and Goldstein. In *Papers in Laboratory Phonology I. Between the grammar and physics of speech*, edited by J. Kingston and M. Beckman. Cambridge: Cambridge University Press. 398-405.

Ladefoged, P. and I. Maddieson. 1986. Some of the sounds of the world's languages. *UCLA Working Papers in Phonetics* 64.

Lass, R. 1984. *Phonology.* Cambridge: Cambridge University Press.

Lehiste, I. 1964. *Acoustical characteristics of selected English consonants.* Bloomington: Indiana University Press.

Lieberman, P. and S. E. Blumstein. 1988. *Speech physiology, speech perception, and acoustic phonetics.* Cambridge: Cambridge University Press.

Lindblom, B. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35 (11): 1773-1781.

Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H & H theory. In *Speech production and speech modelling*, edited by W. J. Harcastle and A. Marshal. Dordrecht: Kluwer. 403-439.

Lisker, L. 1957. Minimal cues for separating /w, r, l, y/ in intervocalic position. *Word* 13: 24-43.

Local, J. 1990. Some rhythm, resonance and quality variations in urban Tyneside speech. In *Studies in the pronunciation of English*, edited by S. Ramsaran. London: Routledge. 286-292.

Local, J. 1995. Making sense of dynamic, non-segmental phonetics. *Proceedings of the 13th International Congress of Phonetic Sciences* 3: 2-9.

Lubker, J. and T. Gay. 1982. Anticipatory labial coarticulation: experimental, biological, and linguistic variables. *Journal of the Acoustical Society of America* 71 (2): 437-448.

Macchi, M. J., M. F. Spiegel and K. L. Wallace. 1990. Modeling duration adjustment with dynamic time warping. *Proceedings the International Conference on Acoustics, Speech and Signal Processing 90*: 333-336.

Magen, H. S. 1984. Vowel-to-vowel coarticulation in English and Japanese. *Journal of the Acoustical Society of America* 75 (1): 541.

Magen, H. S. 1997. The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics* 25: 187-205.

Manuel, S. Y. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America* 88 (3): 1286-1298.

Marslen-Wilson, W. and P. Warren. 1994. Levels of perceptual representation and process in lexical access - words, phonemes and features. *Psychological Review* 101: 653-675.

Maurer, D., B. Gröne, T. Landis, G. Hoch and P. W. Schönle. 1993. Re-examination of the relation between the vocal tract and the vowel sound with electromagnetic articulography (EMA) in vocalizations. *Clinical Linguistics and Phonetics* 7 (2): 129-143.

Meddis, R. 1984. *Statistics using ranks. A unified approach.* Oxford: Blackwell.

Mitton, R. A. 1992. *A computer-usable dictionary file based on the Oxford Advanced Learner's Dictionary of Current English.* Available from ftp://ota.ox.ac.uk/pub/ota/public/dicts /710/tet710.dat.

Moll, K. L. and R. G. Daniloff. 1971. Investigation of the Timing of Velar Movements during Speech. *Journal of the Acoustical Society of America* 50: 678-684.

Mooshammer, C. and P. Hoole. 1993. Articulation and coarticulation in velar consonants. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 31: 249-262.

Mooshammer, C., P. Hoole and B. Kühnert. 1995. On loops. *Journal of Phonetics* 23: 3-21.

Mooshammer, C. and N. O. Schiller. 1996. Coarticulatory effects in kinematic parameters of rhotics in German. Paper read at 1st ESCA Tutorial and Research Workshop on Speech Production Modelling, Autrans, France.

Morgan, D. and C. Scofield. 1991. *Neural Networks and Speech Processing.* Dordrecht: Kluwer.

Mrayati, M., R. Carré and B. Guerin. 1988. Distinctive regions and modes: a new theory of speech production. *Speech Communication* 7: 257-286.

Narayanan, S., A. Alwan and K. Haker. 1997. Towards articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals. *Journal of the Acoustical Society of America* 101: 1064-1077.

Nearey, T. M. 1990. The segment as a unit of speech perception. *Journal of Phonetics* 18 (3): 347-373.

Nearey, T. 1997. Speech perception as pattern recognition. *Journal of the Acoustical Society of America* 101 (6): 3241-3254.

Nguyen, N. 1996. A Matlab package for the analysis of EMA/EPG data. *QMC Working Papers in Speech and Language Sciences* 1: 5-12.

Nguyen, N. and S. Hawkins. 1999. Implications for word recoginition of phonetic dependencies between syllable onsets and codas. *Proceedings of the 14th International Congress of Phonetic Sciences* 1: 647-650.

Nguyen, N., P. Hoole and A. Marchal. 1994. Regenerating the spectral shapes of [s] and [ʃ] from a limited set of articulatory parameters. *Journal of the Acoustical Society of America* 96 (1): 33-39.

Nguyen, N. and A. Marchal. 1993. Assessment of an electromagnetic system for the investigation of articulatory movements in speech production. *Journal of the Acoustical Society of America* 94 (2): 1152-1155.

Nolan, F. 1985. Idiosyncrasy in coarticulatory strategies. *Cambridge Papers in Phonetics and Experimental Linguistics* 4.

Nolan, F. 1999. The devil is in the detail. *Proceedings of the 14th International Congress of Phonetic Sciences* 1: 1-8.

O'Connor, J. D. and G. F. Arnold. 1961. *Intonation of Colloquial English.* London: Longman.

O'Connor, J. D., L. J. Gerstman, A. M. Liberman, P. C. Delattre and F. S. Cooper. 1957. Acoustic cues for the perception of initial /w, j, r, l/ in English. *Word* 13: 256-267.

Ohala, J. J. 1992. The segment: primitive or derived? In *Papers in Laboratory Phonology II, Gesture, Segment, Prosody,* edited by G. J. Docherty and D. R. Ladd, Cambridge: Cambridge University Press. 166-183.

Ohala, J. 1993. Coarticulation and phonology. *Language and Speech* 36 (2-3): 155-170.

Ohala, J. 1997. The relation between phonetics and phonology. In *The Handbook of Phonetic Sciences*, edited by W. Hardcastle and J. Laver. Oxford: Blackwell. 674-694.

Ohala, J. and M. Ohala. 1995. Speech perception and lexical representation: the role of vowel nasalisation in Hindi and English. In *Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV,* edited by B. Connell and A. Arvaniti. Cambridge: Cambridge University Press. 41-60.

Öhman, S.E.G. 1966. Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 24: 979-998.

Olive, J. P., A. Greenwood and J. Coleman. 1993. *Acoustics of American English Speech.* New York: Springer-Verlag.

Palmer, F. R. ed. 1970. *Prosodic Analysis.* Oxford: Oxford University Press.

Perkell, J. 1990. Testing theories of speech production: implications of some detailed analyses of variable articulatory data. In *Speech production and speech modelling*, edited by W. Hardcastle and A. Marchal. Dordrecht: Kluwer. 263-288.

Perkell, J. 1991. Models, theory and data in speech production. *Proceedings of the 12th International Congress of Phonetic Sciences* 1: 182-191.

Perkell, J. and M. Matthies. 1992. Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and across-subject variability. *Journal of the Acoustical Society of America* 91 (5): 2911-2925.

Perkell, J., M. Cohen, M. Svirsky, M. Matthies, I. Garabieta and M. Jackson. 1992. Electromagnetic midsaggital articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America* 92 (6): 3078-3096.

Perkell, J., M. Matthies, M. Svirsky and M. Jordan. 1993. Trading relations between tongue-body raising and lip rounding in production of the vowel /u:/: A pilot "motor equivalence" study. *Journal of the Acoustical Society of America* 93 (5): 2948-2961.

Perkell, J., M. Matthies, H. Lane, F. Guenther, R. Wilhelms-Tricarico, J. Wozniak and P. Guiod. 1997. Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication* 22: 227-250.

Pierrehumbert, J. 1994. Knowledge of variation. *Parasession on Variation and Linguistic Theory, 30th meeting of the Chicago Linguistics Society.*

Polka, L. and W. Strange. 1985. Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. *Journal of the Acoustical Society of America* 78 (4): 1187-1197.

Pompino-Marschall, B. 1991. Assimilation as a continuous coarticulatory process: First articulographic results. *Proceedings of the 12th International Congress of Phonetic Sciences* 3: 2-5.

Rabiner, L.R. and R.W. Schafer. 1978. *Digital processing of speech signals.* Englewood Cliffs, N.J.: Prentice-Hall.

Recasens, D. 1985. Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language and Speech* 28 (2): 97-114.

Recasens, D. 1989. Long-range coarticulation effects for tongue-dorsum contact in VCVCV sequences. *Speech Communication* 8 (4): 293-307.

Recasens, D. 1999a. Lingual coarticulation. In *Coarticulation: theory, data and techniques*, edited by W. Hardcastle and N. Hewlett. Cambridge: Cambridge University Press. 80-104.

Recasens, D. 1999b. Acoustic analysis. In *Coarticulation: theory, data and techniques*, edited by W. Hardcastle and N. Hewlett. Cambridge: Cambridge University Press. 322-336.

Recasens, D., J. Fontdevila and M. D. Pallarès. 1996. Linguopalatal coarticulation and alveolar-palatal correlations for velarized and non-velarized [l]. *Journal of Phonetics* 24: 165-185.

Recasens, D., M. D. Pallarès and J. Fontdevila. 1997. A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America* 102 (1): 544-561.

Recasens, D. and M. D. Pallarès. 1999. A study of /ɾ/ and /r/ in the light of the "DAC" coarticulation model. *Journal of Phonetics* 27: 143-169.

Repp, B. H. 1981. On levels of description in speech research. *Journal of the Acoustical Society of America* 69 (5): 1462-1464.

Repp, B.H. 1985. Can linguistic boundaries change the effectiveness of silence as a phonetic cue? *Journal of Phonetics* 13: 421-431.

Repp, B. H. 1992. Perceptual restoration of a missing speech sound - auditory induction or illusion? *Perception and Psychophysics* 51 (1): 14-32.

Rosner, B.S. and J.B. Pickering. 1994. *Vowel Perception and Production.* Oxford: Oxford University Press.

Rouco, A. and D. Recasens. 1996. Reliability of electromagnetic midsaggital articulometry and electropalatography data acquired simultaneously. *Journal of the Acoustical Society of America* 100 (5): 3384-3389.

Samuel, A. G. 1981. The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology - Human perception and performance* 7 (5): 1124-1131.

Samuel, A. G. 1996a. Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology - General* 125 (1): 28-51.

Samuel, A. G. 1996b. Phoneme restoration. *Language and Cognitive Processes* 11 (6): 647-653.

Schönle, P. W., K. Gräbe, P. Wenig, J. Höhne, J. Schrader and B. Conrad. 1987. Electromagnetic articulography: use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language* 31: 26-35.

Scott, D. 1984. More on the /t/:/d/ distinction in American alveolar flaps. Paper read at 107th meeting of ASA, Norfolk, Virginia.

Slater, A. and J. Coleman. 1996. Non-segmental analysis and synthesis based on a speech database. *Proceedings of the 4th International Conference on Spoken Language Processing* 4: 2379-2382.

Spencer, A. 1996. *Phonology.* Oxford: Blackwell.

Sproat, R. and O. Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21: 291-312.

Stevens, K. 1989. On the quantal nature of speech. *Journal of Phonetics* 17: 3-45.

Stevens, K. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Stevens, K. and S. Blumstein. 1994. Attributes of lateral consonants. *Journal of the Acoustical Society of America* 95 (5): 2875.

Stevens, K. and A. House. 1963. Perturbation of vowel articulations by consonantal context: an acoustical study. *Journal of Speech and Hearing Research* 6 (2): 111-128.

Stevens, K., S. Keyser, and H. Kawasaki. 1986. Toward a phonetic and phonological theory of redundant features. In *Invariance and variability in speech processes*, edited by J. Perkell and D. Klatt. Hillsdale, NJ.: Lawrence Erlbaum Associates. 426-449.

Strik, H. and L. Boves. 1990. A dynamic programming algorithm for time-aligning and averaging physiological signals related to speech. *Journal of Phonetics* 19: 367-378.

Trudgill, P. 1990. *The dialects of England*. Oxford: Blackwell.

Tuller, B., S. Shao and J.A. Scott Kelso. 1990. An evaluation of an alternating magnetic field device for monitoring tongue movements. *Journal of the Acoustical Society of America* 88 (2): 674-679.

Tunley, A. 1999. *Coarticulatory influences of liquids on vowels in English*. Ph.D. thesis, Department of Linguistics, University of Cambridge.

van Santen, J. H. P. 1997. Segmental duration and speech timing. In *Computing Prosody*, edited by Y. Sagisaka, N. Campbell and N. Higuchi. New York: Springer-Verlag. 225-249.

Walsh Dickey, L. 1997. *The Phonology of Liquids*. Amherst: GLSA, University of Massachusetts.

Warren, R. M. 1996. Auditory illusions and perceptual processing of speech. In *Principles of Experimental Phonetics*, edited by N. J. Lass. St Louis: Mosby. 435-466.

Warren, R. M. and C. J. Obusek. 1971. Speech perception and phonemic restorations. *Perception and Psychophysics* 9 (3B): 358-362.

Warren, R. M. and G. L. Sherman. 1974. Phonemic restorations based on subsequent context. *Perception and Psychophysics* 16 (1): 150-156.

Weismer, G. and K. Bunton. 1999. Influences of pellet markers on speech production behaviour: Acoustical and perceptual measures. *Journal of the Acoustical Society of America* 105 (5): 2882-2894.

Wells, J. C. 1982. *Accents of English*. Cambridge: Cambridge University Press.

Westbury, J. R. 1991. The significance and measurement of head position during speech production experiments using the x-ray microbeam system. *Journal of the Acoustical Society of America* 89 (4): 1782-1791.

Westbury, J. R. 1994. On coordinate systems and the representation of articulatory movements. *Journal of the Acoustical Society of America* 95 (4): 2271-3.

Westbury, J. R., M. Hashi and M. J. Lindstrom. 1995. Differences among speakers in articulation of American English /r/: An x-ray microbeam study. *Proceedings of the 13th International Congress of Phonetic Sciences* 4: 50-57.

Whalen, D. 1990. Coarticulation is largely planned. *Journal of Phonetics* 18: 3-35.

Whalen, D. 1991. Perception of the English /s/ - /ʃ/ distinction relies on fricative noises and transitions, not on brief spectral slices. *Journal of the Acoustical Society of America* 90 (4 Part 1): 1776-1785.

Woods, A., P. Fletcher and A. Hughes. 1986. *Statistics in language studies*. Cambridge: Cambridge University Press.

Wrench, A. 1998. Manuscript. *Report on the Speech Production Research Facility.* Edinburgh: Queen Margaret University College.