LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

**Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin**

**Copyright statement**

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

**Liability statement**

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

**Access Agreement**

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

# Analysis of Lingual Articulation using Electropalatography and Electromagnetic Articulography

Liam Fitzpatrick

Centre for Language and Communication Studies

University of Dublin

Trinity College

[illegible]

Submitted for the degree of Ph.D.

October 2000.

TRINITY COLLEGE
06 JUN 2001
LIBRARY DUBLIN
Thesis
6178

# Declaration

I, the author, declare that this thesis, unless otherwise stated, is my own work. It has not been submitted for a degree at any other University.

Signature of Author

Liam Fitzpatrick

.... October, 2000.

## DECLARATION

I give permission for the Library to lend or copy this thesis upon request.

Signed

LIAM J FITZPATRICK
1 MAY 2000.

## Summary

Electromagnetic mid-sagittal articulography (EMA) and electropalatography (EPG) are two techniques for measuring lingual activity during speech production. EMA measures the location of a small number of sensors attached to the midline tongue, and EPG detects contact between the tongue and sensors embedded in an artificial palate which fits over the hard palate. Data from each device holds more information then analytical techniques in general use are capable of extracting. They also provide complementary information. Simultaneous use of the techniques would provide more information than either used alone could. This thesis, through the development of tools and analytical techniques, addresses the better exploitation of EPG and EMA data, and investigates their simultaneous use. The techniques are used to addresses some basic issues in phonetic theory by undertaking two exploratory analytical studies.

In the first study, assimilatory patterns for [ t̪ʸ t̪ʲ c k ] in Irish were examined for one speaker in $VC_1\#C_2V$, using EPG data. Categorisation of assimilations were based on a technique, called differential contact pattern templates, which was developed for this thesis. Palatalised and palatal stops were much more prone to assimilation than velarised and velar stops, and secondary articulation generally resisted assimilation. Many, but not all, effects may be explained in terms of articulatory phonology. Some effects my be explained by mechanical lingual constraints. Segments involving contiguous articulators assimilate more readily, and follow a different assimilation route than the non-contiguous. Fewer assimilations occur when the tongue gesture required for the cluster follows the "preferred" forward trajectory.

The second study concerns Fant's nomograms, which explore the relationship between articulatory settings and acoustic consequences for vowel type productions using an articulatory model. EMA, EPG and acoustic data were obtained for a phonetician attempting to replicate Fant's nomograms for rounded and unrounded high vowel continua. Results are compared with the most closely corresponding nomogram data. One difference was the constancy of the resonance primarily associated with the back cavity, which contrasted with the rising frequency of Fant's nomograms. A more rapid fall in the resonance primarily associated with the front cavity was also seen for articulations in the pre- to midpalatal region. The simplifying assumption of the articulatory model which retains a constant constriction length and shape regardless of stricture location may represent an impossibility in human speech, where constriction length varies considerably. Divergences from Fant's nomograms

may also be due to adjustments such as vertical larynx movement, which speakers can control but which may covary with changes in stricture location.

The EPG and EMA data from the second study was further investigated to evaluate the advantages and potential problems of combining the techniques. The location of the lingual constriction, estimated from EMA, was compared to the narrowest row on corresponding EPG frames. Broad agreement was observed over a range of articulations where the location of the lingual constriction was neither close to nor behind the most posterior row of the EPG electrodes. A further comparison was made where the EMA estimate of constriction location was compared to a EPG based tongue contour estimate which was drawn on the basis of the locations of activated and inactivated electrodes. This was useful in enhancing the understanding of differences in the nature of the data from the two devices, and drawing attention to three dimensional aspects of articulation. However it proved disappointing as a method of estimating the location of the lingual constriction. A third comparison was made between the EMA estimate of constriction location and the EPG centre of gravity index. While the centre of gravity produced similar results over a range of articulations it was observed that the index was not effective for constrictions located close to the back edge of the palate.

Both studies showed that the techniques developed for the thesis provide advantages in the study of lingual activity using EPG and EMA. They provide methodological advantages in the calculation of results and enhance the presentation of both basic data and of results. They facilitated the two studies in this thesis where interesting theoretical questions were addressed. They also provided indications as to how such work aimed at comprehensively addressing the simultaneous analysis of data from the two devices might proceed. These techniques and tools should have a wider applicability than these studies. Techniques such as the tongue circle model should be of use in similar studies, while contact pattern templates, viewing tools and so on should have a more general applicability.

## Acknowledgements

I wish to acknowledge my gratitude to Ailbhe Ní Chasaide for discussions, insights, support and supervision that she has provided to me throughout this work. Thanks also to Christer Gobl, Francis Nolan, and Tine Mooshammer for their valuable insights and suggestions. I am most grateful to Phil Hoole, for his careful work in making many of the recordings used in this thesis and his help in understanding the data, and to Francis Nolan and Noirín Ní Fheinne for being the subjects of those recordings.

Many thanks to Carolyn, Jeff, Tine for proof-reading. And special thanks to all my friends who supported me throughout this endeavour and discouraging me from conceding defeat. They include especially my parents, Ann and William; my siblings Deirdre, John, Mary, and Paul; colleagues Aevanne, John(s), Mike, Peter, and Rosemary and from the Phonetics Laboratory; colleagues from Programming Research Ltd.; those who applied various forms of "encouragement", including Fergal and Teresa, Emma, Ita and Steven; and, certainly not least, Daniëlle.

The research in this thesis was supported by ESPRIT/Basic Research Actions ACCOR I and ACCOR II.

## Contents

**SUMMARY** ..... **I**

**ACKNOWLEDGEMENTS** ..... **III**

**CONTENTS** ..... **IV**

**TABLE OF FIGURES** ..... **VII**

**TABLE OF TABLES** ..... **XIV**

**PART I - INTRODUCTION AND BACKGROUND** ..... **1**

**CHAPTER 0 INTRODUCTION** ..... **2**

0.1 INTRODUCTION ..... 2
0.2 RESEARCH QUESTIONS ..... 3
0.3 OUTLINE OF THESIS ..... 4

**CHAPTER 1 VOCAL TRACT - MEASURING AND MODELLING** ..... **5**

1.1 WHY MEASURE SPEECH? ..... 5
1.2 HOW CAN SPEECH BE MEASURED? ..... 5
1.3 MODELS OF SPEECH ARTICULATORS ..... 6
*1.3.1 Modelling tongue shape* ..... *6*
1.4 MEASURING VOCAL TRACT GEOMETRY ..... 9

**CHAPTER 2 ELECTROPALATOGRAPHY AND ELECTROMAGNETIC-ARTICULOGRAPHY - MEASURING TONGUE MOVEMENT** ..... **13**

2.1 INTRODUCTION TO ELECTROPALATOGRAPHY ..... 13
2.2 INTRODUCTION TO ELECTROMAGNETIC-ARTICULOGRAPHY ..... 15
2.3 COMPARING EPG AND EMA ..... 16

**CHAPTER 3 SPEECH PRODUCTION - ARTICULATORY AND ACOUSTIC RELATIONSHIPS** ..... **18**

3.1 AN ACOUSTIC MODEL OF SPEECH PRODUCTION ..... 19
3.2 ACOUSTIC NOMOGRAMS: REPRESENTING THE RELATIONSHIPS BETWEEN FORMANT FREQUENCIES AND ARTICULATIONS ..... 20
3.3 MEASURING NOMOGRAMS FOR HUMAN SPEECH ..... 27

**CHAPTER 4 COARTICULATION, ASSIMILATION, AND CONNECTED SPEECH** ..... **30**

4.1 COARTICULATION AND ASSIMILATION ..... 30
4.2 THEORIES AND MODELS OF COARTICULATION ..... 32

**PART II - EXPERIMENTAL METHODS FOR ELECTROPALATOGRAPHY AND ELECTROMAGNETIC-ARTICULOGRAPHY** ..... **36**

**CHAPTER 5 ELECTROPALATOGRAPHY (EPG)** ..... **37**

5.1 GENERAL EXPERIMENTAL TECHNIQUE ..... 37
5.2 MEASURING CONTACT LOCATION ON PALATES ..... 38
*5.2.1 Examples of existing practice* ..... *38*
*5.2.2 Measurements in this study* ..... *39*
*5.2.3 Best method developed for this work* ..... *40*
5.3 EPG DATA REPRESENTATION ..... 41
*5.3.1 Enhanced Representation of Single Frames* ..... *42*
*5.3.2 Reduced Representation of a Sequence of Frames* ..... *48*

**CHAPTER 6 ELECTROMAGNETIC-ARTICULOGRAPHY (EMA)....58**
6.1 General Experimental Technique....58
6.2 EMA Data Representation....59
*6.2.1 Single Point Analysis....60*
*6.2.2 Multiple Point Analysis....62*
6.3 Conclusion....66

**CHAPTER 7 COMBINING EPG AND EMA....67**
7.1 General Experimental Technique....67
*7.1.1 Guiding location of EMA sensors on the tongue....67*
*7.1.2 Alignment of the Measurements of the EPG palate and the EMA co-ordinate system....68*
*7.1.3 Interference between EMA and EPG....70*
7.2 Combined Data Representation....71
*7.2.1 Two dimensional representation....71*
*7.2.2 Three dimensional representation....76*

**PART III - ANALYTICAL STUDIES....80**

**CHAPTER 8 AN EPG STUDY OF ASSIMILATION OF IRISH LINGUAL STOPS....81**
8.1 Introduction....81
8.2 The Consonantal System of Irish....82
8.3 Data Acquisition....83
*8.3.1 Materials....83*
*8.3.2 Method....84*
8.4 Results....89
*8.4.1 The prepausal stops: VC#....89*
*8.4.2 The homorganic clusters: VC#CV....91*
*8.4.3 Differential contact pattern templates: $VC_1\#C_1V$ vs. $VC_2\#C_2V$....93*
*8.4.4 Non-homorganic clusters: $VC_1\#C_2V$....97*
8.5 Conclusions....118

**CHAPTER 9 A STUDY OF 'HUMAN NOMOGRAMS'....120**
9.1 Introduction....120
9.2 Data Acquisition....120
*9.2.1 Materials....120*
*9.2.2 Method....121*
*9.2.3 Segmentation....123*
*9.2.4 Representation....124*
9.3 Analysis From EMA and Acoustic Data....125
*9.3.1 Movement of the Two Rearmost Tongue Coils....125*
*9.3.2 Articulatory-Acoustic Correlation....128*
*9.3.3 Lingual Constriction....132*
*9.3.4 Within Utterance Variability....134*
9.4 Discussion....135

**CHAPTER 10 ADDITIONAL INFORMATION FROM EPG IN THE STUDY OF 'HUMAN NOMOGRAMS'....141**
10.1 Introduction....141
10.2 Differences in measurements by EMA and EPG....142
10.3 Comparison of EMA Tongue Circle Model and EPG Data....143
*10.3.1 EPG Top views....143*
*10.3.2 EPG Derived Tongue Contours (Side Views)....145*
*10.3.3 EPG Centre of Gravity Index....149*
10.4 Discussion....152

**PART IV - CONCLUSIONS** ..... **80**

**CHAPTER 11 DISCUSSION AND CONCLUSIONS** ..... **156**

11.1 INTRODUCTION ..... 156
11.2 MOTIVATION ..... 156
11.3 ACHIEVEMENTS ..... 156
11.4 FURTHER WORK ..... 159

**REFERENCES** ..... **161**

## Table of Figures


Figure 1.1 The major muscle groups of the tongue. (Fig. 24 and Fig. 25 from Hardcastle, 1976). ..... 7

Figure 1.2 An outline of the model presented in Mermelstein, 1973, indicating the parameters to that model. The principal positions indicated are: J (jaw), L (lips), B (tongue blade/body join), C (centre of tongue body), V (velum), H (hyoid), K (larynx). ..... 8

Figure 2.1 (a) A palatogram for a dental affricate of Beijing Mandarin where dental contact is indicated by the dark band close to the teeth (from Lee, 1999). In (b) a representation from this thesis for an EPG contact pattern of a dental stop of Irish is shown, where the approximate area of contact is shaded grey and electrode locations shown as small circles. ..... 13

Figure 2.2 a) the distribution of electrical contacts on a Reading EPG palate, viewed from below, for a particular adult female, and b) a schematic layout typically used to present the contact palatogram. ..... 14

Figure 2.3 Typical layout of transmitter and receiver coils for the MoveTrak electromagnetic articulography system. ..... 15

Figure 3.1 A schematised representation of the source-filter model of speech production for a voiced, non-nasalised vowel. (From Monahan, 1996). ..... 19

Figure 3.2 Top left: vocal tract outline traced from x-ray for [i]. Mid left: cross section area of the outline. Bottom left: cross section modelled as many tubular sections (solid lines) and as three sections (dashed lines). Right: cross sections at points along the vocal tract corresponding to numbered lines in the top left part of this figure. (Adapted from Fant, 1970, Fig 2.2-5). ..... 21

Figure 3.3 The four tube model in Badin et al., 1988, showing pharynx cavity: A4=8 $cm^2$, l4 varying ; tongue constriction: A3=0.65 $cm^2$, l3=5 cm ; mouth cavity: A2=8 $cm^2$, l2 varying; lips: A1=0.16 $cm^2$, 4 $cm^2$, or no lips, l1=1 cm. The constriction centre co-ordinate $X_c$ (measured from the glottis position) can vary from -2.5 to 17.5 cm, keeping l2 + l3 + l4 = 15cm. (Adapted from Badin et al., 1988). ..... 22

Figure 3.4 Nomograms for the different cavities of the model in Figure 3.3: (a) back cavity : pharynx (cs-area=8$cm^2$; length varying) + tongue constriction (cs-area=0.65$cm^2$; length=5cm) (b) front cavity : mouth (cs-area=8$cm^2$; length varying) + lip constriction (cs-area=0.16$cm^2$ : dotted-line, 4$cm^2$ : solid-line, or no lips : dashed-line; length=1cm) (c) tongue constriction (cs-area=0.65$cm^2$; length=5cm). $X_c$ is the distance between the constriction centre and the glottis. $X_c$ is the constriction centre co-ordinate, measured from the glottis position. (Adapted from Badin et al., 1988). ..... 23

Figure 3.5 Resonances of individual cavities in Figure 3.4 drawn as dashed lines, overlaid on the overall system response drawn as dark lines. The lip constriction cs-area=4 $cm^2$. (From Badin et al., 1988). ..... 24

Figure 3.6 Four tube model of the vocal tract during high vowel production. $l_1$ is the length of the lip tube, $l_2$ the front cavity tube, $l_3$ the lingual constriction tube, and $l_4$ the back cavity tube. x is the location of the centre of the lingual constriction. (Adapted from Fant, 1970 , Fig. 1.4-8.) ..... 25

Figure 3.7 Nomograms for 4-tube model in Figure 3.6, plotting the first five resonances of the model for different values of x. Five nomograms are presented for five

different lip tube settings: $l_1$ = (0 cm for curve 1, 1 cm for curve 2-5), $A_1$ = 8 $cm^2$, 4 $cm^2$, 2 $cm^2$, 0.65 $cm^2$, 0.16 $cm^2$ for curves 1 to 5 respectively. The length of the lingual constriction tube and its cross section area is kept constant ($l_3$ = 5 cm, $A_3$ = 0.65 $cm^2$). (From Fant, 1970, Fig. 1.4-9.).....26

Figure 3.8 Nomograms corresponding to one set of vowels produced by two speakers (adapted from Ladefoged & Bladon, 1982, Figure 6).....29

Figure 5.1 "Structured light" used horizontal bands of light projected onto a curved surface (subject). These bands appear distorted when the surface is observed from another angle (camera view). The shape of the surface can be determined from these distortions. (From Shadle & Carter, 1993.) .....39

Figure 5.2 Photographs of a view from the side and a view from below of an EPG palate..40

Figure 5.3 illustrates a method of measuring the locations of the electrodes on the EPG palate in 3 dimensions. Distance front-to-back (Y) and sideways (Z) is measured from the image on the left. Depth (X) is measured using a callipers protruded through a sheet of Perspex as shown in the image on the right. The Perspex is balanced on three points on the plastercast which holds the EPG palate, as marked in the left-hand image.....41

Figure 5.4 Standard representation of Reading EPG contact patterns (from the Edit software). The leftmost shows the representation used for printing. The middle and rightmost are used for on-screen displays. .....43

Figure 5.5 Several representations of electrode contact patterns developed: (a) traditional rectangles, (b) rectangles scaled to measured palate, (c) circles, (d) simple 3D display, (e) 3D display with contact region. .....44

Figure 5.6 A sample view of the EPG analysis software developed. Along with a view of the entire waveform for an utterance (top panel) and an expanded section of the waveform (lower left), three representations of the EPG contact pattern are shown: a traditional contact pattern as in Figure 5.5 (a), positioned electrodes when viewed from the side as in a side-view version of Figure 5.5 (c), and a rotatable 3D view as in Figure 5.5 (e).....45

Figure 5.7 A number of representations of frequency of contact for each electrode on the palate. The frequency of contact for an electrode is indicated by, left to right: a numerical percentage, the shade of a circle, the size of a circle, or the shade of a rectangle.....46

Figure 5.8 Frequency of contact patterns for five repetitions of [-aṯ$^{ɣ}$#ṯ$^{ɣ}$a-] over the range from 15 ms before closure to 15 ms after closure. The percentage of repetitions in which there is contact with an electrode is indicated by the colour of that electrode. .....47

Figure 5.9 Representation of the area of lingual contact (left) and highest lingual contact (right) based on frequency of contact of five repetitions of [-at̲$^{ʲ}$].....48

Figure 5.10 Total number of electrodes activated on each of the eight EPG rows for the speech fragment [#t$^{ɣ}$a]. .....51

Figure 5.11 EPG Electrodes grouped into three contact regions - dorsal, palatal, and coronal. (From Barry, 1992.) .....52

Figure 5.12 Contact pattern templates for contact found in all repetitions (top row) and in any repetitions (bottom row) for five repetitions of [-aṯ$^{ɣ}$#ṯ$^{ɣ}$a-]. Frequency of contact displays for these data are shown in Figure 5.8.....55

Figure 5.13 Contact pattern templates for contact found in all repetitions (top row) and in any repetitions (bottom row) for five repetitions of [-aʲ#ʲa-].....55

Figure 5.14 Differential contact pattern templates for [-aˠ#ˠa-] and [-aʲ#ʲa-]. If an electrode is coloured green it is found in all repetitions of [-aˠ#ˠa-] and none of [-aʲ#ʲa-]. If an electrode is coloured blue it is found in all repetitions of [-aʲ#ʲa-] and none of [-aˠ#ˠa-].....56

Figure 5.15 Comparative contact patterns for two repetitions of [-aˠ#ʲa-] with the differential contact pattern template for [-aˠ#ˠa-] and [-aʲ#ʲa-]. The upper repetition is an example with the required activation for the template [-aˠ#ˠa-] and none for template [-aʲ#ʲa-]. The lower repetition does not have the required activation for [-aˠ#ˠa-] but shows most of the activation required for the template [-aʲ#ʲa-].....57

Figure 6.1 Typical experimental setup for recording EMA and, optionally, EPG. The T-bar is used for calibration and is not present throughout the recording. A helmet (not shown) holding three transmitter coils is placed on the speaker's head.....59

Figure 6.2 Three curves show the trajectory of three lingual EMA sensors during the production of / əjə /, with small arrow indicating the direction of movement. Straight lines join the locations of the three sensors at five different time points: during first vowel (V1), acoustic closure (AC), mid-closure (MID), acoustic release (AR), and during second vowel (V2).....62

Figure 6.3 Examples of the match between tongue contours measured using ultrasound and estimated from simultaneously recorded EMA sensor positions. EMA receiver sensors are marked as crosses. (From Honda & Kaburagi, 1993).....64

Figure 6.4 The circle model of the tongue is shown overlaid for two time-points during one speech segment. Each circle of 2 cm radius is fixed to two EMA sensors on the tongue body. The trajectories of EMA sensors over an utterance are shown, as is the corresponding trajectory of the centre of the circle model. Also shown is the curve of the roof of the mouth.....65

Figure 7.1 An example, using software introduced in Figure 5.6, of a recording where it seems that EMA sensors are blocking contact with EPG electrodes at the place indicated.....70

Figure 7.2 One of the physiological data viewers developed. Top left is the acoustic signal; below that are two EPG displays; below that is a display of the EMA (x, y) trajectories for the EMA sensors displayed as raw movement in the 6 windows to the right (amplitude of EMA signals is in 1/100 mm).....72

Figure 7.3 A viewer which shows the acoustic signal in the top window; overlaid EMA (x, y) and EPG palate as seen from the side in the middle window; and the bottom window the EPG contact pattern (viewed from below), and zoomed up displays of the EMA and acoustic data. The lower two displays are aligned to a marker in the top window.....73

Figure 7.4 Representation of tongue contours for EMA with tongue circle model (a) and for EPG (b) for a single time-point. (a) and (b) are overlaid in (c). EMA x-y movement for entire [i–ɯ] indicated, along with that for "pseudo-sensor" for tongue circle centre. Rotated top views for EPG also shown.....75

Figure 7.5 A viewer for EMA, EPG , and acoustic information that incorporates simple representations of that data (e.g. sensor locations, EPG contact patterns,

formants) along with estimates from that data (e.g. location of narrowest lingual constriction, EPG contact region. ....76

Figure 7.6 A 3 dimensional representation of EMA and EPG added to the bottom-right of the viewer presented in Figure 7.3. ....77

Figure 7.7 A time-point where the midline EMA data and the EPG contact are at different heights at the line marked (A). ....78

Figure 7.8 The EPG palate representation from the lower right panel of Figure 7.7 when rotated slightly. The same line (A) is marked. ....78

Figure 7.9 A sketch of the cross-section of the vocal tract at the point marked (A) in Figure 7.7, showing the upper contour of the tongue and the contour of the palate. ....79

Figure 8.1 Photographs of the EPG artificial palate used in this study showing the vertical view (left) and side view (right). EPG Row 1 and EPG Row 8 is marked. The scale of the photographs is shown in centimetres. ....85

Figure 8.2 EPG frequency of contact patterns (left) and approximate tongue contours (right) for stops in VC# at the time-point of closure and at maximum contact. ..90

Figure 8.3 Average positions of EMA sensors (1) and (2) for [ t̪ʸ ] and [ t̪ʲ ] at closure, with representations of the tongue tip impressionistically sketched. ....91

Figure 8.4 Frequency of contact for homorganic clusters, VC#CV, of the stops at closure and for the interval 40 ms before closure to 25 ms after closure. The frequency of contact for an electrode is indicated by the colour of a rectangle representing that electrode, where the range the colour indicates is given in the legend on the right. ....92

Figure 8.5 Differential contact pattern templates for [ t̪ʸ ] vs. [ t̪ʲ ] (top), [ t̪ʸ ] vs. [ c ] (middle), [ t̪ʲ ] vs. [ c ] (bottom) based on the frequency of contact for the corresponding non-homorganic allophones shown in Figure 8.4. Archetypal contact for a particular consonant is where a contact is activated in 100% of the repetitions of that consonant and 0% of the repetitions of the other consonant. ....95

Figure 8.6 Differential contact pattern templates for [ t̪ʸ ] vs. [ k ] (top), [ t̪ʲ ] vs. [ k ] (middle), [ c ] vs. [ k ] (bottom) based on the frequency of contact for the corresponding non-homorganic allophones shown in Figure 8.4. Archetypal contact for a particular consonant is where a contact is activated in 100% of the repetitions of that consonant and 0% of the repetitions of the other consonant. ....96

Figure 8.7 t̪ʸ.vs. t̪ʲ template (top row) and comparative contact patterns for five repetitions of Vt̪ʸ#t̪ʲV. ....101

Figure 8.8 t̪ʸ.vs.c template (top row) and comparative contact patterns for five repetitions of Vt̪ʸ#cV. ....102

Figure 8.9 t̪ʸ.vs.k template (top row) and comparative contact patterns for five repetitions of Vt̪ʸ#kV. ....103

Figure 8.10 t̪ʲ.vs.t̪ʸ template (top row) and comparative contact patterns for five repetitions of Vt̪ʲ#t̪ʸV. ....104

Figure 8.11 t̪ʲ.vs.c template (top row) and comparative contact patterns for five repetitions of Vt̪ʲ#cV. ..... 105

Figure 8.12 t̪ʲ.vs.k template (top row) and comparative contact patterns for five repetitions of Vt̪ʲ#kV. ..... 106

Figure 8.13 c.vs.t̪ˠ template (top row) and comparative contact patterns for five repetitions of Vc#t̪ˠV. ..... 107

Figure 8.14 c.vs.t̪ʲ template (top row) and comparative contact patterns for five repetitions of Vc#t̪ʲV. ..... 108

Figure 8.15 c.vs.k template (top row) and comparative contact patterns for five repetitions of Vc#kV ..... 109

Figure 8.16 k.vs.t̪ˠ template (top row) and comparative contact patterns for five repetitions of Vk#t̪ˠV. ..... 110

Figure 8.17 k.vs.t̪ʲ template (top row) and comparative contact patterns for five repetitions of Vk#t̪ʲV. ..... 111

Figure 8.18 k.vs.c template (top row) and comparative contact patterns for five repetitions of Vk#cV ..... 112

Figure 8.19 Trajectory of the rearmost EMA sensor for a repetition of Vc#cV and of Vk#kV (where V = / a /). The small circles mark the sensor position at the time of closure, and the arrowheads mark the sensor position at the last time-point shown. ..... 118

Figure 9.1 Photographs of a view from the side and a view from below of the speaker's EPG palate. ..... 122

Figure 9.2 Summary of the circle model, discussed in 6.2.2, with the three EMA sensors on the tongue used in this experiment shown as black dots. The position of the circle, which represents the tongue body, is controlled by position of the rearmost sensors. The closest point between this circle and the roof of the mouth can be located, and is taken to equate to the narrowest point of lingual constriction. The distance to the teeth from the narrowest point of the lingual constriction is illustrated. ..... 124

Figure 9.3 Horizontal vs. vertical movement of the two rearmost tongue sensors for each repetition of (A) unrounded utterances [ɯ–i] and [i–ɯ], (B) rounded utterances [u–y] and [y–u]. The same movements are again shown in (C) for back to front vowel sequences and in (D) for front to back vowel sequences. ..... 126

Figure 9.4 EMA sensor positions and associated centre of the circle model for a repetition of the continuum [i–ɯ]. A time-point is marked by the line through the EMA sensor positions. The corresponding circle calculated using the rearmost sensors is also shown. ..... 127

Figure 9.5 F2 and F3 values vs. distance from the narrowest point of lingual constriction to the teeth for all the rounded utterances. Dark lines show formant traces for lip cross sectional areas of 4 $cm^2$ and 2 $cm^2$ from Fant's nomograms. Mid-sagittal contour of roof of mouth from EPG and x-ray shown above. EPG electrodes are shown as circles. Slanted lines connect x-axis grid to points on

this contour. The extent of the lingual constriction and approximate point of F2-F3 proximity is indicated above the contour. ..... 129

Figure 9.6 F2 and F3 values vs. distance from the narrowest point of lingual constriction to the teeth for all the unrounded utterances. Dark lines show formant traces for lip cross sectional areas of 0.65 $cm^2$ and 0.16 $cm^2$ from nomograms in Fant, 1970. Mid-sagittal contour of roof of mouth from EPG and x-ray shown above. EPG electrodes are shown as circles. Slanted lines connect x-axis grid to points on this contour. The extent of the lingual constriction and approximate point of F2-F3 proximity is marked above the contour. ..... 130

Figure 9.7 A spectrogram of one repetition of [iɯ]. Note that the formant F2 before 3.8 s, and formant F3 thereafter, are maintained at a relatively steady frequency of 2.1 kHz. ..... 131

Figure 9.8 The midline distance between the tongue and the roof of the mouth at the point of narrowest lingual constriction plotted against the distance of this constriction to the teeth for all utterances. Mid-sagittal contour of roof of mouth and EPG electrodes shown above, as for Figure 9.5 ..... 132

Figure 9.9 Estimated cross section area at the point of narrowest lingual constriction plotted against the distance from this constriction to the teeth for all utterances. The cross-section area is estimated using data from Baer et al., 1991, and measured data in Figure 9.8 to estimate the cross-section area for each utterance. The dark black line is the cross-section area used in Fant's nomograms (0.65 $cm^2$). Mid-sagittal contour of roof of mouth and EPG electrodes shown above, as for Figure 9.5. ..... 134

Figure 9.10 EMA tongue sensor locations for two repetitions of [u–y], where the jagged lines represent the sensor (x,y) locations for the entire repetition. (a) An example with rather steady changes of vertical and horizontal position. (b) An example where vertical movement changes part way through the utterance. ..... 135

Figure 9.11 Possible vocal tract shapes for [i] in solid curves and for [ɯ] in dashed curves. The midline of the back cavity for each vowel is marked with the arrow-headed lines ..... 136

Figure 9.12 vocal tract shapes from Maeda's model for [ɯ] (top) and for [i] (bottom). The midline of the back cavity is marked with arrow-headed lines. ..... 137

Figure 9.13 Modelled data showing the narrowest point of constriction (X) and its approximate length (shaded in grey). Upper panel shows X at c. 3.5 cm from the teeth and the lower panel shows X at c. 4.5 cm. Also shown are the trajectories for the three tongue sensors, the lower jaw and lip sensors, and the centre of the circle model, for a single unrounded vowel continuum. EPG contact patterns are also shown ..... 139

Figure 10.1 A replica of Figure 9.6 is shown, with a mid-sagittal contour of roof of mouth from EPG and x-ray shown above. Slanted lines connect x-axis gridlines to points on this contour. EPG electrodes are shown as circles. Corresponding EPG top views are shown in the upper part of the figure from a single repetition of [i–ɯ]. On the top views the row with greatest EPG narrowing is coloured grey, and the location of the EMA constriction estimate is marked with an 'x'. ..... 144

Figure 10.2 Representation of tongue contours for EMA with tongue circle model and for the EPG derived tongue contours for three successive time-points from a single repetition of the [i–ɯ] continuum. The time-points are labelled (a), (b), and (c). EMA x-y movement of all sensors for the entire [i–ɯ] are indicated,

along with that for "pseudo-sensor" for tongue circle centre. Rotated top views for EPG also shown. .....147

Figure 10.3 The midline of the vocal tract for the vowel [i] on the left side, and three cross section outlines on the right side for the points labelled (1), (2) and (3). (Adapted from Fant, 1970, Fig 2.2-5). .....149

Figure 10.4 Constriction location estimated on the basis of EPG Centre of Gravity (vertical axis) plotted against the EMA circle model estimate for each individual utterance. The thick black line is the overall trend line for all utterances.....151

Figure 10.5 A replica of Figure 10.1 is shown with an 'O' overlaid on the EPG top views to indicate the EPG centre of gravity value for the EPG contact patterns. .....152

## Table of Tables

Table 8.1 The consonant system of Irish (on the basis of Cooper, 1994).....83

Table 8.2 Classifications of the type of anticipatory coarticulation observed for the non-homorganic clusters.....87

Table 8.3 Number and type of assimilations found for all clusters at the first EPG frame of consonantal closure or, in the case of an initial fricative being produced, the first EPG frame point of maximum fricative-like narrowing. The classifications are given in Table 8.2.....113

# Part I - Introduction and Background

# Chapter 0 Introduction

## *0.1 Introduction*

As the most complex organ involved in shaping the vocal tract during speech production, and as an important contributor to that shape, the tongue is worthy of being studied in our attempts to understand how speech is produced by humans. Unfortunately the tongue is not generally easy to observe during speech. As a result much of the earliest literature on lingual activity was based on introspection by phoneticians on their own tongue movement. However there is an ever evolving development of experimental techniques which make measurements of lingual activity possible.

Experimental techniques for measuring lingual activity fall largely into two groups: invasive and non-invasive (Stone, 1997). Non-invasive techniques are generally imaging techniques, where measurements are made from images of the tongue. Examples of such techniques include the use of x-ray, ultrasound, or magnetic resonance imaging (MRI). Invasive techniques generally involve introducing sensors into the mouth to directly measure some aspect of lingual activity. Examples of such techniques include point-tracking devices, such as electromagnetic mid-sagittal articulography (EMA), where sensors attached to points on the tongue are used to measure the location of those points, and electropalatography (EPG), where sensors embedded in an artificial palate detect lingual contact with the hard palate.

Imaging techniques mentioned above have the advantages of not requiring the placing of anything directly within the vocal tract, and thus they do not directly affect speech production, and of showing considerable detail about tongue shape. However invasive techniques can typically provide more measurements per second which is important for investigations of dynamic aspects of tongue movement.

This thesis focuses on two invasive techniques: EMA and EPG. These techniques measure different aspects of lingual articulation. EMA measures the location of a small number of points that are strictly on the mid-sagittal plane, whereas EPG measures contact with the hard palate which is three dimensional in nature. As EMA and EPG only measure certain aspects of lingual activity the analysis of the data always requires some degree of interpretation. To do so, ideally, all the information provided by each technique would be used, but the general analytical techniques presented in the literature show that neither is exploited to the full. Often analysis of EMA data focuses on individual sensors. And though it has been recognised that the exact anatomical details of the speaker's hard palate those details should be taken into account in the analysis of EPG data (Jones & Hardcastle, 1995), the general analysis

techniques for EPG data ignore such information. Potential for exploitation of the data extends considerably if data from the two devices is available simultaneously. Though both of these devices are well established laboratory techniques, they have rarely been used simultaneously. With simultaneous use the complementary nature of the data would provide better measurements of lingual activity than is available from either device alone.

## *0.2 Research questions*

With the aim of improving upon the current analytical possibilities of the EMA and EPG techniques, the research presented here addresses three principal questions. Firstly, how can more information be extracted from EPG data? Secondly, how can more information be extracted from EMA data? And thirdly, how can EPG and EMA be used as complementary devices? A parallel aim was to address some basic issues in phonetic theory, using these techniques to re-examine aspects of assimilation and articulatory/acoustic relationships in speech production.

Given the dual aims, the approach taken on each question posed above follows two processes. Firstly new analytical techniques and tools are developed to address types of analysis that are either not possible with current techniques described in the literature, or where improvements to current techniques are considered necessary. Secondly, two exploratory analytical studies are undertaken where analysis of the data would benefit from, or indeed require, improved analytical tools. The two processes proceeded in parallel. The development of tools is at all times undertaken within the context of trying to maximise the information yield for researchers whose primary objective is trying to understand the speech production process. At the same time, as tools evolved they allowed additional questions to be asked in the analytical studies.

The first analytical study is a re-analysis of data from an experiment which examined assimilation in connected speech for Irish lingual stops abutting across a work boundary. This was essentially an EPG study, and involves four stops: two coronals and two dorsals. The original experiment encountered methodological limitations which are inherent in the analysis techniques currently in general use. The original study (Slattery, 1993) raised interesting questions, but underlined the difficulty of adequately differentiating between the pair of coronal and dorsal stops, and in particular of detecting fine differences in degree and type of articulation. It was expected that with techniques developed for this thesis reanalysis would provide better access to the information in the data, allowing for more detailed descriptions of the assimilations (and of similar phenomena). With such descriptions it was possible to examine the compatibility of assimilation observed for Irish with current models of

coarticulation, and in particular with the predictions of articulatory phonology (Browman & Goldstein, 1989; Browman & Goldstein, 1992).

The second analytical study concerned acoustic/articulatory relationships in speech, and involved an analysis of EMA data was recorded of a trained phonetician producing high vowel continua. The aim of the study was to locate the centre of the lingual constriction for vowels along the continua from EMA, and plot those locations against the formant frequencies produced for those vowels. This could then be compared to similar plots presented in Fant, 1970, which were for a model of a vocal tract instead of a human speaker. For this study EPG data was recorded simultaneously with EMA, and so the study also allows for investigations of how simultaneous EMA and EPG data can be used, and explore the correlations between the two types of data in a way that would enhance the analysis of either.

## *0.3 Outline of thesis*

This thesis can be considered to be made up of three major parts. The first part, which is composed of Chapter 1 to Chapter 4, discusses the literature on both the measuring and modelling of tongue movement, and on the topics of the analytical study. The second part, composed of Chapter 5 to Chapter 7, deals primarily with the development of techniques and tools for recording and analysing EPG, EMA, and simultaneous EMA and EPG data. The final part, composed of Chapter 8 to Chapter 10, presents the results and discussions of the two short studies that were carried out for EPG data, EMA data, and simultaneous EMA and EPG data. Finally, overall conclusions are presented in Chapter 11.

# Chapter 1 Vocal Tract - Measuring and Modelling

## *1.1 Why measure speech?*

The human speech process is complex. It is a chain of events which can broadly be categorised as production, transmission, reception/hearing, and comprehension. This chain starts and ends with the brain but involves many other human organs, and the environment, in the process. Within this chain there are many areas where aspects of the process may be studied. These include: neural activity while speaking and listening, muscle activity in controlling the speech organs, the physiological arrangement and movement of speech organs in the vocal tract during speech production, the acoustic signal produced, and the response of the ear to this acoustic signal.

Being able to measure speech, and being able to understand these measurements, would support many ways of improving speech communication. For example, speech communication with machines would allow us to operate them without using our hands, without being in the physical presence of the machine, and allow feedback from the machine that we easily understand. Being able to convert speech into a small set of measurements for transmission would yield lower communication costs. Medical problems with speech organs could be detected earlier by observing changes in speech from 'normals' for a given speaker. Assistance in learning the speech sounds used in a given language/accent could be provided for people who are learning to speak a new language or who are relearning to speak after, for example, a stroke.

There are many examples of the usefulness of having good and well-understood measurements of human speech. Unfortunately there are considerable limitations on the measurements we can make at present. There are also limitations in our ability to interpret the measurements we can make in the overall context of speech production.

## *1.2 How can speech be measured?*

Perhaps the best-measured part of the speech process is the acoustic signal. This signal can be recorded with simple, inexpensive equipment that is inherently non-invasive. Studying acoustics has yielded a wealth of information, allowing many applications based on measurements of the acoustic signal. But exploring such measurements is always restricted by knowledge of the production processes that yield the acoustic signal, and how this acoustic signal relates to the configurations and dynamic movement of speech organs. Aspects of this relationship are discussed in Chapter 3.

However, this thesis is concerned with measuring aspects of speech organ configuration during speech production. Of primary interest is the tongue, as an important organ involved in speech production. Instruments used to measure aspects of the tongue ideally need to be able to see through the cheeks, teeth and jaw. Otherwise they have to be introduced into the mouth to make measurements. If they must be introduced into the mouth then it is important that they are designed to minimise any perturbation they cause to the movement of the tongue during speech. Minimising the level of perturbation is vital if the measurements are to reflect natural speech. The fact that the tongue is made of soft tissue, is enclosed in the vocal tract, and is capable of having parts of it moving rapidly while other parts are relatively steady makes the task of developing effective instrumentation difficult.

To be able to understand measurements made with such instruments is also challenging. The tongue in itself has considerable freedom of movement as its shape is determined by many collections of muscles. However analysis of the tongue's movement during speech production shows that the curved shapes the tongue adopts are related in some kind of ordered way (Harshman, Ladefoged, & Goldstein, 1977). To use this knowledge to more effectively analyse measurements of tongue movement, analytical methods such as modelling need to be applied.

## *1.3 Models of Speech Articulators*

In modelling the articulatory settings of the vocal tract, the aim is to form simplified representations of the infinite variations of its surface shape due to changes such as tongue shape and position, the position of the jaw, or the shape of the lips. As already stated, it is known that during speech production organs, such as the tongue, are not moved in all the ways they are capable of being moved. A very effective way to capture the constraints on movement, and thus to describe articulatory settings suited to speech production, is through modelling.

There are many models of articulations in the literature. In general these models either focus on control of an individual articulator or divide into sub-models of the individual articulators. A model may allow for some interaction between sub-models. E.g., between a sub-model of the tongue and of the hyoid bone.

### 1.3.1 Modelling tongue shape

One of the simplest yet most complete methods of representing a tongue shape is to use a picture of that shape, such as an x-ray trace. Such a picture contains far more information than is essential to describe the shape (Harshman et al., 1977). This makes it difficult to quantify changes in tongue shape for analysis, or to predict tongue shape from existing knowledge.

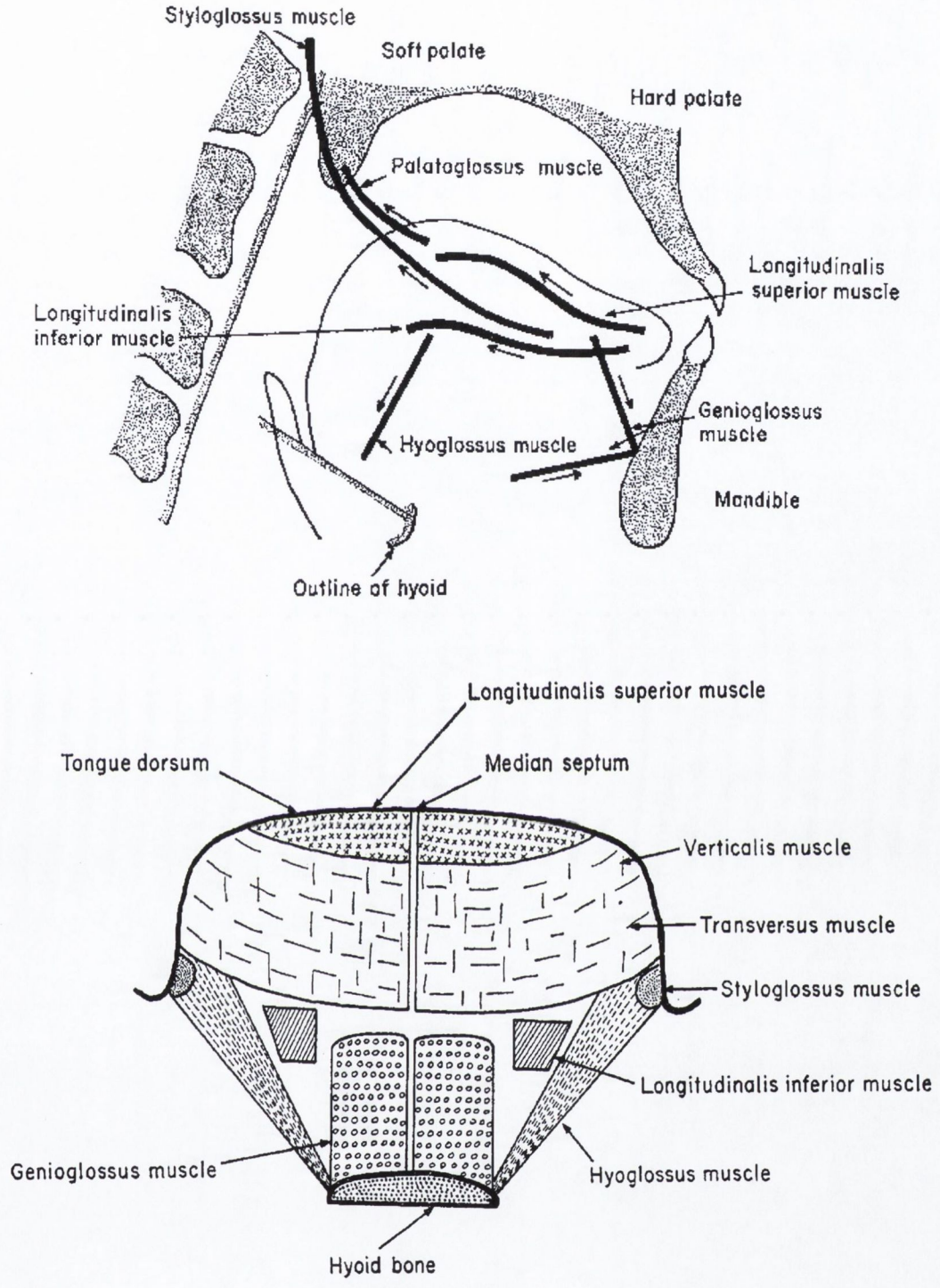



**Figure 1.1** The major muscle groups of the tongue. (Fig. 24 and Fig. 25 from Hardcastle, 1976).

Tongue shape is controlled by a number of major muscle groups, as shown in Figure 1.1. These act together to control the overall tongue shape. Each group has primary effects on

certain aspects of shape, such as protruding the tongue tip, raising the tongue upwards and backwards, or giving the tongue blade a convex shape (Gray, 1918).

The lower jaw/mandible and hyoid bone, to which some of the tongue muscles are attached, also move. This movement causes the tongue shape to alter and causes the location of the tongue within the mouth to alter, and these changes also need to be accounted for.

While the movements of the tongue can be numerous and complicated (Gray, 1918), the movements observed in speech seem to be limited. For example, Harshman et al., 1977, showed that control of two factors was sufficient to describe differences between x-ray traces of ten English vowels. A primary modelling technique used to capture these movements has been to observe the principle factors of tongue shape change in data such as x-ray recordings. The means of extracting the factors differs among researchers, with some using ad-hoc models (Mermelstein, 1973, Coker, 1976) or statistical methods (Harshman et al., 1977, Maeda, 1990), producing representations of the tongue and vocal tract which represent different vocal tract settings with different degrees of accuracy.

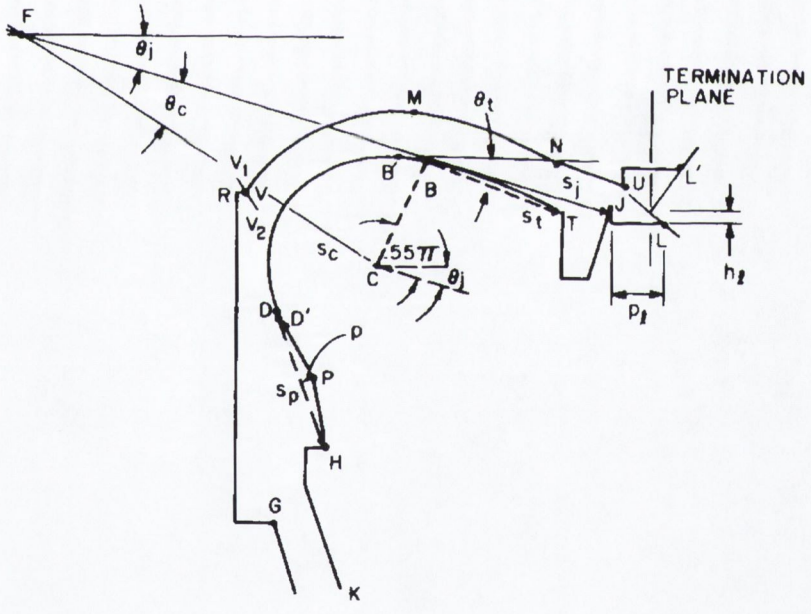



**Figure 1.2** An outline of the model presented in Mermelstein, 1973, indicating the parameters to that model. The principal positions indicated are: J (jaw), L (lips), B (tongue blade/body join), C (centre of tongue body), V (velum), H (hyoid), K (larynx).

The model that is most often referred to in this thesis is that developed by Mermelstein, 1973. This is an ad-hoc model which represents the entire mid-sagittal vocal tract outline. The model is shown in Figure 1.2. Specifically it was designed to represent the jaw, the hyoid bone, the tongue body, the tongue blade, the lips, the velum, and the maxilla and rear

pharyngeal wall. For each item modelled there are one or more parameters which allow for much freedom of movement. The restrictions on the parameters in real speech were determined by finding suitable settings for the model parameters to produce a vocal tract shape to match x-ray data of numerous utterances. The tongue body was represented as a circle of a fixed 2 cm radius. This value for the radius was based both on Coker, 1976, and on the fact that good matches to most x-ray tracings were found with that constant radius.

The parameters for the model presented in Maeda, 1990, are also determined from the examination of many x-ray traces of human speech. However where the parameters in Mermelstein, 1973, were selected based on his expectations and experience, the parameters in Maeda, 1990, were determined through statistical analysis of measurements of the vocal tract shape taken from x-ray. Of the two subjects analyzed, 88-90% of the variance of tongue shape could be accounted for by the parameters called jaw position, tongue-dorsal position, tongue-dorsal shape, and tongue-tip position.

The model in Mermelstein, 1973, and the model in Maeda, 1990, are both models of 2 dimensional, mid-sagittal, changes of shape of the vocal tract. There have more recently been considerable developments in 3 dimensional modelling of the various articulators. For example, Bateson & Ostry, 1995, present a model of jaw movement which allows for 3 dimensional movement rather than the restricted movement within the mid-sagittal plane of Mermelstein. However some of the devices presented in the next section cannot measure 3 dimensional movement and for data from those devices models such as Mermelstein's and Maeda's are useful for interpreting the data. For example, the representation of the tongue body as a circle of 2 cm radius as seen in Mermelstein, 1973, is also used in this thesis to assist in interpreting EMA data (see §6.2.2 and Chapter 9). The model in Maeda, 1990, is used during discussion of results in Chapter 9.

## *1.4 Measuring Vocal Tract Geometry*

To model speech production we need to measure the movement of the speech organs and know the shape of the parts of the vocal tract that do not move. Stone, 1997, discusses the difficulties with developing instruments to make these measurements, saying that the difficulties arise because the articulators differ widely in speed of movement, structural composition, dimensionality of movement, degree of complexity, shape, and location. For example, the lip, the jaw, the tongue and the velum may all move during speech to adapt the shape of the vocal tract for different sounds. Movement of the jaw and lips can be easily observed. The tongue in general cannot be seen, and nor can the velum unless the mouth is wide open. And though the jaw and the lips can both be observed, the fact that the jaw is a solid bone structure and the lips are soft tissue means that the type of movement they make

and the ways that movement can be measured means that different instruments may be needed.

There are several types of instruments that can be applied to taking measurements of lingual articulation. Such instruments may be categorised as being either invasive (i.e., involve transducers being introduced into the vocal tract) or non-invasive. With invasive techniques there is a risk of distorting speech through physical interference or by causing the speaker discomfort. Non-invasive techniques may also distort speech by requiring, for example, speakers to have their heads in unusual positions. A brief description of some of these instruments is given here, but a more thorough discussion can be found in Stone, 1997.

An important category of non-invasive instruments is that of imaging techniques. Such techniques can "see through" the walls of the vocal tract to the organs within. Examples include x-ray, ultrasound, and magnetic resonance imaging (MRI).

X-ray is a category of radiation that has the ability to penetrate through human issue. However different types of tissue block different amounts of the radiation. In standard x-ray a photograph of the amount of radiation passing through is taken. This gives good images of jaw and reasonable images of tongue, velum, and lips. Films showing movement of these organs can be made at good frame rates. With projection x-ray structures can be obscured by other structures (e.g. the teeth may block x-ray from much of the tongue), and soft tissue is generally not clearly defined. Computed tomography (CT) is a technique where thin x-ray image slices are made at various points around the body being studied and a composite image is computed from the many scans. This gives clearer images but at slower rates than with standard x-ray film making it unsuitable for measuring continuous speech. However, little new data for speech research is being collected with x-ray, primarily because prolonged exposure to x-ray radiation is harmful so unnecessary x-rays are generally avoided.

Magnetic resonance imaging (MRI) is a technique which involves exciting the hydrogen atoms in tissue and measuring the position of these atoms in a non-harmful way. It provides good images of soft tissue in the vocal tract. However, a single high-quality frame takes several seconds to record. Quality can be traded for scan rate, but the technique is not yet suited to measuring dynamic aspects of speech.

Ultrasound is a technique which makes use of the reflection of sound at acoustic boundaries, such as the boundary between the flesh of the tongue and air. A high frequency sound transmitter and receiver is placed against the outer surface of the skin. By measuring the time from when a sound is emitted until it is reflected off a boundary and detected at a receiver, the distance from the sound source to the boundary can be determined. A section of tissue can be scanned by repeating this at multiple points (by using multiple transducers and/or moving the

transducer) and a composite image can be computed from the combined measurements. The technique has been effectively used for measuring tongue movement, though the tongue tip cannot in general be imaged. Dynamic recording can be made with a scan rate as fast as 60 scans per second.

Imaging techniques have the potential to provide complete information about the configuration of the vocal tract when techniques such as those just mentioned improve in scan rates and scan resolution. Interpretation of the images they produce so as to identify the individual articulator configurations is another inhibitor of wide-scale use, as currently considerable human interpretation is required. This area too is being developed and greater automation is to be expected.

Invasive techniques, such as electropalatography and point tracking techniques discussed below, cannot provide such complete information due in part to constraints on their shape and the material they are made from which arise from the requirement to avoid interfering with speech production. But the measurements they provide are generally directly representative of some aspect of articulatory activity.

Electropalatography (EPG), uses a thin artificial palate with many small electrodes embedded into its surface (see §2.1). Contact between the tongue and any individual electrode can be detected by seeing if a small (harmless) current can pass through the body and through the electrode. By continuously monitoring these electrodes in this way, the contact between the tongue the part of the roof of the mouth that the artificial palate covers can be continuously recorded.

Point tracking techniques measure the position of individual flesh-points on articulators such as the tongue by affixing pellets/sensors to those points. Electromagnetic mid-sagittal articulography (EMA or EMMA) and x-ray microbeam are point tracking techniques, suitable for measuring lingual articulation. EMA involves attaching sensor coils to the midline of the tongue, lips, jaw, or velum. Each sensor is connected by a thin wire to a measuring device. The movement of these individual sensors can be continuously measured with reference to transmitter coils located around the head (see §2.2). (The safety of EMA has not been fully explored, but see Hasegawa-Johnson, 1998, for some analysis of the safety of the Carstens articulograph AG100. In the author's experience some subjects have experienced headaches with the Carstens EMA headset in place. The headaches went away when the headset was removed. These headaches may have been caused by the radiation, by the physical construction of the headset, or by a combination of these effects. The frequency at which transmitters radiate is not, however, considered to be harmful.) X-ray microbeam uses small

gold pellets instead of sensor coils and low-dosage, and focused x-ray beams are used to monitor movement of the pellets.

There are other techniques which measure aspects of lingual activity. One such technique is electromyography (EMG) where sensors are attached or embedded into muscles to measure the electrical signals that are produced during muscle contraction. Difficulties with this technique arise in part from the complex arrangement of muscles in the tongue, and it does not provide a direct, or easily interpretable, indication of the resulting change in shape of the articulator.

When measuring vocal tract geometry, the primary concerns are the accuracy of the equipment and the capacity of the equipment to accurately measure the targeted parts of the vocal tract. The measurements made are not normally affected by noise in the area of the recording unless this affects the measuring equipment itself. Most measurements made are of a localised part of the vocal tract and are relatively simple in nature. However, analysis of these measurements is considerably limited by lack of existing practices and by the fact that a change in the location of an articulator such as the tongue may be caused in several ways (different tongue muscles, the jaw, the hydroid bone).

Thus, a number of techniques could be used to study lingual activity, each with its own set of advantages and disadvantages. For example, x-ray is too harmful, ultrasound does not accurately monitor the apical part of the tongue, and MRI is not suited to continuous speech. Though they are not without restrictions, electropalatography (EPG) and electromagnetic-articulography (EMA) can provide lot of information on tongue movement. They do not harm the speaker and can be used to make measurements during continuous speech production. They are the primary devices used for this thesis.

# Chapter 2 Electropalatography and Electromagnetic-Articulography - Measuring Tongue Movement

The principle work of this thesis looks at two devices to measure lingual articulation: electropalatography (EPG) and electromagnetic-articulography (EMA). The specific implementations considered throughout this study are the Reading EPG palate (Hardcastle, Jones, Knight, Trugeon & Calder, 1989) and the Carstens Medizinelektronik AG100 EMA system (Hoole, 1996), though much of that discussed would equally apply to other implementations. This chapter presents an introduction to EPG in §2.1, and to EMA in §2.2, along with a brief comparison of the techniques in §2.3. Further details are provided in Chapter 5, Chapter 6, and Chapter 7 respectively.

## *2.1 Introduction to Electropalatography*

Palatography is the measurement of the amount of contact the tongue makes with the roof of the mouth, specifically with the hard palate. This can provide very useful information on tongue activity during, for example, lingual stops and fricatives. The earliest palatograms were made by coating the surface of the subject's palate with a coloured coating such as chocolate powder. The subject would then produce a speech gesture which, if the tongue made contact with the palate, would remove the powder where contact was made. A photograph of the palate was then made, showing the region of lingual contact. An example of a palatogram is shown in Figure 2.1(a).

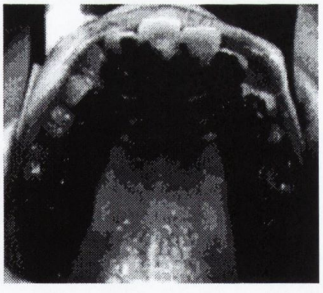
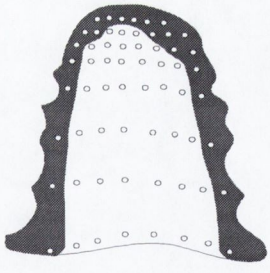
(a)  (b) 

**Figure 2.1** (a) A palatogram for a dental affricate of Beijing Mandarin where dental contact is indicated by the dark band close to the teeth (from Lee, 1999). In (b) a representation from this thesis for an EPG contact pattern of a dental stop of Irish is shown, where the approximate area of contact is shaded grey and electrode locations shown as small circles.

Such palatograms could not illustrate dynamic change during the gesture, and were difficult to set-up and to make measurements with. With the advent of inexpensive computing an

electrical version of palatography was created. This used an artificial palate, similar to those used for dentures, in which were mounted many electrical contacts (electrodes). A schematic representation of such a palate is shown in Figure 2.1(b) where small circles represent the location of the electrodes, and a grey region represents the approximate region of contact with the EPG palate that was observed for an Irish dental stop. The artificial palate is worn throughout an experiment. At intervals of, for example, 5ms, the contact with the tongue or lack thereof is recorded for each electrode. This results in sequences of palatograms showing the dynamic change of lingual contact.

There are a number of different electropalatography systems in use (Hardcastle, Gibbon & Nicolaidis, 1991). The main differences between the systems are the configuration of the contacts on the artificial palate, and, to a lesser extent, the shape of the artificial palates. The system used in this work is the Reading EPG2 system. This system uses 8 rows of contacts, with rows closer together at the front of the mouth than at the back. There are eight contacts in each row except for the front-most on which there are six contacts. The exact position of the electrodes is different on each palate as it depends on factors relating to the shape of the roof of the mouth that the artificial palate covers and on the skill of the person who makes the artificial palate. An example of the layout of the contacts is shown in Figure 2.2 (a). The representation often presented for this system consists of 62 regular rectangular cells, each cell representing an electrical contact (Figure 2.2 (b)).

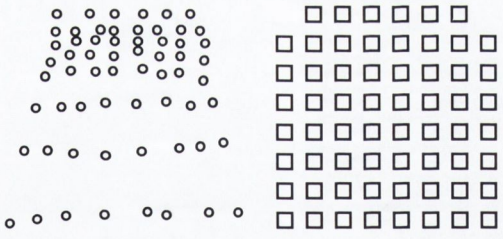


**Figure 2.2** a) the distribution of electrical contacts on a Reading EPG palate, viewed from below, for a particular adult female, and b) a schematic layout typically used to present the contact palatogram.

There are a number of improvements that could be made to the EPG palate. The artificial palate is made of a hard plastic. Therefore it cannot extend over the velum which is a soft tissue that can move during speech. The fact that the artificial palate can only cover the region to the back edge of the hard palate is limiting as it gives little information for post-palatal articulations. Research in the area of palatography includes developing palates of different materials which can cover the soft palate and/or more of the teeth. Also of interest is the development and use of new types of electrodes and sensors capable of measuring contact

pressure between the tongue and the palate or measuring the proximity of the tongue to the palate (see, for example, Wakumoto, Masaki, Honda & Ohue, 1998).

## 2.2 *Introduction to Electromagnetic-Articulography*

Articulography refers to a representation of an articulation, and in this context it refers to an articulation of the tongue. EMA makes use of electromagnetic radiation to measure the articulation. An EMA system consists of a set of transmitter and receiver coils. The transmitters are located at fixed points around the head, and each transmits an electromagnetic signal of fixed frequency and strength, with a different frequency for each transmitter. Each of the receiver coils, or sensors, is attached to a point on the tongue or another articulator and picks up the transmitted signals. The strength of a received signal depends on the distance the receiver coil is from the transmitter of that signal. The relationship between signal strength and distance between a transmitter and a receiver is a well defined property of electromagnetic radiation. Using this, the distance between each receiver coil and the transmitters can be computed.

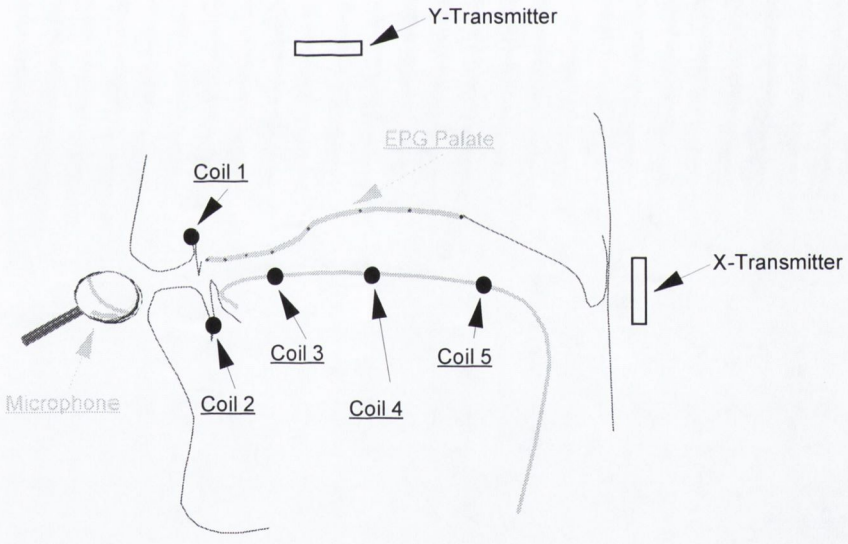



**Figure 2.3** Typical layout of transmitter and receiver coils for the MoveTrak electromagnetic articulography system.

There are a number of EMA systems in use (Hoole & Nguyen, 1999; Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I. & Jackson, M., 1992). The systems in most common use are capable of measuring movement in one plane (2-dimensions) only. The chosen plane for measurement is usually the midline of the vocal tract. The main differences between the systems are the number of transmitters used, which affects tolerance of transmitter/receiver coil misalignment, and in the frequencies and signal strengths they use, which effects the size of the receiver coils. The one represented in Figure 2.3 is called MoveTrak (Branderud,

1985). It uses two transmitter coils arranged at 90° to each other to allow for measuring horizontal and vertical movement. Using just two coils can result in significant errors if the cores of the transmitter and receiver coils rotate relative to each other, or if the receiver coil is moved off the midline. This will change the measured strength resulting in a false measurement of the position as the sensor is considered to be on the mid-sagittal plane.

To lessen this problem, other systems such as Carstens AG100 use three transmitter coils and a method called triangulation to determine the relative locations of the receiver coils. Triangulation is less effected by such rotations, as the horizontal and vertical position of the coil is determined by relative distance from each of the three receivers. So long as all three signals are reduced in proportion, the receiver coil position determined will be the same. Rigorous calibration is used before each recording to maintain the accuracy of the results as the properties of the coils can change over time (Hoole, 1996).

Current research to develop the technique is to make a system which can also measure movement parallel to a plane (Zierdt, 1993), giving three-dimensional measurement of the movement of receiver coils. This would allow the coils to be located on articulators anywhere in the mouth, reduce the risk of errors through side-to-side movement of articulators, and the information on the side-to-side movement may provide additional insight into the operation of articulators. Carstens Medizinelektronik GmbH have developed a system called the AG500 which performs multi-dimensional measurement, giving the three-dimensional location of the coil and the angle to which the coil is rotated.

## *2.3 Comparing EPG and EMA*

The measurements made with these two techniques are in many ways complementary.

EPG shows areas of the roof of the mouth with which the tongue makes contact. The area for which this is possible is constrained by the area of the EPG palate: that is an area bounded by the back of the front teeth, the back of the hard palate, and the side teeth. Also EPG measurements cannot show what parts of the tongue are making contact, nor can they show how close other parts of the tongue are to the roof of the mouth, nor can they show what shape the tongue is forming. To some extent this can be inferred from the contact patterns (Hardcastle et al., 1989) but the reliability of this inference depends on being able to clearly identify the position of some points on the tongue. This cannot be determined for the general case from EPG contact patterns.

In contrast, EMA measures the position of a small number of points on the midline of the tongue. With a sufficient number of these points the shape of the midline of the tongue can be

determined (Kaburagi & Honda, 1994). However EMA does not give direct information on what the sides of the tongue are doing.

If the measurements from both measurement systems are used together the potential to infer the tongue's shape increases. By knowing where certain points on the tongue are, as EMA indicates, the parts of the tongue that are making contact with the EPG palate can be determined. Therefore it is known from EMA where points on the midline of the tongue are and from EPG where parts of the tongue across its width are.

Such information should provide much of the data that is required to determine the tongue's shape. The task is not without its challenges. These include questions about methodology during recording (as discussed in Chapter 7), limitations of spatial resolution of EPG (as discussed in Chapter 10), and general difficulties in understanding how the data relate (also discussed in Chapter 10).

# Chapter 3 Speech Production - Articulatory and Acoustic Relationships

Critical to using measurements of acoustics and articulations together is understanding the relationship between them. This relationship is not linear, and several different articulatory settings can result in the same acoustic output. The acoustic signal emitted at the lips is determined by a large number of physiological factors. These include the activity of the lungs and of the larynx, the shape of the vocal tract above the larynx, and the materials from which the vocal tract is composed.

To help understand the relationship, greatly simplified models of aspects of speech production have been used. The acoustics of speech production is typically modelled by separately considering the source of spectral energy, such as vibrations of the vocal folds, and filtering of that source caused by the shape of the vocal tract. This modelling of the acoustics of speech production as a source and a filter is discussed in §3.1. This allows for separate modelling of the source and the filter.

Reviews of models of the voice source can be found in, for example, Ljungqvist & Fujisaki, 1985. More relevant to this thesis is the modelling of the filter. Models that represent the vocal tract configurations as a small number of tubes of different dimensions joined in series are used to present the relationship between vocal tract shapes and their acoustic effects. Fant, 1970, builds on such models, using a greater number of tubes to better represent the shape of the vocal tract. He produces graphs called nomograms which show the acoustic filtering effect for sequences of modelled vocal tract configurations which he hoped would be a useful reference for interpreting human productions. A discussion of the vocal tract filter used in the Fant, 1970, nomograms is presented in §3.2.

To see how close these simple models of articulation in Fant, 1970, are to articulations a human speaker makes during vowel production, data are required. In Chapter 9 a small study of a human speaker producing high vowel continua is presented. These vowels should be roughly equivalent to those modelled in Fant, 1970, where the model had a small cross section area at the constriction. EMA, EPG, and acoustic data were recorded so that nomograms could be produced for the human speaker. Another investigation of the equivalence between human speaker productions and Fant's monograms is presented in Ladefoged & Bladon, 1982. Though their work is very much limited by a lack of articulatory data they still present an interesting discussion, and this is outlined in §3.3.

## 3.1 An Acoustic Model of Speech Production

The acoustics of speech production is generally based on the concept of a source and filter function (Fant, 1982), and acoustic models primarily represent the acoustics of the source and the resonance characteristics of the filter. The acoustics of the source describe the characteristics of the possible acoustic energy sources which can be used. For the production of vowels the principal acoustic energy source used is a quasi-periodic source generated by the vibrating vocal folds. The resonance characteristics of the filter are determined by the configuration of the supraglottal vocal organs. A given configuration of these organs creates a filter which imparts a particular pattern of resonances and anti-resonances on the source signal. Many different configurations may be distinguished based on the different resulting patterns on the acoustic signal.

An illustration of the source and filter model that is representative of a typical voiced, non-nasalised vowel is shown in Figure 3.1. On the left of this figure the (idealised) acoustic energy source is represented both by a temporal waveform of the airflow through the glottis, $U_g(t)$, and by the spectrum of the acoustic energy of the signal. Next to this is shown the transfer function for a particular vocal tract (VT) configuration. Harmonics in the source spectrum that are close in frequency to peaks in this transfer function are amplified by the filter, while harmonics at other frequencies are not amplified and may even be attenuated. The frequencies of the peaks in the filter are called formant frequencies and the first five are marked with F1 to F5 in the figure. Next to this again is shown an approximate representation of the sound pressure radiating at the lips. Finally on the right is shown the resulting speech output, both by a temporal speech pressure waveform, P(t), and by the spectrum of the acoustic energy in the signal.

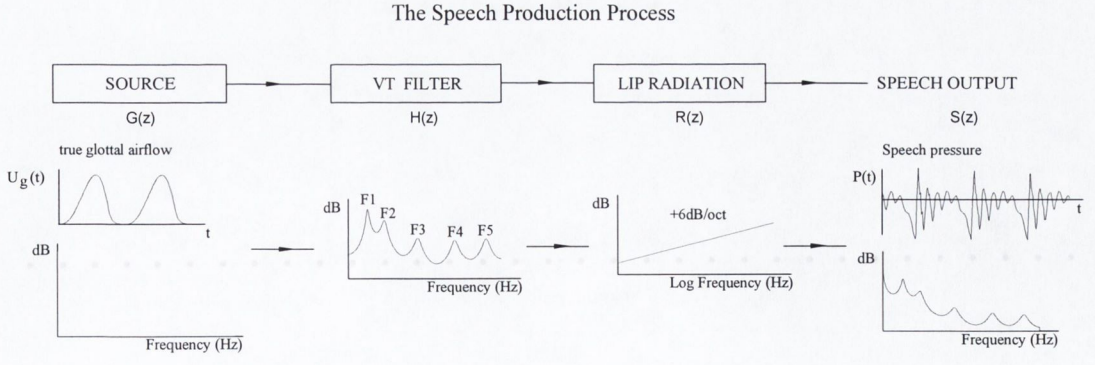



**Figure 3.1** A schematised representation of the source-filter model of speech production for a voiced, non-nasalised vowel. (From Monahan, 1996).

Modelling the vocal tract as a set of formants works well for non-nasal sounds in voiced speech such as the vowels that are studied in Chapter 9. For other speech sounds the source

and filter can be quite different. With nasals, for example, the effect of having a second pathway changes the filter by introducing additional resonances and anti-resonances, which have a major effect in the F1 region of the filter function (Hawkins & Stevens, 1985). With fricatives, for example, the turbulent noise produced at the constriction for the fricative acts as a source of spectral energy with very different characteristics to those shown in Figure 3.1. With voiced fricatives more than one source of spectral energy and associated filters have to be considered. While these result in more complicated computations the source filter model is still applicable.

## 3.2 *Acoustic Nomograms: representing the relationships between formant frequencies and articulations*

Fant, 1970, presented a number of models of the human vocal tract, along with representations of their acoustic output in the form of nomograms.

The vocal tract models presented were composed of a series of uniform tubes of various cross-sectional areas. An example is shown for a vowel-type phonation in Figure 3.2 where a mid-sagittal section of the corresponding vocal tract is presented. Beneath the mid-sagittal section the approximate cross section area along the vocal tract is drawn in a graph called an area function. Beneath this again are two area functions of approximations of the vocal tract shape as a series of short tubes (solid lines) and as three tubes (dashed lines). The three tubular cavities represent the cavity between the lingual constriction and the opening of the mouth, the lingual constriction itself, and the cavity between the constriction and the larynx. If this had been a rounded vowel then an additional tubular cavity could have been added to represent the protruded lips.

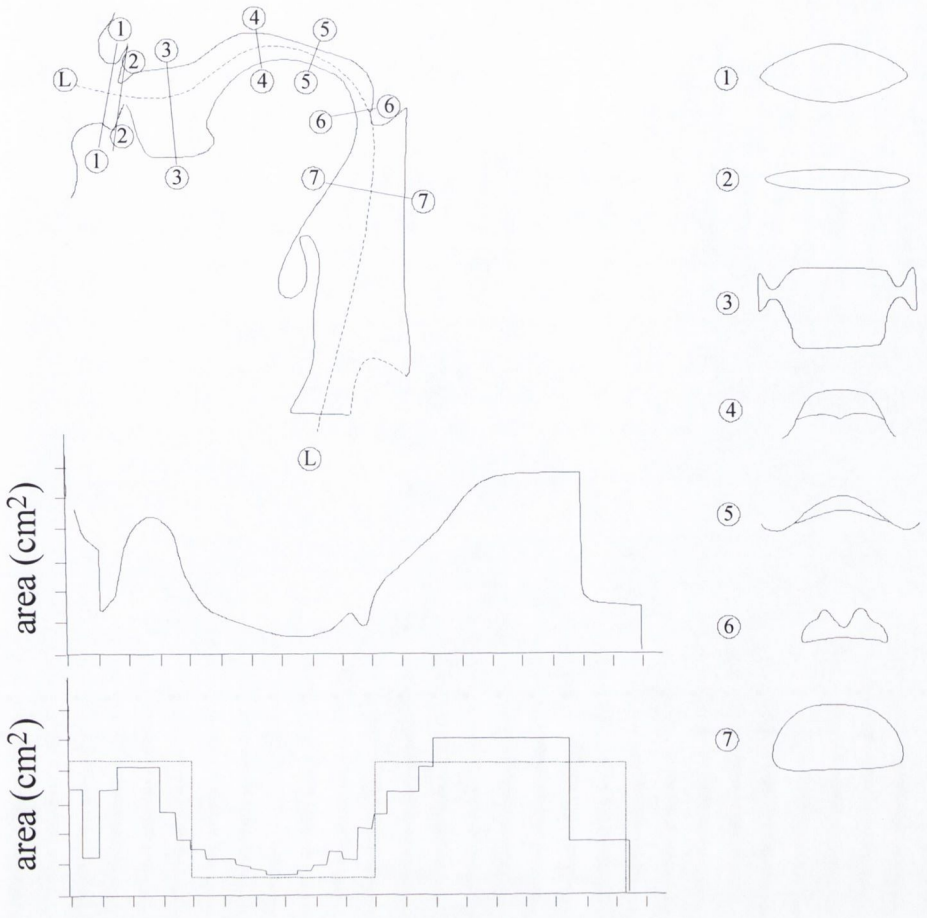



**Figure 3.2** Top left: vocal tract outline traced from x-ray for [i]. Mid left: cross section area of the outline. Bottom left: cross section modelled as many tubular sections (solid lines) and as three sections (dashed lines). Right: cross sections at points along the vocal tract corresponding to numbered lines in the top left part of this figure. (Adapted from Fant, 1970, Fig 2.2-5).

The resonances of such a tube complex can be determined by calculating the resonances of each tubular cavity and determining the overall response of the complex by considering the coupling between the cavities. In a uniform tube of length L that is open at one end, the air within will vibrate at quarter wavelength resonance frequencies (odd integer multiples of c/4L). In a uniform tube of length L that is either open or closed at both ends, the air within will vibrate at half wavelength resonance frequencies (integer multiples of c/2L). Another resonance, called the Helmholtz resonance, occurs where there is a "bottle-shaped" cavity (e.g. between the back cavity and constriction neck). The resonance is given by the equation $(c/2\pi)*\sqrt{(A/(LV))}$, where A and L are the cross-sectional area and length of the "neck", and V is the volume of the "bottle". These resonances can be used to calculate the overall resonance of a tubular complex like that in Figure 3.3.

Figure 3.3 shows a tube complex of four tubes that are used to approximate the vocal tract in Badin, Perrier, Boé & Abry, 1988. They present graphs that show the resonances of the individual cavities that would be produced for different values of $X_c$. Those graphs are presented here in Figure 3.4. They also present a graph, reproduced in Figure 3.5, that shows the resonances that the individual cavities would have if they were not connected/coupled, which are drawn as dashed lines, overlaid with the overall resonance of the coupled tube complex, drawn as solid lines. The graphs are for a constriction area of 0.65 $cm^2$ and an area at the lips of 4 $cm^2$.

Badin et al., 1988, use these nomograms to illustrate the "affiliation" between formants and certain cavities. They demonstrate that a given formant may be primarily due to a resonance in a particular cavity. In the overlay in Figure 3.5, where a solid line comes very close to a dashed line, the resonance of the entire system is mostly due to the resonance represented by the dashed line. When the solid line is not close to any particular dashed line, the formant is due to coupling between cavities (as the tube complex is a coupled system, two resonances/formants cannot coincide - this could only happen if the constriction were completely closed as then the cavities would not be coupled).

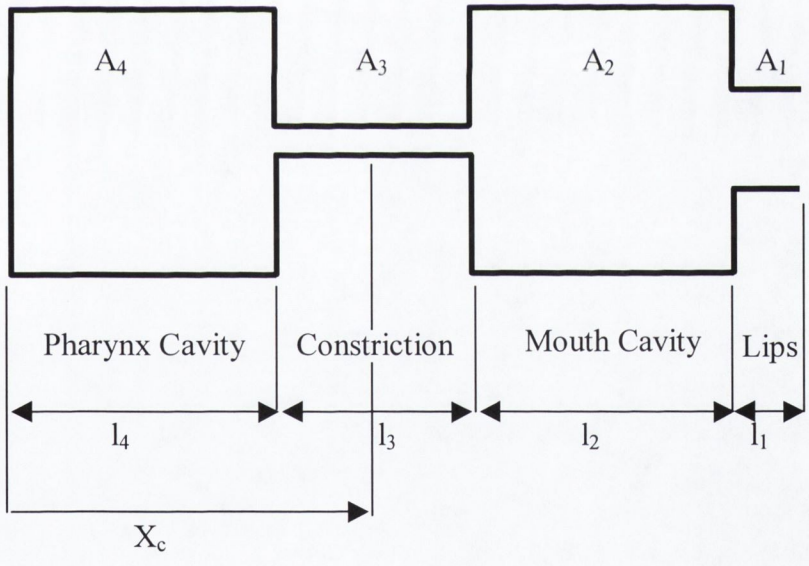



**Figure 3.3** The four tube model in Badin et al., 1988, showing pharynx cavity: A4=8 $cm^2$, l4 varying ; tongue constriction: A3=0.65 $cm^2$, l3=5 cm ; mouth cavity: A2=8 $cm^2$, l2 varying; lips: A1=0.16 $cm^2$, 4 $cm^2$, or no lips, l1=1 cm. The constriction centre co-ordinate $X_c$ (measured from the glottis position) can vary from -2.5 to 17.5 cm, keeping l2 + l3 + l4 = 15cm. (Adapted from Badin et al., 1988).

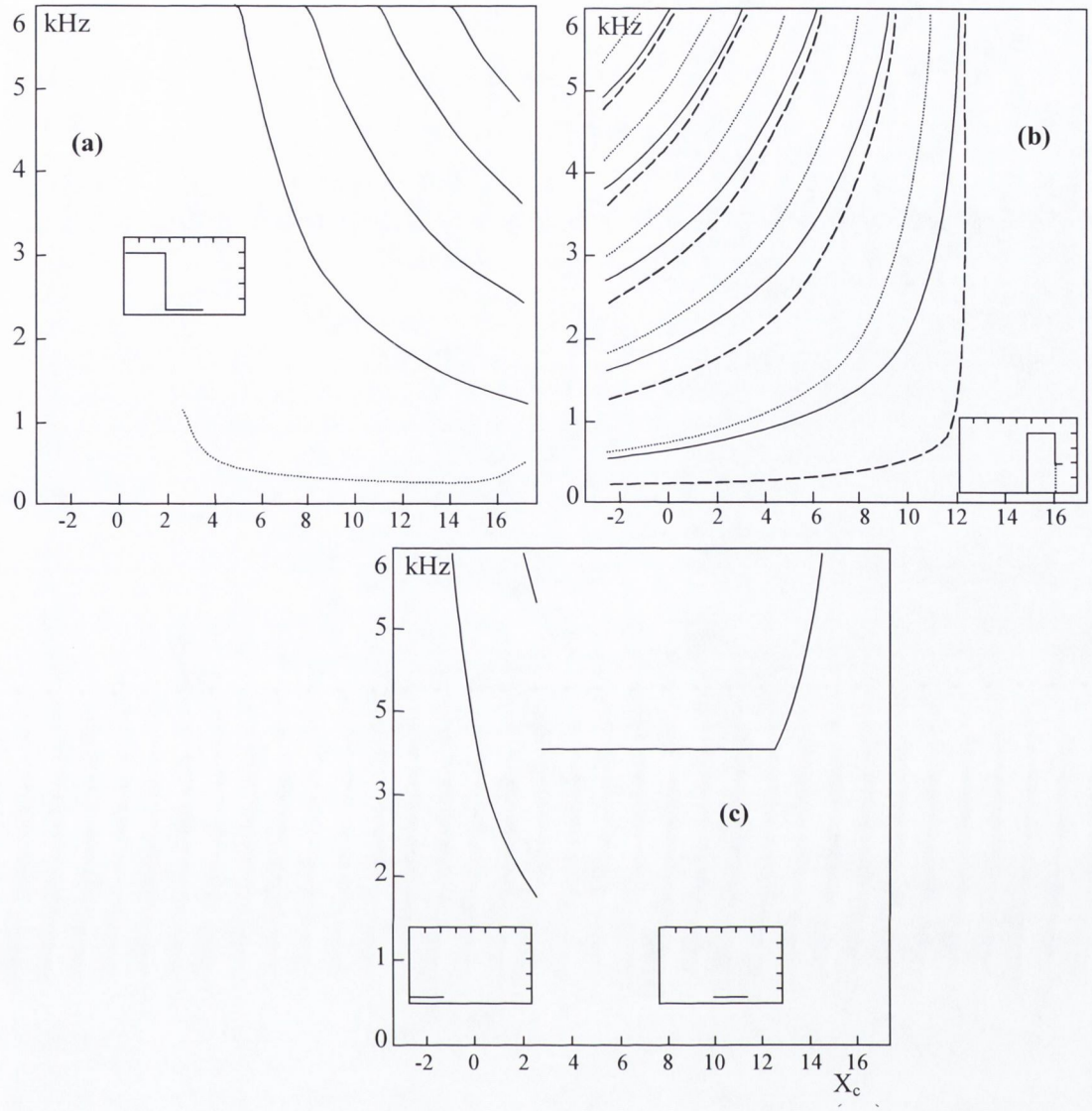



**Figure 3.4** Nomograms for the different cavities of the model in Figure 3.3: (a) back cavity : pharynx (cs-area=8cm$^2$; length varying) + tongue constriction (cs-area=0.65cm$^2$; length=5cm) (b) front cavity : mouth (cs-area=8cm$^2$; length varying) + lip constriction (cs-area=0.16cm$^2$ : dotted-line, 4cm$^2$ : solid-line, or no lips : dashed-line; length=1cm) (c) tongue constriction (cs-area=0.65cm$^2$; length=5cm). $X_c$ is the distance between the constriction centre and the glottis. $X_c$ is the constriction centre co-ordinate, measured from the glottis position. (Adapted from Badin et al., 1988).

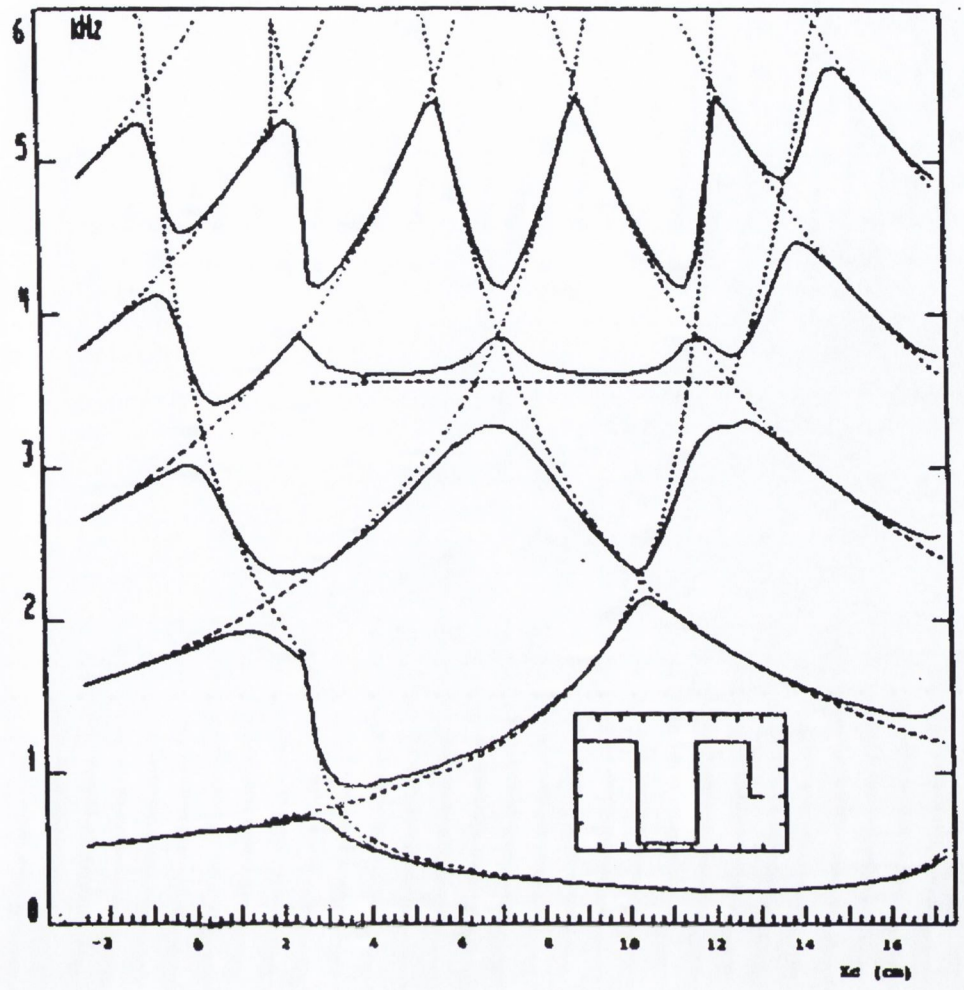



**Figure 3.5** Resonances of individual cavities in Figure 3.4 drawn as dashed lines, overlaid on the overall system response drawn as dark lines. The lip constriction cs-area=4 $cm^2$. (From Badin et al., 1988).

From Figure 3.5, one can see that formants are not affiliated to the same cavity for all articulatory settings. F1 is affiliated to the first resonance of the front cavity for the furthest back settings, but switches to the Helmholtz resonance of the back cavity for further forward settings, and F2 becomes strongly affiliated to the first front-cavity resonance. Further forward, F2 switches affiliation to the first half-wavelength resonance of the back cavity.

In the region around Xc=3.8 cm, F1 is not strongly affiliated to either resonance. This is the region where the resonances of the uncoupled front cavity resonance and back cavity resonance cross each other. In this region too, F1 and F2 are closest to each other. A similar exchange can be seen at Xc=10 cm, where F2 and F3 are largely due to the coupling between the first front-cavity resonance and the first half-wavelength back-cavity resonance. The coupling between the two resonant systems spreads apart the natural frequencies of the two systems. The smaller the coupling the closer the formants will be (Badin et al., 1988).

Fant, 1970, also presents nomograms for models based on tube complexes. The simplest of the models is similar to that in Badin et al., 1988, and is shown in Figure 3.6. The generation of the nomogram can be thought of as a process of moving the constriction, centred at x, within the model of the vocal tract by shortening and lengthening the front and back cavities ($l_2$, $l_4$), while keeping the overall length (l) the same. This models the output of the human vocal tract for a series of possible (and some impossible) high vowels, and the resonances of the model are the equivalent of formants.

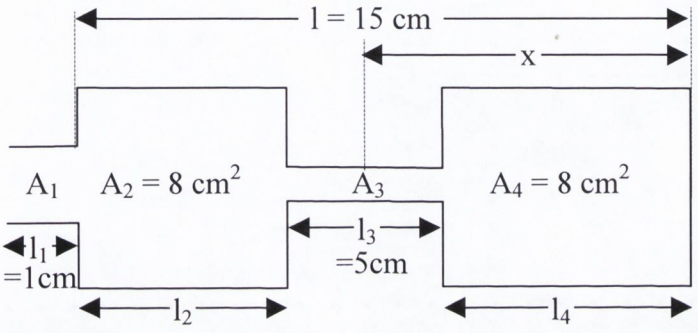



**Figure 3.6** Four tube model of the vocal tract during high vowel production. $l_1$ is the length of the lip tube, $l_2$ the front cavity tube, $l_3$ the lingual constriction tube, and $l_4$ the back cavity tube. x is the location of the centre of the lingual constriction. (Adapted from Fant, 1970 , Fig. 1.4-8.)

The length of the model in Figure 3.6 is 15 cm when the lips are not considered (as for unrounded speech). A 1cm tube is added when lip rounding is modelled. The cross section area of each of the large cavities is 8 $cm^2$, the cross section area of the tongue constriction is 0.65 $cm^2$, and its length is 5 cm. Nomograms for different configurations are shown Figure 3.7, where the first five overall resonances of the system (labelled F1 to F5) are drawn for five different lip-tube settings. The different settings of the lips are: curve 1 is for the unrounded case (i.e., no lip tube in the model), and curves 2 to 5 are for a lip tube of cross section areas 4, 1, 0.65 and 0.16 $cm^2$ respectively.

Fant presents a similar discussion on the origins of particular formants for different articulatory settings to that given above. He states that "if an advance of the tongue causes a resonance frequency to rise it can be assumed that the resonance is mainly influenced by a cavity of decreasing length, in this instance, the front cavity." Similarly, when the frequency is decreasing it is mainly influenced by a cavity of increasing length, in this instance, the back cavity. He describes four classes of resonance frequencies for uncoupled resonances:

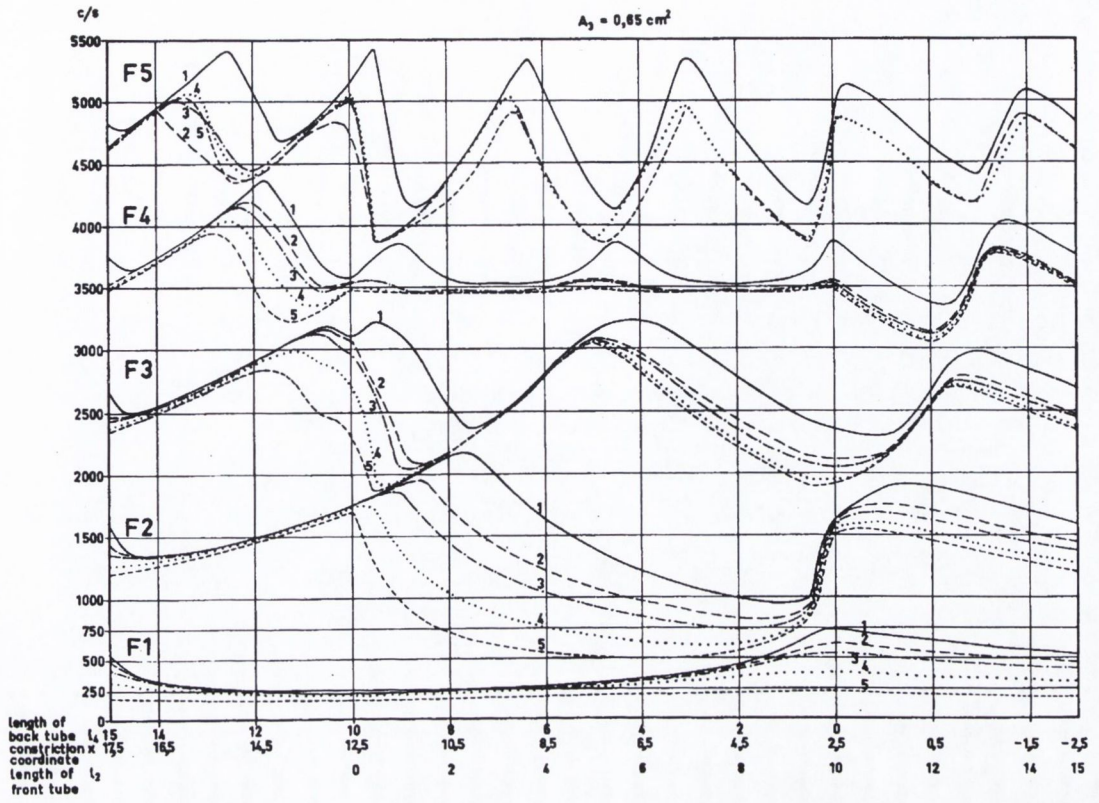



**Figure 3.7** Nomograms for 4-tube model in Figure 3.6, plotting the first five resonances of the model for different values of x. Five nomograms are presented for five different lip tube settings: $l_1$ = (0 cm for curve 1, 1 cm for curve 2-5), $A_1$ = 8 cm$^2$, 4 cm$^2$, 2 cm$^2$, 0.65 cm$^2$, 0.16 cm$^2$ for curves 1 to 5 respectively. The length of the lingual constriction tube and its cross section area is kept constant ($l_3$ = 5 cm, $A_3$ = 0.65 cm$^2$). (From Fant, 1970, Fig. 1.4-9.)

I Helmholtz resonances - front cavity and lip section; back cavity and constriction

II standing wave resonances in cavities terminated differently at two ends - front cavity in unrounded conditions; constriction when front cavity is zero length and lip rounding such that the lip section is significantly narrower than the constriction; constriction when the back cavity length is zero

III standing wave resonances in cavities with equal termination at two ends - front cavity with considerable lip rounding; back cavity; constriction when front and back cavities are of non-zero lengths

IV resonances of a character intermediate between the above types, due to one end of the resonating system being half-open.

The F1 curve, posterior (left) of the F1-F2 proximity, is affiliated with a resonance of the front cavity; the quarter wavelength without lip rounding (type II), and changing with lip rounding (type IV), eventually being due to a Helmholtz resonance in the front cavity (type I).

Anterior of the proximity, F1 is most affiliated to a Helmholtz resonance in the back cavity, and the front cavity resonance becomes affiliated to F2 up to the F2-F3 proximity. Thereafter, the front cavity resonance becomes affiliated with F3, and F2 becomes affiliated to the half wavelength of the back cavity.

The F3 curve, between the middle F3 maximum and the posterior F2-F3 proximity, is affiliated to a resonance of the front cavity, being the three-quarter wavelength resonance without lip rounding (type II) and approaching the half wavelength resonance as lip rounding increases (type IV). Between the middle F3 maximum and the anterior F2-F3 proximity, F3 is affiliated to the half wavelength of the back cavity.

In the model, F4 is essentially determined by the half wavelength resonance of the tongue constriction, except where the front cavity is very short, when it becomes affiliated to the first resonance of that cavity. However, both Fant, 1970, and Badin et al., 1988, point out that in real speech F4 may be heavily influenced by the quarter wavelength of the larynx tube and the resonances of other cavities posterior and anterior of the constriction. F5 is, in the model, associated with various standing wave resonances of higher order. Fant considers the F5 data to be more indicative of the model than natural speech and thus of little or no phonetic interest.

Fant also presents nomograms for other configurations of the model, but these settings are most similar to high vowel continua discussed elsewhere in this thesis (Chapter 9). He also presents nomograms for a more complex model of short tube sections that better approximate the actual shape of the vocal tract. Fant considers nomograms from this model to be more useful that the simpler model as a reference for estimates of formant frequencies from articulatory data. It is data from this model that is used for comparative purposes in Chapter 9.

## *3.3 Measuring nomograms for human speech*

Discussion of nomograms from models is useful for investigating the factors determining the system resonances of a vocal tract like system. To be able to apply the results and conclusions derived from such modelling it is necessary to check that human speakers operate their vocal tract in a manner which is comparable to the model. It is also necessary to check that the output of the human vocal tract is sufficiently similar to that of the model, so that the simplifications made in the model can be taken to be unimportant.

Ladefoged & Bladon, 1982, attempt to produce nomograms for a range of vowel articulations. Each author produced a series of vowels for which the location of the minimum aperture was moved progressively further from the glottis. The location of the constriction was moved progressively, they feel, from just below that for a low back vowel [ɑ] to high back [ɯ]

and from there progressively to slightly beyond that for a high front vowel [i], passing through a large number of points between. In all cases the important feature was that they should exhibit the most extreme degree of constriction compatible with vowel articulation. Similar series were produced with three different degrees of lip aperture, yielding four series for each speaker. Lip aperture was controlled through the use of bite blocks and by the speaker observing himself in a mirror, and some synchronised lip photographs were taken. Larynx height was monitored by the speaker holding his neck so that he could feel the location of the thyroid cartilage. Both speakers maintained a fairly constant larynx position when producing all vowels with the exception of very low back vowels. This was done to prevent variations in larynx height which would affect vocal tract length.

The sounds were nearly, but not quite, fricative, which the authors considered sufficient to justify assuming the whole set had a fairly uniform minimum aperture. Ultrasonic and radiographic methods were used to estimate the area of the minimum aperture. For one speaker published x-ray data was available for the vowel [ɑ] and the vowel [i]. The formant patterns from the published data matched sufficiently well in the new data. The vocal tract configuration of the x-ray seemed appropriate for comparing to Fant's nomograms. For other vowels the configuration was not verified. For the second speaker, an ultrasound sector scan of 90°, showing the tongue surface from the blade to the root, was made for half of the speaker's vowels. This was done at a later time to obtaining the acoustic data used in the paper, but the speaker used the same controls. The original recording were replayed to the speaker while making the ultrasound measurements. The distance between the ultrasound transducer and the tongue surface at the approximate place of lingual constriction was measured, and subtracted from the distance between the transducer and the upper vocal tract surface (measured from separate ultrasound record). Using this cross dimension an estimate of the area of minimum aperture of the vocal tract was made using the relationship in Ladefoged at al. (1971). They considered the vowels to be highly comparable to the area (0.65 $cm^2$) simulated by Fant in Figure 3.7.

The formant patterns of the vowels were determined from wide-band spectrograms. As the exact location of constriction was unknown, the formants were plotted so a line drawn through F2 would be relatively smoothed, similar to the curves in Fant's nomograms.

A number of differences were found between their data and Fant's nomograms. The range of vowels they could produce, while maintaining the convex tongue contour Fant modelled, was limited to a range where the constriction was between 4 cm and 14.5 cm from the glottis. While they consider the data outside these limits to be of some interest, they recommend caution as the tongue shape modelled may not be feasible for human speech production.

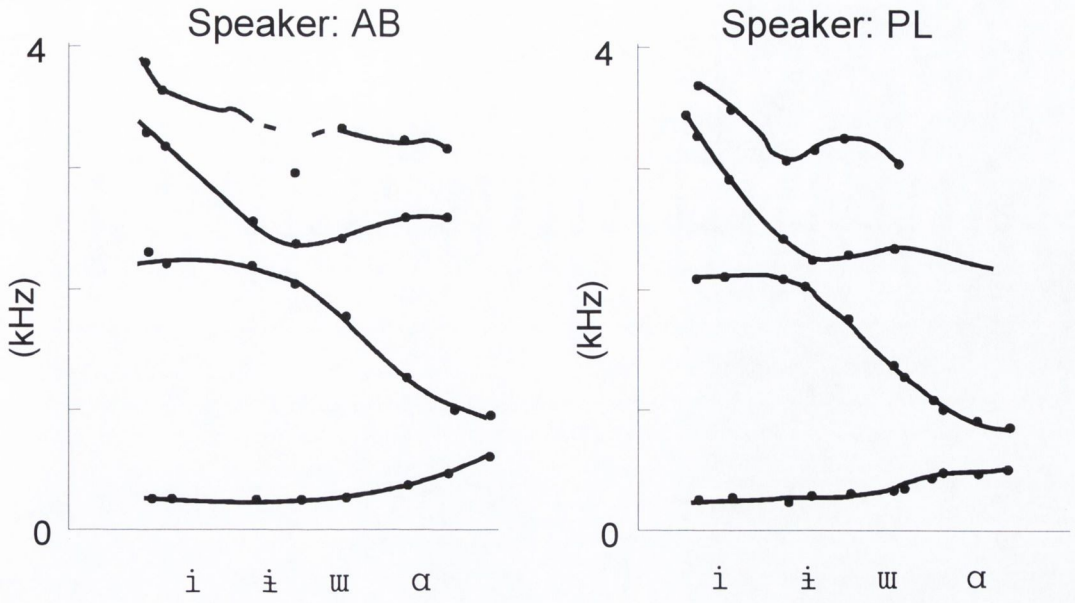



**Figure 3.8** Nomograms corresponding to one set of vowels produced by two speakers (adapted from Ladefoged & Bladon, 1982, Figure 6).

Within this region, some discrepancies were seen between the data and Fant's nomograms. In the region corresponding to 11-14 cm from the glottis on the nomogram, F2 becomes lower as the constriction moves forward. This did not happen in the data. A number of possible explanations are proposed. One is that in the region of 10-12 cm, the high vowel models have an improbably large lip aperture. Also suggested is that, as F2 is primarily affiliated with the back cavity in this region, the model may not be accurately reflecting changes to this cavity's length. Because of the curvature of the vocal tract, the length of this cavity does not increase when the constriction moves closer to the alveolar ridge and the tongue body flattens. Fant's model assumes this cavity increases in length as the constriction moves forward, while the overall vocal tract length remains constant. However the front cavity length will change. This is seen in the data as F3 rises in the region where F2 is constant.

As the study in Ladefoged & Bladon, 1982, was largely based on acoustics with small amounts of reference data from x-ray and ultrasound their results are limited by a lack of articulatory information. In Chapter 9 EMA and EPG are used to record data for a human speaker producing high vowel continua, and a comparison is made between the articulatory-acoustic relationships found for this data and that in Fant's nomograms. With the availability of articulatory data both the similarities and differences between the nomograms for the human speaker and those from Fant can be explored in more depth. These discussions are in §9.4

# Chapter 4 Coarticulation, Assimilation, and Connected Speech

As discussed in Chapter 3, the production of an isolated speech-like sound is made by way of a configuration of the articulators to produce a certain vocal tract shape. Speech is a continuum of sounds, where the articulations that would be used to produce a particular sound are in some ways modified by the influence of the context in which they are produced. The observed change may be the result of cognitive planning (often termed phonological rules) or of physical properties of the vocal tract (often termed implementation or phonetic constraints).

This contextual variability of segments of speech is most generally referred to as coarticulation. In Chapter 8 a study of coarticulation of Irish lingual stops in connected speech is presented. The explanation of the causation and the significance attached to coarticulation in human speech is very much a matter of debate in the literature. A brief discussion of some of the terms used in discussing coarticulation is given in §4.1, and some theories and models of coarticulation are discussed in §4.2. These form an important background to discussions of coarticulation of Irish lingual stops in Chapter 8 where coarticulatory phenomena observed are discussed in light of the articulatory phonology model in Browman & Goldstein, 1992.

## *4.1 Coarticulation and assimilation*

In the literature there is considerable variation in the definitions associated with terms used in discussing coarticulation. Indeed the term "coarticulation" itself is often used with different meanings.

In their discussion of the origin of coarticulation, Kühnert and Nolan, 1999, say that the term coarticulation refers to the fact that a phonological segment is not realised identically in all environments, but is often varied to become more like an adjacent or nearby segment. This description has the advantages of simplicity and of drawing on both the domains of phonology and phonetics.

Barry, 1984, presents a synthesis of findings and observations from a survey of the literature on connected speech processes. He says these processes describe ways in which the production of words and phrases in speech can vary from, for example, citation forms presented in dictionaries. He considers the result when a phoneme is produced in connected speech, if a different phoneme results the variation is called assimilation; if the result is an allophone of the phoneme under consideration then the variation is called coarticulation. However Barry considers this to be true only if the phoneme takes on characteristics of an

adjacent phoneme. For other changes to a phoneme he uses the term lenition. Within this he has two classifications: consonantal weakening and deletion, such as a stop phoneme being articulated as a fricative, and vowel reduction, such as a back central vowel being articulated as a mid central vowel. He also allows for suppression of boundaries where, for example, a word boundary can lose its prominence.

Farnetani, 1997, presents a summary description of coarticulation and connected speech processes. She states that the articulatory configuration that is invoked to produce a given speech sound is not produced in isolation but is instead invoked in the context of preceding and following articulatory configuration targets. Thus the vocal tract configuration is, at any one time, influenced by the configuration requirements of more than one target. It is to this multiplicity of influences to which the term coarticulation applies. Similarly, a speech sound produced in speech is not identical to that produced in isolation. That is, phonetic properties associated with a sound in isolation are modified and become similar to those of adjacent speech sounds. This contextual variability of speech sounds is called assimilation. Thus coarticulation refers to influences on articulatory configuration targets and assimilation to influences on sounds. She questions though whether (the extremely similarly described processes of) assimilation and coarticulation refer to qualitatively different processes or to similar processes described with different terms, and says the answers to this question are many and controversial. She gives an example of Chomsky & Halle, 1968, who say that assimilation is accounted for by phonological rules whereas coarticulation results from physical properties of the speech mechanism. Solé & Ohala (1991) propose such a distinction in their study of nasalization in American English and peninsular Spanish. They found that in American English vowel sequences followed by a nasal are nasalised as a proportion of varying vowel duration, whereas in Spanish the portion of the preceding vowel that is nasalised is of a constant duration. They argue that nasalization is Spanish is an unintended effect resulting from properties of the speech mechanism as duration is largely unaffected by speech rate, whereas in American English nasalization is planned and accounted for by phonological rules as the duration of nasalization varies with speech rate. However Clumeck, 1976, found that velar coarticulation differed in temporal extent across all the six languages analyzed. As there was no clear constant duration of nasalization of a preceding vowel across languages, this raises the question as to the criterion that can be used to say in which languages the patterns are unintentional and automatic as opposed to being ascribed by phonological rules.

As considerable differences arise in modelling of the processes that result in coarticulation due to assigning processes to one or other domain, the use of the term coarticulation in the next section will be the same as that given in Kühnert and Nolan, 1999.

## 4.2 Theories and models of coarticulation

There are many theories regarding possible processes that could give rise to, and control, contextual variability such as coarticulation in speech, and models to represent the working of these theories. In her review of coarticulation, Farnetani, 1997, presents theoretical accounts of coarticulation in four major groupings: one for historical or pioneering studies and three to represent the major current theories. The headings she uses for the major current theories are coarticulation as speech economy, coarticulation as a "creature" of the language grammar, and coarticulation as coproduction.

Underlying theories of coarticulation as speech economy is that the speech motor system is governed by the principle of economy. With the assumption that different communicative situations require different perceptual contrasts in acoustic output, this principle means that no more effort is made by the speech motor system than is necessary to produce the required perceptual contrast. Important work in this area has been performed by Lindblom, 1983, in developing his theories of "adaptive variability" and "hyper-hypo-speech". In this work he makes a clear distinction between coarticulation and assimilation. Coarticulation he views as a continuous motor process whereas assimilation he views as a categorical change (that is a consequence of coarticulation as language adapts to the speech constraints in the development of grammatical rules).

Theories which view coarticulation as a "creature" of the language grammar propose that coarticulatory processes are applied to timeless phonological segments. Features and properties associated with these segments are the parameters that control the coarticulatory processes that result in plan for the speech production system to articulate. In the model of coarticulation presented in Henke, 1966, the input is a sequence of segments that are specified for articulatory targets in terms of binary phonological features (+/-). A segment may be unspecified for a feature, in which case that segment is assigned that feature from the next segment for which it is specified yielding anticipatory coarticulation. This assignment is called feature spreading. Apart from not being able to account for carryover coarticulation, both the use of specified versus unspecified features for blocking or allowing coarticulation and the use of feature spreading to account for temporal aspects of coarticulation are questioned.

The model presented in Keating, 1985, also has as its input specified in terms of binary features, but also relates the input to the physical reality of speech through a phonetic component of windows. Unspecified segments in this model though may be left unspecified or may be acquired through language specific phonological rules. For a given articulatory or acoustic dimension a feature value is associated with a range of values called a window. A

window is specified in terms of duration and a width which is the range of contextual variability of a segment. Paths, which are constrained by requirements of smoothness and minimal articulatory effort, join adjacent windows. Such paths then represent articulatory and acoustic variations over time. So wide windows have little effect on the contour of the path, allowing considerable coarticulation, while narrow windows control the path, thus inhibiting coarticulation.

Fowler, 1980, presents objections to theories of coarticulation which do not consider time as a dimension of phonological units. She presents a theory where coarticulation is characterised as the coproduction of canonical forms of phonological segments, where the phonological segments are defined by the coordinative structures that are invoked in their realisation. Coordinative structures are described as functional organisations of muscles, where the purpose of the organisation is to produce a particular skilled activity. As phonological segments are defined by coordinative structures, the phonological and phonetic representations of an utterance are inherently linked without the need for translation. Another implication is that timing is intrinsically specified in the phonological representation.

Browman & Goldstein, 1992, present a model of coarticulation as coproduction. In their model, phonological units are gestures, and utterances are modelled as organised patterns of gestures. Gestures have their own intrinsic timing patterns which are not altered to cause coarticulation. Instead coarticulation is a direct result of gestures overlapping. A gesture is specified using a set of related tract variables. A tract variable characterises an aspect of vocal tract constriction. It is the control of tract variables, and not of the articulators used to form the constriction, that is characterised dynamically. However the control of a tract variable is specified in terms of the set of articulators used to achieve a constriction and the values of the parameters in the dynamic equation describing its movement: target position, stiffness, and damping. When a gesture is effected the result is movement of speech articulators. An example of a tract variable given is lip aperture, which is affected by the action of three articulators: the upper lip, the lower lip and the jaw. An example of a set of related tract variables that a gesture would be specified in terms of is lip aperture and lip protrusion, as they are two dimensions of the same constriction. Other gestures are defined with respect to the tongue tip (tongue tip constriction location and degree), the tongue body (tongue body constriction location and degree), velum (velic aperture), and glottis (glottal aperture).

To describe an utterance, patterns of gestures are co-ordinated with respect to one another in time. They describe general principles as to how classes of gestures are organised/phased. For example they suggest the phasing may be stated as dynamically defined points within gestures at which gestures may be synchronised. For a gesture classified as consonantal points such as the beginning of movement away from the target, achievement of the target or, perhaps, the

beginning of movement towards the target could be considered. Gestures with a similar extent in time may overlap. Three categories of overlap are presented: minimal, partial, and complete. If two overlapping gestures are for distinct tract variables then the targets for those tract variables should be achieved. If two overlapping gestures are for the same tract variable then their contribution towards articulation will be given by a blend of their dynamic specifications. For example, if the tongue tip gestures from an abutting dental stop and alveolar stop overlap then the lingual constriction should not be identical to a dental or an alveolar but fall somewhere in between.

The many and varied theories of coarticulation provide useful frameworks within which the results of analysis of experimental data can be discussed. For example: in the production of casual speech Browman & Goldstein, 1992, make a strong hypothesis that all examples of fluent speech modification are due to two gradient modifications to the gestural structure: an overlap of gestures and a decrease in gesture magnitude. Holst and Nolan, 1995, discuss this hypothesis by way of seeing if it accounts for observed assimilation between word final [s] and an abutting word initial [ʃ]. In articulatory phonology both of these involve a gesture specified by tongue tip tract variables. In their data they categorised observed assimilation based on observed spectra of fricative energy, and the duration that frication is present for, in many repetitions of such an utterance. Based on the observations of the energy spectra of the acoustic signal, they detail a number of categories of production for different repetitions. In one group of repetitions spectra for [s] and [ʃ] were distinctly produced. These could be accounted for by non-overlapping gestures (no assimilation). Another group of repetitions showed patterns of energy which start as [s]-like energy spectra and end as [ʃ]-like energy spectra, with a gliding pattern between the two. The duration of frication is shorter than the previous group. This could possibly be accounted for by the overlap of their respective gestures, resulting in a blending of the dynamic specifications of their tongue tip tract variables. Another group showed only [ʃ]-like spectra, but with a duration of frication greater than that seen for a canonical [ʃ], but shorter than repetitions in the previous categories. The absence of [s]-like spectra could be accounted for by a decrease in magnitude of the gesture for [s] along with overlap of gestures to account for reduction of duration. However the apparent replacement of an [s]-like spectrum with a [ʃ]-like spectrum cannot be accounted for by either modification.

In the short study of assimilation of Irish lingual stops across word boundaries in Chapter 8, observed assimilation is discussed in terms of a representation using articulatory phonology. As Irish features a contrastive secondary place of articulation then clusters of gestures are

involved in the production of the stops. In articulatory phonology these gestures are time aligned so when overlap occurs for one gesture corresponding overlap should be seen for the other gestures. While the assimilation observed for the Irish data was generally amenable to descriptions compatible with articulatory phonology, there are examples where time-alignment of gestures did not appear to be maintained, and there are examples where a description in articulatory phonology terms seems rather complex. These descriptions are presented in Chapter 8.

# Part II - Experimental Methods for Electropalatography and Electromagnetic-Articulography

# Chapter 5 Electropalatography (EPG)

Electropalatography as a device for collecting articulatory data on the articulatory characteristic of speech production is well established (Shibata, 1968). The data recorded provides patterns that represent the contact between the tongue and an area of the roof of the mouth. That area is enclosed by the teeth and the back edge of the hard palate. These contact patterns are recorded by electrodes that are embedded in a thin acrylic palate. From these patterns of lingual contact inferences can be made about both the movement and overall shape of the tongue.

The shape of the roof of the mouth varies from speaker to speaker, and thus so too do EPG palates. Therefore when analysing EPG data the morphology of each EPG palate must always be always considered. Software developed for this study assists in making this consideration by making use of the exact locations of electrodes on the palate. Though recording the exact locations of these electrodes is not yet standard practice it is becoming so. A number of methods have been developed to measure the locations (Jones & Hardcastle, 1995; Carter, Shadle & Davis, 1996; Chiu, Shadle & Carter, 1995). In §5.2 a summary of methods to measure electrode locations, including a method developed for this study, is presented.

While EPG contact patterns are conceptually simple, each contact pattern consists (for a Reading EPG palate) of 62 values. A new contact pattern is recorded every 5 ms - 10 ms. In §5.3 ways in which this data can be represented to assist with analysis are discussed. The discussion focuses on two ways in which analysis can be assisted. Firstly, improvements to the presentation of an individual contact pattern by representing the EPG palate both schematically and with exact electrode positions are examined. Secondly, ways in which the contact pattern can be "summarised" are presented. For example in a particular example the number of electrodes in particular groups that are activated can indicate whether a particular articulation is formed or not.

## *5.1 General Experimental Technique*

The EPG system used to produce data discussed in this thesis is the Reading EPG2 system, as introduced in Section I. The artificial palate is produced from a dental impression of the roof of the mouth of the subject speaker. The dental impression covers the region from the teeth to the back of the hard palate. The artificial palate is made from a thin sheet of acrylic with electrodes mounted flush with its surface on the basis of anatomical reference points. According to Hardcastle, Jones, Knight, Trudgeon & Calder, 1989, the front row of electrodes is placed along a line along the palatal junctures of the upper incisors and the back row is

placed along the juncture between the hard and soft palates. The remaining rows are arranged so that the spacing between the front four rows is half that of the back four rows. Electrodes are equally spaced across a row with the most lateral ones just inside the side teeth. It is a recommended practice to produce a second artificial palate without embedded electrodes for the subject to wear before recording to become familiar with speaking with such a palate in place. If a second palate is not produced then the palate with embedded electrodes is used for familiarisation. This has a disadvantage that there is then an increased risk of damaging the fragile wires used on that palate.

As the EPG system is not very sensitive to electronic noise, the recording conditions are generally constrained by other factors such as the required quality of the accompanying acoustic signal. The data is typically recorded directly to computer and results in frames of 64 bits, of which 62 represent electrodes. The recording rate can be up to 200Hz, and 200Hz was the rate used for any experimental data presented here.

## *5.2 Measuring Contact Location on Palates*

A step performed in EPG recording for this thesis (that is not standard practice) is measuring the location of the contacts/sensors on the EPG palate. These locations are used extensively in much of the work in this thesis. Though making such measurements has been performed by others (see, for example, Jones & Hardcastle, 1995), there is no well established standard practice for making the measurements, and the measurements are, unfortunately, not provided by the manufacturers of the artificial palate. In §5.2.1 a summary of methods discussed in the literature is presented, and in §5.2.2 methods attempted for this thesis are briefly discussed. A rather successful method that was developed, which made use of easily accessible tools including an image scanner and a callipers, is presented in some detail in §5.2.3.

### 5.2.1 Examples of existing practice

While the literature on measuring of EPG palates is not extensive, a number of methods have been used. These have varied from simple and inexpensive methods to roughly estimate the location of EPG electrodes, to more expensive and technically challenging methods which measure the shape of the entire surface of the EPG palate. An example of each is presented here.

An example of a simple measuring method was carried out by Nolan and Pandeli (personal communication). The distance between electrodes along the surface of the palate was measured by arranging a piece of fine thread along each row and column of electrodes and marking the electrode positions on each string. This allows a flattened palate to be drawn, giving a better representation of the actual surface areas between the electrodes.

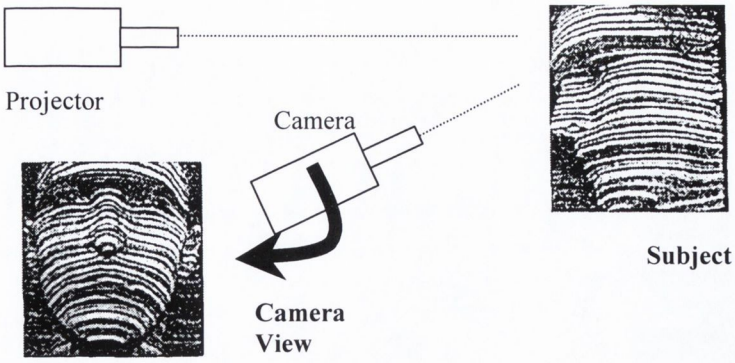



**Figure 5.1** "Structured light" used horizontal bands of light projected onto a curved surface (subject). These bands appear distorted when the surface is observed from another angle (camera view). The shape of the surface can be determined from these distortions. (From Shadle & Carter, 1993.)

Shadle & Carter, 1993, and Chiu et al., 1995, describe the use of a technique called "structured light" to measure the shape of the palate. In this method, horizontal lines of light of different colours are shone onto the curved surface of the palate. A camera, at an angle to the light source, sees resulting curved bands of light when the light strikes a curved surface (see Figure 5.1 above). Using image processing techniques the curved surface of the palate can be determined. By colouring the palate with a reflective paint and the electrodes with light-absorbing paint the locations of the electrodes can be found as black patches on the surface. This method, which the authors call enhanced EPG (eEPG), produces good quality palate surface traces for palates which do not feature very steep slopes. For palate shapes with a number of such slopes it is difficult to produce a good image with this method. Special image capturing and processing equipment is required to measure the surface images.

### 5.2.2 Measurements in this study

Several methods were used in this thesis to try to recover the locations of the electrodes. One attempt involved forming a cast of the palate. A drop of coloured stain was then applied to each electrode on the palate. The cast was reinserted into the palate so that the stain would be transferred to the cast. The electrode positions could then, in principle, be measured more easily from the cast. In practice this was found to be a difficult method. Making measurements of the locations of the ink marks with accuracy depended on careful cutting of the casts. While such methods have been used to measure the shape of part of the vocal tract in Ladefoged et al., 1971, it does not seem practical for casts as small as the one involved here.

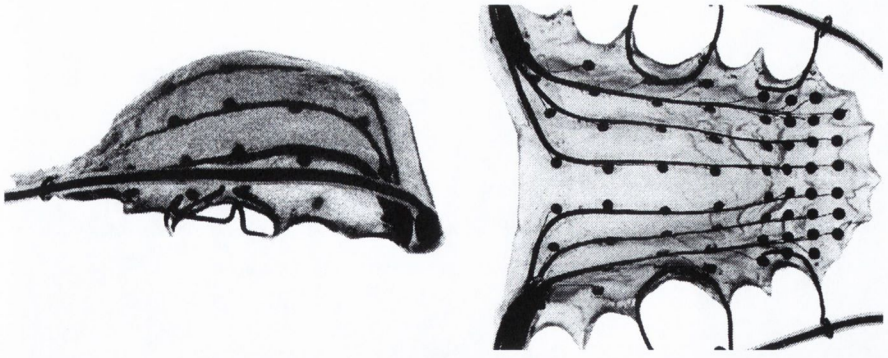


**Figure 5.2** Photographs of a view from the side and a view from below of an EPG palate.

Another method developed in this thesis made use of a professional photographer. The palate was illuminated from within and photographs were taken of the palate from each side and from below. For the side-view photographs the palate was illuminated from within, and for the photograph from below the palate was illuminated from above. A ruler was included in each photograph to provide a scale. Examples of these photographs are shown in Figure 5.2. In general the electrode positions could be easily measured from these photographs, though for some palates it was difficult to determine the locations of all of the electrodes from side views. This was because the plastic material of the palate was thick enough in places to make it difficult to see the exact position of electrodes behind the plastic.

### 5.2.3 Best method developed for this work

The most successful method that was developed in this study made use of a flatbed scanner for a computer, a sliding callipers, and a sheet of clear Perspex. This method is shown in Figure 5.3 below, and is a two part procedure.

The first part of this procedure is to capture a view of the palate from below. This involved placing the EPG palate in its plastercast underneath-side down on a scanner in such a way that it rested firmly on three points on the plastercast. The points it rested on were noted as in Figure 5.3. A ruler was placed beside it as a scale and an image was scanned. The location of the electrodes was determined from the scanned image using a software package called Corel Draw, but any software package that allows measurements of positions on an image to be made could be used. Alternatively a photograph could be made as just discussed, and the locations could be measured directly from that.

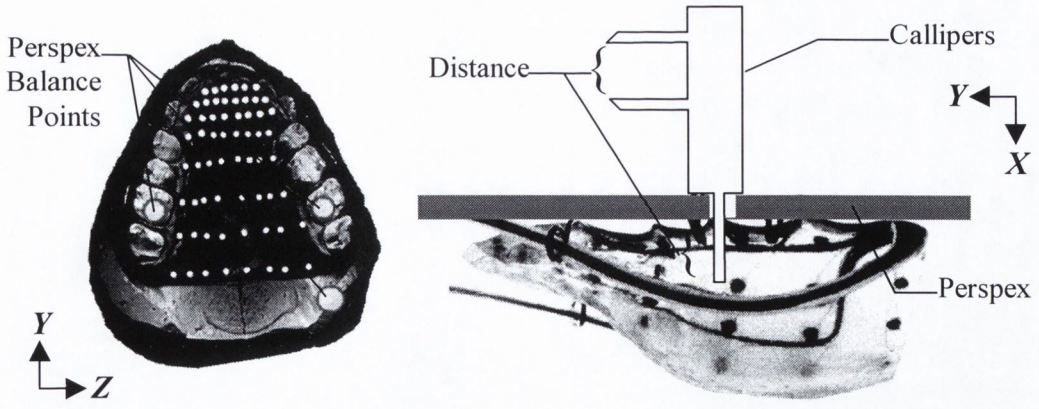



**Figure 5.3** illustrates a method of measuring the locations of the electrodes on the EPG palate in 3 dimensions. Distance front-to-back (Y) and sideways (Z) is measured from the image on the left. Depth (X) is measured using a callipers protruded through a sheet of Perspex as shown in the image on the right. The Perspex is balanced on three points on the plastercast which holds the EPG palate, as marked in the left-hand image.

The second part of the procedure is to find the vertical distance between the electrodes and the plane that the EPG palate/plastercast rested on while being scanned. A thick sheet of clear Perspex was placed on the plastercast so that it rested on the same three points as in the first part. The sheet of Perspex had a small hole drilled through it into which the stem of a sliding callipers fitted tightly. The Perspex was moved so that the small hole was directly above an electrode. The length of the calliper's stem was then adjusted so it just touched the electrode and the depth was noted. This was repeated for each electrode.

A similar method is discussed in Jones & Hardcastle, 1995. They measured two dimensions from an enlarged photograph instead of from a scanned image, and the depth is measured using a workshop clock depth gauge instead of a sliding callipers. It is not clear how the plane from which the depth was measured was controlled, but similar accuracy should be obtained with a depth gauge or a callipers for this type of measurement.

## *5.3 EPG Data Representation*

At its most basic representation, a Reading EPG frame (contact pattern) is a set of 62 values, stating that contact was or was not made with each of the 62 electrodes on the EPG palate. As raw values they are difficult to associate with the actual tongue contact against the palate. Visualisation tools are needed to help in this task.

There are two principal types of EPG data representation. One type of representation is used to enhance the illustration of a given frame. A second type of representation is one that

summarises the contact seen in several frames, reducing the amount of that that needs to be presented.

Considering a single frame, the visualisation can be enhanced to show the likely contact area more clearly. Examples include displaying electrodes in their measured positions as 2D and 3D images or estimating the likely contact region for the tongue. This is discussed in §5.3.1.

Where a sequence of frames are to be considered over time then the detailed contact information that is contributed from each frame becomes difficult to present in a succinct manner. There are many ways this information can be presented in a reduced form by focusing on particularly relevant aspects of the contact within each frame. For example, indices such as the centre of gravity index and the anteriority index represent information about the layout of the 62 electrode values in a frame as a single value. Such reductions by necessity lose information. In §5.3.2 the trade-offs involved in reduction methods are discussed and recommendations are presented as to how reduced representations can be selected for a particular data set.

### 5.3.1 Enhanced Representation of Single Frames

There are many ways of representing a single EPG frame. In §5.3.1.1 visualisations which vary from very schematic representations to representations that closely match the shape and electrode layout of the artificial palate the frame is for are considered.

Most representations in some way indicate that when the tongue is in contact with an electrode it is also likely to be in contact with an area of the palate around that electrode. When repetitions of utterances are recorded then different forms of visualisation are necessary to make use of the additional information available across representations.

#### 5.3.1.1 Palate Visuals

**Standard representation**

The primary representation of an EPG frame has traditionally been to show it as 8 rows of 6 to 8 equally spaced markers (6 in the front row, and 8 in the other rows). The marker was dark if there was contact with the electrode it represented, or light otherwise. This was the representation provided by the standard 'Edit' software provided with the Reading system. Examples from the standard software are shown in Figure 5.4 below.

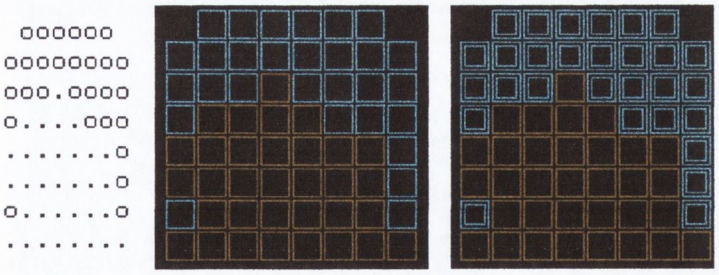


**Figure 5.4** Standard representation of Reading EPG contact patterns (from the Edit software). The leftmost shows the representation used for printing. The middle and rightmost are used for on-screen displays.

When compared to the photograph of the palate shown earlier in Figure 5.2 it can be seen that such a representation is quite crude. It can be misleading if the experimenter is not careful to keep the true layout of electrodes in mind. The Edit software allowed the locations of the electrodes to be entered and showed the electrodes on screen as small dots. This is more accurate in representing electrode positions, but the dots are very small and do not update dynamically when different parts of the recording are selected. Also it is difficult to visualise the contact area from such displays.

### Using electrode location information

Simplified/schematic representations are useful when looking for information that does not vary with EPG palate shape. For example, if the time-point of consonantal closure of an alveolar stop is of interest this can be determined without considering the palate's electrode layout. But improved visualisation gives the experimenter a much better understanding of the articulation of the tongue.

A primary activity in this work has been to develop improved visualisation tools for EPG and other data. Several different representations were provided, as it was understood that the simplifications used in other presentations helped when working with a more abstract concept of the EPG palate. These representations are shown in Figure 5.5 below, where the visualisations on the right more closely match the EPG palate than those on the left. All of these visualisations are available as displays in viewing tools developed, where the EPG displays automatically update to match the EPG frame corresponding to a movable marker in a waveform window (as seen Figure 5.6).

Figure 5.5 (a) is a traditional display whose purpose is to allow those familiar with such layouts to draw on their experiences with a familiar presentation. Its greatest drawback is that it fails to show anatomical differences between rows for a given palate and between different palates.

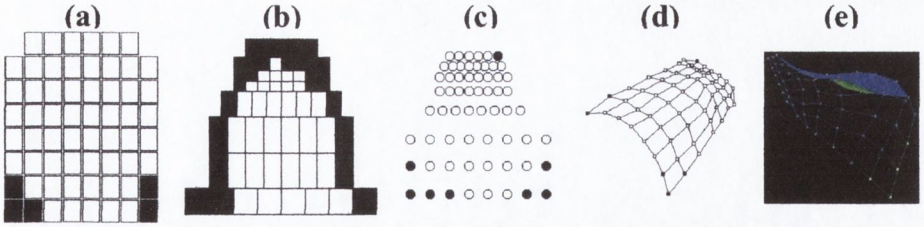



**Figure 5.5** Several representations of electrode contact patterns developed: (a) traditional rectangles, (b) rectangles scaled to measured palate, (c) circles, (d) simple 3D display, (e) 3D display with contact region.

Figure 5.5 (b) uses a similar presentation, but scales the rectangles so that the rows have the same relative heights as on the real palate. This retains the familiar row layout while enhancing the anatomical relevance of the rectangles in that they more accurately reflect the area an electrode could be said to represent. There are some problems with this display. In some cases the measured values need to be adjusted to ensure the rectangles for the front-most rows are large enough to be easily seen, requiring the experimenter to still make some mental adjustments when viewing. It also fails to allow for unevenness in the electrodes on a given row. Nonetheless it provides a reasonable, simple representation of the palate that is easily understood and prints well.

Figure 5.5 (c) represents the electrode layout as seen when the palate is viewed from below, but the circles used to represent each electrode are greatly enlarged so that it can be clearly seen whether the electrode is active or not. Unfortunately this can mean that the representations of the electrodes in the front-most rows can overlap, reducing the legibility of the display. However this disadvantage would seem minor compared to the advantage of having an accurate electrode layout. The accuracy of the layout depends, of course, on the accuracy with which the electrode positions were measured. In the particular example shown the measurements were made using pieces of thread as discussed in §5.2.1, and are certainly not exact.

Figure 5.5 (d) is a simple three-dimensional (3D) display, based on ideas from a static 3D viewer discussed by Barry (personal communication). This attempts to represent the position of the electrodes as balls in a view which allows them to be rotated. The electrode are joined to neighbouring electrodes by straight lines. The advantage is that it allows the experimenter to observe the contact patterns from below, from the side, or from any other angle chosen. A disadvantage is that such a rotatable view takes considerable practice to use as it can be quite difficult to understand from which direction the palate is being observed. In the EPG3 system (see Jones & Hardcastle, 1995) wire frame displays like this are also made available from two viewing positions: a side view and a view from beneath.

Figure 5.5 (e) extends display (d) in a number of ways. An enlarged version of this representation, viewed from a different angle of rotation, can be seen in the lower right panel of Figure 5.6. When viewing the palate from above the view is quite similar, showing the electrodes and lines joining the electrodes. When viewed from beneath, as in Figure 5.6, the palate appears solid. This makes it easier for the experimenter to understand the orientation of the palate when the display is rotated. It also makes it easier in a side view to distinguish electrodes on the near side and far side of the palate. Contact is represented by using different colours: green for activation, blue for no activation. When an electrode is activated its colour spreads towards the surrounding electrodes. If the neighbouring electrode is also activated the area between them is a solid colour as the tongue must be making contact with the palate in that entire area. If the neighbouring electrode is not activated then the colour fades out as an indication that contact in this area is unknown. This view seems easier to use due to the indications of the side of the palate being viewed and the display of the contact area.

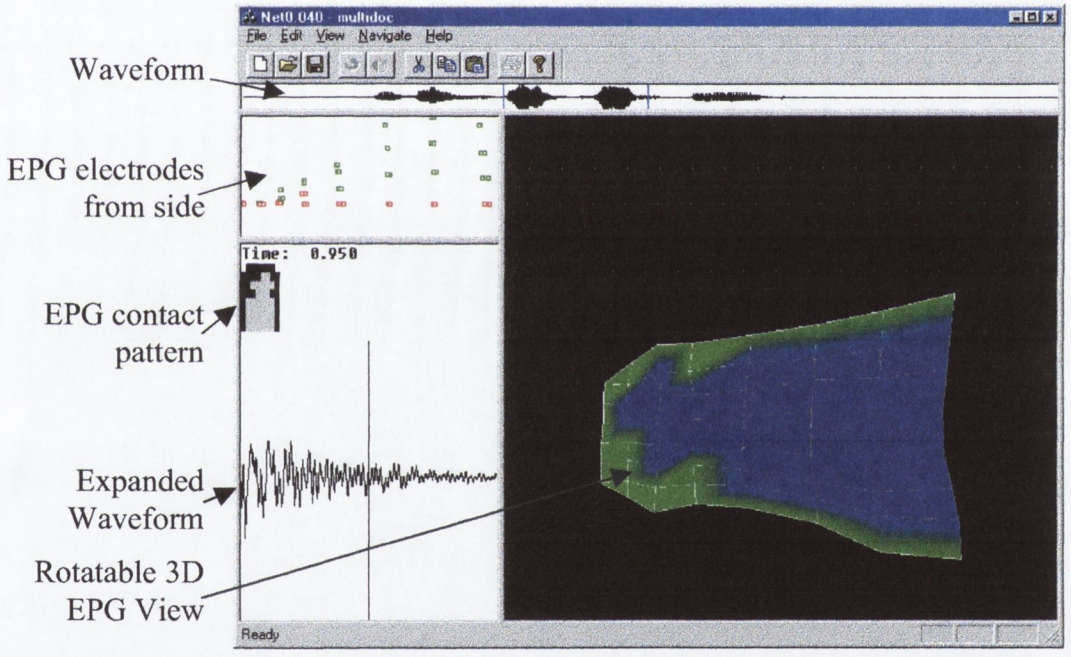



**Figure 5.6** A sample view of the EPG analysis software developed. Along with a view of the entire waveform for an utterance (top panel) and an expanded section of the waveform (lower left), three representations of the EPG contact pattern are shown: a traditional contact pattern as in Figure 5.5 (a), positioned electrodes when viewed from the side as in a side-view version of Figure 5.5 (c), and a rotatable 3D view as in Figure 5.5 (e).

## Conclusion

In practice it seems best to have a number of representations visible simultaneously, so the experimenter can refer to them in turn for different types of information. Representations as in

Figure 5.5 (a), (b), and (c) facilitate discussion in traditional terms of activation on given rows, with (b), and (c) giving some guidance as to the particular arrangement of electrodes on a given palate. Representations as in Figure 5.5 (d) and (e) help in understanding what that means on a particular palate.

### 5.3.1.2 Frequency of activation

Where multiple repetitions of an utterance are available it is important to be able to effectively use the additional information found by comparing corresponding EPG frames in different repetitions. To do this it is important to have some "landmarks" in the repetitions which are comparable. Examples are, for stops, the points of closure, release, and maximum contact. If the experimenter is satisfied that the repetitions at a given point are comparable then a number of different representations of the typical contact pattern can be used.

#### Standard Representations

A commonly used representation is a frequency of activation display. These are presented in many ways in the literature, and a number of representations are shown in Figure 5.7. All of the displays represent in some way the number of frames in a set of EPG frames for which each electrode is activated. The number of frames that each electrode is activated in is represented in one of many ways: a numerical value, a graded shading or colouring of a circle/rectangle, or a graded sizing of a circle/rectangle. The representative presentations of electrodes are typically arranged in a grid-like manner.

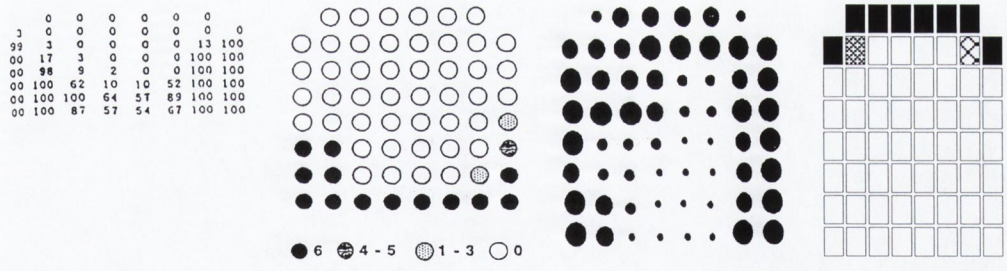



**Figure 5.7** A number of representations of frequency of contact for each electrode on the palate. The frequency of contact for an electrode is indicated by, left to right: a numerical percentage, the shade of a circle, the size of a circle, or the shade of a rectangle.

Typically in a frequency of activation display the set of frames that provides the data is made up of frames which represent some articulatory event, such as a consonantal closure, from several repetitions of an utterance. Another example is where the frequency of contact effectively represents a long term contact pattern, where the data set is of a series of frames

from a single utterance. Such analysis was not used in this work, but discussion of it is found in Hardcastle et al., 1991.

### Enhancing the representation

The standard representations provide an excellent summary of the information from several representations. However if the electrode locations are used the experimenter can again be assisted in considering the EPG palate shape. This can be done by presenting the information using visualisations similar to those from Figure 5.5.

Software was developed to display the frequency of contact for each repetition of an utterance at, for example, closure, and for a number of frames to each side of closure (typically 40 ms before and 25 ms after closure). The output of the software is shown in Figure 5.8. This shows seven schematised EPG frames representing contact around closure for five repetitions of [-a̪tʸ#t̪ʸa-]. The EPG frame at closure is labelled 'C'.

The schematic layout is similar to that in Figure 5.5 (b), and makes use of information about the positions of the electrodes to scale the rectangles representing each electrode. The number of repetitions for which contact occurred with an electrode is indicated by the colour of the rectangle. The percentage of repetitions in which contact is found is represented by a colour. The colours and the ranges the colours represent is shown in the legend.

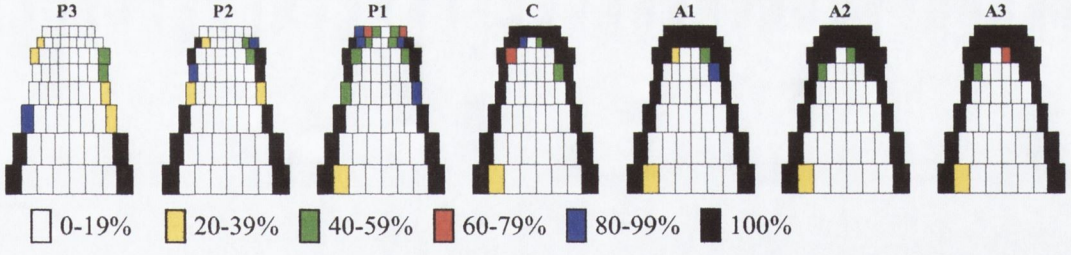



**Figure 5.8** Frequency of contact patterns for five repetitions of [-a̪tʸ#t̪ʸa-] over the range from 15 ms before closure to 15 ms after closure. The percentage of repetitions in which there is contact with an electrode is indicated by the colour of that electrode.

The frequency of contact information can also be used to infer additional information about the areas of the palate that the tongue is likely to be in contact with. For example, in Figure 5.5 (e) if the electrode neighbouring an activated electrode is not activated then a first estimate of the contact area is that the tongue is in contact with the palate to half-way between the activated and inactivated electrodes. When analysis across multiple repetitions is made, stronger inferences about the area of contact is possible. For example, if an electrode is activated in all repetitions and a neighbouring electrode is activated in four out of five

repetitions then typically contact should be represented as being closer to the neighbouring electrode than if that neighbouring electrode is activated for only one out of four repetitions.

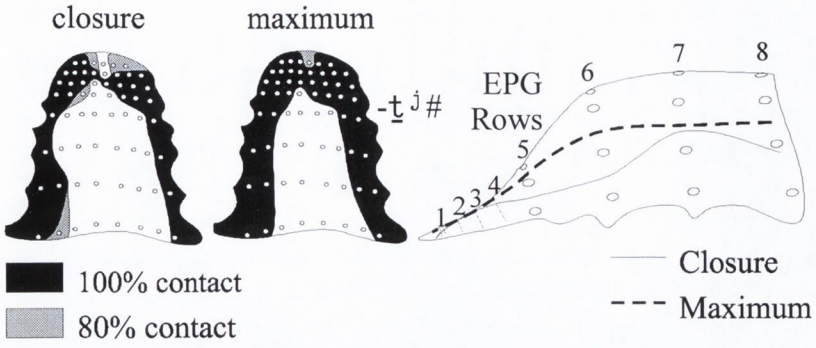



**Figure 5.9** Representation of the area of lingual contact (left) and highest lingual contact (right) based on frequency of contact of five repetitions of [-at̪ʲ].

Figure 5.9 is drawn, using the assumptions just mentioned, on the basis of frequency of contact data for consonantal closure and maximum contact during consonantal closure for five repetitions of [-at̪ʲ]. On the left a top view of the EPG palate at closure and at maximum is drawn. The area of contact in all repetitions is shown as black in these views. The additional area of contact that is found in four out of five repetitions is shown in grey. The black area passes very near to electrodes in the yellow area, indicating that contact will typically be near to that electrode. The black area does not come so close to electrodes with lower frequencies of activation. On the right of Figure 5.9 is shown a side view of the EPG palate. An outline marking the highest expected contact between the tongue and the EPG palate is drawn for closure and maximum using a similar method to the top views.

Such displays are useful for investigating the range of contact patterns used by a speaker when targeting a particular production (and are used to this end in Chapter 8). Displaying a number of frames before and after a particular production event makes it easier to investigate the potential differences in the way the production is made. The displays are also useful in the development of important regions of palatal activity for a particular target production. They are used in the following discussion of "reduced representations" (see §5.3.2.4).

## 5.3.2 Reduced Representation of a Sequence of Frames

To be able to process the data in sequences of frames, and from repetitions of utterances, it is important to be able to reduce or summarise the information. When a reduced representation can adequately represent an important characteristic in the EPG data, it can be used to display the change to this characteristic over time. It can be used to compare the differences seen in different repetitions of an utterance, by the same speaker or by different speakers, and be used

as a basis for other computations such as statistical analysis. Ensuring that the representation is adequate is the greatest challenge involved in developing and using such methods.

There are a number of standard formulae presented in the literature which produce a value that in some way represents the contact pattern of an EPG frame. Such approaches are discussed in §5.3.2.1. The resulting representations are, necessarily, abstractions of the actual contact patterns. If such standard reductions do not suit, or if the experimenter is not sure that they identify important characteristics of the contact patterns, other reduction methods are required.

With the aim of developing a method of extracting important characteristics of contact patterns several possible representations were investigated. In §5.3.2.2 information that was yielded by plotting simple counts of activated electrodes on EPG rows was considered. These totals are informative, but typical contact patterns involve electrodes on different rows. In §5.3.2.3 considerations involved in the selection of regions of electrodes which are suitable to total for a given analytical question are discussed. Such regions seem useful for characterising contact patterns. In §5.3.2.4 a method for region selection developed for this work is offered, and a way in which such regions can be used is presented in §5.3.2.5. This method is used in Chapter 8.

### 5.3.2.1 Standardised reduction methods

A number of data reduction methods are presented in the literature. For example, Hardcastle et al., 1991, present a summary of a number of methods. Some methods count the electrode activations in a particular region of the EPG palate, either for a single frame (e.g. centre of gravity) or for more than one frame (e.g. coarticulation index), and compute an index value based on these regions.

One index discussed by Hardcastle et al., 1991, is the centre of gravity index. This index gives an indication of where on the EPG palate the bulk of electrode contact is located. It is calculated with a simple formula ((½R8+ 1½R7 + 2½R6 + 3½R5 + 4½R4 + 5½R3 + 6½R2 + 7½R1) / total number of activated electrodes, where R8, R7, etc. is the number of activated electrodes on row 8, row 7, etc.). The result is a number that indicates how anterior the concentration of activated electrodes is for a frame, where a COG of 0.5 represents the most posterior articulation, and 7.5 the most anterior articulation, that EPG can detect. A similar index, discussed in Faber, 1989, is called the anteriority index. In Fontdevila, Pallarès & Recasens, 1994, similar formulae are presented to indicate anteriority, posteriority, and centrality of the contact concentration.

An index to indicate lateral asymmetry of contact is discussed in Farnetani, 1988. This index gives the ratio of the count of activated electrodes in the one half of each EPG row to the count of activated electrodes in the other half of each EPG row.

Farnetani Hardcastle & Marchal, 1989, present an index called the coarticulation index, which is based on differences in electrode activation on rows from two frames. The frames being compared would be, for example, the frames at consonantal closure in two VCV utterances where the vowel is different in the two utterances.

These indices provide information about the contact pattern seen in EPG frames. There are, however, a number of problems with using them. First of all it is not always obvious what the resulting value represents. For example the centre of gravity index will yield the same index number regardless of the number of electrodes activated on a particular row when there are no activated electrodes on any other row. Secondly, the indices do not account for variations in the arrangement of electrodes on an EPG palate. For example, the asymmetry index will not correctly indicate asymmetry unless the electrodes are arranged on the palate with perfect symmetry. Thirdly because the indices are not customised for individual EPG palates, if the values they provide are directly compared across speakers they may result in irrelevant differences that can be accounted for simply by differences in palate shape.

These problems are not insurmountable. Pompino-Marschall & Mooshammer, 1997, presented a variant of the centre of gravity index which made use of the location of EPG electrodes in determining the centre of activated electrode concentration on individual EPG palates. However understanding how general-purpose indices can be applied to a given experiment is a difficult task. It is for this reason that this work focuses on the development of specific indices for individual experiments. These involve selecting regions of electrodes to calculate an activation count/total in, and deriving a measure from these totals.

#### 5.3.2.2 Row totals

The electrodes on the Reading EPG palate are arranged in eight rows of six to eight electrodes per row. The scheme for electrode placement (Hardcastle et al., 1989) aims at ensuring these are placed according to anatomical landmarks of a palate. A typical use of this design feature is to divide the EPG palate into places of articulation such that row 1 and row 2 mark the alveolar region, row 3, row 4, and row 5 the palatal region, and row 6, row 7, and row 8 the dorsal region (though in practice the place that a row represented can vary to some extent from EPG palate to EPG palate). Another simple measure totals the front four rows as the anterior region, and the back four as the posterior region.

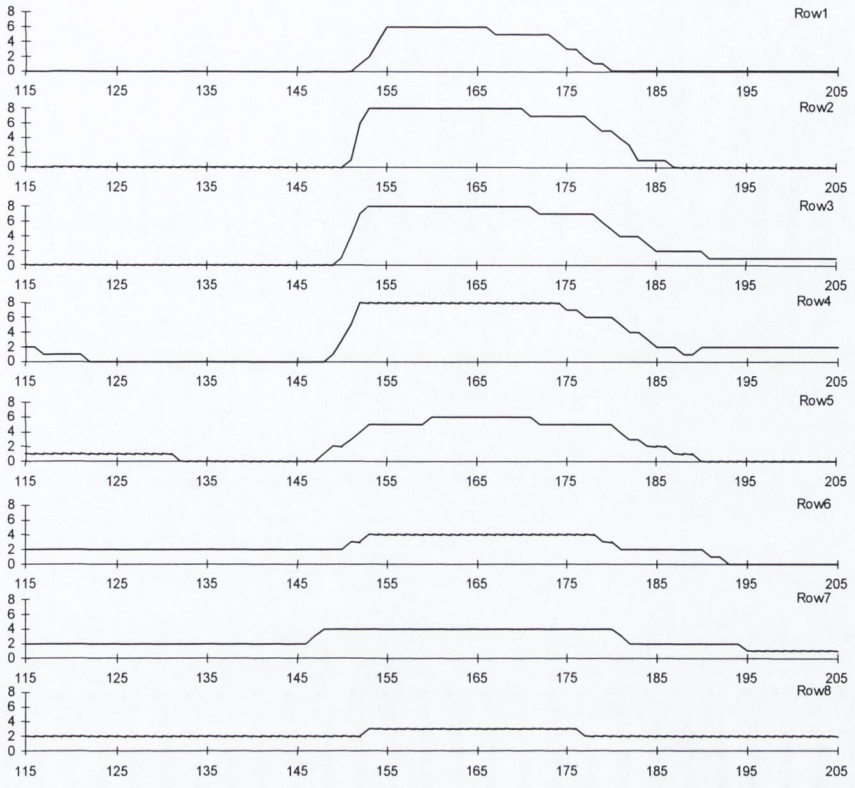



**Figure 5.10** Total number of electrodes activated on each of the eight EPG rows for the speech fragment [#tʸa].

While a finer granularity for discrimination was needed for the assimilation experiment in Chapter 8, such simple total choices seemed a useful starting point for analysis. The simplest such region is a row of electrodes. For each frame in each utterance, the number of electrodes with which the tongue makes contact was counted for each row. A typical display of the eight row totals is shown in Figure 5.10 above. From this, for example, the build-up of contact on each row before consonantal closure (around EPG frame number 152) can be seen. Here we can see that complete closure is achieved on row 2, row 3, and row 4, and it can be assumed that the primary place of articulation is located at these rows. A build-up of contact can also be seen at row 7, which it was presumed is due to secondary articulation. The build-up of contact begins earlier on row 7 than it does on row 3. Such information invites questions such as whether this indicates that the gesture for secondary articulation begins before the gesture for primary articulation or if the difference is due to the arrangement of electrodes on these rows resulting in the secondary articulation gesture being detected earlier.

If several rows are totalled as an index, the selection of rows to total can be based on the arrangement of electrodes on individual palates. In Nolan, Holst & Kühnert, 1995, an index was developed to characterise EPG contact patterns as being [s]-like and [ʃ]-like. The rows were chosen for each of three subjects so as to maximally discriminate between that subject's

canonical [s] and [ʃ] articulations. For one subject the contact on rows 1, 2, and 3 was contrasted with contact on rows 4,5, and 6. For the other two speakers the contrast was between contact on rows 1 and 2 and contact on rows 3 and 4.

However the dependence on rows is by nature crude as information regarding an articulation tends to be spread across several rows and may not involve all electrodes on a given row. Also while some account can be made for variations in the layout of EPG rows it is difficult to account for the exact electrode layout.

#### 5.3.2.3 Subjective region totals

Though the scheme used in placing electrodes is directed towards placing electrodes at anatomical landmarks, it is not sufficient to separate the palate into simple rows or groups of rows. Experimenters have observed that by grouping electrodes that include those where contact is expected for particular places of articulation, contact patterns for different articulations can be more easily discriminated. An example is found in Barry, 1992, and is shown in Figure 5.11 below. Here the experimenter observed that for dental, palatal, and velar articulations in two Russian speakers there were regions of the palate which the tongue came into contact with for only one type of articulation and not for the other two.

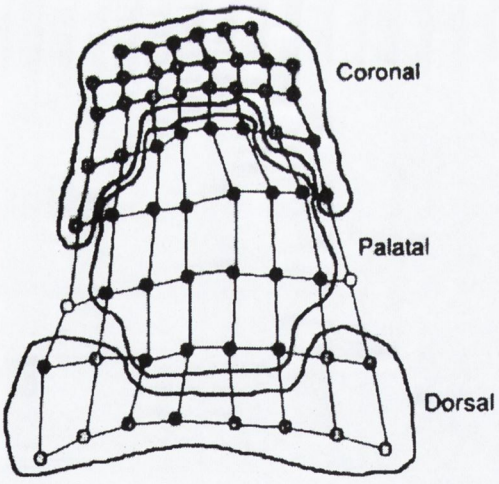



**Figure 5.11** EPG Electrodes grouped into three contact regions - dorsal, palatal, and coronal. (From Barry, 1992.)

Such contact regions are attractive when using the EPG palate to determine the type of articulation occurring as they are designed around contact patterns that speakers produce as opposed to artefacts of the palate layout. Selection of the regions though requires interpretation of the data by the experimenter, and there are issues to be considered. These include selection of data to derive regions from, whether the regions are to be selected for

each speaker or if regions are to be used across speaker, and if it is acceptable for electrodes to appear in more than one region.

For the assimilation experiment in Chapter 8 a selection of discriminatory contact regions for the four lingual stops was investigated. The study involves investigating assimilation of lingual stops involving articulations where the contact patterns have many activated electrodes in common (e.g. for [ t̪ʸ ] and [ t̪ʲ ]). It was found that it is inherently difficult to select data to use to determine contact regions that can in turn be used to discriminate between the contact patterns of stops in assimilatory contexts. Ideally, prototypical examples of each stop being investigated would be used to decide the regions, but by nature such examples cannot be obtained. In the study, approximations to prototypical examples were taken from sentences with each stop in sentence initial, in sentence final position, and in a homorganic cluster across a word boundary.

In deciding whether regions can be used across speaker or should be limited to a particular speaker the actual shape of EPG, and human, palates needs to be considered. Human palates vary in shape, and EPG palates for different speakers are different as a result. The palate manufacturer chooses the exact layout on a palate by palate basis. While the manufacturer is following guidelines, the layout calls for human judgements to be made and the electrode positioning is performed by hand. As a result two different palates can have rather different distances between rows, rows may be crooked, and electrodes on rows can be placed at different distances apart. If the same regions are to be used for multiple speakers then the experimenter must be sure that the places on the palate the regions represent are sufficiently similar. Also, as speakers may use slightly different articulations to produce a given utterance, the selected regions need to take account of this. Thus a certain degree of coarseness is required in selecting regions. On the other hand if the regions are selected for each speaker in a cross speaker study, care must be taken to ensure that those selected for one speaker are not selected on the basis of particular artefacts of speech for that speaker that will not apply for another speaker. (In the study in Chapter 8 only one speaker was involved, so this issue was not considered.)

Whether regions being selected may or may not overlap is determined by how the regions are to be used in analysis. There are two major possibilities in selecting a region to characterise the contact pattern of a particular stop. One is to choose only the electrodes that are activated in every repetition of that stop, (i.e. electrodes that appear to be required for the production of that stop for that speaker). The other is to select all electrodes activated in all repetitions (i.e. electrodes which the tongue may contact in producing that stop). Where many repetitions are involved standard statistical practice of ignoring outlying cases can be used as appropriate.

Choosing regions to be used to discriminate between two stops also involves a number of possibilities. An EPG palate contact pattern which one wants to classify for its similarity to the either of two stops could be compared to regions as just described, and the closeness of match weighted. In that case the regions chosen to describe each stop may overlap. Another possibility, if the intention is to discriminate between two stops, is to select sub-regions from the regions just described and take only those where contact is found for one stop but not for the other. For sub-regions calculated in this way there is no overlap.

A method for selection of contact regions is discussed in Byrd, Flemming, Mueller & Tan, 1995. The authors investigated assimilation of lingual consonants [s], [d], and [g] across word boundaries. The regions selected were based on ten repetitions of [d#d], [s#s], and [g#g] in a carrier phrase "Say baC Cab again". This produced two regions: a front region based on electrodes activated for [s] and [d], and a back region based on electrodes activated for [g]. All electrodes activated during frames of maximum contact which were not activated during the preceding vowel and which would not be considered part of both regions were included. Other electrodes which were never contacted in the control sequences were included in the region that they were physically closest to. Byrd et al., 1995, consider that the contact found in these regions is attributable to an upcoming consonant produced in that region.

In the assimilation study in Chapter 8 many combinations of region selection methods were investigated to determine suitable regions to use when investigating the data. For the final analysis, a method termed "differential contact pattern templates" was developed. This method, discussed below, computes regions for sequences of frames around a time-point such as consonantal closure, thus maximising the discriminatory possibilities of the regions.

### 5.3.2.4 Template regions

One way of defining important contact regions for a particular production is to consider frequency of activation of electrodes for a number of repetitions and to select electrodes showing contact in a minimum percentage of these repetitions as just described. By considering the frequency of contact data presented in Figure 5.8 above, the template regions in Figure 5.12 can be produced. The upper template in Figure 5.12 shows in black the electrodes that are activated in all repetitions of [-a$\underset{\cdot\cdot}{t}^{\gamma}$#$\underset{\cdot\cdot}{t}^{\gamma}$a-].The lower template in Figure 5.12 shows in black the electrodes that are activated in any repetition of [-a$\underset{\cdot\cdot}{t}^{\gamma}$#$\underset{\cdot\cdot}{t}^{\gamma}$a-].The difference between the two templates is useful to illustrate the variation that is seen between the repetitions. This variation is illustrated by the fact that the lower template, which shows electrodes that are activated in any repetition, has considerably more activated electrodes than the upper template. Figure 5.13 shows similar contact pattern templates for [-a$\underset{\cdot}{t}^{j}$#$\underset{\cdot}{t}^{j}$a-].

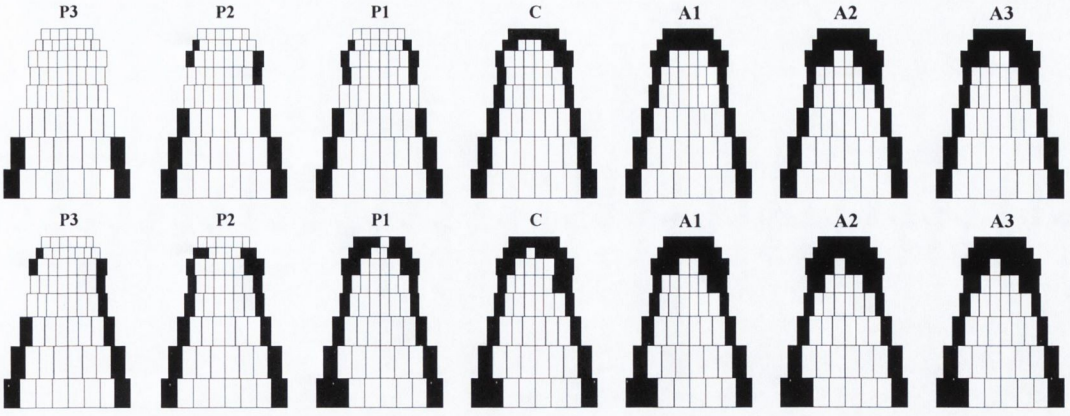



**Figure 5.12** Contact pattern templates for contact found in all repetitions (top row) and in any repetitions (bottom row) for five repetitions of [-a̤tʸ#t̤ʸa-]. Frequency of contact displays for these data are shown in Figure 5.8.

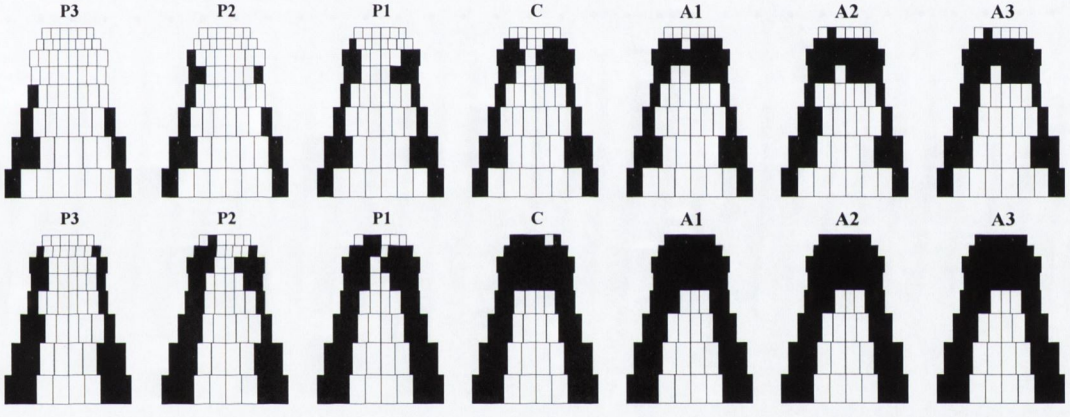



**Figure 5.13** Contact pattern templates for contact found in all repetitions (top row) and in any repetitions (bottom row) for five repetitions of [-a̱tʲ#ṯʲa-].

These template regions serve to describe the contact used in a production by a speaker. They can be used to determine how acceptable a particular EPG contact pattern would be as representing a given production i.e. how similar it is to known productions of a given type. They are also the foundation data for differential contact pattern templates.

### 5.3.2.5 Differential contact pattern templates

When comparing the contact patterns of two different productions there may be regions where these contact patterns overlap. In such cases, the overlapping electrodes cannot be used to discriminate between the two productions. In defining templates which are suited to the task of comparing an EPG contact pattern for its similarity to two prototypical productions, two regions of electrodes are identified: those electrodes that are unique to one prototypical

production and those that are unique to the other prototypical production. These regions are called differential contact pattern templates. Figure 5.14 below shows differential contact pattern templates for [-aṯʸ#ṯʸa-] and [-aṯʲ#ṯʲa-]. This template is defined using the requirement that to be considered archetypal for a given utterance, an electrode must be activated in all repetitions of that utterance and in no repetition of the other utterance. Thus the archetypal [ ṯʲ ] contains all the activated electrodes from the upper contact pattern template in Figure 5.13 that are not activated in the lower contact pattern template in Figure 5.12.

In Figure 5.15 the differential contact pattern templates are used to categorise two repetitions of the utterance [-aṯʸ#ṯʲa-]. At the frame of closure (C), the upper repetition shows activation on the one electrode that is in the archetypal contact for [-aṯʸ#ṯʸa-] and no activation in any of the electrodes in archetypal [-aṯʲ#ṯʲa-]. The second repetition shows no contact matching archetypal [-aṯʸ#ṯʸa-] and 86% contact matching archetypal [-aṯʲ#ṯʲa-].

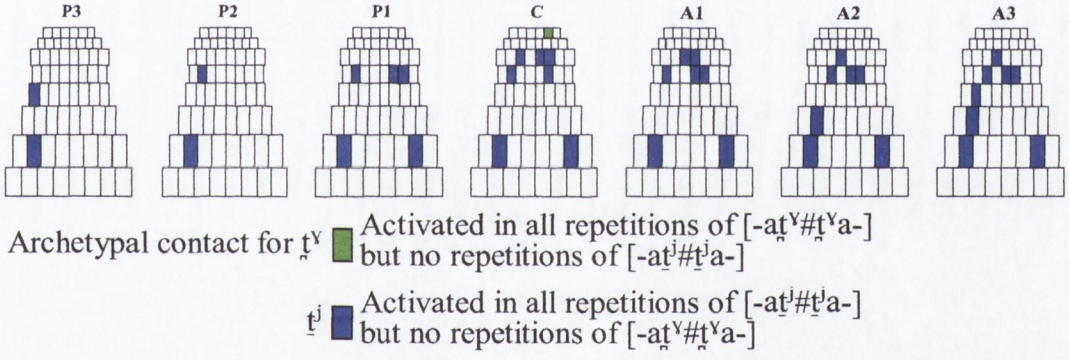



**Figure 5.14** Differential contact pattern templates for [-aṯʸ#ṯʸa-] and [-aṯʲ#ṯʲa-]. If an electrode is coloured green it is found in all repetitions of [-aṯʸ#ṯʸa-] and none of [-aṯʲ#ṯʲa-]. If an electrode is coloured blue it is found in all repetitions of [-aṯʲ#ṯʲa-] and none of [-aṯʸ#ṯʸa-].

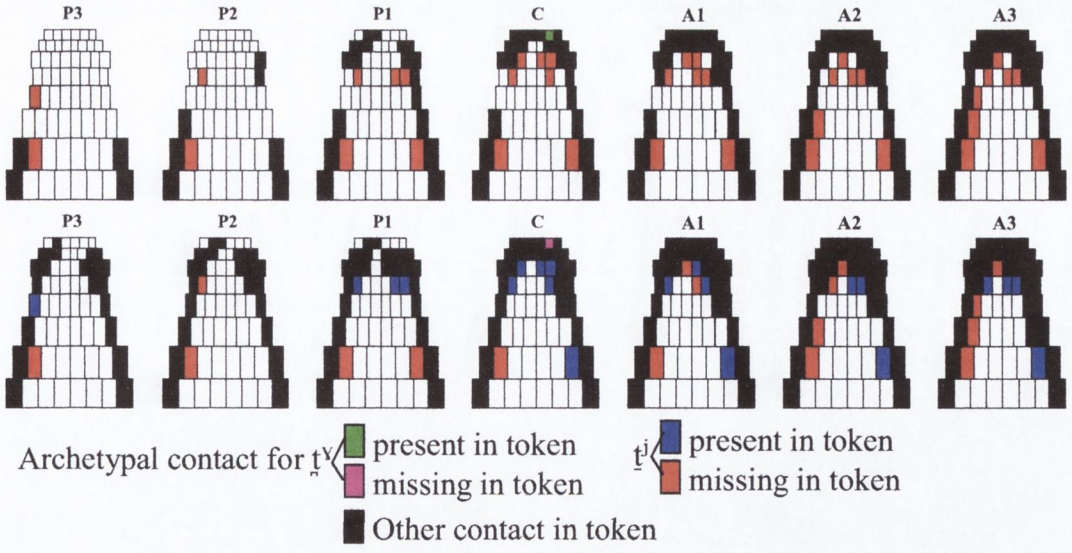



**Figure 5.15** Comparative contact patterns for two repetitions of [-at̪ʸ#t̪ʲa-] with the differential contact pattern template for [-at̪ʸ#t̪ʸa-] and [-at̪ʲ#t̪ʲa-]. The upper repetition is an example with the required activation for the template [-at̪ʸ#t̪ʸa-] and none for template [-at̪ʲ#t̪ʲa-]. The lower repetition does not have the required activation for [-at̪ʸ#t̪ʸa-] but shows most of the activation required for the template [-at̪ʲ#t̪ʲa-].

The proportion of contact in the different contact regions can be used to determine the closeness of the production to other productions. It can also be used to guide decisions as to the type of assimilation, for example, seen in the data. Thus closure for the first repetition is a classical [ t̪ʸ ], whereas closure for the second repetition almost, but not entirely, matches a classical [ t̪ʲ ].

### 5.3.2.6 Conclusion

Having a standard method to represent the important information in an EPG contact pattern as a single number to facilitate further analysis would be valuable. Standard indices do not account for differences in electrode layout on EPG palates. As a result such indices may not capture the information relevant to a particular experiment. Row totals are easy to understand, but relevant electrodes for analysis do not fall into simple rows. Use of regions of relevant electrodes for a particular research question does not suffer from these limitations. Selection of these regions is difficult though. Methods are needed to determine these regions. In Chapter 8 the method of differential contact pattern templates is found to be successful in discriminating between contact patterns for similar productions.

# Chapter 6 Electromagnetic-Articulography (EMA)

EMA is a method of recording the movement of flesh-points on organs such as the tongue. The experimental technique is relatively new (see discussion in Schönle, 1993). The general usage of this experimental technique is discussed in §6.1. Representations of the data can draw on methods developed for other flesh-point techniques like x-ray microbeam. Nevertheless there is need for development of further representations. These issues are discussed in §6.2, along with suggestions and methods developed for this thesis.

## *6.1 General Experimental Technique*

The EMA system used here was the Carstens AG100. A study of the experimental accuracy and setup techniques are presented in Hoole et al., 1993. An example of a recording setup using a 10 sensor Carstens AG100 is shown in Figure 6.1 below. Before each recording the sensors are typically calibrated to compensate for changes over time to the material in the sensors.

In this setup four sensors are dedicated to co-ordinate system references: two for reference throughout the recording and two for a reference t-bar used at the start and end of recording. The two reference sensors recorded throughout the recording session are located on the gum between the upper incisors and on the bridge of the nose. They are used to reduce the measured error caused by the speaker's head moving within the EMA helmet to which the transmitter coils are attached. While the EMA helmet is designed to restrict head movement, some movement is usually detected over the duration of the recording. The post-processing determines, for each individual recorded time-point, the repositioning/translation and the rotation that is needed to map the reference sensors onto the location they were in at the start of the recording. This translation and rotation is then applied to the data for each other sensor at that same time-point, effectively cancelling any effect of head movement (or at least movement of the head along the mid-sagittal plane).

The t-bar is a T-shaped rigid surface with two EMA sensors attached (as illustrated in the bottom right of Figure 6.1). It is custom made for a given speaker to rest steadily and repeatedly at given dental points, usually in the middle of the upper incisors at the front and on chosen molar teeth at the back. Its purpose is to give fixed reference points across multiple recording sessions of a given speaker. The EMA data is, typically, translated so that the sensor at the front of the t-bar is considered to be located at (0, 0), and rotated so that the other sensor on the t-bar is also located on the x-axis. It also gives an excellent reference for

matching separately measured EPG electrode positions, as is discussed under combined analysis.

The other six sensors are available for measurement of points such as movement of the lips, lower jaw, the tongue, or the velum. One sensor may be reserved for tracing the curve of the roof of the mouth.

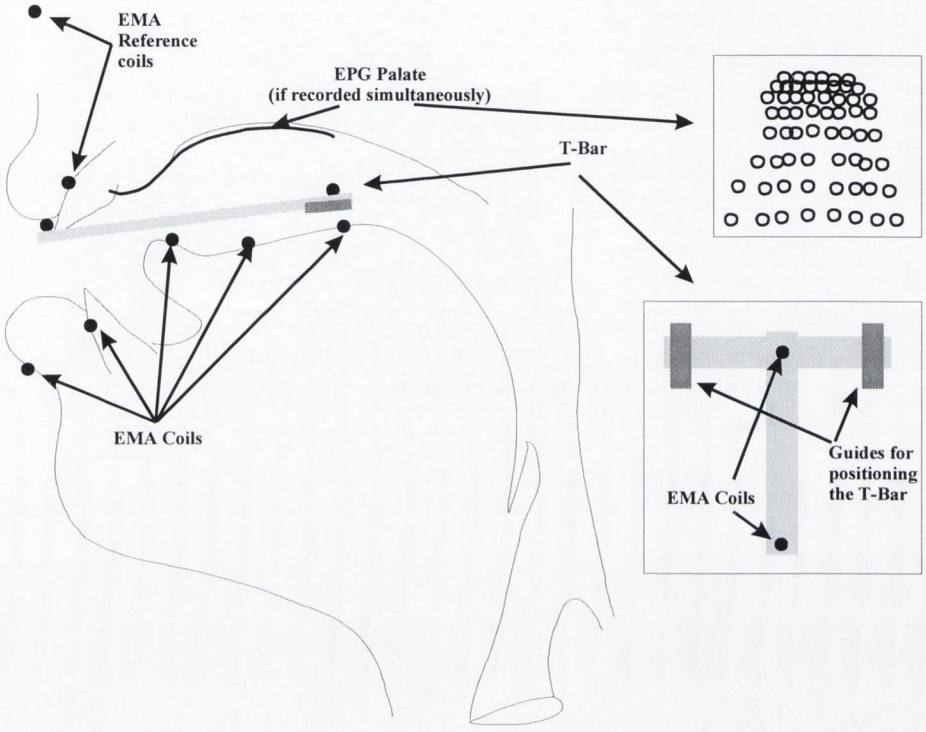



**Figure 6.1** Typical experimental setup for recording EMA and, optionally, EPG. The T-bar is used for calibration and is not present throughout the recording. A helmet (not shown) holding three transmitter coils is placed on the speaker's head.

The movement data from each sensor is recorded directly to computer, and is typically represented as horizontal and vertical movement. Also recorded is an indication of expected error levels which is calculated based on the strength of the signal received from the three transmitter coils. Typical data rates that the data is stored at are around 250 Hz, but this can be varied to match the recording rate of other equipment. For example, when EPG is recorded simultaneously this rate is often reduced to the same rate as EPG (typically 200 Hz).

## *6.2 EMA Data Representation*

A primary concern when discussing analytical practices is to establish an understanding of the limitations of flesh-point analysis. When attached to an articulator such as the lips, whose degrees of freedom of movement are limited, analysis based on the movement of a carefully chosen single point on a lip can be sufficient to discuss the movement of the lip during

speech. However the same is not necessarily true for an articulator that is as complex as the tongue.

Much work has been done by studying the movement of the individual flesh-point sensor that is believed to be the closest sensor to a point of articulation being studied. The experimenter always has to consider that whenever the flesh-point and centre of the place of articulation of interest are not co-located, the movement of the flesh-point will not necessarily match the movement of the tongue at the place of articulation. Nonetheless where the data is considered appropriate there are several methods that are used to aid analysis presented in the literature. Some of these methods are discussed in §6.2.1.

It is difficult to understand the movement of the tongue by observing individual sensor trajectories without some support from additional representations or models. Some work has been done on recovering the contour of the midline of the tongue from a number of flesh-points (Kaburagi & Honda, 1994). For the nomogram study in Chapter 9 the data was not suitable for applying the method discussed in Kaburagi & Honda, 1994, as there were too few sensors recorded. Instead a method was developed, based on established articulatory approximations, to determine tongue settings from EMA sensors on the tongue dorsum. Both the method in Kaburagi & Honda, 1994, and the method developed for Chapter 9 are discussed in §6.2.2.

### 6.2.1 Single Point Analysis

Much analysis of EMA data focuses upon analysing the movement pattern of an individual sensor. The sensor analysed is chosen specifically to suit the study of a particular production. Ideally the sensor studied is at, or very close to, the place of articulation of interest. For example, Kroos, Hoole, Kühnert & Tillmann, 1997, use EMA to investigate the production of German vowels in a CVC context. Where C is /p/, a sensor on the lower lip is analysed; where C is /t/, a sensor on the tongue tip is analysed; where C is /k/, a sensor on the tongue dorsum is analysed.

Both static and dynamic features of the data recorded for a sensor can provide insights into the activity of the tongue. The static position of a sensor is generally examined at a time-point which is annotated based on dynamic features of the EMA data or using another data source (such as a corresponding acoustic signal). For example, Hoole et al., 1993, examine variability on the horizontal (x) and vertical (y) positions of individual tongue sensors for productions of lingual fricatives at a time-point that is considered to be the centre of the fricative. The selection of the time-point was made using the acoustic signal.

To examine dynamic aspects of the movement of a sensor, measures derived from changes in the horizontal position and vertical position of the sensor are typically used. These typically relate to how the position changes in either direction or in both directions simultaneously.

The most easily understood measures are velocity and acceleration of a sensor in a given direction. Velocity is the rate of change of position, and can be simply expressed as $x'_n = x_n - x_{n-1}$, i.e. the rate of change of position at the current time point is the distance between where the sensor is positioned now and where it was positioned at the last sampled time point. Similarly acceleration is the rate of change of velocity which can be simply expressed as $x''_n = x'_n - x'_{n-1}$.

Lövqvist, Gracco & Nye,1993, discuss the difficulties of studying a single dimension of movement of a sensor. They point out that a sensor's velocity can change at different rates and at different time-points in different dimensions. To make use of both the horizontal and vertical movement simultaneously, measures such as tangential velocity and curvature can be used. These depend directly on movement in both directions, and can be described mathematically as: tangential velocity: $v = \sqrt{(x'^2 + y'^2)}$, and curvature: $c = (x'y'' - x''y) / v^3$, where x' is velocity in the horizontal dimension and x'' is acceleration in the horizontal dimension.

If an utterance is annotated using EMA data then the annotation points are typically defined in terms of zero-crossings, local minima and local maxima on a graph of a derived measure against time. For example a zero-crossing of a velocity graph means a sensor was moving (increasingly slowly) in one direction. At the time point where the graph is at zero there is no movement. After crossing through zero this sensor is moving again but in a different direction. A local maximum of velocity, which would show as a zero-crossing on an acceleration graph, is the time-point where a sensor is moving at its fastest for a certain period within an utterance.

It should be noted that in practice the expressions given for velocity and acceleration are not used as they are too sensitive to noise inherent in the digital sampling. There are two approaches to reduce the sensitivity to this noise. One approach is to use different expressions where the sensitivity to such noise is reduced making a calculation based on a greater number of sample points to either side of the time-point of interest. While this results in a value that is less sensitive to noise, it is also not an exact value of the velocity at any given time-point. Another approach is to use sensor positions averaged over several repetitions and to calculate the velocity/acceleration from this average. Hoole et al., 1993, suggest that averaged data is more easily interpreted but that the averages give a somewhat idealised result.

### 6.2.2 Multiple Point Analysis

Where more than one flesh-point is measured simultaneously, it is possible to infer far more about the overall tongue movement than can be inferred from a single sensor. Figure 6.2 below shows the trajectories of three sensors attached to points on the tongue. Joining the trajectories are five lines where each line joins the position those sensors/flesh-points were at for a particular time-point during the utterance / əjə /. Those time-points are: during the preceding vowel, just at acoustic closure, a point between closure and release, just at acoustic release, and a point in the following vowel.

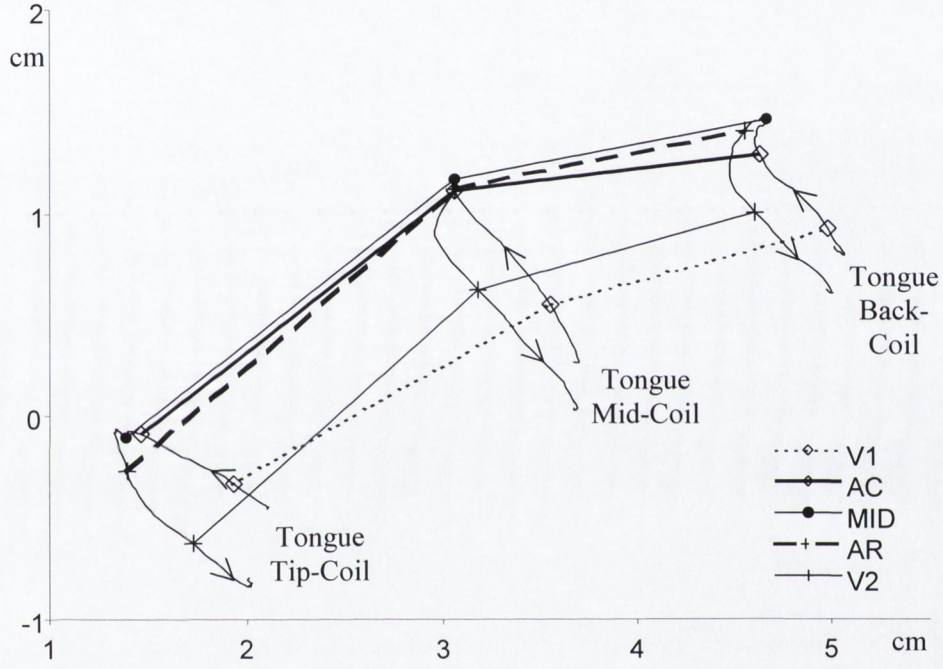



**Figure 6.2** Three curves show the trajectory of three lingual EMA sensors during the production of / əjə /, with small arrow indicating the direction of movement. Straight lines join the locations of the three sensors at five different time points: during first vowel (V1), acoustic closure (AC), mid-closure (MID), acoustic release (AR), and during second vowel (V2).

Concentrating on the point of acoustic closure it can be seen that there is still movement of all three flesh-points just after this time-point. The movement of the mid-sensor is quite small, but movement of the other sensors is initially at a similar rate to that before closure. However the movement of these flesh-points change rather abruptly some time after closure, and the trajectories that they are following changes. An explanation for this continued movement could be that the point of closure is not located precisely at any of the flesh-points. While the tongue at the exact point of closure will have stopped moving towards the palate due to

contact with the palate, other parts of the tongue will continue to move as they are not yet constrained. Gradually points that are near the point of closure will also be stopped through contact with the palate. Points that are further away will slow as there will no longer be muscle activity to cause their movement towards the palate (and due to their being slowed by their being connected to parts of the tongue that are already in contact with the palate).

The three flesh-points do not reach this turning point synchronously, as can be seen in the figure by examining the movement after the time-point after closure. Here the middle flesh-point has come to rest very soon after closure while the other two continue to move towards the palate. This indicates that the middle flesh-point is closer to the place of articulation of the closure than the other two flesh-points as the middle flesh-point is affected more immediately by the closure. The next most affected flesh-point is the rearmost one, indicating that the place of articulation is probably behind the middle flesh-point rather than in front of it.

Such qualitative analysis is useful for insight into flesh-point data for the tongue. It is not useful for quantitative analysis such as inferring the point of articulation. Such measures require a more complete contour of the tongue that the sensors are attached than that yielded by straight lines. Using only flesh-point measurements, recovery of this contour requires a model which can be used to successfully guess the rest of the contour given only a small number of points. Two such models are presented here.

The first is that presented in Honda & Kaburagi (1993). It uses the constraint that the surface of the tongue is a smooth curve that must pass through a number of flesh points, and applied regression analysis to predict the tongue contour. The paper presents comparisons of contours of the midline of the tongue estimated in such a manner from four EMA sensors to contours measured using ultrasound. Figure 6.3 below shows the EMA sensor positions as crosses, the tongue contour simultaneously measured with ultrasound as a dashed line, and the estimated contour calculated using regression analysis as solid curves. The calculated curves match the curves measured with ultrasound very closely. The accuracy of the estimated contour calculated from EMA sensor positions depends on both the arrangement and the number of sensors used. The calculation involved requires ultrasound information being available for similar data for a given subject. This method can thus only be applied where these requirements are fulfilled, but it should be useful in a suitably run experiment.

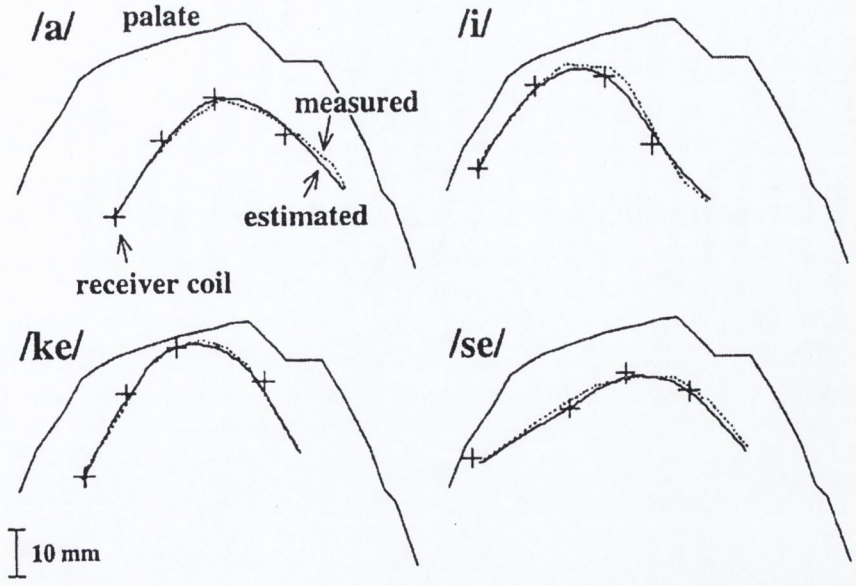



**Figure 6.3** Examples of the match between tongue contours measured using ultrasound and estimated from simultaneously recorded EMA sensor positions. EMA receiver sensors are marked as crosses. (From Honda & Kaburagi, 1993).

Where there are insufficient sensors for regression analysis as just described it is necessary to look to other known constraints and models of the tongue in speech production. For the study presented in Chapter 9 only three tongue sensors were recorded: one near the tip and two further back at approximately 2 cm intervals. To construct nomograms it was necessary to determine the point on the tongue body that was closest to the roof of the mouth. In personal correspondence with Dr. Francis Nolan, the gross representation of the tongue body as a ball during high vowel production was discussed. Such a representation was also used as part of a model of the vocal tract that had been verified by its ability to produce an outline to the midline of the tongue that matched quite closely to x-ray data (Mermelstein, 1973).

In this model a circle of a 2 cm fixed radius, where that radius is the value given in Mermelstein, 1973, is used to represent the tongue body. Where there are two EMA sensors attached to the tongue body, the circle also is considered to be correspondingly attached to the two EMA sensors. Whenever the sensors move the position of the circle can be calculated and the edge of the circle corresponding to the tongue body can be determined. From this we can make measurements of interest such as the location of the nearest point on the tongue body to the roof of the mouth.

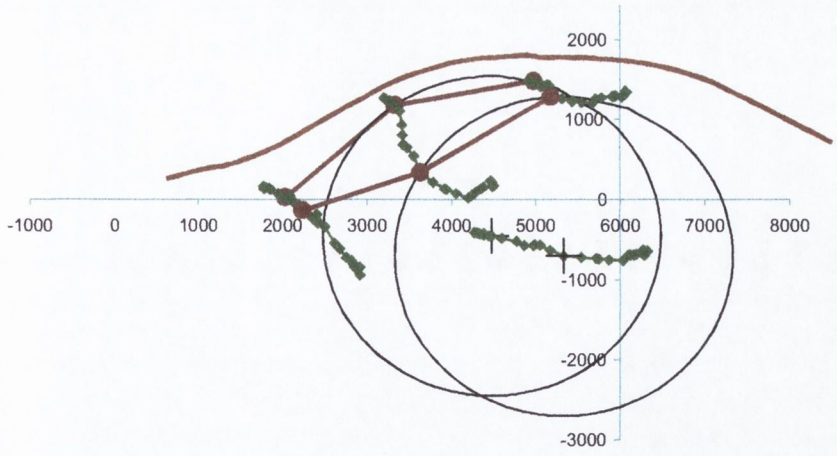



**Figure 6.4** The circle model of the tongue is shown overlaid for two time-points during one speech segment. Each circle of 2 cm radius is fixed to two EMA sensors on the tongue body. The trajectories of EMA sensors over an utterance are shown, as is the corresponding trajectory of the centre of the circle model. Also shown is the curve of the roof of the mouth

In Figure 6.4 the trajectories of 3 lingual EMA sensors are shown for a single high vowel continuum. This is an extract from data presented in Chapter 9 where the aim was to estimate the location of the centre of the lingual constriction from the EMA data. Two time-points on the trajectories are indicated by straight lines joining the sensor positions at those time-points. Considering individual sensors, there is a small amount of horizontal (x) displacement for all three sensors between the two time-points. There is also a striking vertical height difference seen for the middle sensor. As high vowels are produced throughout the utterance (this was verified from the acoustic signal) this vertical movement can be assumed to relate in some way to the overall horizontal movement of the tongue body. While determining this contribution to the horizontal movement is not an easy task when the raw data is thus examined, the circle model inherently considers the contribution. Two circles are drawn attached to the two rear sensor positions for the two time-points. The trajectory for the centre of corresponding circles at every time-point in the data is shown, and a cross highlights the centre of the two illustrated circles. Note that there is greater horizontal than vertical displacement of the centres of the circles, which corresponds well with what would be expected for the tongue body.

This circle model is used extensively in Chapter 9, and further discussion of these particular trajectories is found in §9.3.1. The results that were found using the circle model are further compared to results of the analysis of corresponding EPG data in Chapter 10, where broad compatibility of results is observed.

## 6.3 Conclusion

The movement of a localised part of the tongue can be investigated by studying the movement of an individual EMA sensor. Methods for studying such movement are well established in the literature. Where a sensor is not exactly located at areas of the tongue that are of interest, information about these areas can to some extent be inferred. If a sufficient number of sensors are mounted on the tongue, the overall shape of the surface contour of the tongue can be calculated. However if the number of sensors recorded are not sufficient then other inferences must be made. A method based on modelling as a circle the mid-sagittal section of the tongue body, during the production of high vowels, is a way of estimating the position of the tongue body from two EMA sensors attached to the tongue body.

# Chapter 7 Combining EPG and EMA

EPG and EMA record different aspects of lingual activity. EPG records contact, or lack thereof, between the tongue and 62 points on the roof of the mouth, whereas EMA records the movement of a small number of points on the midline of the tongue. The two devices can be used simultaneously with the experimental techniques and analysis methods such as those discussed in Chapter 5 and Chapter 6. Doing so does not directly take advantage of additional benefits of complementary information available from their simultaneous use. This chapter investigates those how those benefits may be extracted.

Separately EPG and EMA have been used for some time. As a result good experimental techniques to use them have been developed. When they are used together additional steps, such as alignment of the devices, can be taken to allow for better analysis of the resulting data. These steps are discussed in §7.1. Having data from both sources available simultaneously allows for the development of more informative representations. These are presented in §7.2.

## *7.1 General Experimental Technique*

The two measurement devices just discussed, EMA and EPG, yield measurements that can complement each other (see §2.3). EMA gives information about the position of points on the midline of the tongue, regardless of how raised/lowered or retracted/advanced the tongue is. EPG shows contact between the tongue and the artificial EPG palate only when the tongue is sufficiently raised to make contact with the EPG palate. But it gives information about contact that is made towards the sides as well as the midline of the EPG palate.

The recording procedure incorporates the procedures already seen for the individual techniques. But some additional procedures are required to ensure the two sets of measurements can be aligned. For this thesis a number of different recordings of simultaneous EMA and EPG were made using different aids for calibration. The progression of methods are presented here.

### 7.1.1 Guiding location of EMA sensors on the tongue

An EPG palate can be used to assist in the positioning of EMA sensors on the tongue if it is considered advantageous to locate the sensors near certain articulatory targets. The artificial palate, which can be easily inserted into the mouth, can be marked with a non-toxic stain along rows that are considered important for a particular research question. When inserted into the mouth and the speaker makes a swallowing gesture, the stain transfers onto the

tongue. Thus the part of the tongue which is beneath the marked place on the EPG palate is located (at least for this swallowing gesture).

This method of marking the tongue can also be used to help avoid placing EMA sensors at certain locations. If a sensor is located too close to the back of the EPG palate then there is an increased risk of the sensor coming loose during a recording session due to its being dragged across the ridge at the back of the EPG palate. It may also be useful to try to avoid placing a sensor at exactly the place of closure for stop articulations as the sensor may interfere with electrode contact on the EPG palate.

Note this last point is a direct conflict with the general aim of wishing to have an EMA sensor as close as possible to the part of the tongue which is directly involved in producing an articulation which is to be studied. Trade-offs have to be made in the positioning of EMA sensors to allow for the additional effect of the EPG palate being present in the mouth.

### 7.1.2 Alignment of the Measurements of the EPG palate and the EMA co-ordinate system

In the systems described earlier, the locations of the electrodes on the EPG palate were determined outside of the mouth, and separate from any recording. It is necessary therefore to make measurements during a recording session to be able to align the measured EPG electrode positions with the recording of EMA sensor positions. There are several possible ways of doing this, and a combination of methods may be used to allow double-checks to be performed on the alignment.

During recording the midline of the EPG palate can be traced in place in the same way as a normal palate tracing is made. That is by affixing a sensor to the tip of a finger and tracing the sensor along the roof of the mouth while the subject is wearing the EMA helmet. An indication in the trace can be made to mark the location of the back edge of the EPG palate by either resting the sensor at that point for some extra time, or by moving the sensor away from the EPG palate and back again at that point.

Such a trace is known to be difficult to produce with great accuracy and without the speaker's head moving within the helmet. An alternative method that has been used is to trace the EPG palate while mounted in its plaster-cast rather than in the speaker's mouth. This requires a calibration method similar to that for a normal EMA recording. The reference t-bar is placed onto the dental impression at the same dental points as are used on the speaker for whom the EPG palate is designed (see discussion in §6.1). With this reference, the trace can be aligned to the data from a normal EMA recordings.

If the measured locations of the EPG electrodes are accurate, and if they can be located with reference to the plaster-cast dental impression, then EMA and EPG can be easily aligned. The locations of the reference sensors on the t-bar are also known with respect to a number of dental points on the plaster-cast. The position of a number of dental points can also be measured in the same way as the EPG electrode positions are measured, using a sliding callipers as discussed in §5.2.3. The electrode locations can then be geometrically rotated and translated so that the measured dental points have the same co-ordinates as they have in the EMA recording. This is the preferred method of alignment.

As an additional step, the calibration can be verified by examining the EPG and EMA data to see that they match reasonably well together. This can be done by inferring the position of the tongue from data from the two devices. If the calibration was effective then the results from both methods should match closely. A comparison can easily be made if any of the EMA sensors prevent contact between the tongue and an EPG electrode during, for example, a stop closure. This too gives a reasonable correlation point as the EMA sensor location should be directly beneath the blocked EPG electrode. Alignment based on such comparisons would not be as accurate as the measurement methods described above due to the limited resolution of EPG. However if a t-bar was not used during recording, or if accurate palate traces were not made, then such comparisons represent the only option for alignment. If enough correlation points can be found then a reasonable alignment accuracy can result.

To assist the task of aligning the data an application was developed using Microsoft Excel. This displays a number of related pieces of information which may have been measured using different co-ordinate systems. Examples include traces from EMA, contours of the surface of the EPG palate and EPG electrode positions measured using a callipers (see §5.2.3) and additional sources such as traces from x-ray. It provided controls to allow the user to move and rotate the positions of measurements of EPG to visually verify that the adjustments seemed acceptable. It has two possible outputs: suitably rotated EPG electrode locations and a trace of the roof of the mouth. This trace of the roof of the mouth allows for the palate being in place. Traces of the midline of the roof of the mouth may also be taken from x-rays to extend along the velar region that the trace of the EPG palate does not extend over. For a study of vowels as given in Chapter 9 this seems reasonable as the velum is likely to stay quite steady. The resulting trace, which is used in figures such as Figure 7.4 and Figure 9.3, is created using mathematical matching of a polynomial curve to the points given by the various displayed curves. Points on the curve can be produced with the required proximity for later analysis. The curve used for Chapter 9 was an eleventh order polynomial whose roots were determined by least squares. This polynomial order was chosen by visual inspection to give a reasonably smooth curve.

### 7.1.3 Interference between EMA and EPG

In Rouco & Recasens, 1996, it was reported that little significant electrical interference was detected between EPG (Reading) and EMA (Carstens) systems. However the authors also reported that the presence of EMA sensors could physically interfere with lingual contact with the EPG palate. Physical interference has also been observed in data for this thesis. Figure 7.1 shows an example of an unlikely EPG contact pattern that was observed in data from a simultaneous EMA and EPG recording of a palatal stop. Two electrodes on both row 6 and row 7 do not show contact even though there is full contact on row 5 and row 8. This contact pattern is not reasonable. However the two rear EMA sensors on the tongue are located close to EPG row 6 and row 7 during this production.

It seems likely that the EMA sensors are causing the unreasonable contact pattern. Even though the EMA sensors are very small, they may be thick enough to keep the tongue from touching the palate in the area immediately surrounding the sensor. Such contact patterns were not unusual in the data from a simultaneous EMA and EPG recording, but they occurred only when an EMA sensor was close to the electrode(s) effected.

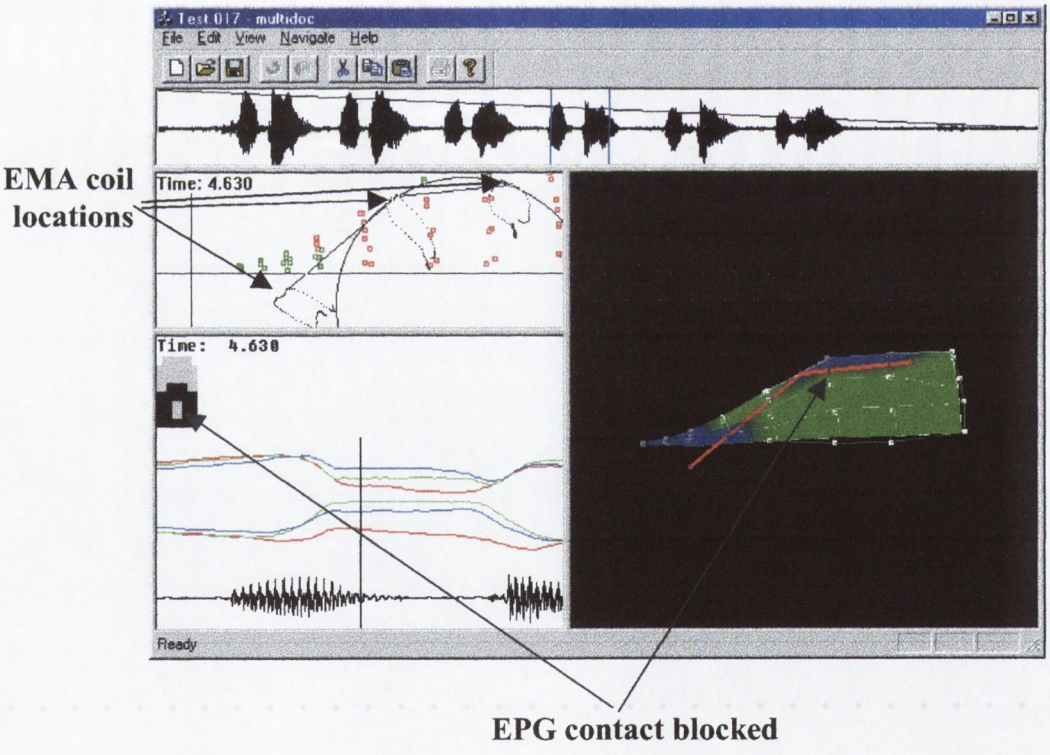



**Figure 7.1** An example, using software introduced in Figure 5.6, of a recording where it seems that EMA sensors are blocking contact with EPG electrodes at the place indicated.

This apparent blocking is unfortunate as it means that data must be carefully (and possibly manually) inspected to check for cases where there is interference with the EPG contact

pattern in an area of interest. However because it only occurs for electrodes near the midline and because EMA sensors are involved, the choice of data to inspect can be guided by EMA sensors coming very close to the EPG palate. It is possible that such an inspection could be automated and a correction to the EPG contact pattern could be applied automatically, but such automation was not attempted in this thesis.

## *7.2 Combined Data Representation*

Both EMA and EPG represent aspects of tongue movement. Presenting an overlay of the data from both sources can help visualise the information available from the data. This may be represented in two dimensions as a side-view or in three dimensions to allow viewing from various angles.

### 7.2.1 Two dimensional representation

A very flexible viewer was developed to display and annotate physiological data. This viewer is capable of presenting several display windows to represent different types of data. The viewer supports data that was acquired simultaneously or that is from different utterances to be displayed side-by-side, or overlaid, thus supporting both intra-utterance and inter-utterance analysis.

The display windows of the viewer are grouped by the utterance for which they are displaying data. An annotation can be created for a given utterance in any waveform window which is displaying data for the utterance. When an annotation is created or moved in one window the annotation is updated in all other windows for the same utterance. The windows can be arranged on screen by the user. The positions of the windows as they are displayed can be saved, and the same layout can be applied to other utterances. The viewer can show waveforms for signals such as acoustics, airflow, and EMA. Filtered versions of such signals can also be displayed to show, for example, velocity or acceleration of an EMA sensor or the centre of gravity of EPG data. EPG frames can also be shown in several ways, with electrodes shown as rectangles, positioned circles, or as balls on a three dimensional wire-frame. The trajectories of EMA sensors can also be displayed with the vertical location of a sensor plotted against its horizontal position (this is called an (x, y) plot in the following discussion).

An example of using the viewer is shown in Figure 7.2. This shows a number of display windows for a single utterance of / əjə / and annotations marking closure (AC) and release (AR) of the consonant as determined from the acoustic signal.

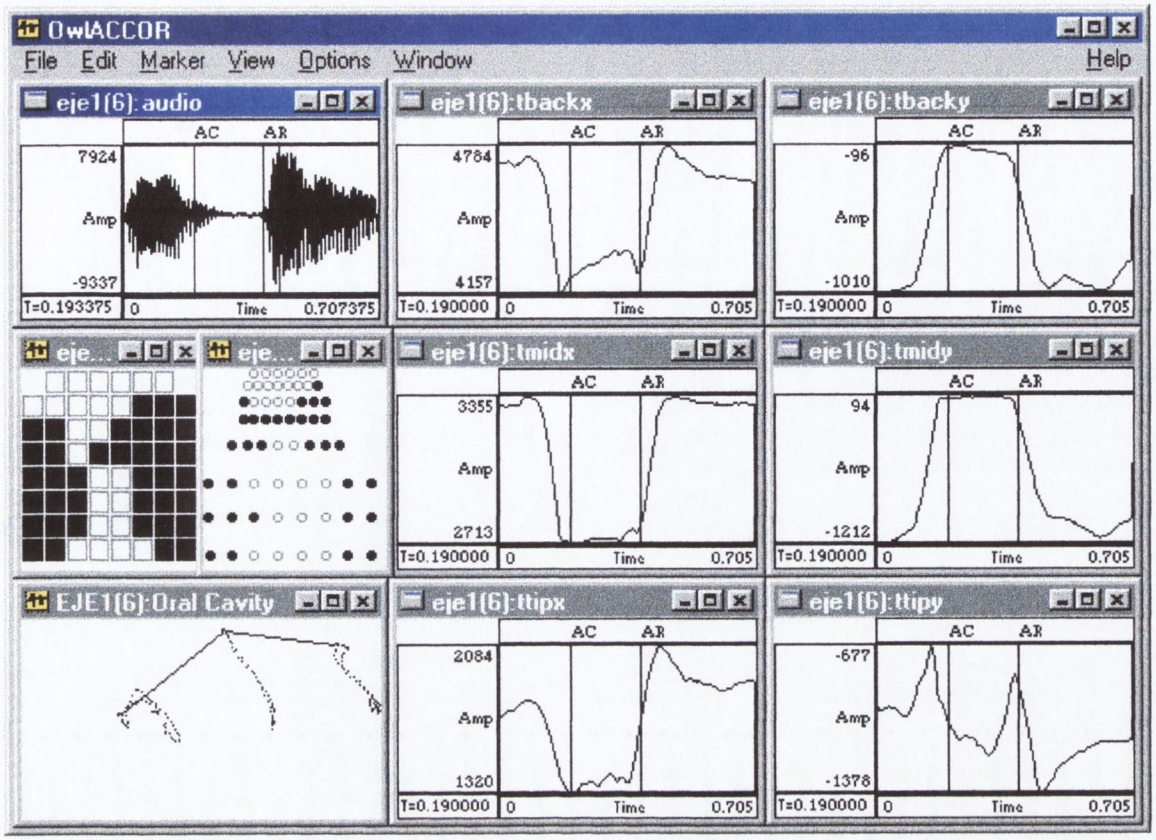



**Figure 7.2** One of the physiological data viewers developed. Top left is the acoustic signal; below that are two EPG displays; below that is a display of the EMA (x, y) trajectories for the EMA sensors displayed as raw movement in the 6 windows to the right (amplitude of EMA signals is in 1/100 mm).

The display windows that are shown for this utterance can be divided roughly into three columns. The window at the top of the first column shows the acoustic waveform for the utterance. Immediately beneath this are two windows showing EPG frames. The one on the left is a simple rectangular display which is locked to, and is displaying the EPG frame at, the annotation AC. The one on the right uses small circles to represent electrodes and places them at points approximate to their location on the palate; the EPG frame shown is for the annotation AR. In the window beneath these is an (x, y) plot of the three sensors on the tongue. The position a sensor holds at any time during the recording is shown as a dot. Straight lines join the sensor locations for the annotated time-point AR.

The second column of windows show the horizontal positions of each of the three sensors. The y-axis in each window gives the distance that the sensor is behind the front teeth. The third column of windows show the vertical positions of each sensor. The y-axis in each window of this column gives the distance that the sensor is below the point where the front teeth meet the gum (that is, the point where one of the reference coils is located).

This viewer gives excellent access to all the data for a recording. It does, however, suffer from usability difficulties as there can be too many displays to easily watch them all. It is useful for repetitive analysis work. But for quick viewing of data another viewer was developed.

This second viewer presents the data in a similar way but makes more use of overlapped displays to reduce the number of places the user needs to look at when browsing data. This is shown in Figure 7.3.

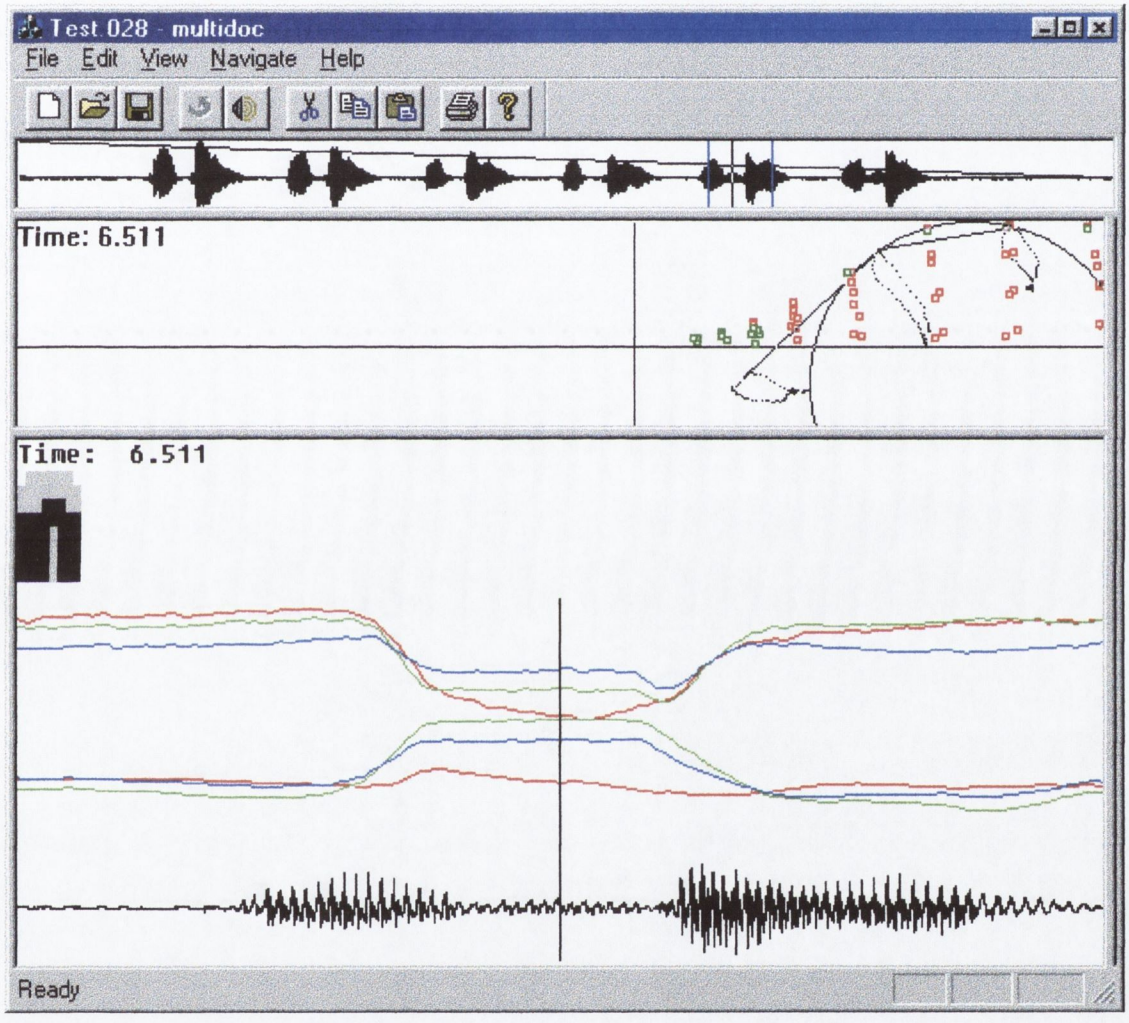



**Figure 7.3** A viewer which shows the acoustic signal in the top window; overlaid EMA (x, y) and EPG palate as seen from the side in the middle window; and the bottom window the EPG contact pattern (viewed from below), and zoomed up displays of the EMA and acoustic data. The lower two displays are aligned to a marker in the top window.

The top windowed area shows the entire audio waveform, where both an active time-point and an active time-span of interest can be selected. Beneath this is a display of the EPG electrodes presented in a side-on view, along with the EMA (x, y) movement for the active

time-span. The EPG contacts colour to show activation and a line joins EMA sensor locations, for the active time-point. In the lower display EMA waveforms are presented as overlays: all x-movement is overlaid in waveforms of different colours. Similarly all y-movement is overlaid. Acoustic data is again shown. The marked time-point in the middle is the active time-point in the audio waveform, and the duration shown is configured for that view. A rectangular view of the EPG electrodes is also presented in the top left of this display.

All these displays can be resized and certain aspects, such as which waveforms should be overlaid, can be set and saved. The data display can be animated. While the acoustic signal is playing through the computer's sound system the screen is continuously updated to show corresponding physiological data. Annotations can be made and saved. A limitation is that data from different utterances cannot be displayed at the same time, but several viewers can be run simultaneously to achieve a similar result.

Along with these viewers, Microsoft Excel was often used for analysis work for this thesis. Tools were developed to extract data from recordings and transfer the information into Excel for analysis. For example, data at annotated points could be transferred to Excel for graphing. To facilitate analysis many Excel macros, formulas, and charts were developed. These were necessary to support the many calculations involved in analysis of EPG, and to allow the data to be easily displayed.

Extracts of a Microsoft Excel based viewer that was developed to represent EPG and EMA data is shown in Figure 7.4. This viewer is used in Chapter 9 and Chapter 10 to present data for discussion.

Figure 7.4(a) represents EMA data, showing the locations of the EMA sensors over an utterance and a trace of the corresponding centre of the tongue circle. Points on the sensor traces are joined to mark a particular time-point. An 'X' marks the point on the roof of the mouth that is closest to the tongue circle at that time-point. Discussion regarding the basis on which the roof of the mouth is drawn is presented in §7.1.3.

Figure 7.4(b) represents EPG contact at the same time-point. An EPG side view in which contact is represented by shaded circles and lack of contact by empty circles. An "EPG tongue contour" is overlaid on the EPG side view. This contour is drawn on the basis of electrode activation. The highest activated electrode in either half of the palate was used as an initial guide: points were calculated that were half way between those electrodes and the next innermost (inactivated) electrodes. A smooth curve (thin black line) was then fitted using the method of least squares to those points, with an end point anchored at the approximate location of the epiglottis. The epiglottis location was taken from an x-ray of the speaker. To the left of the figure is a rotated EPG top view.

Figure 7.4(c) is an result of overlaying Figure 7.4(b) on top of Figure 7.4(a). This complex figure is the corner-piece of the full viewer shown in Figure 7.5.

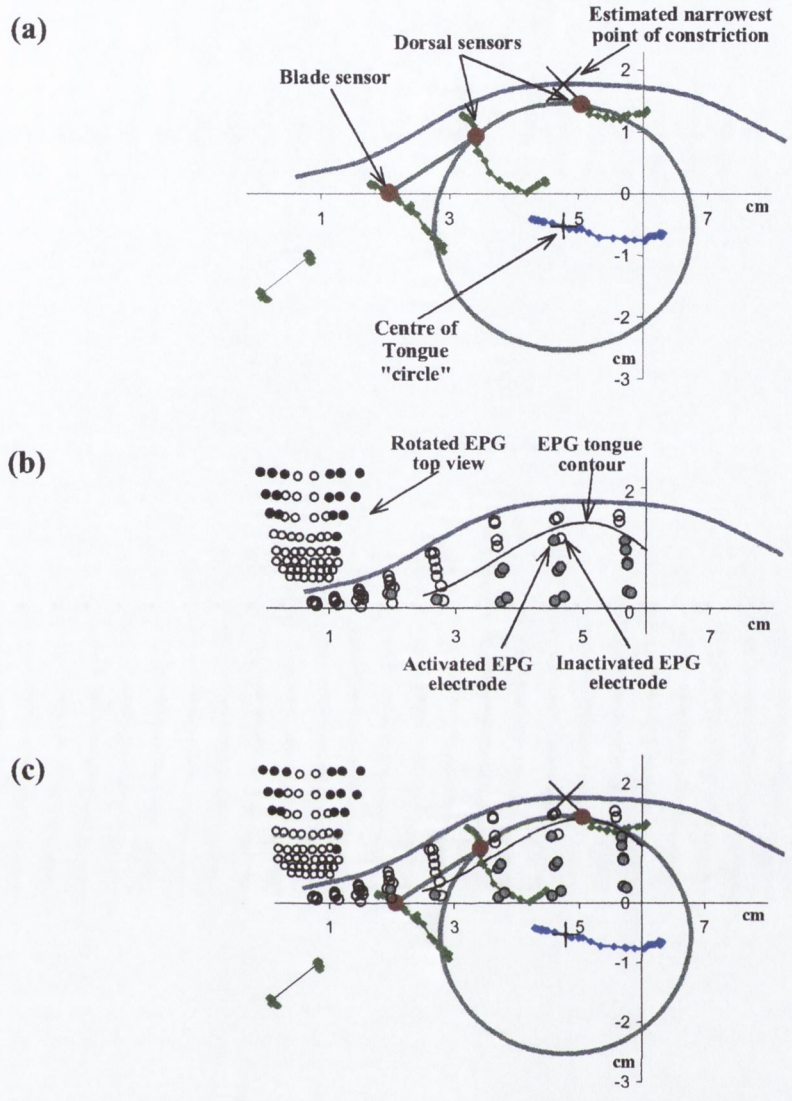



**Figure 7.4** Representation of tongue contours for EMA with tongue circle model (a) and for EPG (b) for a single time-point. (a) and (b) are overlaid in (c). EMA x-y movement for entire [i–ɯ] indicated, along with that for "pseudo-sensor" for tongue circle centre. Rotated top views for EPG also shown.

Figure 7.5 shows a viewer developed in Microsoft Excel. The presentation is dense with information which at first is difficult to follow but with experience proves to be very useful. This viewer shows different aspects of information regarding the acoustic, EMA, and EPG data for a single utterance (in this case of a high vowel continuum between [i] and [ɯ]). The top of Figure 7.5 shows a chart of the first three formants plotted against a selected articulatory reference - in this case the horizontal position of the centre of the tongue body

circle (see §6.2.2) is used. Beneath this is a complex representation of the EMA and EPG data that includes the display just discussed for Figure 7.4, but also includes a tongue contour estimated from EPG (see §5.3.1.1), an indication of the EPG centre of gravity, and an outline of the roof of the mouth if the EPG palate were removed. At the bottom of the display is drop-down list box and a scroll-bar, which allow selection of the utterance and the time-point that is to be displayed in the upper parts.

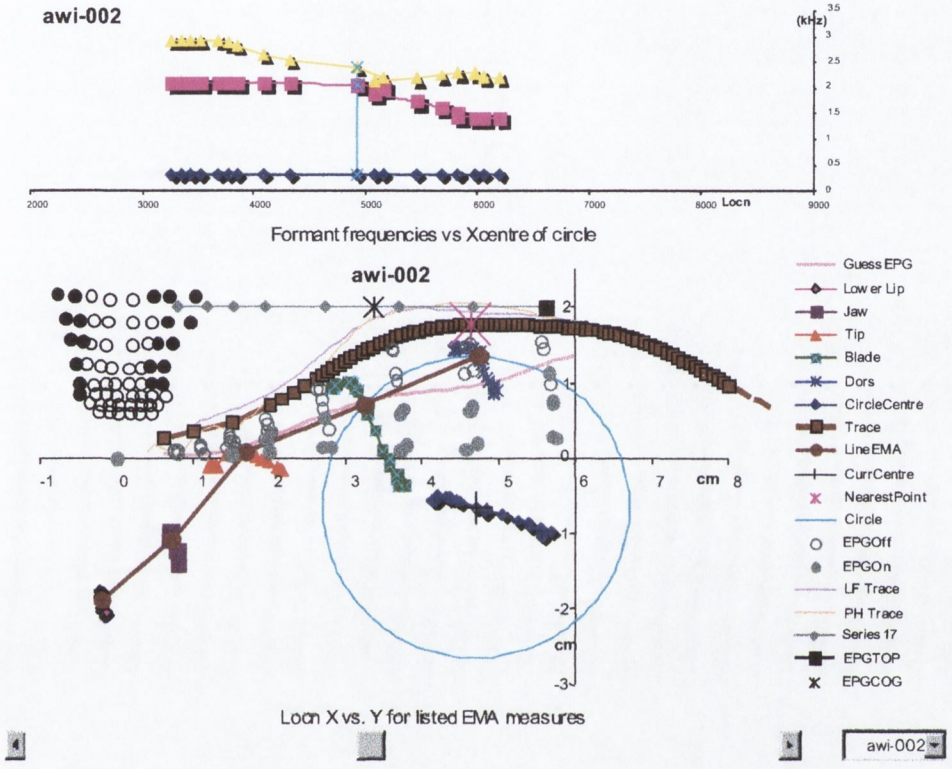



**Figure 7.5** A viewer for EMA, EPG , and acoustic information that incorporates simple representations of that data (e.g. sensor locations, EPG contact patterns, formants) along with estimates from that data (e.g. location of narrowest lingual constriction, EPG contact region.

## 7.2.2 Three dimensional representation

In the discussion of representations of EPG contact patterns in §5.3.1.1, the use of three dimensional displays of EPG data was introduced. When trying to combine the information from EPG and EMA, understanding how the data from both systems must be represented in three dimensions is very important.

Figure 7.6 shows the viewer presented in Figure 7.3 with the addition of a three dimension view of EMA and EPG data added to the bottom right of the figure. This additional display shows a thick line, which joins the three lingual EMA sensors for the given data, in the same

space as a representation of the EPG palate. This display can be freely rotated. To ensure that the line joining the EMA sensors can be seen at all times, and to help the user understand which way the palate is oriented, the palate is shown as a wire-frame whenever the user is seeing the outside/topside of the EPG palate. In contrast to this, the underside of the EPG palate is represented as a solid surface. The user can see through wire-frame and thus see the line that joins the EMA sensor positions along with the solid surface that represents the underside of the EPG palate. The EPG palate is coloured to display the contact pattern. A green colour is used to show contact, and blue colour is used to represent no contact. Where there is contact at an electrode the area half way to the next electrode is also considered to have contact. If a neighbouring electrode does not have contact then the colour fades to the no-contact colour.

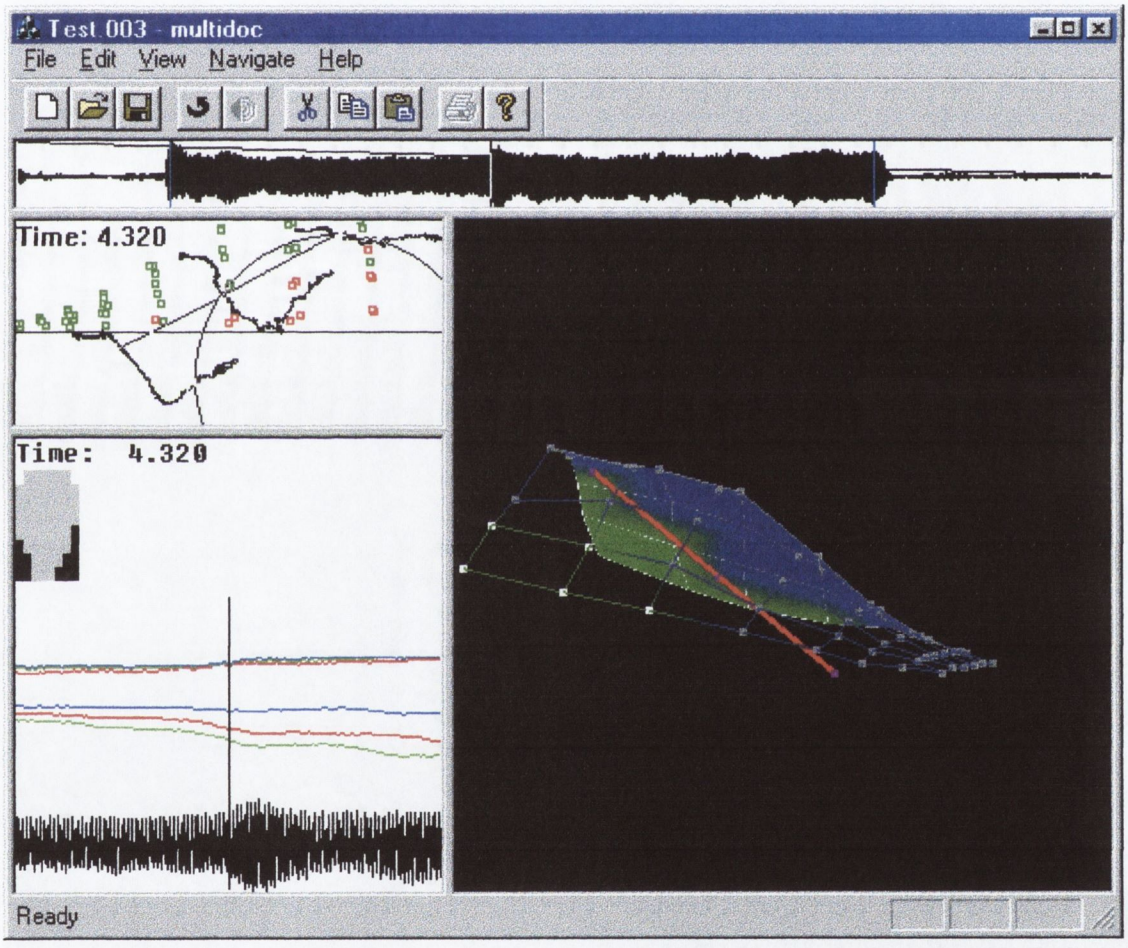



**Figure 7.6** A 3 dimensional representation of EMA and EPG added to the bottom-right of the viewer presented in Figure 7.3.

Figure 7.7 below shows an example which illustrates the three dimensional information contained in the combined EMA and EPG data. Comparing the EMA curve which is on the midline of the tongue with the curve which shows the highest contact with the EPG palate

reveals considerable differences. The midline is considerably higher than the highest contact with the palate. This may appear unreasonable unless one considers the fact that the EMA sensors are on the tongue midline whereas the EPG activated EPG electrodes are in contact with the sides of the tongue. Therefore the EPG electrodes are some distance to the side of the EMA sensors. When the representation in the lower right panel of Figure 7.7 is rotated, as in Figure 7.8, this may be easier to visualise.

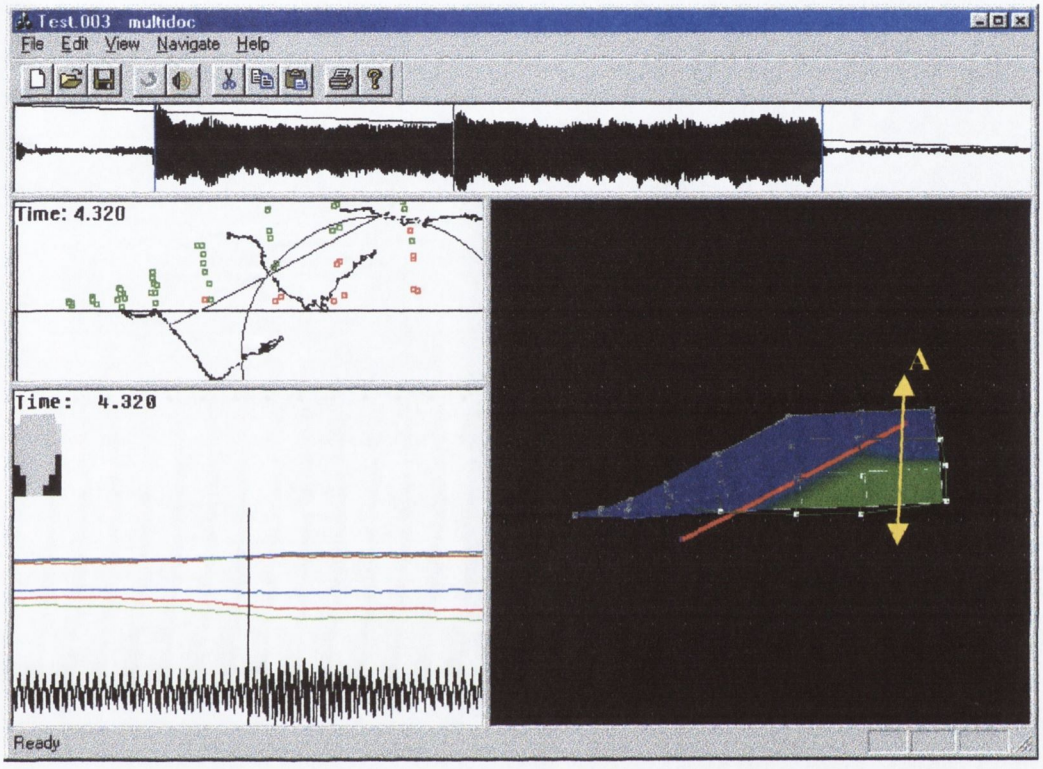



**Figure 7.7** A time-point where the midline EMA data and the EPG contact are at different heights at the line marked (A).

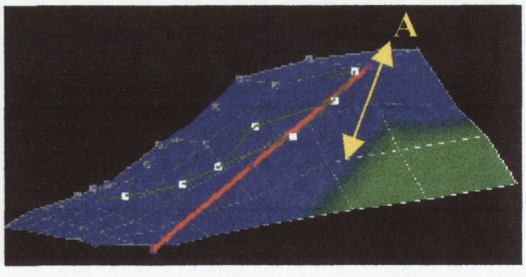



**Figure 7.8** The EPG palate representation from the lower right panel of Figure 7.7 when rotated slightly. The same line (A) is marked.

**Figure 7.9** A sketch of the cross-section of the vocal tract at the point marked (A) in Figure 7.7, showing the upper contour of the tongue and the contour of the palate.

Figure 7.9 above is an impression of a cross section slice through the mouth at (A) in the previous figures, showing the upper contour of the tongue and the contour of the palate. The illustration is based on the EMA and EPG data and on figures from the literature for similar cross-sections. Here we see that the natural tongue shape curves upwards towards the middle, and away from the side-walls of the palate. In Chapter 10 there is further discussion of such phenomena in light of observations of simultaneous EMA and EPG data for the experiment presented in Chapter 9.

# Part III - Analytical Studies

# Chapter 8 An EPG Study of Assimilation of Irish Lingual Stops

## *8.1 Introduction*

There is a four way opposition of lingual stops in Irish, involving differences in secondary articulation (palatalisation and velarisation) as well as differences in the primary place of articulation (see the description of the sound system of Irish in §8.2). The four voiceless lingual stops in Irish are [ t̪ˠ t̪ʲ c k ]. These stops may form clusters across a word boundary. In such clusters assimilation is likely to occur.

The experiment presented here is essentially a re-analysis of data presented in Slattery, 1993. Slattery presents a description of the Irish voiceless lingual stops and an analysis of their propensity to assimilate in $VC_1\#C_2V$ sequences. The data for her study was recorded with Reading EPG as most of the lingual area of interest can be captured using the device. A re-recording of the same materials and subject which included EMA data is used to supplement the description. It was hoped that EMA data would be useful for specific questions regarding the contour of the tongue body, particularly for posterior articulations. However physical interference between the EMA sensors and tongue contact with the EPG palate meant the EMA data is not analysed here but is instead used to supplement some discussion.

Slattery describes the stops using printouts of sequences of EPG frames for individual utterances, frequency of contact for the time-point of maximum contact during the production of the consonants (see §5.3.1) and row totals over the period of production of the consonants (see §5.3.2). For each stop a "representative" row (or rows) was identified where consonantal closure first occurred. In the analysis of non-homorganic clusters contact was examined on the two rows that were representative for the two consonants in the cluster around the time-points of initial closure (to investigate anticipatory coarticulation) and of final release (to investigate carryover coarticulation). The comparison of contact on representative rows was used to classify the similarity of the production of the non-homorganic cluster to the production of the homorganic clusters and thus classify the type of coarticulation observed. However some of the stops have primary places of articulation in overlapping EPG rows. This reduced the effectiveness of classification by representative rows in Slattery, 1993, and required the examination of other EPG rows as well as individual EPG electrode differences to make a classification.

The data in Slattery, 1993, is reanalysed in this study. The stops are described in §8.4.1 using frequency of contact for the time-points of closure, of maximum contact, and of sequences of

frames just before and after closure. The classification in §8.4.2 of any coarticulation observed in the non-homorganic clusters is performed using methods based on differential contact pattern templates (see §5.3.2). Such templates facilitate the comparison of EPG patterns from sets of utterances at the level of individual electrodes and assist in coherent classifications of the type required for such analysis. The reanalysis resulted in some re-classification of the assimilations found, and this is discussed in §8.5. Part of this work was presented at ICPhS 1995 in collaboration with Ní Chasaide (Ní Chasaide & Fitzpatrick, 1995).

## *8.2 The Consonantal System of Irish*

The informant for this study is a female teacher in her twenties from the area of Cois Fhairrge in Co. Galway. This brief outline of the sound system of Irish focuses on the consonant system of the dialect of Irish spoken in that area.

The consonant system of Irish is defined in terms of two distinct sets of consonants: palatalised and velarised. The palatalised set contains both palatal segments and palatalised segments. A palatalised segment is one with a primary articulation at a place other than the palate and a secondary articulation where the front of the tongue is raised towards the hard palate. The velarised set contains both velar segments and velarised segments, where a velarised segment has a secondary articulation where the tongue body is raised towards the soft palate.

The distinction between velarised and palatalised consonants has grammatical significance as well as differentiating the meaning between pairs of words. For example, [ fˠoːdˠ ] and [ fˠoːdʲ ] represent the nominative and genitive cases respectively of the Irish word "fód".

The consonant system of Irish is shown in Table 8.1. Note that the Celticist convention is adopted in Table 8.1 here of showing palatalisation with the diacritic /'/ and leaving the velarised segment unmarked. Also of note is that Note also the use of /L/, /L'/, /N/ and /N'/, which are commonly used for description of Irish but are not IPA symbols:

| | | |
|---|---|---|
| /L/ | voiced velarised dental lateral | [l̪ˠ] |
| /L'/ | voiced palatalised alveolo-palatal lateral | [l̠ʲ] |
| /N/ | voiced velarised dental nasal | [n̪ˠ] |
| /N'/ | voiced palatalised alveolo-palatal nasal | [n̠ʲ] |

The designations in Table 8.1 are based primarily on the EPG based descriptions by Cooper, 1994, for this dialect. Note also the earlier work by de Bhaldraithe, 1975, for an account

which includes further segments which are arguably allophonic realisations of some of the phonemes in Table 8.1.

| | Labial | Dental | Alveolar | Alveolo-palatal | Palatal | Velar | Glottal |
|---|---|---|---|---|---|---|---|
| **Plosives** | b' b<br>p' p | d t | | d' t' | g' k' | g k | |
| **Fricatives** | v' w<br>f' f | | s | ʃ | j ç | ɣ x | h |
| **Nasals** | m' m | N | n' | N' | ŋ' | ŋ | |
| **Laterals** | | L | l' | L' | | | |
| **Vibrants** | | | r' r | | | | |

**Table 8.1** The consonant system of Irish (on the basis of Cooper, 1994).

## *8.3 Data Acquisition*

### 8.3.1 Materials

The data analysed were the four voiceless lingual stops of Irish. In Table 8.1 these are written as / t t' k' k / but in the following discussion the phonetic symbols [ t̪ˠ t̪ʲ c k ] are used. The were looked at in both unassimilating and assimilating contexts. The unassimilating contexts that were elicited were #CV (word initial), VC# (word final), and ˈVC#ˈCV (homorganic cluster where the same stop occurs at either side of a word boundary). The assimilating contexts elicited were ˈ$VC_1$#ˈ$C_2$V. In all cases V = / a / and C is one of [ t̪ˠ ], [ t̪ʲ ], [ c ], or [ k ]. As only anticipatory coarticulation is considered in this analysis the stops elicited in #CV are not used.

The words containing initial variants of the stops are all verbs, and the words containing final variants are nouns. The word pairs containing clusters were noun + verbal past participle. The sentence frame was changed as little as possible resulting in some semantically odd sentences, but helping to ensure constant rhythm and stress patterning.

These utterances were elicited in sentences, and five randomised repetitions of each sentence were recorded. The list of sentences used was the following (with the words containing the stop or cluster of interest underlined):

| | |
|---|---|
| #t̪ʸ- | Tachtann sé |
| -t̪ʸ# | Dúirt sé slat |
| -t̪ʸ#t̪ʸ- | Nach raibh'n slat tachtaithe a'd |
| | |
| #t̪ʲ- | Teagasc é |
| -t̪ʲ# | Dúirt sé plait |
| -t̪ʲ#t̪ʲ- | Nach raibh'n plait teagascaithe a'd |
| | |
| #c- | Ceadaigh é |
| -c# | Dúirt sé faic |
| -c#c- | Nach raibh faic ceadaithe a'd |
| | |
| #k- | Casann sé |
| -k# | Dúirt sé mac |
| -k#k- | Nach raibh'n mac casta leat |
| | |
| -t̪ʸ#t̪ʲ- | Nach raibh'n slat teagascaithe a'd |
| -t̪ʸ#c- | Nach raibh'n slat ceadaithe a'd |
| -t̪ʸ#k- | Nach raibh'n slat casta leat |
| | |
| -t̪ʲ#t̪ʸ- | Nach raibh'n plait tachtaithe a'd |
| -t̪ʲ#c- | Nach raibh'n plait ceadaithe a'd |
| -t̪ʲ#k- | Nach raibh'n plait casta leat |
| | |
| -c#t̪ʸ- | Nach raibh faic tachtaithe a'd |
| -c#t̪ʲ- | Nach raibh faic teagascaithe a'd |
| -c#k- | Nach raibh faic casta leat |
| | |
| -k#t̪ʸ- | Nach raibh'n mac tachtaithe a'd |
| -k#t̪ʲ- | Nach raibh'n mac teagascaithe a'd |
| -k#c- | Nach raibh'n mac ceadaithe a'd |

### 8.3.2 Method

For the main recording on which this analysis is based, the subject was recorded using the Reading/IBM multi-channel's EPG and audio modules. EPG data was recorded at 200 Hz (i.e. one EPG frame every 5 ms) and acoustic data at 20 kHz. The speaker had experience of speaking with the EPG palate in place. A vertical and a side view of the speaker's EPG palate are shown in Figure 8.1. The eight EPG rows have the following approximate places. Due to an asymmetry of the two incisors, row 1 electrodes are either on the teeth (i.e. on a part of the EPG palate partially covers the rear of the incisors) or at the border between the teeth and the alveolar ridge: the precise location of the electrodes is shown in Figure 8.1. Row 2 and row 3 are alveolar. Row 4 would correspond to the junction of the alveolar ridge and hard palate, and we will refer to it as post-alveolar. Row 5 to row 8 cover the hard palate. Row 5 is prepalatal. Row 6 and row 7 are in the mid-palatal region. Row 8 is close to the junction of the hard and soft palate, and it is referred to as post-palatal.

A separate recording was also made which included EMA, EPG and audio data. This was with the same materials, the same speaker, and with 10 randomised repetitions. The EMA data was recorded at 200 Hz with the Carstens system using four lingual sensors. The four lingual EMA sensors were positioned: (1) at 0.5 cm behind the tip; (2), (3) and (4) respectively at about EPG row 5, row 7 and 0.3 cm behind the EPG palate (during a swallow). EPG was recorded at 200 Hz using the Reading EPG2 system. Audio was recorded to DAT and later down-sampled to 16 kHz. The only direct use of this material in this analysis though is for the EMA illustrations.

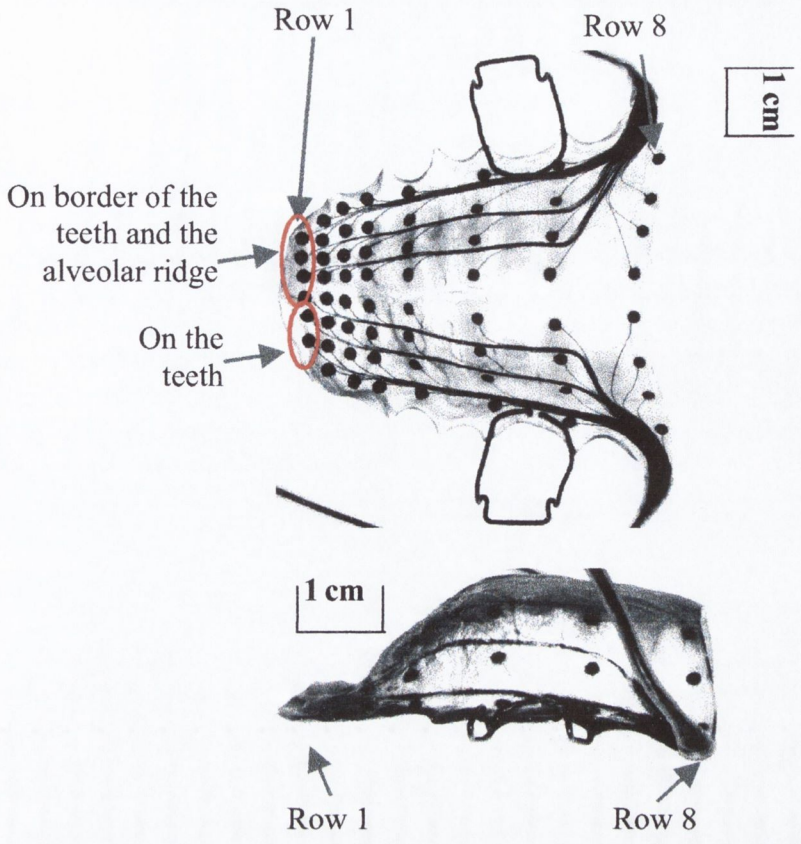



**Figure 8.1** Photographs of the EPG artificial palate used in this study showing the vertical view (left) and side view (right). EPG Row 1 and EPG Row 8 is marked. The scale of the photographs is shown in centimetres.

Returning to the first EPG recording, each utterance was annotated on the basis of EPG contact patterns. The initial location of closure for velar stops was sometimes behind the artificial palate. For these stops the time-point of closure and of release could not readily be identified from EPG contact patterns. In these cases the annotations for closure and release were made on the basis of the acoustic signal.

All repetitions in unassimilating contexts were annotated for the time-points of initial consonantal closure, C, of consonantal release, R, and of maximum lingual contact with the EPG palate during the consonantal closure, M. For repetitions in assimilating contexts an initial inspection was made for each utterance to ascertain whether two distinct primary articulations were present. For cases where two distinct primary articulations were present, the first articulation resulted in either complete closure or in a fricative-like constriction. Where complete closure was observed, time-points of consonantal closure, $C_C$; of consonantal release, $R_C$; and of maximum lingual contact with the EPG palate during the consonantal closure, $M_C$, were annotated. (The second letter, '$_C$', was a single letter code which identified

the first consonant in the target cluster.) Where a fricative-like narrowing was observed the following corresponding time-points were used: the first frame where the maximal narrowing for the constriction was attained was annotated $N_C$; the last time-point with maximal narrowing before widening was annotated $W_C$; maximum lingual contact was annotated $M_C$. The second articulation was annotated the same way. For cases where only one distinct primary articulation was observed the same annotations as for unassimilating contexts were used. It turned out that many of the annotations were the same as those found by Slattery, 1993. However there were a small number of corrections, and additional annotations were made for cases of incomplete closure not considered by Slattery.

The descriptions of the unassimilated stops are based on the VC# context and on the VC#CV context. This choice of contexts was made on the basis that all assimilations were anticipatory. The description in the VC# context is produced from a frequency of contact analysis of the stops at (C) and at (M). The representations as shown in Figure 8.2 are as described in §5.3.1.2, where the outline of the contact areas encloses all electrodes with the given percentage of activation and its proximity to neighbouring electrodes depends on their frequency of activation. The description of the VC#CV context is produced from analysis of the frequency of contact patterns at the time-point of closure and for the preceding 40 ms and following 25 ms.

For the analysis of assimilation non-homorganic consonantal clusters were examined. To decide on the extent to which a given non-homorganic cluster $C_1\#C_2$ had assimilated, homorganic clusters ($C_1\#C_1$ and $C_2\#C_2$) were used as a basis for comparison. Focus was given to the dynamic patterning in the interval from 40 ms before the initial closure (C), or initial narrowing (N), of the cluster to 25 ms after. Differential contact pattern templates (also described in §5.3.2) were used for this comparison. An "essential" contact template was calculated for each EPG frame over this interval, describing that part of the contact pattern which was found in all repetitions of $C_1\#C_1$, but never in any repetition of $C_2\#C_2$. Similarly, it was established on a frame-by-frame basis the "essential" pattern for $C_2\#C_2$ never present in $C_1\#C_1$. This interval for each repetition of the non-homorganic $C_1\#C_2$ was then compared to the "essential" templates. The ratio of activated to inactivated electrodes in the essential contact pattern of $C_1$ and of $C_2$ was used to categorise the observed coarticulation. This process was carried out for every combination of $C_1$ and $C_2$, using the classifications in Table 8.2. Informal listening tests were used at intervals to check that the measurements being made corresponded to potentially important acoustic results.

| Type | Contiguous Articulators | Number of consonants | EPG patterning observed at (first) closure |
|---|---|---|---|
| | | *Complete Assimilation* | |
| **Full** | - | 1 | None of the electrodes in the uniquely $C_1$ pattern activated and all in the uniquely $C_2$ pattern activated. |
| **Weak-ened $C_2$** | Y | 1 | None of the electrodes in the uniquely $C_1$ pattern activated and some uniquely $C_2$ pattern activated. |
| | | *Partial Assimilation* | |
| **Blend** | Y | 1 | Some of the electrodes in the uniquely $C_1$ pattern activated and some uniquely $C_2$ pattern activated. |
| **Coprod-uced** | N | 1 | Some of the electrodes in the uniquely $C_1$ pattern activated but all in the uniquely $C_2$ pattern activated. |
| **Len** | N | 2 | Some of the electrodes in the uniquely $C_1$ pattern activated but the stricture never becomes a stop. **Len1** is a fricative-like stricture held for a relatively long period. **Len2** is a wider stricture of shorter duration. |
| **Hint** | N | 2 | All of the electrodes in the uniquely $C_1$ pattern activated and some uniquely $C_2$ pattern activated. |
| | | *No Assimilation* | |
| **None** | - | 2 | All of the electrodes in the uniquely $C_1$ pattern activated and none uniquely $C_2$ pattern activated. |

**Table 8.2** Classifications of the type of anticipatory coarticulation observed for the non-homorganic clusters.

The goal of these classifications was to reduce the data to make it more amenable to interpretation. Given that the focus is on variability, and that a fine-grained analysis is being used, the challenge is to order and structure the results in a way that can yield generalisations concerning the different types of assimilatory phenomena that can be observed with even this fairly limited data set. The categories as outlined in Table 8.2 were formulated after careful qualitative analysis of the data.

Two things should be noted concerning this table. First of all these classifications refer only to the primary aspects of the consonants in question. In the case of secondary articulation the

behaviour is noted and discussed as a separate issue. Secondly the ultimate decision concerning the classification of any particular utterance was based on the contact pattern for the frame, which is called the focus frame, corresponding most closely to the instant of closure in the first consonant presenting. That is following the notation outlined earlier, the time-point $C_C$ or $N_C$ (in the case where there are clearly two consonants present) or C (for cases where there is only one). The column "number of consonants" states explicitly whether there was one or two consonants clearly present.

It is also of note that the four lingual consonants are not equally independent of each other as two are coronal and two are dorsal. As it has implications for what would be predicted in terms of coarticulation, the column "contiguous articulators" indicates with "Y" the given classification pertains to clusters of coronals or of dorsals only.

The classifications fall into more general groups based on the type of anticipatory articulation they relate to. The groups are complete assimilation, partial assimilation, and no assimilation.

**Complete Assimilation**

**Full**: Contact on the focus frame has a contact pattern with all the essential contact of $C_2$ and none of $C_1$.

**Weakened $C_2$**: This is essentially the same as full, but some of the essential contact for $C_2$ is absent. However the absence of essential $C_1$ contact suggests $C_1$ has been completely assimilated.

**Partial Assimilation**

**Blend**: This is applied only to clusters produced with contiguous articulators. Where both $C_1$ and $C_2$ are coronals this classification is used where some, or all, of the essential $C_1$ and $C_2$ contact is present. As initial contact for [k] is typically behind the palate there are no essential [k] electrodes in the contact template for -c#k- (see Figure 8.6, bottom), so where both $C_1$ and $C_2$ are dorsals the classification is used where some but not all of the essential [c] contact is present.

**Co-produced**: This is applied only to clusters produced with non-contiguous articulators. It is used where the focus frame is for $C_2$ closure and where in that frame there is some essential $C_1$ contact present as well as all the essential $C_2$ contact.

**Hint**: This is applied only to clusters produced with non-contiguous articulators where all the essential $C_1$ contact is present in the focus frame but there is a small amount of the essential $C_2$ contact present also.

**Len**: This is applied only to clusters produced with non-contiguous articulators, where the focus frame is for a fricative-like stricture for $C_1$ (i.e., it was annotated as $N_C$). Len1 is used where a narrow fricative-like stricture is produced for $C_1$, and Len2 is used where a wider stricture is seen.

**No Assimilation**

**None**: Contact on the focus frame has a contact pattern with all the essential contact of $C_1$ and none of $C_2$.

## *8.4 Results*

### 8.4.1 The prepausal stops: VC#

Figure 8.2 presents EPG contact patterns for the unassimilated stops in VC# (i.e., at the end of the utterance) at the points of consonant closure and maximum EPG contact. The figures are drawn using frequency of contact information. The way they are drawn is discussed in some detail in relation to Figure 5.9 of §5.3.1.2. On the left of the figure are frequency of contact displays showing in black the area of contact present in all repetitions of an utterance and in grey the additional contact area for which there is contact in 80% of the repetitions (i.e. where there is contact in four out of five repetitions). For these displays the term "top views" is adopted. On the right are displays that are also drawn using frequency of contact information. The EPG palate is drawn as though viewed from the side, and dark lines are drawn to represent the typical upper limit of contact for the utterance. The lines are drawn to pass through points between electrodes with 100% contact and those with less that 100% contact at a distance that is proportional to the latter contact. The solid line is for the time-point of closure and the dashed line for the point of maximum contact. For these displays the term "side views" is adopted.

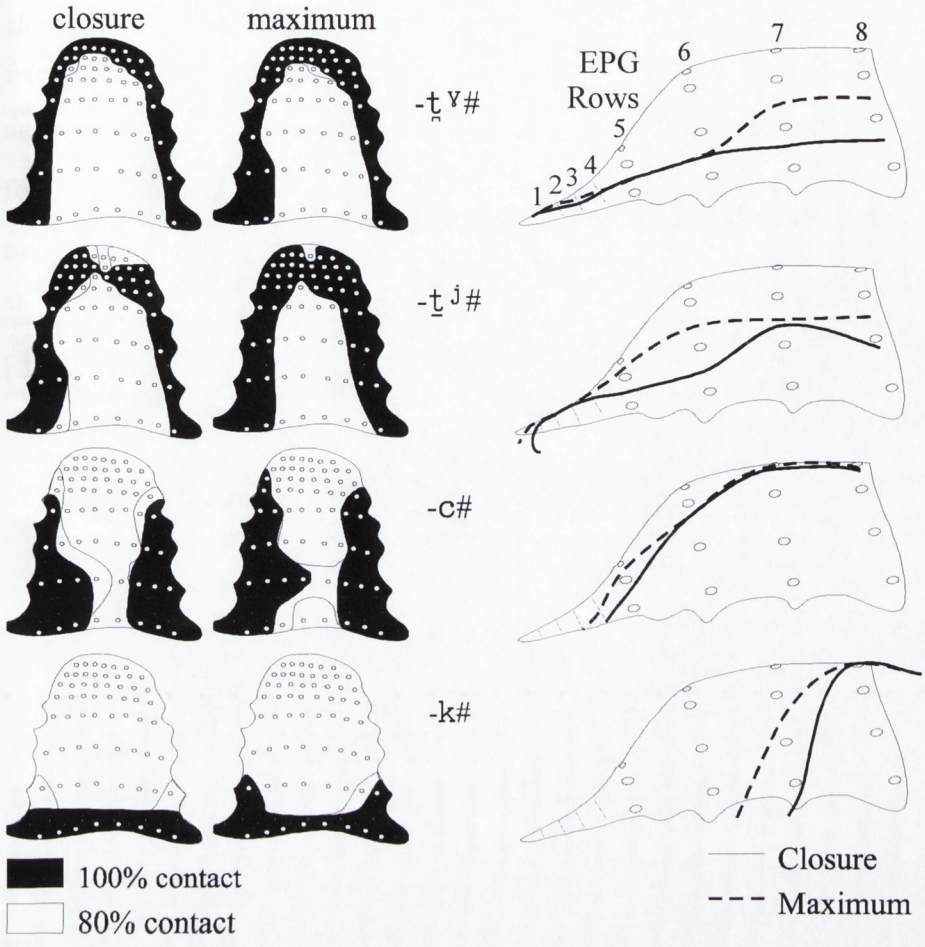



**Figure 8.2** EPG frequency of contact patterns (left) and approximate tongue contours (right) for stops in VC# at the time-point of closure and at maximum contact.

For the coronal stops [ t̪ʸ ] and [ tʲ ] there is considerable overlap in the area of contact at the maximum time point. At the time point of closure though the contact patterns are quite different. The initial point of contact for [ t̪ʸ ] is in the dental region (row 1). During the stop closure this contact spreads onto the alveolar ridge, so that at maximum there is full contact on row 2. It has been pointed out by Catford, 1977, that dentals tend to involve additionally some alveolar contact. For [ tʲ ] the initial contact pattern at closure is more posterior and clearly alveolar (row 2 and row 3) with contact spreading after closure forward and backwards of this onto row 1 and row 4, yielding a very distributed lingual contact. Of note also is the notch-like reduced level of contact at the centre of row 1 and row 2 at closure and persisting in row 1 through to maximum contact. Pandeli, 1993, observed similar patterns in her data and suggests they result from a laminar articulation with lowered tongue-tip. The parallel EMA recordings (made of the same subject and material as this study) offer corroboration of Pandeli's interpretation. The illustration in Figure 8.3 is drawn using the EMA data, and allows the different relative positions of tip and blade for the two stops at

closure to be visualised. Here the average positions (across ten repetitions) of the two frontmost EMA sensors at the time of closure are plotted for the two stops. Lines have been drawn to connect the EMA sensors as a solid line for the averaged [ t̪ʸ ] and as a dashed line for averaged [ t̪ʲ ]. The frontmost EMA sensor is located approximately 1 cm behind the tongue tip. Forward of the frontmost EMA are sketched impressions of where the tongue tip should be located, showing for [ t̪ʸ ] the tongue tip touching the back of the teeth whereas for [ t̪ʲ ] the tip would appear to be lowered. Note also the raising of the tongue blade for [ t̪ʲ ] relative to [ t̪ʸ ].

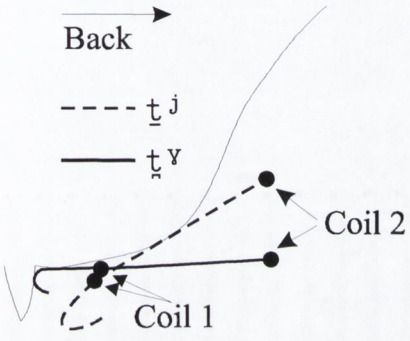



**Figure 8.3** Average positions of EMA sensors (1) and (2) for [ t̪ʸ ] and [ t̪ʲ ] at closure, with representations of the tongue tip impressionistically sketched.

As concerns secondary articulation, the greater raising of the tongue front for [ t̪ʲ ] is evident from the tongue contours on the side views and top views of Figure 8.2. At closure there is greater lateral contact on row 7 and row 8 for [ t̪ʲ ] than for [ t̪ʸ ]. At maximum contact there is some increase of contact on those rows for [ t̪ʸ ] but the contact is even greater for [ t̪ʲ ].

For the dorsal stops, [ k ] and [ c ], one cannot easily differentiate primary and secondary aspects of the articulation. The main stricture for [ c ] is formed with a raised tongue front and a more anterior point of contact than for [ k ]. There was some variability in the precise location of the occlusion for [ c ]. This gives the impression in Figure 8.2 of incomplete closure in row 7 of the EPG palate. For [ k ], much of the contact is likely to be behind the artificial palate.

### 8.4.2 The homorganic clusters: VC#CV

Sequences of frequency of contact patterns for the period from 40 ms before closure to 25 ms after closure for the four homorganic utterances are shown in Figure 8.4.

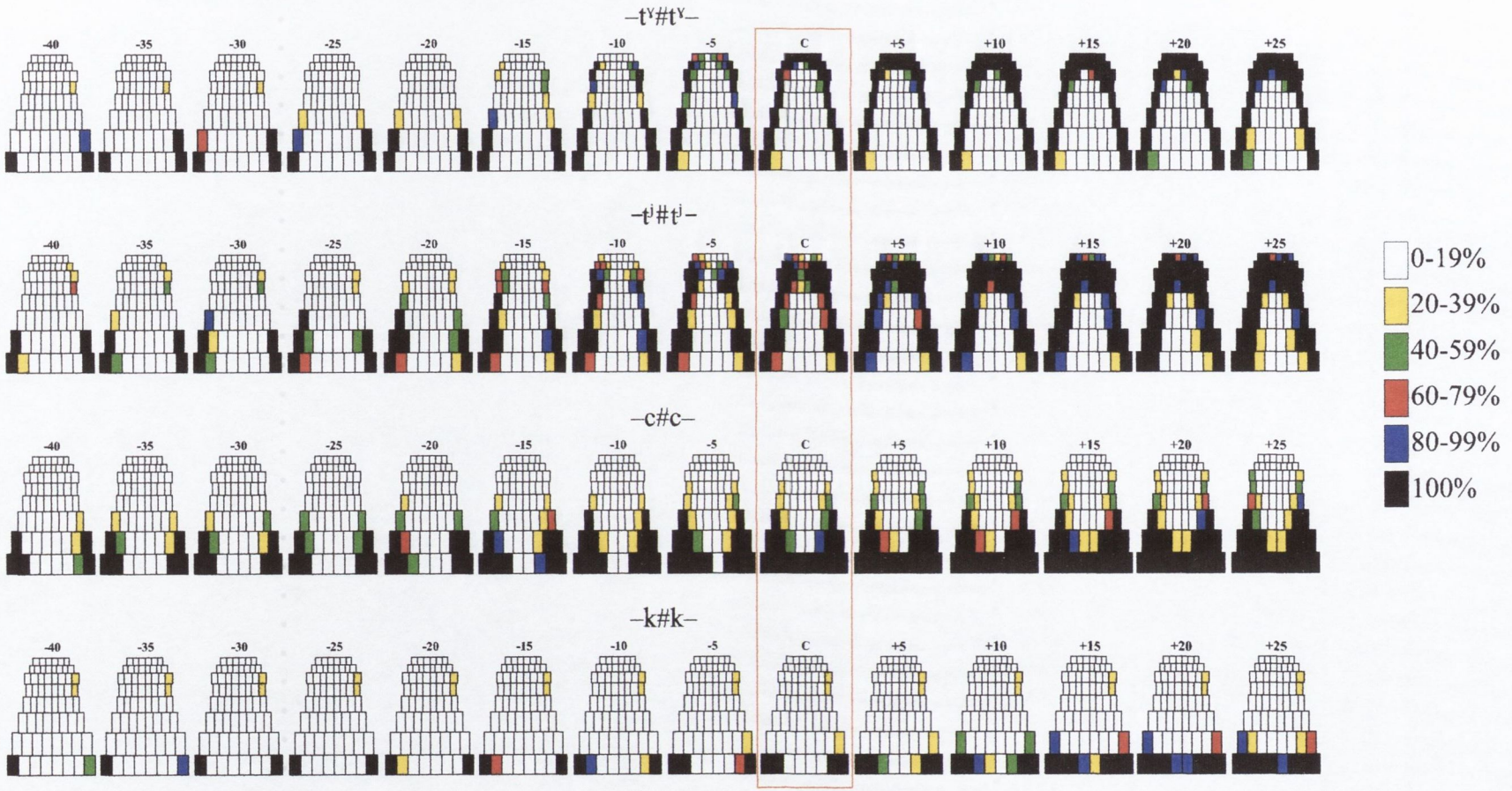



**Figure 8.4** Frequency of contact for homorganic clusters, VC#CV, of the stops at closure and for the interval 40 ms before closure to 25 ms after closure. The frequency of contact for an electrode is indicated by the colour of a rectangle representing that electrode, where the range the colour indicates is given in the legend on the right.

For [ t̪ʸ ] initial closure is always on row 1, spreading onto row 2 and row 3 after closure. This suggests an apical dental articulation, at least initially. The contact is very like that just described for the prepausal allophone. There is contact with the outer electrodes of all rows at closure.

For [ t̲ʲ ] closure occurs on either row 2 or row 3 with considerable contact on both these rows. Contact spreads onto row 1 and row 4 after closure. There is less contact towards the centre of row 1 than towards the sides both at closure and for frames shown after closure. As mentioned earlier, both this notch-like contact and the distributed nature of the contact suggest a laminar articulation (Pandeli, 1993). Again the contact is very like that for the prepausal allophone. Considering secondary articulation, note the greater lateral contact on row 6, row 7, and row 8 for [ t̲ʲ ] than for [ t̪ʸ ]. Note the build-up of contact on these rows 20 - 25 ms before closure compared to the build-up of contact for the primary articulation on row 2 and row 3 which occurs 10 - 15 ms before closure.

[ c ] and [ k ] have somewhat more posterior contact at closure in the homorganic clusters than in the prepausal stops. Closure for [ c ] occurs on row 8 with considerable contact on row 7. Closure for [ k ] is behind the artificial palate with little contact on any row other than row 8.

### 8.4.3 Differential contact pattern templates: $VC_1\#C_1V$ vs. $VC_2\#C_2V$

The differential contact pattern templates in Figure 8.5 and Figure 8.6 are drawn for each possible combination of homorganic clusters using the data presented in Figure 8.4. Electrodes which are activated in all repetitions of $VC_1\#C_1V$ and in no repetition of $VC_2\#C_2V$ are coloured green. Electrodes which are activated in all repetitions of $VC_2\#C_2V$ and none of $VC_1\#C_1V$ are coloured blue. Each template is labelled as $C_1$.vs.$C_2$.

The template t̪ʸ.vs.t̲ʲ (Figure 8.5, top) shows that the production of [ t̪ʸ ] and [ t̲ʲ ] overlap considerably in their contact patterns on row 1 and row 2. At closure there is a single electrode at the centre of row 1 which is unique to [ t̪ʸ ]. There is unique contact for [ t̲ʲ ] on row 3 and row 4, i.e., a more posterior primary articulation. There is also unique contact for [ t̲ʲ ] on row 7 - clearly indicative of front of tongue raising for the secondary articulation/palatalization.

In t̪ʸ.vs.c (Figure 8.5, middle) unique contact for [ t̪ʸ ] at closure is on row 1 and to the sides of row 2 and row 3. Unique contact for [ c ] is on all except the outermost electrodes of row 8, with some spread onto row 7.

In t̪ʲ.vs.c (Figure 8.5, bottom) the contrast at closure is seen in the unique contact for [ t̪ʲ ] on row 2, row 3, and row 4, and unique contact for [ c ] on the inner four electrodes of row 8.

In t̪ʸ.vs.k (Figure 8.6, top) unique contact for [ t̪ʸ ] at closure is on row 1 and to the sides of row 2 and row 3. As closure for [ k ] is behind the artificial palate the only unique contact for [ k ] is a single electrode on row 8.

In t̪ʲ.vs.k (Figure 8.6, middle) the contrast at closure is seen in the unique contact for [ t̪ʲ ] on row 2, row 3, and row 4. There are no unique contact electrodes for [ k ].

In c.vs.k (Figure 8.6, bottom) the contrast at closure is seen in the unique contact for [ c ] on the inner four electrodes of row 8 and electrodes on row 7 and row 6. There is no unique contact for [ k ].

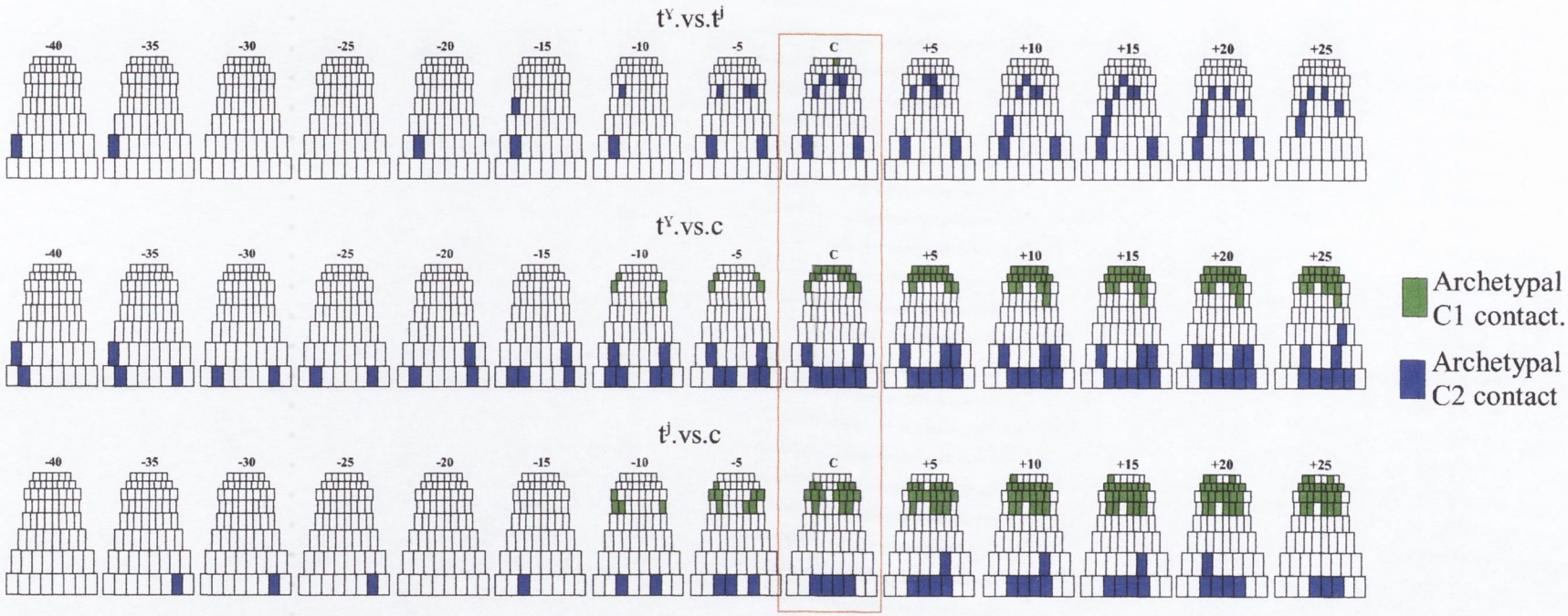



**Figure 8.5** Differential contact pattern templates for [ t̪ʸ ] vs. [ t̪ʲ ] (top), [ t̪ʸ ] vs. [ c ] (middle), [ t̪ʲ ] vs. [ c ] (bottom) based on the frequency of contact for the corresponding non-homorganic allophones shown in Figure 8.4. Archetypal contact for a particular consonant is where a contact is activated in 100% of the repetitions of that consonant and 0% of the repetitions of the other consonant.

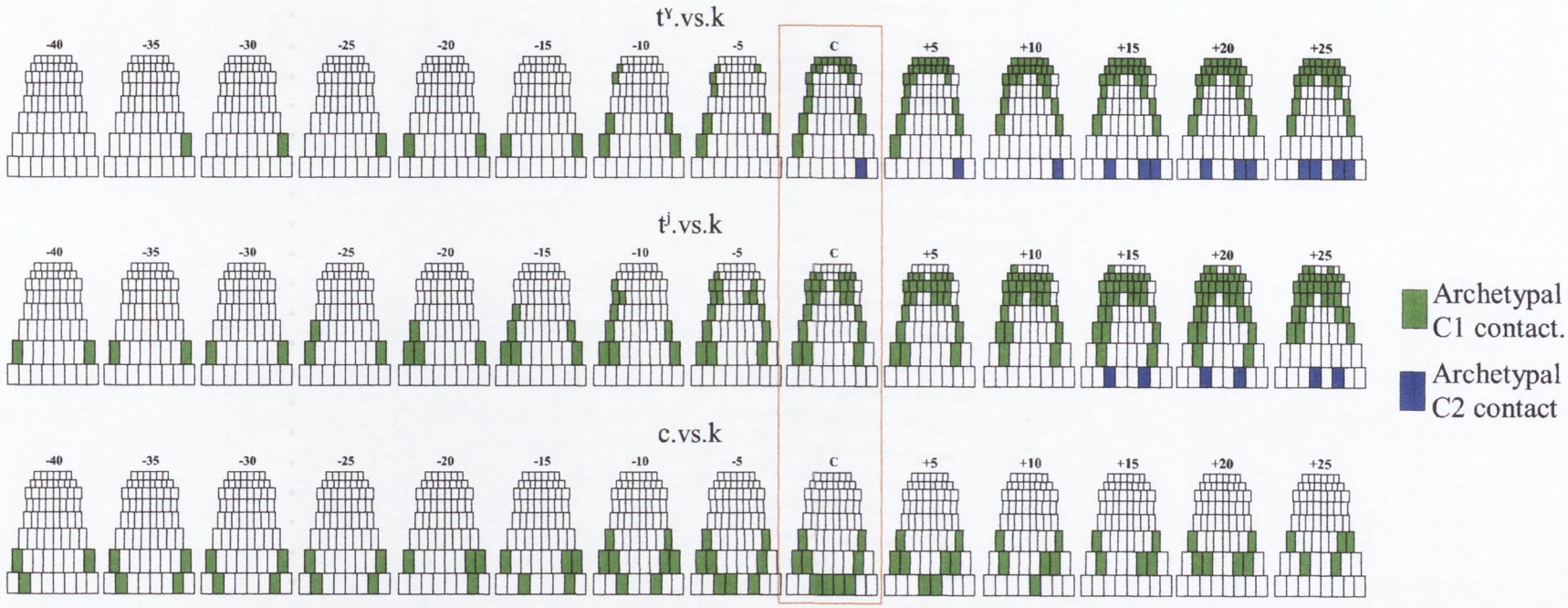



**Figure 8.6** Differential contact pattern templates for [ t̪ʸ ] vs. [ k ] (top), [ t̪ʲ ] vs. [ k ] (middle), [ c ] vs. [ k ] (bottom) based on the frequency of contact for the corresponding non-homorganic allophones shown in Figure 8.4. Archetypal contact for a particular consonant is where a contact is activated in 100% of the repetitions of that consonant and 0% of the repetitions of the other consonant.

### 8.4.4 Non-homorganic clusters: $VC_1\#C_2V$

This analysis focuses on primary articulation at closure. For each repetition of a non-homorganic utterance the corresponding template described in §8.4.3 is applied. This shows which of the unique electrodes for $C_1$ and $C_2$ are present/absent in the contact pattern for the repetition. In Figure 8.7, for example, the differential contact pattern template $\underset{\sim}{t}^{\gamma}$.vs.$\underline{t}^{j}$ is presented at the top of the figure. Beneath it are five repetitions of $V\underset{\sim}{t}^{\gamma}\#\underline{t}^{j}V$ with the template applied. That is for a given repetition colours are used to show how the presence or absence of electrode activation matches the template. In any given repetition, electrodes that correspond to uniquely $C_1$ electrodes (here [ $\underset{\sim}{t}^{\gamma}$ ]) in the differential contact pattern template are coloured green if activated, otherwise they are coloured yellow. Similarly electrodes that correspond to uniquely $C_2$ electrodes (here [ $\underline{t}^{j}$ ]) in the differential contact pattern template are coloured blue if activated, otherwise they are coloured red. Activated electrodes that do not correspond to electrodes in the differential contact pattern template are coloured black. (Note that such electrodes may be activated for some repetitions of the homorganic $C_1$ or $C_2$ clusters at closure, but do not meet the criteria used for generating the differential contact pattern templates. They may also be electrodes that are never activated in any repetition of homorganic $C_1$ or $C_2$ clusters.)

The principal frame focused on here is the instant of closure (C), boxed in orange in this figure, and it is on the basis of this frame that individual repetitions are classified using the schema in Table 8.2. Observations regarding secondary articulation and contact at other time-points are discussed where noteworthy.

**$V\underset{\sim}{t}^{\gamma}\#\underline{t}^{j}V$ (Figure 8.7)**

The stops involve the contiguous articulators of the tongue tip/blade. In all repetitions there is only one clear point of closure. At closure, repetition 1 and repetition 5 do not have the unique [ $\underset{\sim}{t}^{\gamma}$ ] feature and have all the uniquely [ $\underline{t}^{j}$ ] features for the primary articulation. They are thus classified as **fully assimilated**. Note however that in frames before and after closure unique [ $\underline{t}^{j}$ ] contact is missing. So although these repetitions are treated as full assimilations it is clearly not a straight substitution of [ $\underline{t}^{j}$ ] for [ $\underset{\sim}{t}^{\gamma}$ ]. In particular the secondary articulation aspects of [ $\underline{t}^{j}$ ] are missing in the frames -20 to C. In contrast repetition 2 has the unique [ $\underset{\sim}{t}^{\gamma}$ ] feature at closure and does not have any of the uniquely [ $\underline{t}^{j}$ ] features. This is classified as **unassimilated**. Repetition 3 also has the unique [ $\underset{\sim}{t}^{\gamma}$ ] feature but also has 2/5 of the uniquely [ $\underline{t}^{j}$ ] features. As archetypal contact from both

consonants is present at closure this is classified as a **blend**. Repetition 4 does not have the unique [ t̪ʸ ] feature at closure but has 3/5 of the uniquely [ t̲ʲ ] features. This is classified as a **weakened** [ t̲ʲ ] .

**Vt̪ʸ#cV (Figure 8.8)**

For all repetitions two consonantal closures are produced, and all the uniquely [ t̪ʸ ] features are present at the first closure. All except repetition 4 show none of the uniquely [ c ] features at closure and are thus classified as **unassimilated**. Repetition 4 shows 2/7 of the uniquely [ c ] features, indicating a somewhat raised tongue body. This indicates some degree of anticipatory coarticulation: note also that by time-point +25 ms there is considerable narrowing on row 8 not seen in the other repetitions. This repetition is classified as having a **hint** of assimilation.

**Vt̪ʸ#kV (Figure 8.9)**

At closure all repetitions show most of the uniquely [ t̪ʸ ] features and none of the uniquely [ k ] features and are all classified as **unassimilated**. Note that repetition 4 lacks some of the uniquely [ t̪ʸ ] features on row 2, row 3, and row 4 indicating a lower than usual tongue blade and front. A smaller number of features are lacking in repetition 1 and repetition 3.

**Vt̲ʲ#t̪ʸV (Figure 8.10)**

The stops involve the contiguous articulators of the tongue tip/blade. In repetition 1 a fricative is produced before full consonantal closure is attained. The focus frame has all the uniquely [ t̲ʲ ] features and none of the [ t̪ʸ ] features and is thus classified as a lenition (**len1** as the fricative gap is narrow). In repetitions 2 through 5 there is only one clear point of closure and all have the uniquely [ t̪ʸ ] feature. Repetition 2, repetition 3, and repetition 5 have some of the uniquely [ t̲ʲ ] features of primary articulation: 2/5, 1/5, and 4/5 respectively. These are classified as **blends**, though there is obviously some difference in the degree of blending. Repetition 4 does not have any of the uniquely [ t̲ʲ ] features of primary articulation and is classified as a **full assimilation**. Regarding secondary articulation it is noteworthy that palatalization is retained in the three blends whereas for the lenition and the full assimilation the uniquely [ t̲ʲ ] features on row 7 are reduced at closure/the focal frame and in the preceding frames.

**Vt̲ʲ#cV (Figure 8.11)**

In repetition 1 and repetition 2 two consonantal closures are produced. In both repetitions the first closure has all the uniquely [ t̲ʲ ] features and none of the [ c ] features and are classified

as **unassimilated**. In repetition 3 a narrowing is produced with 10/15 of the uniquely [ t̠ʲ ] features and 1/4 of the uniquely [ c ] features, and is classified as a lenition (**len2**). In repetition 4 and repetition 5 closure is clearly produced in the palatal region. However there is also considerable activation of uniquely [ t̠ʲ ] features. These are classified as **co-produced**. It should be noted though that the term co-produced here draws attention particularly to the simultaneous timing. The degree of stricture is not the same: full closure is achieved for $C_2$, along with different degrees of stricture for $C_1$, though not full closure.

**Vt̠ʲ#kV (Figure 8.12)**

In repetition 2 two consonantal closures are produced, and the first closure has all the uniquely [ t̠ʲ ] features. It is thus classified as **unassimilated**. In repetition 1, repetition 3, and repetition 5 there is a narrowing in the area of the primary articulation of [ t̠ʲ ] without achieving full closure. The fricative-like gap is narrow in repetition 1 and this is classified as **len1**, whereas in the other two repetitions the gap is wider and they are classified as **len2**. Repetition 4 has none of the uniquely [ t̠ʲ ] features in the primary place of [ t̠ʲ ] articulation and is classified as **fully assimilated**. Considering the secondary articulation for [ t̠ʲ ], corresponding contact is seen in all repetitions, even the "fully assimilated" repetition 4, suggesting that the secondary articulation is more resistant to assimilation that the primary place for [ t̠ʲ ].

**Vc#t̠ˠV (Figure 8.13)**

In repetition 1 through to repetition 4 two consonantal closures are produced, and in all cases the first closure has all the uniquely [ c ] features and none of the [ t̠ˠ ] features. They are therefore classified as **unassimilated**. Repetition 5 has most of the uniquely [ c ] features and none of the [ t̠ˠ ] features, but with a fricative-like gap on row 8 (with fricative-like noise observed in the acoustic signal). This is classified as **len1**.

**Vc#t̠ʲV (Figure 8.14)**

In repetition 3 and repetition 5 two consonantal closures are produced, and in both cases the first closure has all the uniquely [ c ] features and none of the [ t̠ʲ ] features. They are therefore classified as **unassimilated**. In repetition 2 two consonantal closures are produced, and the first closure has all the uniquely [ c ] features and 2/15 of the uniquely [ t̠ʲ ] features and is classified as having a **hint** of assimilation. It is noteworthy that full closure in the primary articulation region for [ t̠ʲ ] is achieved 20 ms after the first closure. Repetition 1 has most of the uniquely [ c ] features and none of the [ t̠ʲ ] features, but with a fricative-like gap

on row 8 (with fricative-like noise observed in the acoustic signal). This is classified as **len1**. Repetition 4 has 1/4 of the uniquely [ c ] features and all of the [ t̠ʲ ] features and is classified as **co-produced**.

**Vc#kV (Figure 8.15)**

The stops involve the contiguous articulators of the tongue front/body. In all repetitions there is only one point of closure that was identified from the acoustic signals. Because closure was always behind the palate this suggests [ k ] like productions. However all repetitions have uniquely [ c ] features so they are classified as **blends**. Note however that repetition 1 has 6/9 uniquely [ c ] features whereas the other four repetitions only have 1/9 uniquely [ c ] features. So although these are all characterised as blends it is worth noting that in all but the first repetition the balance of the blend is weighted heavily towards $C_2$ (i.e., [ k ]).

**Vk#t̠ʸV (Figure 8.16)**

In all repetitions two consonantal closures are produced, and the first closure has all the uniquely [ k ] features and few of the [ t̠ʸ ] features. The row 3 contact in repetition 1 and repetition 3 appear to be spurious EPG contact (e.g. caused by a saliva thread). The [ t̠ʸ ] feature on row 7 of repetition 1 and repetition 5 seem to be too posterior to be considered indicative of [ t̠ʸ ] coarticulation. The five repetitions are thus classified as **unassimilated**.

**Vk#t̠ʲV (Figure 8.17)**

In all repetitions two consonantal closures are produced, and the first closure has all the uniquely [ k ] features and few of the [ t̠ʲ ] features. The uniquely [ t̠ʲ ] features in all repetitions are on row 7 suggesting assimilation of palatalisation. However none of the uniquely [ t̠ʲ ] features in the area of primary articulation of [ t̠ʲ ] are activated so all five repetitions are classified as **unassimilated**.

*Vk#cV (***Figure 8.18***)*

The stops involve contiguous articulators of the tongue front/body. In all repetitions there is only one point of closure that was identified from the acoustic signals. Repetition 4 has 1/9 uniquely [ c ] features on row 7 and is classified as a **blend**, although the balance is strongly towards [ k ] (note also that there is little uniquely [ c ] contact in frames before or after closure). The other four repetitions have no uniquely [ c ] features and are considered **unassimilated** at closure.

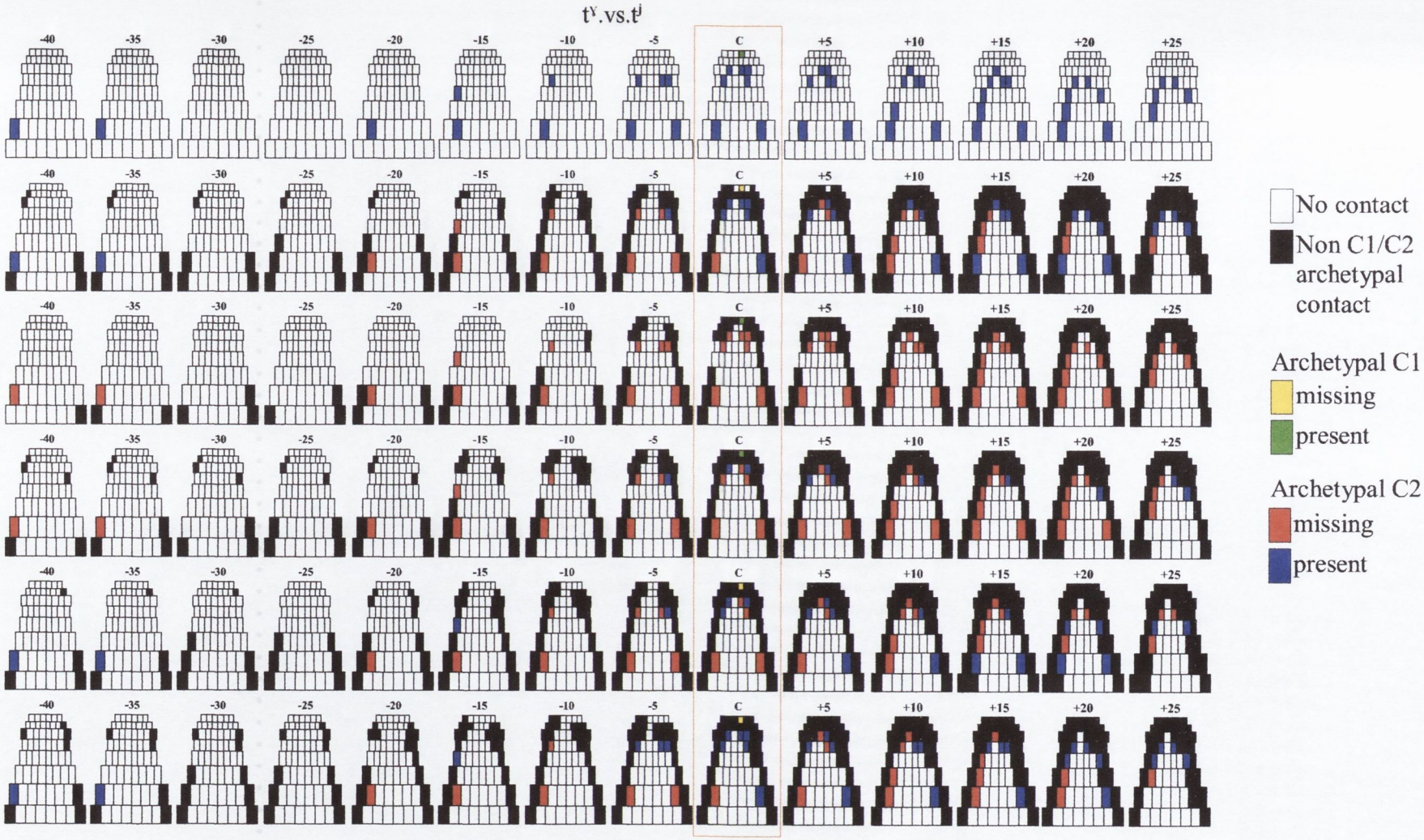



**Figure 8.7** t̪ʸ.vs. t̪ʲ template (top row) and comparative contact patterns for five repetitions of Vt̪ʸ#t̪ʲV.



– 101 –

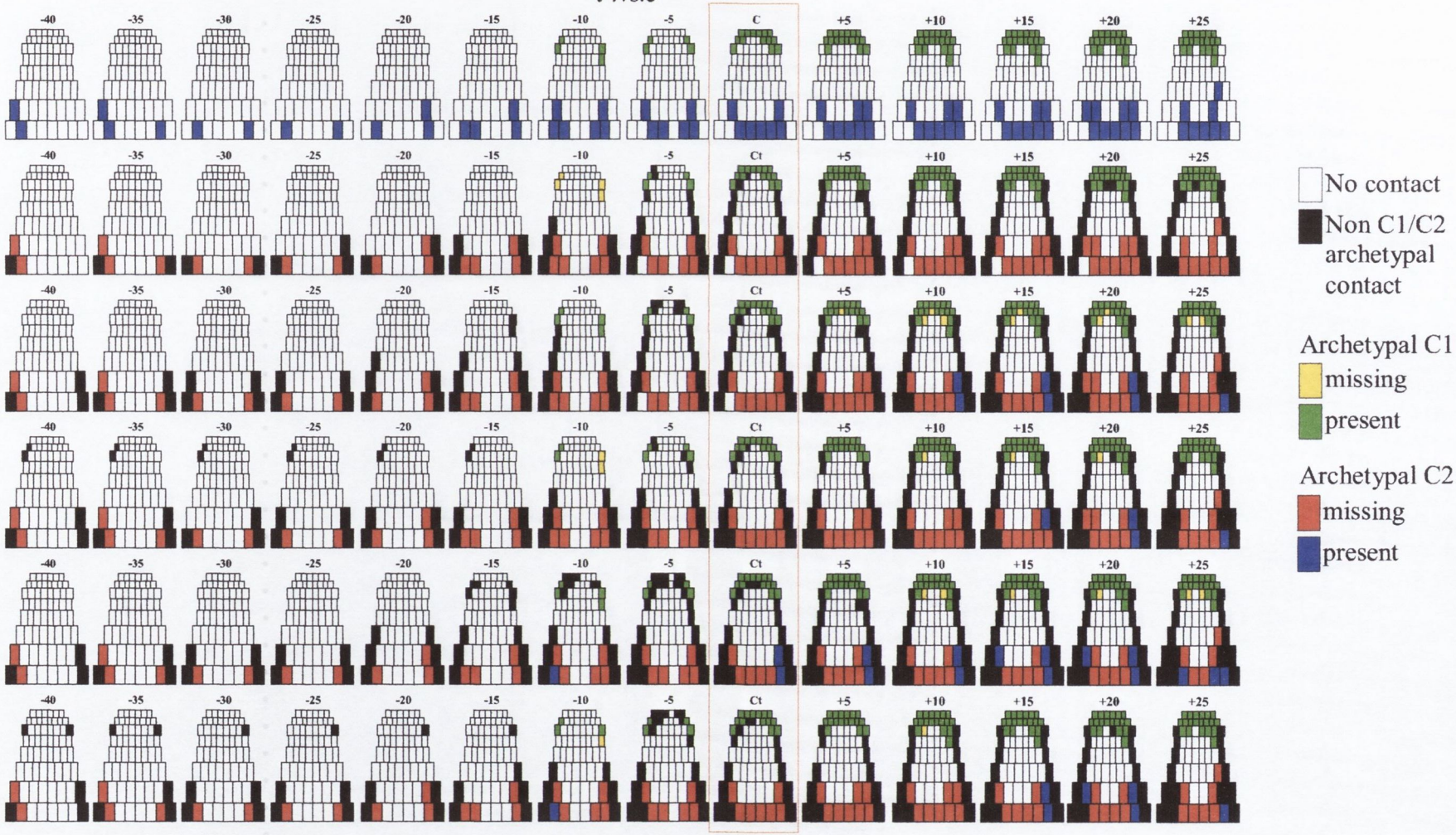



**Figure 8.8** t̪ʸ.vs.c template (top row) and comparative contact patterns for five repetitions of Vt̪ʸ#cV.



– 102 –

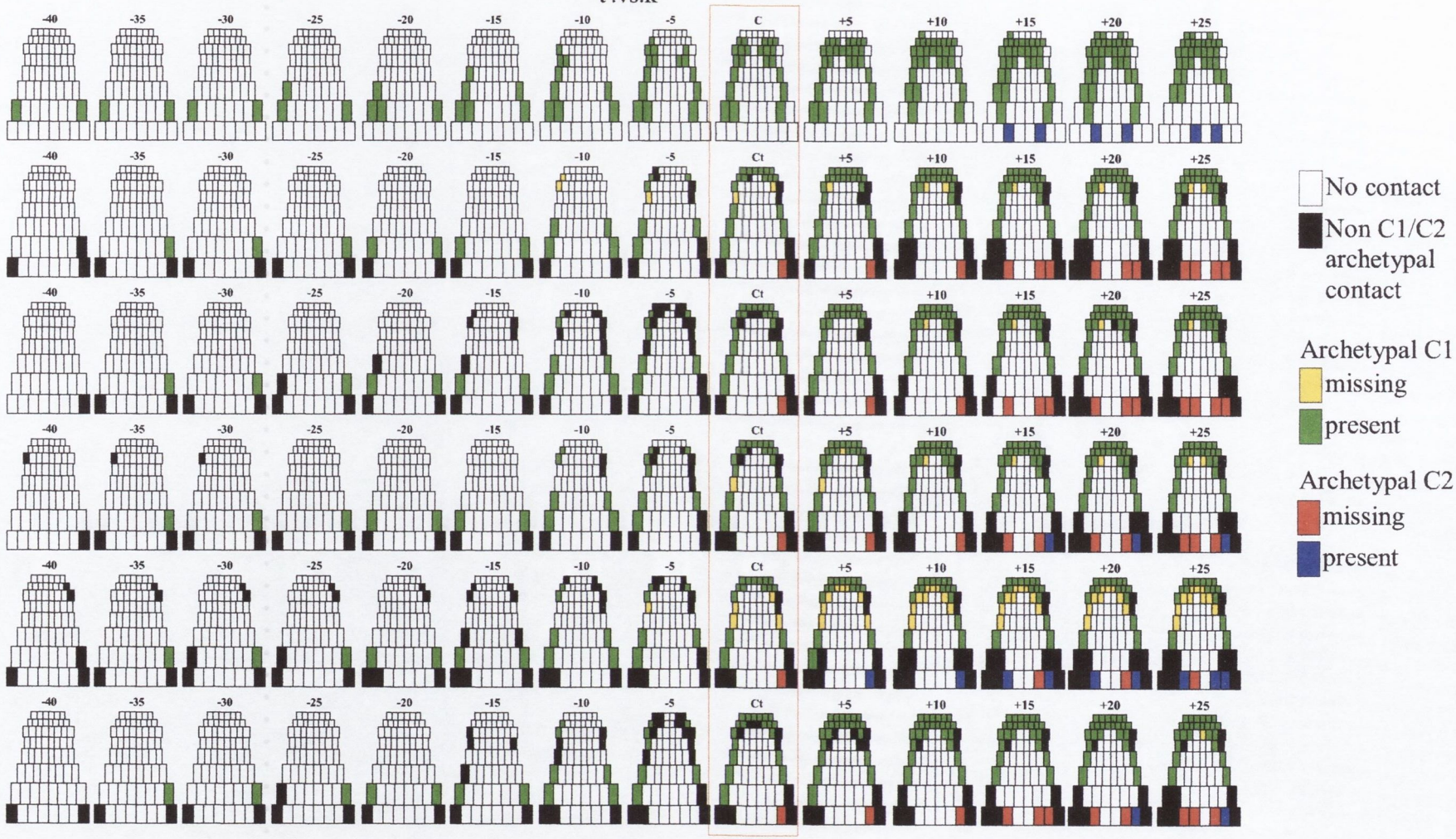



**Figure 8.9** t̪ʏ.vs.k template (top row) and comparative contact patterns for five repetitions of Vt̪ʏ#kV.



– 103 –

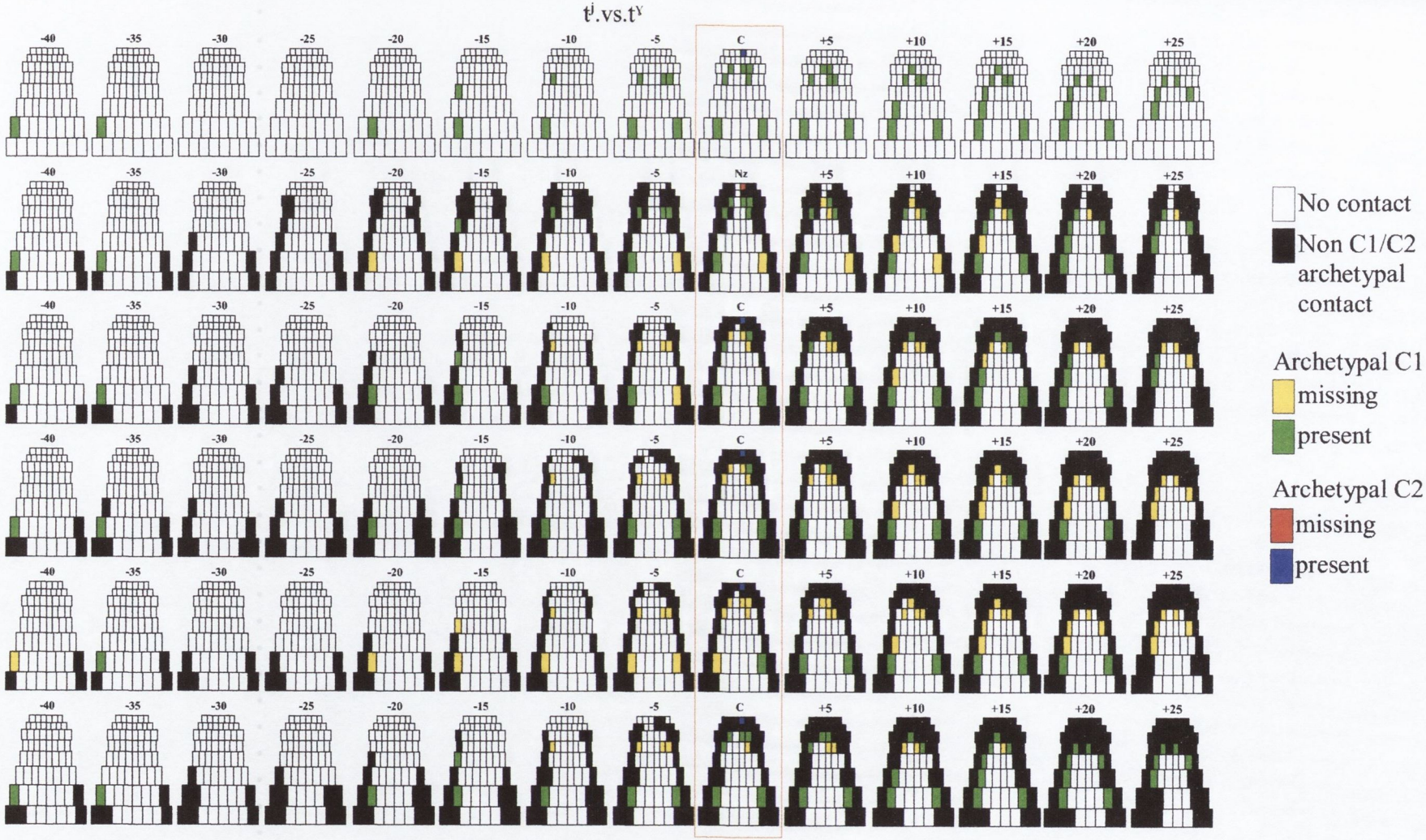



**Figure 8.10** t̪ʲ.vs.t̪ʸ template (top row) and comparative contact patterns for five repetitions of Vt̪ʲ#t̪ʸV.



– 104 –

**Figure 8.11** t̲ʲ.vs.c template (top row) and comparative contact patterns for five repetitions of Vt̲ʲ#cV.



– 105 –

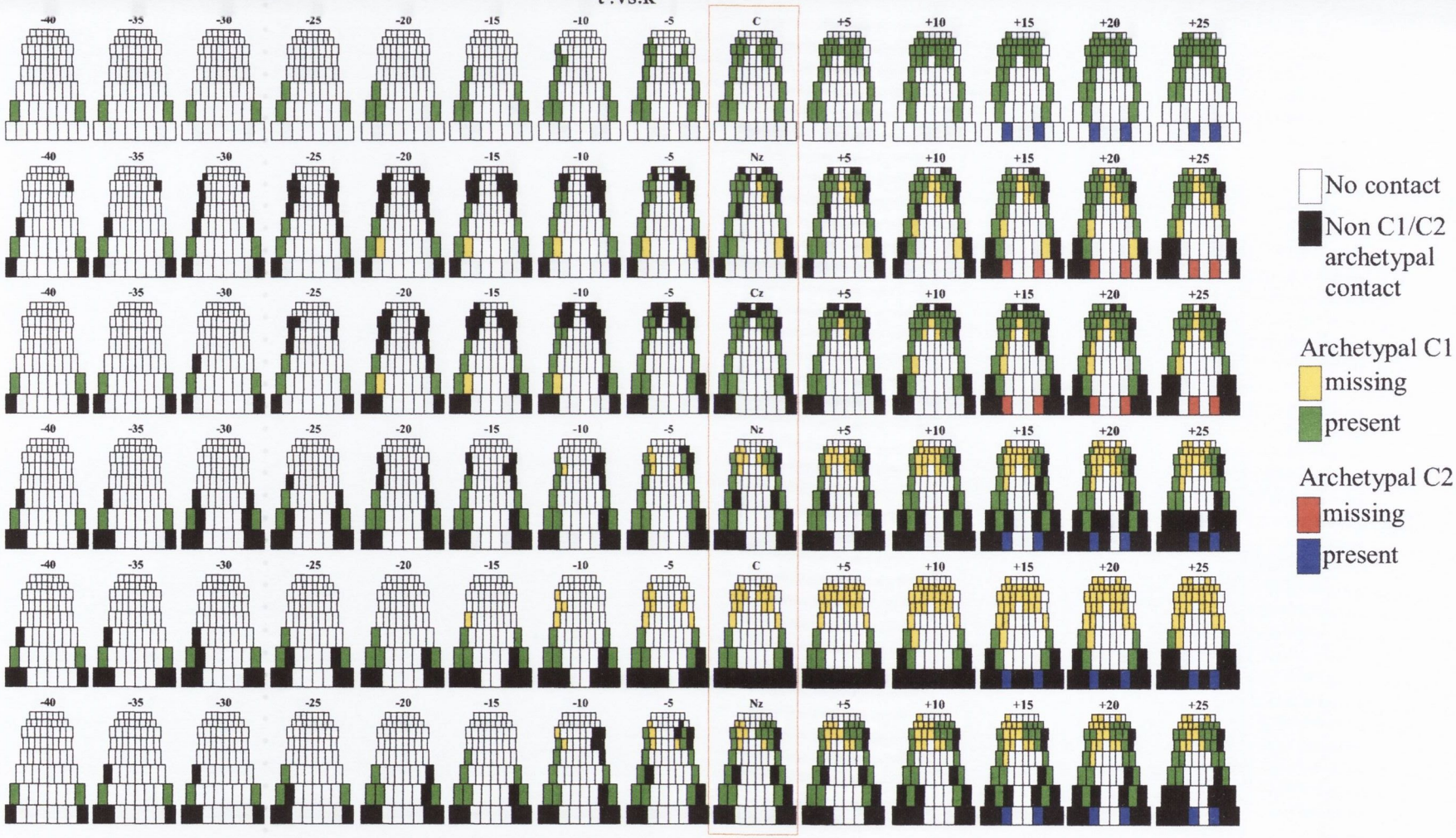



**Figure 8.12** t̲ʲ.vs.k template (top row) and comparative contact patterns for five repetitions of Vt̲ʲ#kV.



– 106 –

**Figure 8.13** c.vs.t̪ˠ template (top row) and comparative contact patterns for five repetitions of Vc#t̪ˠV.



– 107 –

**Figure 8.14** c.vs.t͡ʲ template (top row) and comparative contact patterns for five repetitions of Vc#t͡ʲV.



– 108 –

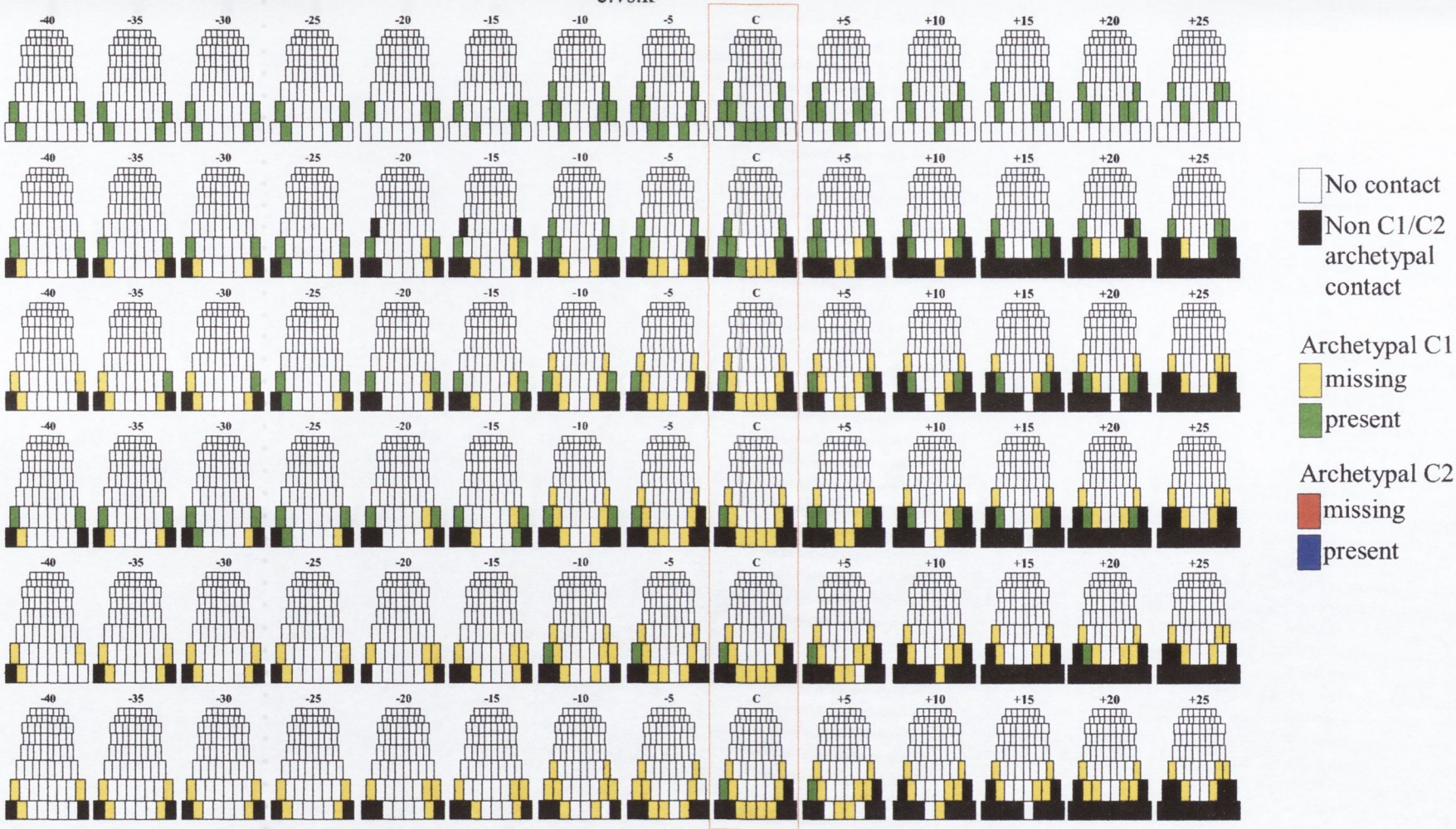



**Figure 8.15** c.vs.k template (top row) and comparative contact patterns for five repetitions of Vc#kV.



– 109 –

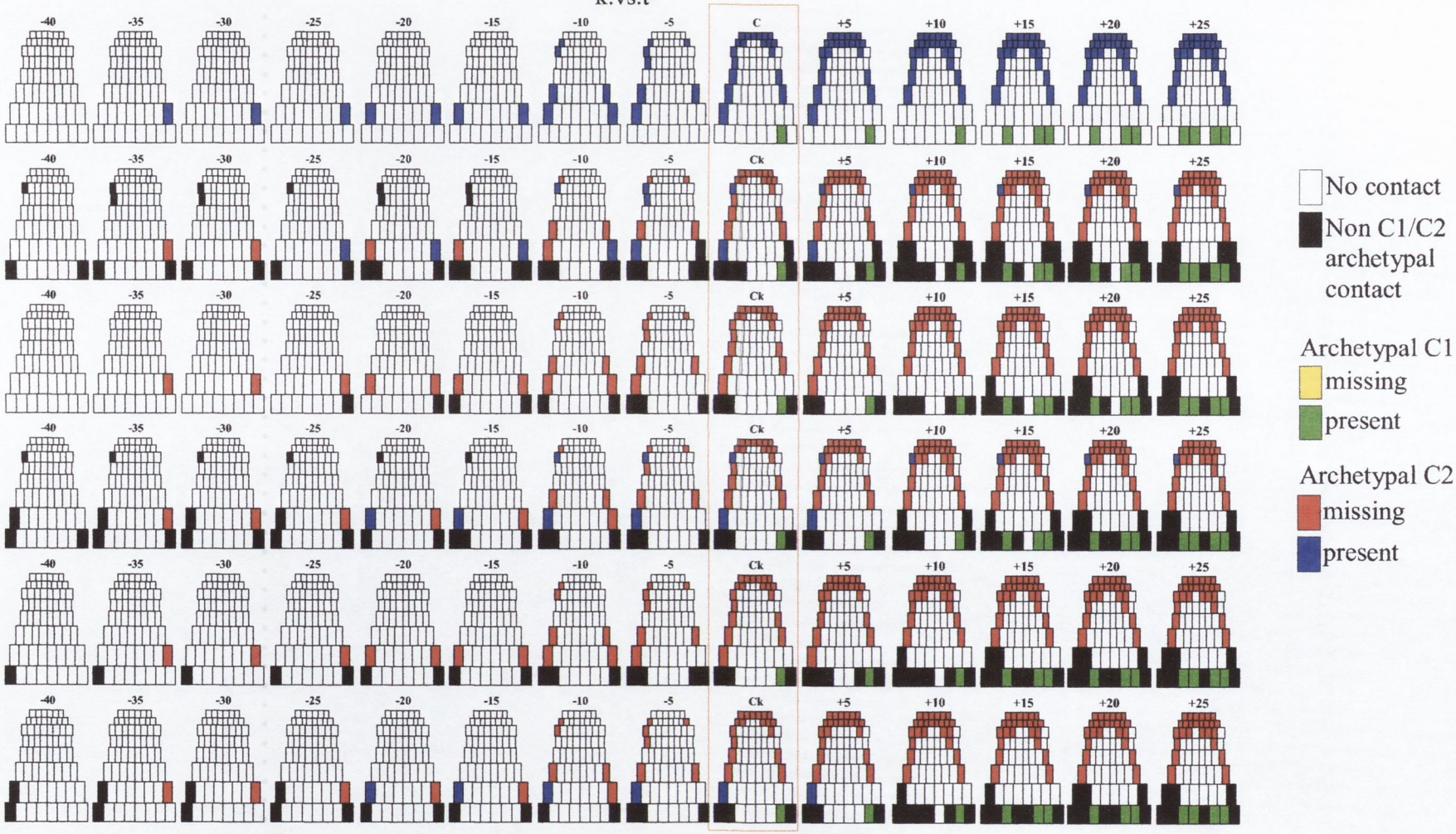



**Figure 8.16** k.vs.t̪$^{y}$ template (top row) and comparative contact patterns for five repetitions of Vk#t̪$^{y}$V.



– 110 –

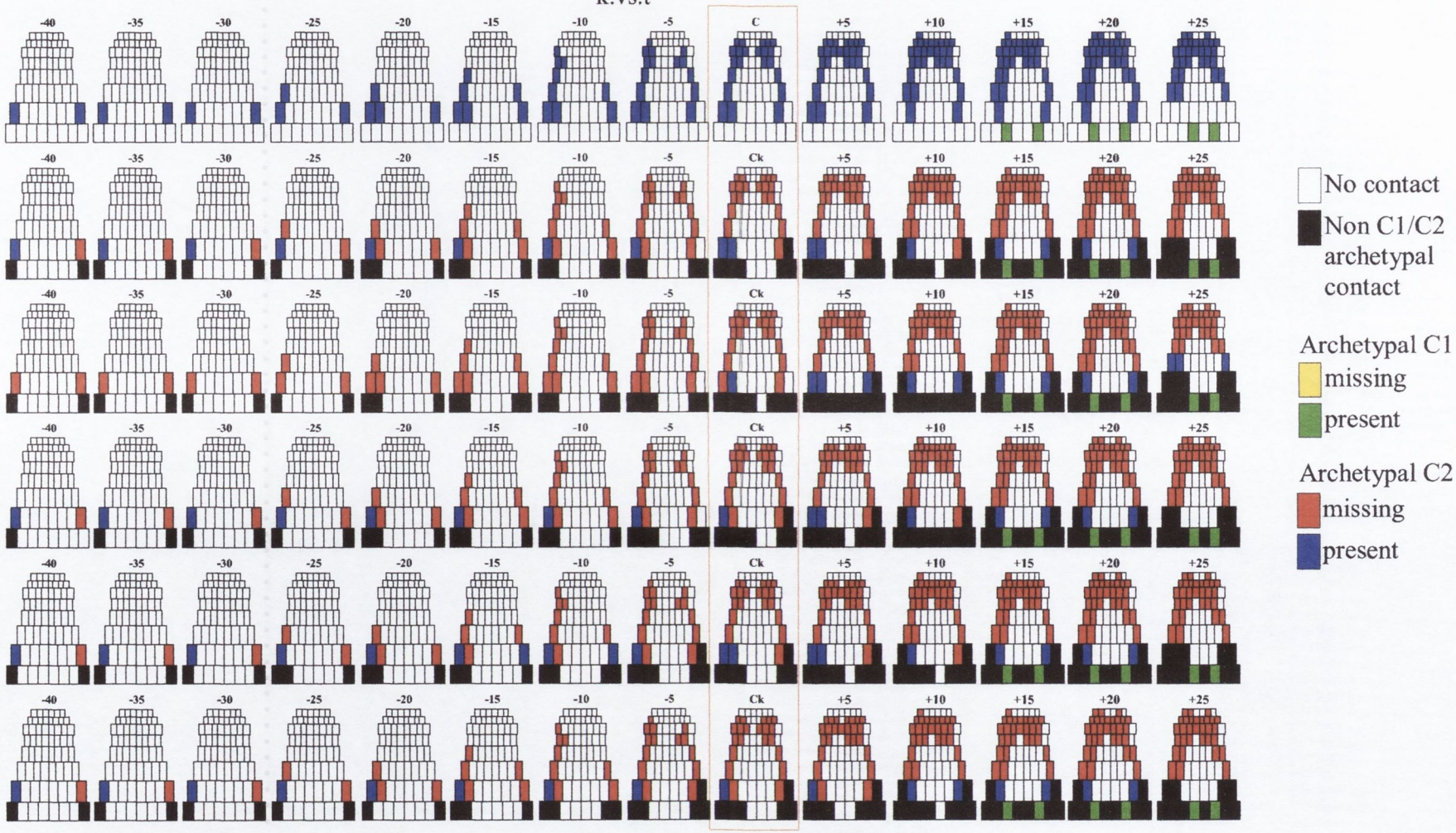



**Figure 8.17** k.vs.t̲$^{j}$ template (top row) and comparative contact patterns for five repetitions of Vk#t̲$^{j}$V.



– 111 –

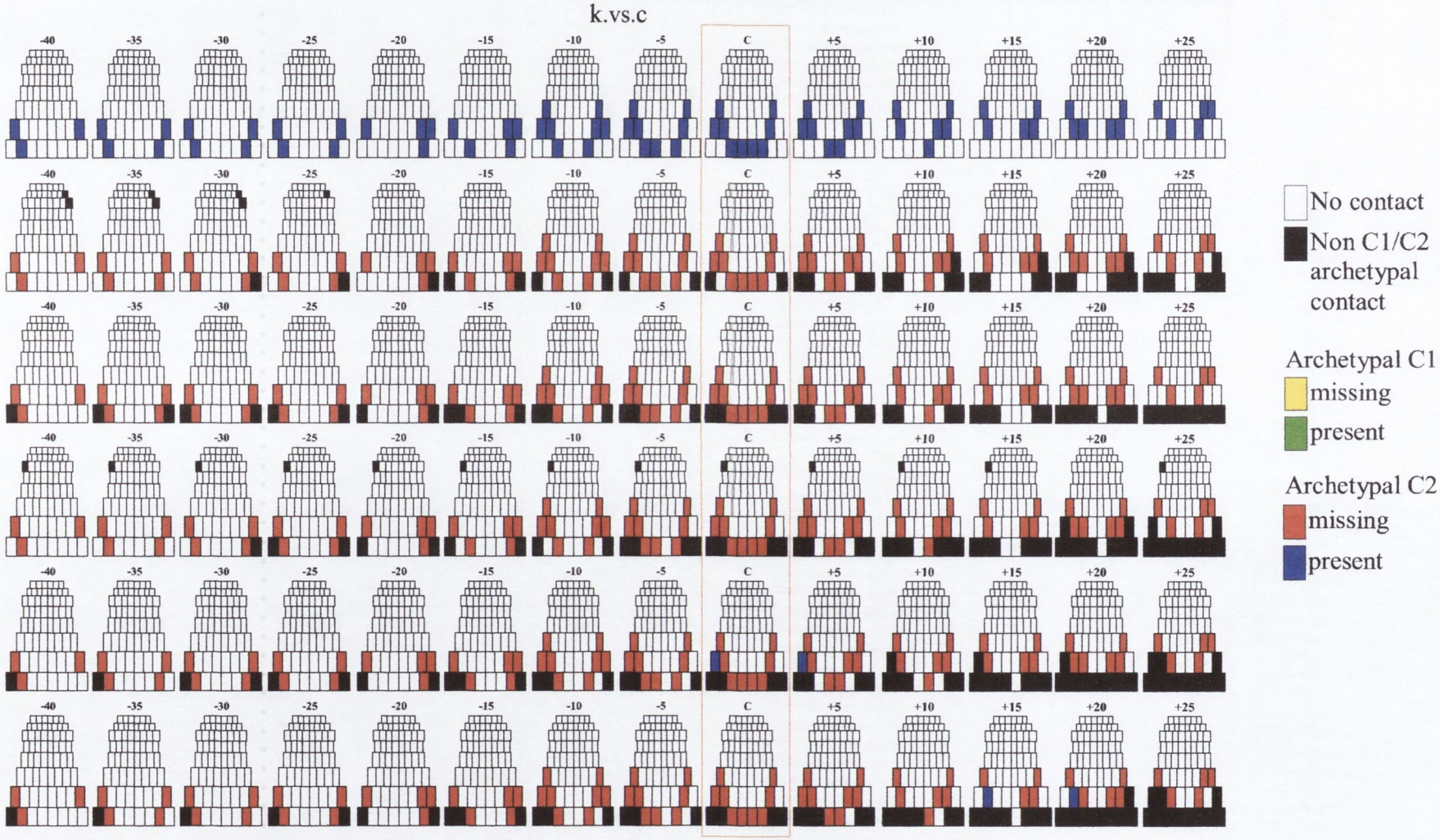



**Figure 8.18** k.vs.c template (top row) and comparative contact patterns for five repetitions of Vk#cV.



– 112 –

These results are summarised in Table 8.3. In terms of resistance to assimilation (based on the numbers in the **None** column), the following hierarchy: k > t̠ʸ > c > t̲ʲ was found. The stops in the palatalised series emerge as the "weakest". Note that a rather analogous phenomenon has been discussed for Russian, where palatalisation is tending to disappear in segments which precede velarised consonants in word-internal clusters (Barry, 1992).

| | None | Len1 | Len2 | Hint | Co-prod. | Blend | Weak $C_2$ | Full |
|---|---|---|---|---|---|---|---|---|
| Vt̠ʸ#t̲ʲV | 1 | | | | | 1 | 1 | 2 |
| Vt̠ʸ#cV | 4 | | | 1 | | | | |
| Vt̠ʸ#kV | 5 | | | | | | | |
| Vt̲ʲ#t̠ʸV | | 1 | | | | 3 | | 1 |
| Vt̲ʲ#cV | 2 | | 1 | | 2 | | | |
| Vt̲ʲ#kV | 1 | 1 | 2 | | | | | 1 |
| Vc#t̠ʸV | 4 | 1 | | | | | | |
| Vc#t̲ʲV | 2 | 1 | | 1 | 1 | | | |
| Vc#kV | | | | | | 5 | | |
| Vk#t̠ʸV | 5 | | | | | | | |
| Vk#t̲ʲV | 5 | | | | | | | |
| Vk#cV | 4 | | | | | 1 | | |

**Table 8.3** Number and type of assimilations found for all clusters at the first EPG frame of consonantal closure or, in the case of an initial fricative being produced, the first EPG frame point of maximum fricative-like narrowing. The classifications are given in Table 8.2.

For each of the stops, the likelihood of assimilation appears to be greatest when it abuts a stop whose primary articulation involves a contiguous, mechanically linked part of the tongue (e.g. tip and blade). Thus, abutting coronals or abutting dorsals are more prone to assimilation than clusters which involve a combination of coronal and dorsal.

When $C_1$ and $C_2$ involve non-contiguous articulators, assimilation seems to involve two changes. One change is a weakening in the production of $C_1$ involving a gradual lenition until no production of $C_1$ is detected in the EPG data. An example is the sequence of repetitions of Vt̲ʲ#kV in Figure 8.12: repetition 2 (**none**), repetition 1 (**len1**), repetition 3 (**len2**), repetition 4 (full). The other change is an overlapping in time of the production of $C_1$ and $C_2$.

For example in Vc#t̲ʲV, Figure 8.14, repetition 2 (**hint**), at the point of closure for [ c ] the build-up of electrodes on row 3 and row 4 for the production of [ t̲ʲ ] is clearly present but the gesture for $C_2$ is still phased to be later than that for $C_1$. A more extreme overlap in timing is found in Vt̲ʲ#cV, Figure 8.11, repetition 4 (**co-produced**), where at closure for [ c ] the build-up of electrodes on row 2/row 3 for [ t̲ʲ ] has not yet achieved maximum narrowing making it appear as though $C_2$ (i.e., [ c ]) is phased to be produced before $C_1$ (i.e., [ t̲ʲ ]). Thus it can be seen that what has been classified as hint differs from co-produced essentially in terms of phasing. The relatively minuscule presence of $C_2$ at the time of closure in $C_1$ appears to be the an indicator that the overlap of the two gestures is considerably less.

When $C_1$ and $C_2$ involve contiguous articulators, the assimilation route seems to involve a blending of the $C_1$ and $C_2$ gestures in the first part of the cluster. In Vt̪ʸ#t̲ʲV, Figure 8.7, repetition 3 is an example of **blend**ing where the contact pattern at closure has the characteristic full closure on row 1 of [ t̪ʸ ] yet with contact extending to row 3 that is more characteristic of [ t̲ʲ ]. However in repetition 4 at closure there is notch-like contact on row 1 which suggests laminar rather than apical (i.e. [ t̲ʲ ]-like and not [ t̪ʸ ]-like) closure, but yet uniquely [ t̲ʲ ] contact is missing on row 3 and row 4. The category given to this in the study was that of a **weakened $C_2$** rather than a blend as it is clearly different in nature to repetition 3. Not all repetitions categorised as blends feature the same degree of assimilation. In Figure 8.15, repetition 1, the **blend** has most of the uniquely [ c ] features, whereas in Figure 8.15, repetition 2, the **blend** only has 1/9 of the uniquely [ c ] features. This may indicate that there is gestural weakening of the $C_1$ gesture as well as a temporal overlap of the gestures.

In general the secondary articulation of $C_1$ is retained regardless of the degree of assimilation of the primary place of articulation of $C_1$. This can only be sensibly discussed with this data for abutting coronals as seen in Figure 8.7 for Vt̪ʸ#t̲ʲV and Figure 8.10 for Vt̲ʲ#t̪ʸV. In Figure 8.7 repetition 2 (**none**), repetition 3 (**blend**), and repetition 4 (**weakened $C_2$**) do not have any of the secondary features of [ t̲ʲ ] at closure, and in repetition 1 (**full**) and repetition 5 (**full**) only 1/2 of those features are present. In all repetitions the secondary features of [ t̲ʲ ] are absent in all frames in the 20 ms preceding closure. In Figure 8.10 repetition 2 (**blend**), repetition 3 (**blend**), and repetition 5 (**blend**) generally have all the secondary features of [ t̲ʲ ] in all frames between -20 ms and closure. Repetition 1 (**len1**) and repetition 4 (**full**) have only 1/2 of those features are present, and those features are generally

absent in all frames in the preceding 20 ms. Of note is that these are the only two repetitions of the ten discussed here where the contact patterns in the region of the secondary articulation in the 20 ms preceding contact does not match $C_1$. However immediately after closure in repetition 4, and at +15 ms for repetition 1, the secondary features of [ t̪ʲ ] are activated.

### 8.4.4.1 Data interpretation in terms of Articulatory Phonology

These findings just discussed are broadly compatible with predictions of articulatory phonology (Browman & Goldstein, 1992) where the basic premise is that fluent speech modification is due to an overlap of gestures and a decrease in gesture magnitude, and that if two overlapping gestures are for the same tract variable then their contribution towards articulation will be given by a blend of their dynamic specifications. It is useful here to review the various classifications with examples and examine their compatibility with the articulatory phonology model. (See §4.2 for a brief description of articulatory phonology).

**Partial assimilation: non-contiguous articulators**

The categories of assimilation chosen for clusters involving non-contiguous articulators (or different tract variables in articulatory phonology) can indeed be explained in such a way. The classifications used for such clusters are **Len1**, **Len2**, **hint**, and **co-produced**. For example, a reasonable explanation of differences in assimilation in Figure 8.14, repetition 4 (**co-produced**) and repetition 2 (**hint**), is that that the gesture for $C_1$ (i.e., [ c ]) is weakened in repetition 4 while the gesture for $C_2$ (i.e., [ t̪ʲ ]) is produced with similar magnitude in both repetitions, and that the $C_1$ and $C_2$ gestures overlap in time more in repetition 4 than in repetition 2. In repetition 1 (**len1**) of the same figure the gesture for [ c ] may be reduced in magnitude leading to a failure to produce full closure. In Figure 8.11, repetition 3 (**len2**), the $C_1$ (i.e., [ t̪ʲ ]) and $C_2$ (i.e. [ c ]) gestures overlap partially in time and the [ t̪ʲ ] gesture is reduced in magnitude.

**Partial assimilation: contiguous articulators**

For clusters involving contiguous articulators (or the same tract variable in articulatory phonology), **blend**ing, as seen in Figure 8.7, repetition 3, could be explained as a merging of the gesture for $C_1$ (i.e., [ t̪ʸ ]) and $C_2$ (i.e. [ t̪ʲ ]) resulting in a place of articulation different to that typically found for either stop.

**Complete assimilation**

**Full** assimilation could be explained as a full overlap in time for the gesture for $C_1$ and $C_2$ and a complete reduction in magnitude of the gesture for $C_1$.

The category **weakened $C_2$** is a form of complete assimilation, where in addition to the complete absence of uniquely $C_1$ contact there appears to be a reduction in the contact pattern that would be expected for $C_2$ (see, for example, Figure 8.7, repetition 4). This example is less readily explained using articulatory phonology. In that repetition uniquely $C_1$ (i.e., [ t̪ʸ ]) contact is completely absent and the contact at closure appears to be laminar as for $C_2$ (i.e., [ t̪ʲ ]). Note however that the place of articulation appears to be more forward than is typical for [ t̪ʲ ]. Thus it would appear that both place of articulation and tongue configuration are separately controlled. The only gestural modification within articulatory phonology that might come near to explaining the contact pattern here appears to be gestural weakening of $C_1$, as the change in place of articulation indicates that there is some influence from $C_1$ on the contact pattern at closure. However that influence is only on the place of articulation and not on the tongue configuration. It is not obvious how such separate control of different aspects of the same gesture can be achieved within articulatory phonology.

**Assimilation of secondary articulation**

If secondary articulation is considered in the above discussion of the example of a **weakened $C_2$** in Figure 8.7, repetition 4, then another possible argument could be considered. The uniquely [ t̪ʲ ] features for palatalization are absent at closure for this repetition, suggesting that the gesture for velarisation for [ t̪ʸ ] is present and blending with the [ t̪ʲ ] gesture for palatalization. As there may be some degree of coupling between the tongue body and the tongue blade this could arguably be a reason why the tongue blade is lower than that typical for [ t̪ʲ ] and so contact on row 3 and row 4 could also be less than is typical. A description in articulatory phonology which took advantage of this argument would be that the gestures for [ t̪ʸ ] and [ t̪ʲ ] overlap and that the tongue tip gesture for [ t̪ʸ ] is reduced to zero magnitude, and that the weakening of $C_2$ results from blending at the level of the secondary articulation.

In all repetitions of Vt̪ʲ#t̪ʸV clusters (Figure 8.10) the features for palatalization for $C_1$ (i.e., [ t̪ʲ ]) are generally present regardless of assimilation of the primary articulation of $C_1$. In repetition 3 where the primary articulation of $C_1$ has given way almost entirely the palatalization associated with $C_1$ is present at closure and the interval preceding it. As the production of these coronals of Irish involve two lingual gestures this particular example can be explained within articulatory phonology as involving a weakening of the primary gesture but no weakening of the secondary. This primary (weakened) gesture is blended with a strong primary gesture for $C_2$. There is no evidence however of weakening/blending of the

secondary gesture. A description of the gestural scores for the secondary articulations for the cluster in articulatory phonology would require the secondary articulation to also be regarded as a blend where the $C_2$ secondary gesture has additionally been weakened to zero. However this line of reasoning does not quite work as palatalization is not seen for this repetition at the point of consonantal release (not shown in figure).

Particularly problematic examples are seen in Figure 8.10. In repetition 4 (**full**) features associated with the secondary articulation of $C_1$ (i.e. [ t̪ʲ ]) are absent before closure but are present after closure (and are present in part at closure). An explanation in articulatory phonology would be perhaps that the $C_1$ gestures, where the $C_1$ primary gesture has been weakened to zero, are overlapped to such an extent that $C_2$ gestures are being produced before $C_1$.

In Figure 8.10, repetition 1 (**len1**) features associated with the primary articulation of $C_1$ are clearly present at closure, while features associated with the secondary articulation of $C_1$ are entirely absent until the frame at -5 ms. From -5 ms to +10 ms 1/2 the features are present, and from +15 ms onwards are all present. Perhaps this is the result of a weakened gesture for the secondary articulation of $C_1$, but the fact that the maximum typical contact is achieved at a later time-point is puzzling and might instead suggest that the gestures for the primary and secondary articulators are not time-locked.

### 8.4.4.2 Coronals and Dorsals

On a general note, the two coronal stops are extremely different in terms of their propensity to assimilation. [ t̪ʲ ] shows a high propensity to assimilation, but [ t̪ˠ ] emerges as very resistant to assimilation. Note that in other languages coronals have been found to be more prone to assimilation than dorsals. Clearly not all coronals are equally prone, and in the data presented here [ t̪ˠ ] is less prone to assimilate than the dorsal [ c ].

It is striking that when either coronals or dorsals abut, the order of occurrence appears to matter greatly. Note that only 1 assimilation was found for Vk#cV, whereas for Vc#kV there were 4 assimilations (in the 5 repetitions). This asymmetrical behaviour that results from the juxtaposition of two dorsals might possibly be explained in terms of tongue dynamics. In Figure 8.19 the path that the EMA sensor on the tongue body moves through during the production of a consonantal cluster is shown for a single repetition of Vc#cV (dashed line) and of Vk#kV (solid line). The location of the sensor as the time-point (C) is shown as a small circle, and the roof of the mouth above the area of interest is sketched. The latest time-point of the trajectory is indicated with an arrowhead. For both utterances the trajectory

progresses in a back-to-front elliptical gesture. Similar observations have been made in the literature (see, for example, discussion in Mooshammer, Hoole & Kühnert, 1995).

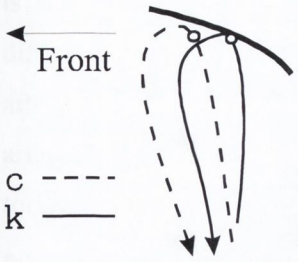



**Figure 8.19** Trajectory of the rearmost EMA sensor for a repetition of Vc#cV and of Vk#kV (where V = /a/). The small circles mark the sensor position at the time of closure, and the arrowheads mark the sensor position at the last time-point shown.

It is proposed that this basic gesture may need little modification to produce Vk#cV: it might simply be a matter of allowing the tongue forward movement to be more extensive. However, the articulation of Vc#kV may require movement that runs counter to the preferred tongue body gesture. It should be noted that even in the (potentially) assimilating environments of Vc#kV no examples of a front-to-back elliptical gesture were found. The robustness of $C_1$ secondary articulation for clusters combining [ t̪ʸ ] and [ t̪ʲ ] does not yield to an account based on a preference for back-to-front elliptical gestures of the tongue body.

## *8.5 Conclusions*

The assimilation observed in the data can, in general, be described using overlap of gestures and a decrease in gesture magnitude as suggested in articulatory phonology (Browman & Goldstein, 1992). However there are examples where the descriptions required are not compatible with other restrictions of articulatory phonology, or where the descriptions appear contrived and unreasonable. An example that is not compatible with articulatory phonology is seen in Figure 8.10, repetition 1 (**len1**), where the gestures for the primary and secondary articulators for [ t̪ʲ ] ($C_1$) do not have the same phase relationship/time-locking as is generally seen for [ t̪ʲ ]. Articulatory phonology does not allow the relative timing of gestures for a phonological segment to change. If that rule is not to be broken here then articulatory phonology would seen to require that [ t̪ʲ ] must be a different phonological segment in this example. An example for which descriptions based on gesture seem somewhat contrived is Figure 8.7, repetition 4 (**weakened $C_2$**) of V t̪ʸ#t̪ʲ V, where the type of contact of the primary articulator is clearly laminar as opposed to apical, i.e., clearly [ t̪ʲ ] like rather

than [ t̠ʸ ] like, and yet contact is more anterior than is archetypal for [ t̲ʲ ]. An explanation which considers the gestures for the primary articulator alone does not seem satisfactory as it is not clear that articulatory phonology supports the separate control of the magnitude of different parameters of tract variables that would be required to achieve the observed articulation. An explanation in terms of linkage between the primary and secondary articulators, where the presence of velarisation at closure may cause sufficient lowering of the tongue front to achieve the observed reduction in contact, is not entirely convincing as it is not clear from the other repetitions of Vt̠ʸ#t̲ʲ V and Vt̲ʲ#t̠ʸ V that such linkage exists.

The different propensities to assimilation of the consonants may well have their origin in mechanical and dynamic properties of the tongue. For example, the palatal consonant is much more prone to assimilation than is the velar consonant. It is hypothesised that this asymmetry arises because tongue-body back-to-front elliptical gestures are preferred to gestures in the opposite direction. So a sequence of [ k ] + [ c ] may represent an easy target, less likely to assimilate, than the sequence [ c ] + [ k ].

Given that the secondary articulation is retained regardless of the assimilation of the primary articulation of $C_1$ suggests a level of control above that of overlapping gestures of varied magnitude and above constraints from mechanical and dynamic properties of the tongue.

# Chapter 9 A Study of 'Human Nomograms'

## *9.1 Introduction*

Fant, 1970, used a number of models to help in the interpretation of the relationships between articulatory settings and acoustic consequences. For each model he presented data for vowel-type productions. The articulatory variables for the models were: the location of the vocalic constriction, the cross-sectional area of the constriction, lip rounding (in terms of length and cross-sectional area). The resulting acoustic measurements were formant frequencies. He presented the data as a series of charts which are called nomograms which plot formant frequencies for the model against the location of the vocalic constriction.

It would be useful if these nomograms could be used as a reference when considering the articulatory-acoustic relationships in speech production. To do this it must be understood how the model corresponds to human speech production and what additional constraints the human speech system imposes.

The aim of this experiment was to record and analyse data which could show relationships between articulatory settings and resulting acoustics, and to compare the measured relationship to those presented in Fant. It was hoped that such an articulatory-acoustic mapping could be used as a reference in future work involving the speaker in this experiment and, more generally, to test predicted relationships between articulation and resonance.

For a continuum of high vowels, produced with consistent lip rounding, the vocalic constriction was located approximately from EMA data. This was plotted against corresponding formant frequencies measured from spectrograms. These plots are compared to nomograms in Fant's work. Differences are discussed, as are difficulties with the techniques used.

## *9.2 Data Acquisition*

### 9.2.1 Materials

A trained phonetician was asked to produce continua of high vowels, covering the range between high front and high back, both with and without lip rounding. The target continua were [ɯ] → [i], [i] → [ɯ], [u] → [y], and [y] → [u]. Three repetitions of each continuum were recorded. The subject was instructed to maintain a constant lip configuration for each series.

As the continua are of high vowels the tongue should, in principle, be always as close to the roof of the mouth as possible without producing frication. Larynx height was not controlled.

### 9.2.2 Method

Articulatory measurements were recorded simultaneously with EMA and EPG systems at the Institut für Phonetik und Sprachliche Kommunikation der Universität München by Dr. Phil Hoole.

The EMA system used was a Carstens AG100 system with 10 receiver sensors. This records directly to computer. The measured data were sampled at 200 Hz using a 40 Hz low pass filter. Three sensors were located on the tongue, the front-most being 0.5-1.0 cm from the tongue tip, the rearmost approximately 4 cm further back, and the mid sensor roughly halfway between the two. Another sensor was attached to the lower central incisors and one on the lower lip. Reference sensors were located on the upper central incisors and on the bridge of the nose. Two other sensors were attached to the reference t-bar, which was used at the start and end of recording. The recorded data was post-processed to reduce errors due to head movement. Figure 6.1 shows the experimental setup, while a description of the recording process and post-processing is give in §6.1.

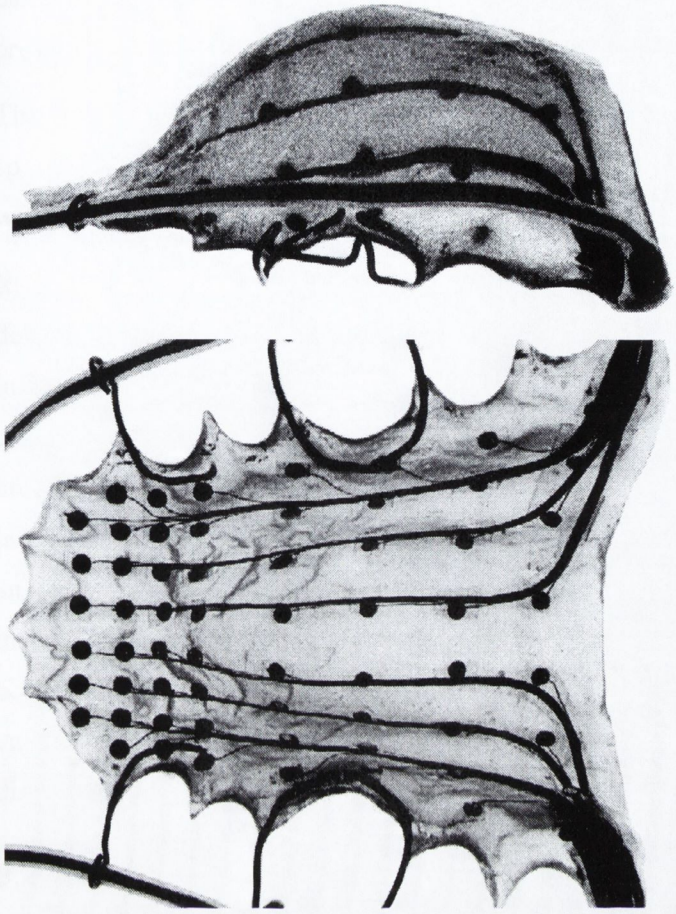


**Figure 9.1** Photographs of a view from the side and a view from below of the speaker's EPG palate.

The EPG system was a Reading EPG2 system, recorded at 200 Hz directly to computer. Photographs of the EPG palate used are shown in Figure 9.1. The acoustic signal was recorded, using a B&K microphone, to a Sony DAT recorder at 48 kHz. This was later down-sampled to 16 kHz, using a 7.5 kHz low pass filter and transferred to computer. The EPG and acoustic data were aligned to the EMA data by use of synchronisation pulses recorded simultaneously, with an estimated accuracy of alignment of ±3 ms (Nguyen & Hoole, 1993).

Additional measurements were used to help locate the EMA and EPG co-ordinate systems with respect to each other. A single EMA sensor was used to trace the contour of a plaster-cast of the speaker's palate. The plaster-cast was the same as that used to hold the speaker's EPG palate. This palate and plaster-cast was aligned using the same t-bar as in the recording above. The midline of the plastercast, the midline of the artificial palate, and the locations of the EPG electrodes were measured using a high resolution scanner and a sliding callipers as described in §5.2. The two traces of the midline of the plastercast were used as part of the data alignment method as described in §7.1.2. The measurements of the midline of the artificial

palate and measurements from an x-ray of the subject, taken for previous work, were used to produce the contour of the roof of the mouth using the method described in §7.1.2.

The area of lip aperture was measured separately by tracing the reflected image of the speaker's lips on a mirror while the speaker produced the target utterances. These apertures were noted to help in the comparisons with Fant's nomograms shown later.

It was intended to use the EMA and EPG measurements in parallel in the analysis to determine the location of lingual constriction. In the analysis presented below, however, information from EPG is not used directly in calculating the location of the narrowest constriction. The reasons for this are discussed in Chapter 10. EPG was very much used for broader, qualitative analysis. As EPG electrodes are placed at anatomically important landmarks, the EPG palate provided for valuable reference information both during analysis and in understanding the results. The EPG data provided a continuous source of additional information to check if the tongue contour being predicted from EMA was plausible. Having both midline information from EMA and lateral information from EPG allowed for many valuable insights that would not have been available from EMA alone. These are also discussed in Chapter 10.

### 9.2.3 Segmentation

The recording above resulted in voiced utterances of duration between 5.3 s to 6.9 s. From these voiced utterances, acoustic and articulatory data were extracted at 100 ms intervals.

The extracted acoustic data were formant frequencies for F1, F2, and F3. These were found by visual inspection of high resolution spectrograms produced using the "Gray Spectrogram Plotter" v2.10 by Soundswell Music Acoustics AB. The frequencies were checked by comparison with formant frequencies determined using the semiautomatic inverse filtering software developed at Trinity College, Dublin. This software is fully described in Monahan, 1996. Where frequencies found by both methods differed the appropriate frequency was selected on a case by case basis.

The extracted articulatory data consisted of the EMA X-Y sensor locations for the three tongue sensors. EPG palate contact frames, and the X-Y sensor locations of the lower lip and lower incisor sensors were also extracted, but these data were not used in this experiment.

With further analysis, the data set was reduced by removing points at the start and at the end of each recording. The points that were removed were those for which articulatory movement indicated the speaker was preparing for, or had already finished, the required production as it was clear that these were not articulations relevant to this study.

### 9.2.4 Representation

Several representations of the data were used in the analysis. All channels were examined at time aligned points with the visualisation tool discussed in §7.2.2 to help in understanding the articulatory information, how it changes, and what the resulting audio signal looks like. The articulatory data was viewed in a three-dimensional representation, facilitating simultaneous viewing of tongue midline movement from EMA and tongue contact against the sides of the palate from EPG. Measurements from x-ray were used to extend the represented region of the vocal tract, as discussed in §7.1.2. As a "circle model" data reduction method (see §6.2.2) was applied to estimate movement of the vowel constriction, this representation was simultaneously presented to the experimenter.

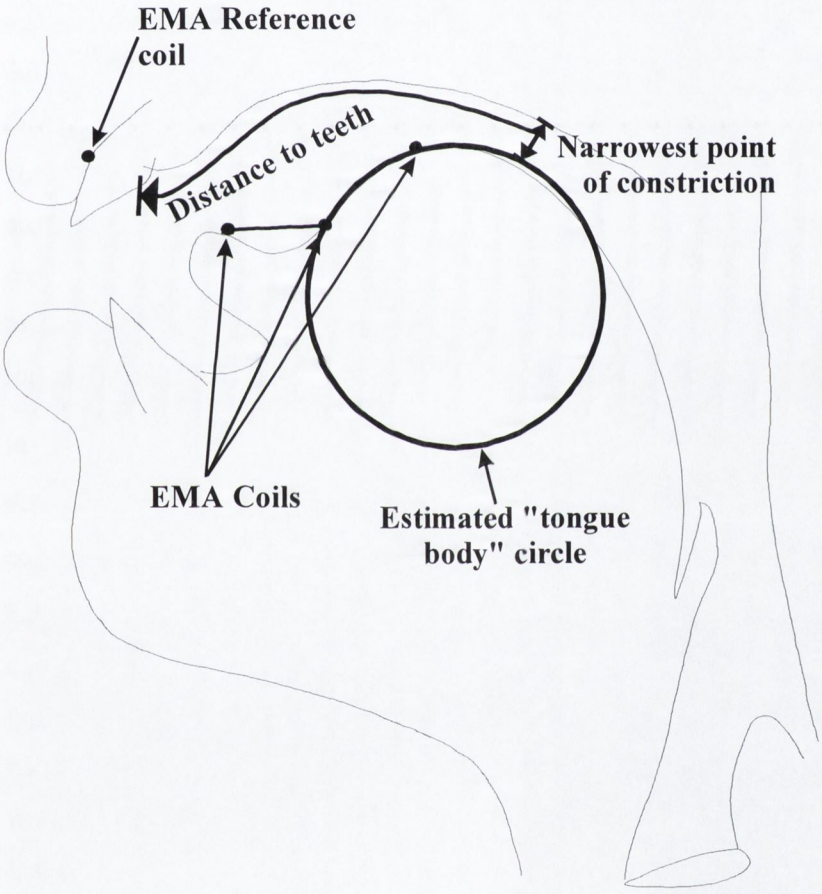



**Figure 9.2** Summary of the circle model, discussed in 6.2.2, with the three EMA sensors on the tongue used in this experiment shown as black dots. The position of the circle, which represents the tongue body, is controlled by position of the rearmost sensors. The closest point between this circle and the roof of the mouth can be located, and is taken to equate to the narrowest point of lingual constriction. The distance to the teeth from the narrowest point of the lingual constriction is illustrated.

The "circle-model" models the back of the tongue as a circle/ball for high vocalic constrictions. Figure 9.2 above provides a summary description of the model, showing how the two rearmost EMA sensors are considered to be attached to the tongue body. The narrowest point of constriction between the circle and the measured roof of the mouth was taken to be the narrowest point of the lingual constriction. This was assumed to correspond to the centre of constriction presented in Fant's nomograms, and is used to produce the data for Figure 9.5 and Figure 9.6.

## *9.3 Analysis From EMA and Acoustic Data*

### 9.3.1 Movement of the Two Rearmost Tongue Coils

Figure 9.3 below plots the total movement of the two rearmost tongue sensors for all utterances. That is, for each repetition each absolute location of a sensor at a time-point during the recording is connected to absolute location of the same sensor at the following time-point. The horizontal and vertical axis show distance from the reference sensor. EPG electrode locations, represented by empty circles, and the curved contour of the roof of the mouth are overlaid for reference. This figure does not immediately yield the location of the narrowest constriction, nor does it directly present any distance from the teeth along the roof of the mouth. Nonetheless it is useful in investigating trends in the movement of both sensors for all repetitions of an utterance and for all utterances.

In general the sequences for front vowel to back vowel continua ([i–ɯ] and [y–u]) show a different extent of movement of both sensors than is seen for back vowel to front vowel continua ([ɯ–i] and [u–y]). Movement of the rearmost sensor, for example, is furthest back at around 5.5 cm for back to front continua, whereas 6.5 cm is typical for front to back continua. The difference in furthest forward position for this sensor is less striking, being typically around 4.5 cm for back to front continua and between 4.7 cm and 4.9 cm for front to back continua. This would seem to indicate that the speaker is more comfortable producing vowels at the extremes of a continuum when he has been progressing towards that extreme than when he is selecting his starting vowel.

The rearmost sensor is generally moved horizontally, showing relatively little vertical movement. An exception to this is all repetitions of [ɯ–i] where the frontmost part of the trajectory is mostly horizontal, but the rest of the trajectory shows similar horizontal and vertical movement. The contrast with the movement of the sensor for the other utterances suggests that the production of the initial [ɯ] may not be as high as, for example, [ɯ] in the continuum [i–ɯ]. Another possibility is that the speaker's velum is slightly lowered for the initial [ɯ]. Data on velar movement was not recorded.

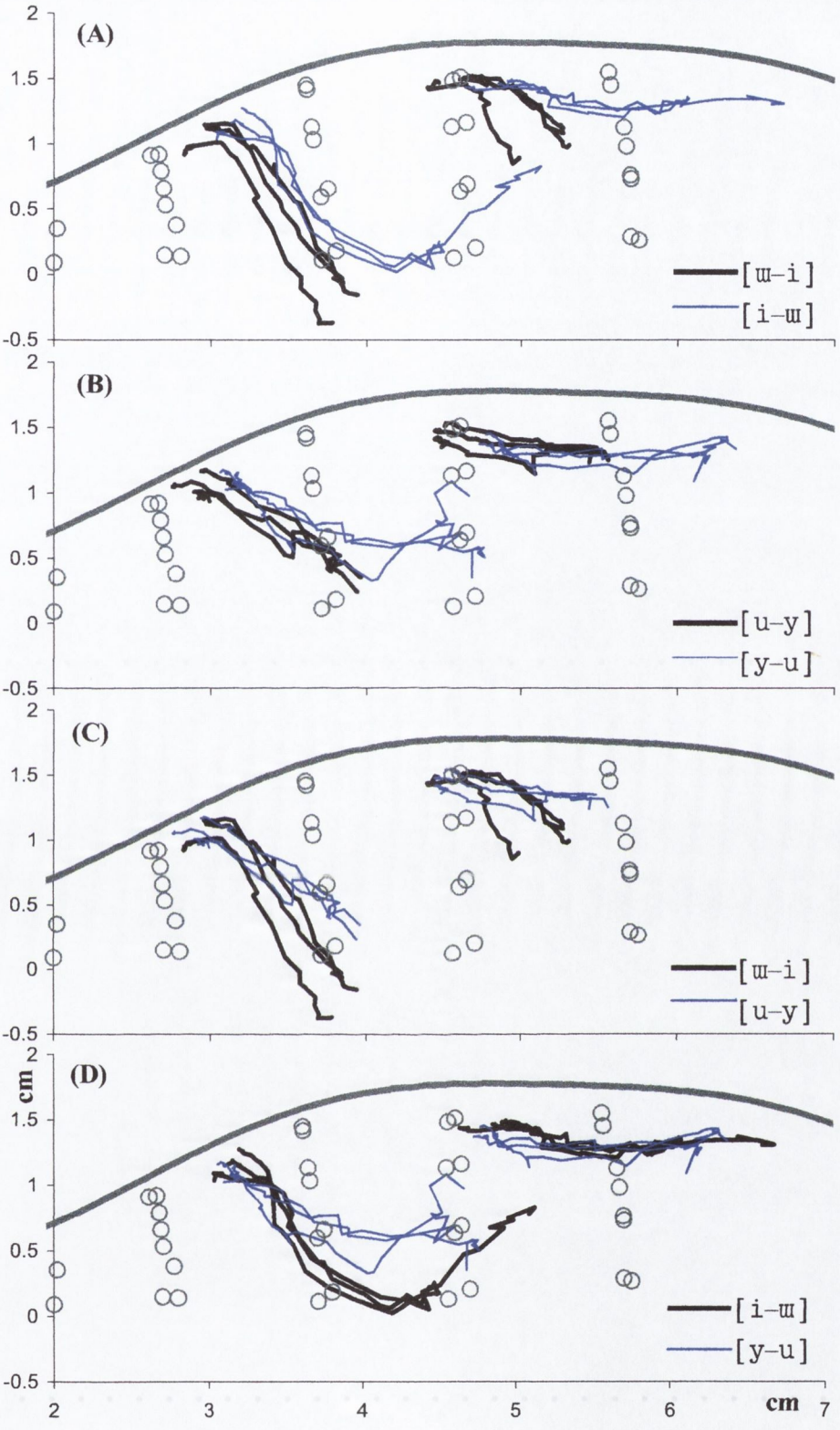



**Figure 9.3** Horizontal vs. vertical movement of the two rearmost tongue sensors for each repetition of (A) unrounded utterances [ɯ–i] and [i–ɯ], (B) rounded utterances [u–y] and [y–u]. The same movements are again shown in (C) for back to front vowel sequences and in (D) for front to back vowel sequences.

Looking at the other sensor, the middle sensor on the tongue, it can be seen that for more retracted vowels this sensor has typically a lower position than more forward vowels when this sensor is in the region of 3 cm to 4 cm from the reference sensor. For back to front continua the sensor does not reach positions more posterior than 4 cm. For front to back continua, as seen in Figure 9.3 (D), this sensor typically has a range as far back as 4.5 cm to 5.1 cm. As the sensor is retracted behind 4 cm there is a change in direction of vertical movement, with a tendency towards higher positions of this sensor.

A possible explanation of the movement of the two sensors may be illustrated using the circle model. A repetition of the continuum [i–ɯ] is shown in Figure 9.4. Here the centre of the circle, calculated using the pair of co-ordinates for the rearmost EMA sensors, is shown for each time-point. There is comparatively little vertical difference over the continuum compared to the horizontal difference. Indeed the trajectory is similar to that of the rearmost sensor, but with greater horizontal displacement. This extra horizontal displacement may be contributed by a rotation of the circle. The middle sensor, attached to a point forward of the centre of the circle, contributes this rotation by its vertical displacement, causing a "pivot" of the circle at the rearmost sensor. The change in direction of the vertical displacement may be due to the tongue approaching the velum. Due to the curve of the velum the tongue may, perhaps, rotate in the opposite direction to facilitate controlling the lingual constriction cross-section area.

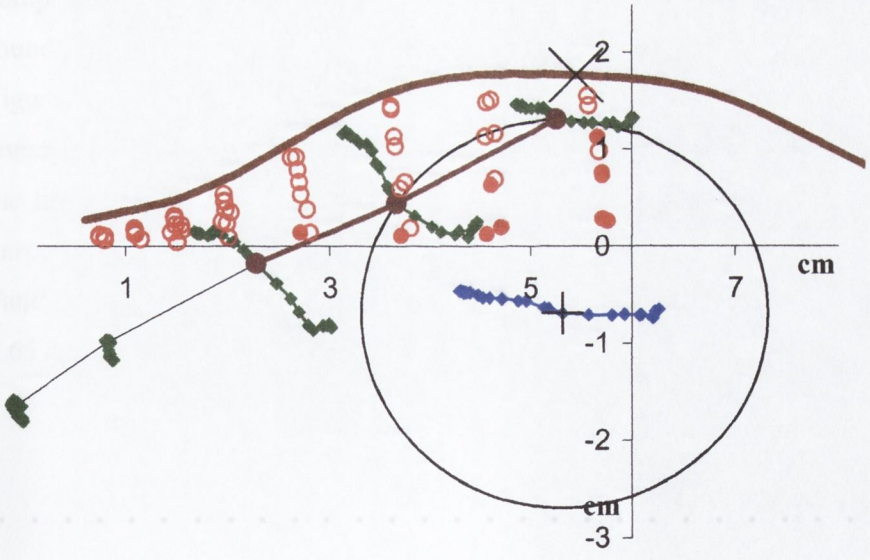



**Figure 9.4** EMA sensor positions and associated centre of the circle model for a repetition of the continuum [i–ɯ]. A time-point is marked by the line through the EMA sensor positions. The corresponding circle calculated using the rearmost sensors is also shown.

### 9.3.2 Articulatory-Acoustic Correlation

To produce nomograms in the style of those in Fant, 1970, the location of the lingual constriction between the tongue body and the roof of the mouth has to be determined from EMA tongue sensor positions. As just discussed, it seems that no individual sensor can indicate the position of the tongue body within the mouth. The circle model is used here to combine the information from two sensors on the tongue body to estimate the overall position of the tongue body. This model is a simplification and will lead to some inaccuracies. But as discussed in §1.3, such a model was found useful in modelling x-ray images of speech.

Figure 9.5 and Figure 9.6 plot the formant frequencies against the distance to the teeth for all utterances, where that distance was calculated using the circle model. In the upper part of these figures is shown the mid-sagittal outline of the roof of the mouth from the front of the artificial palate to a point approximately 2 cm behind the back of the artificial palate. Rows of EPG electrodes are shown as circles. Points on the palate at increments of 1 cm from the front of the teeth are joined by slanted lines to the gridlines in the lower graphs. The x-axes in the lower graphs are marked in centimetres from the front of the teeth. This is equivalent to the axis marked in centimetres from the unrounded lip opening in Fant's nomograms (see an example in Figure 3.7). Note that in the latter, "rounded" effectively means that a 1 cm lip tube is added to the length of the vocal tract: by omitting this there is a measure that is comparable to the point used, i.e., the front of the teeth. Overlaid on top of the formant values found in this analysis, shown as thin lines, are comparable formant values taken from Fant's Figure 1.4-11a, shown as heavy dark lines. These relate to vowels with a narrow lingual constriction of 0.65 $cm^2$ cross sectional area, and with several different cross sectional areas at the lips. As lip aperture for the data in this experiment was approximately 2.5 $cm^2$ for the unrounded series and 0.25 $cm^2$ for the rounded series, the formant traces included here from Fant's nomograms were for lip cross sectional areas of 4 $cm^2$ and 2 $cm^2$ in Figure 9.5, and 0.65 $cm^2$ and 0.16 $cm^2$ in Figure 9.6.

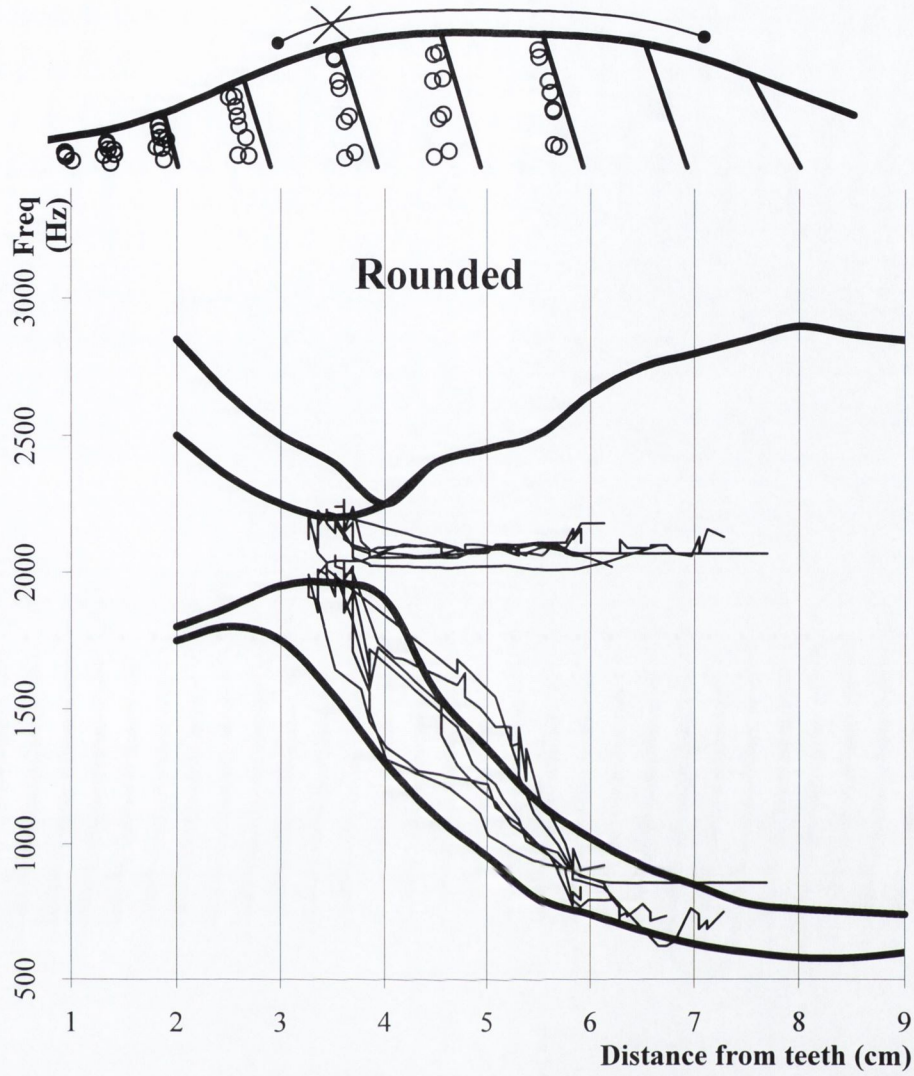



**Figure 9.5** F2 and F3 values vs. distance from the narrowest point of lingual constriction to the teeth for all the rounded utterances. Dark lines show formant traces for lip cross sectional areas of 4 $cm^2$ and 2 $cm^2$ from Fant's nomograms. Mid-sagittal contour of roof of mouth from EPG and x-ray shown above. EPG electrodes are shown as circles. Slanted lines connect x-axis grid to points on this contour. The extent of the lingual constriction and approximate point of F2-F3 proximity is indicated above the contour.

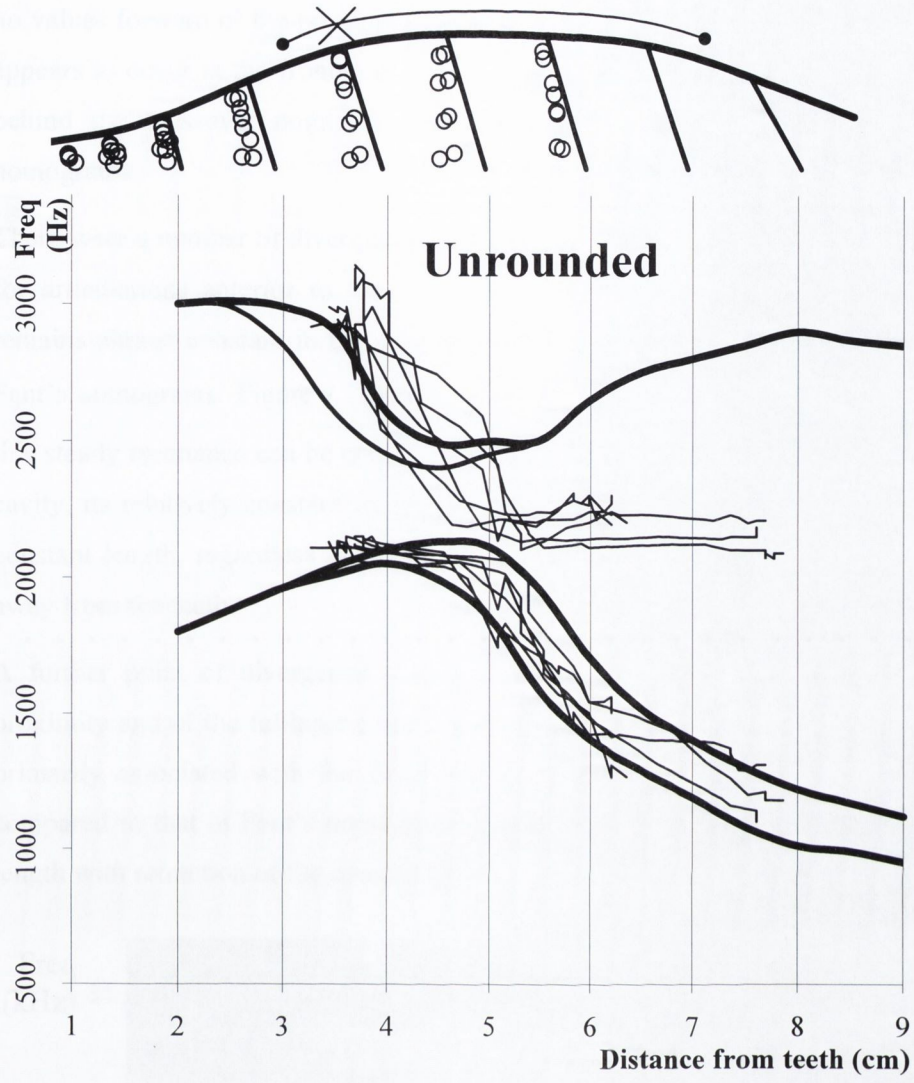



**Figure 9.6** F2 and F3 values vs. distance from the narrowest point of lingual constriction to the teeth for all the unrounded utterances. Dark lines show formant traces for lip cross sectional areas of 0.65 $cm^2$ and 0.16 $cm^2$ from nomograms in Fant, 1970. Mid-sagittal contour of roof of mouth from EPG and x-ray shown above. EPG electrodes are shown as circles. Slanted lines connect x-axis grid to points on this contour. The extent of the lingual constriction and approximate point of F2-F3 proximity is marked above the contour.

The broad agreement between the experimental data and Fant's can be seen from the extent that the experimental data lines correspond to the darker lines in Figure 9.5 and Figure 9.6. Note first of all that the crossover (F2/F3 proximity) occurs slightly under 5 cm from the teeth for the unrounded series of vowels in Figure 9.6, a point which accords very well with Fant's at circa 4.5 cm. For the speaker in this experiment this is the mid palatal region. For the rounded series in Figure 9.5 the crossover is at about 3.5 cm from the teeth, a value almost

identical to Fant's and which corresponds to the prepalatal region for this speaker. There are no values forward of the proximity for the rounded series of vowels in the present data, as it appears to occur at the frontmost limit of this speaker's vowel articulation. For articulations behind the crossover point, the measured F2 values follow the broad trend of Fant's nomograms.

There were a number of divergences from Fant's nomograms. The back cavity resonance (F2 for articulations anterior to the crossover point, and F3 for articulations posterior to it) remains almost constant in the experimental data at about 2200 Hz, rather than rising as in Fant's nomograms. Figure 9.7 below shows a spectrogram of one repetition of [iɯ] where this steady resonance can be clearly seen. If the resonance is indeed associated with the back cavity, its relatively constant frequency could imply that the back cavity maintains a rather constant length, regardless of whether the location of the constriction is moved towards or away from the teeth.

A further point of divergence concerns the rate of change of the falling F3 before the proximity and of the falling F2 just behind the proximity. This resonance is considered to be primarily associated with the front cavity, and the sharper drop in the experimental data compared to that in Fant's nomograms could suggest more extreme changes of front cavity length with retraction of the constriction.

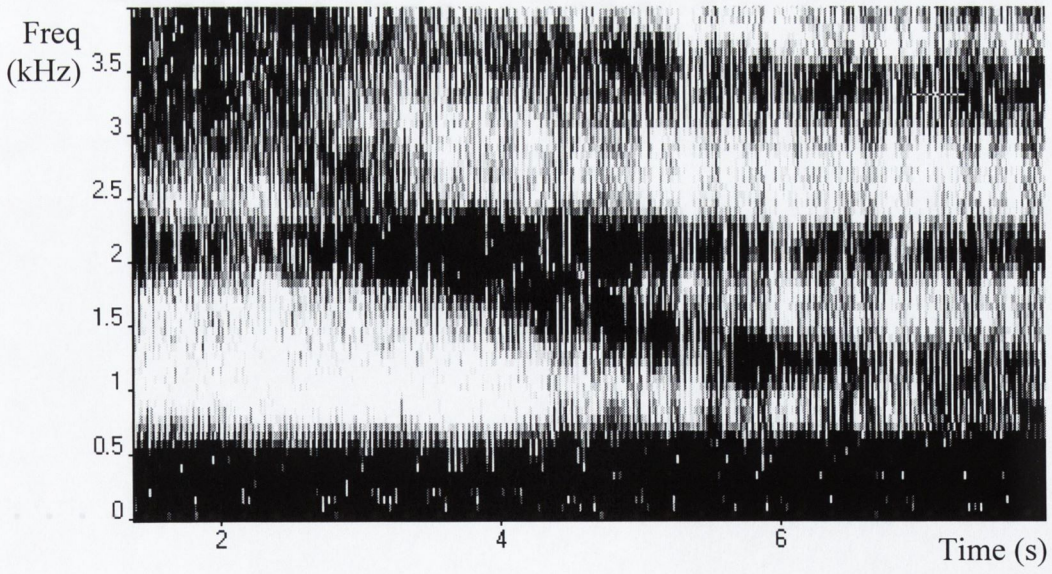



**Figure 9.7** A spectrogram of one repetition of [iɯ]. Note that the formant F2 before 3.8 s, and formant F3 thereafter, are maintained at a relatively steady frequency of 2.1 kHz.

### 9.3.3 Lingual Constriction

The width of the lingual constriction was also estimated. It was taken to be the distance from the circle to the roof of the mouth at the point where this had the smallest value. Figure 9.8 below shows this estimated constriction width plotted against distance to the teeth for all utterances.

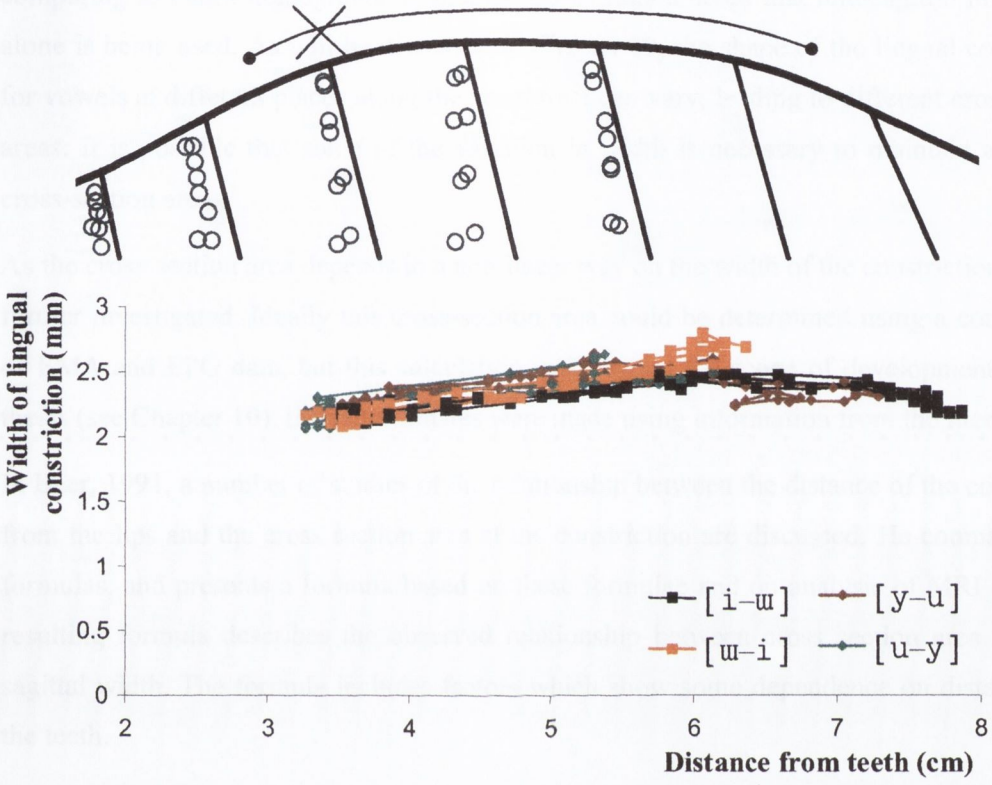



**Figure 9.8** The midline distance between the tongue and the roof of the mouth at the point of narrowest lingual constriction plotted against the distance of this constriction to the teeth for all utterances. Mid-sagittal contour of roof of mouth and EPG electrodes shown above, as for Figure 9.5.

It can be seen that the width is not constant. Looking at productions where the narrowest point of lingual constriction is between 3.2 cm and 6 cm from the teeth more retracted productions have a greater width than those further forward. For production more posterior than this the width of the constriction narrows again. There may be a number of factors contributing to this trend. It may, of course, indicate errors with the method of estimation. It can be seen that the curve of width variation in general reflects the curve of the roof of the mouth. The model may have systematic errors along this curve. Where the lingual constriction is beneath the velum it must be remembered that the trace of the roof of the mouth is based on a single x-ray. The position of the velum may be different to that seen in the x-ray. It can be seen that the [ɯ–i]

continua show a greater constriction width than the other continua for the most retracted productions. This reflects observations regarding [ɯ–i] continua made from Figure 9.3.

However, if the width is not constant this might indicate a further mismatch in attempting to reproduce Fant's nomograms. Constant cross-section area is the important factor when comparing to Fant's nomograms. It should also be remembered that mid-sagittal information alone is being used. As will be discussed in Chapter 10, the shape of the lingual constriction for vowels at different places along the vocal tract can vary, leading to different cross-section areas. It is possible that some of the variation in width is necessary to maintain a constant cross-section area.

As the cross-section area depends in a non-linear way on the width of the constriction this was further investigated. Ideally this cross-section area could be determined using a combination of EMA and EPG data, but this calculation was beyond the scope of developments for this thesis (see Chapter 10). Instead estimates were made using information from the literature.

In Baer, 1991, a number of studies of the relationship between the distance of the constriction from the lips and the cross section area at the constriction are discussed. He compares these formulas, and presents a formula based on these formulae and on analysis of MRI data. The resulting formula describes the observed relationship between cross section area and mid-sagittal width. The formula includes factors which show some dependence on distance from the teeth.

Using data from Baer, 1991, Table VIII, the data in Figure 9.8 was re-plotted as in Figure 9.9. Here the vertical axis is the cross-section area predicted using Baer's formula. The horizontal axis is distance from the teeth. This shows a little consistent variation in cross section area for constrictions in the range from 3.2 cm to 7 cm. An exception to this is for [ɯ–i] continua where rearmost constrictions appear to have larger cross-section areas. This may indeed be due to the vowels being slightly lower or due to the velum being different to that traced from the single x-ray. For constrictions further back than 7 cm the cross-section area is smaller. This is likely to be also due to the traced shape of the velum not matching that of the speaker during production of these vowels.

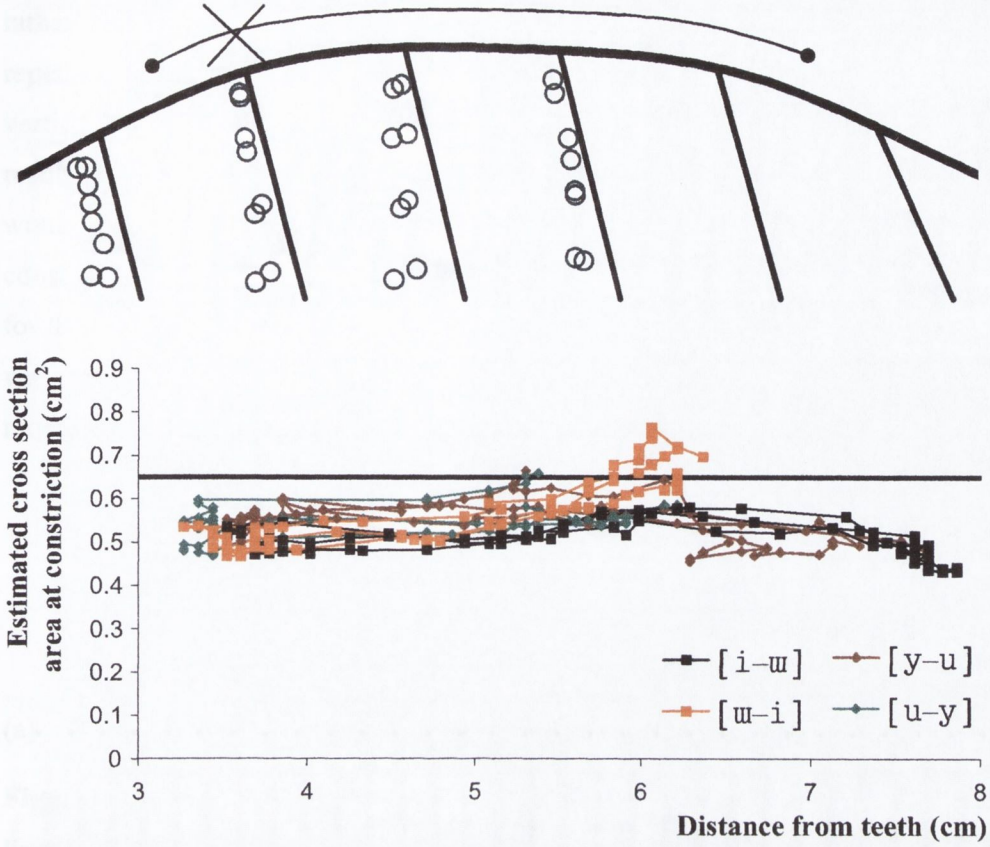


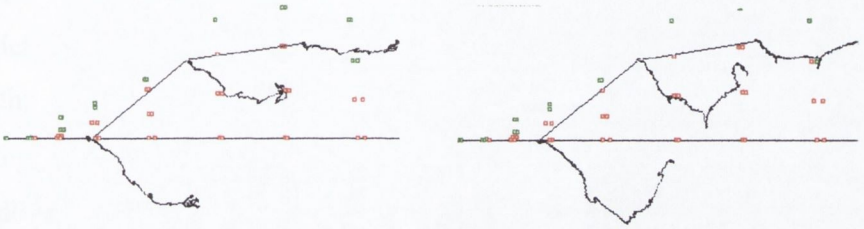

**Figure 9.9** Estimated cross section area at the point of narrowest lingual constriction plotted against the distance from this constriction to the teeth for all utterances. The cross-section area is estimated using data from Baer et al., 1991, and measured data in Figure 9.8 to estimate the cross-section area for each utterance. The dark black line is the cross-section area used in Fant's nomograms (0.65 $cm^2$). Mid-sagittal contour of roof of mouth and EPG electrodes shown above, as for Figure 9.5.

### 9.3.4 Within Utterance Variability

The figures above in some way capture the variability in production of the different utterances, but in a summary manner. Individual utterances were also studied to gain a better understanding as to how the production of those utterances can vary. The estimated lingual constriction was also examined in some detail for individual utterances to see if the circle model performed equally well across utterances and for the entire duration of a given utterance.

Figure 9.10 below illustrates a variation in the speaker's production of an utterance. It presents two repetitions of the utterance [ u–y ]. The illustrations are taken from one of the viewing tools presented in §7.2.1 where the jagged lines represent the locations of the tongue sensors throughout the utterance. Figure 9.10 (a) shows a repetition where the speaker has kept a

rather steady movement of the tongue sensors from start to finish. Figure 9.10 (b) shows a repetition where the speaker would appear to have controlled the tongue differently, as vertical movement of the tongue sensors is more varied in than in (a). While the expected result of retracting the tongue body is achieved, it could be expected that this difference would have some consequence in data presented earlier. Indeed when the width of the lingual constriction for these two repetitions was produced for Figure 9.8 a smooth curve was found for the first repetition, but for the second there are spikes where it would appear that the width was adjusted over a short time interval. These spikes correspond to identifiable changes in the movement of the tongue sensors.



(a) (b)

**Figure 9.10** EMA tongue sensor locations for two repetitions of [u–y], where the jagged lines represent the sensor (x,y) locations for the entire repetition. (a) An example with rather steady changes of vertical and horizontal position. (b) An example where vertical movement changes part way through the utterance.

## *9.4 Discussion*

Returning to the main point regarding articulatory and acoustic mapping, it is recalled that this data broadly agreed with that from Fant, 1970. However there were a number of points where they differed. The speaker exhibited a back cavity resonance which remains almost constant in the experimental data at about 2200 Hz, whereas in Fant's nomograms this resonance is clearly higher for more posterior vowel-like productions than for anterior ones. Also, in the experimental data the front cavity resonance exhibits a greater rate of change of frequency around the F2 F3 proximity than is seen in Fant's nomograms. The following text explores possible reasons for these differences.

In Fant's model the constriction maintains a constant length and cross-sectional area. Other than by the possible addition of a lip section, the entire vocal tract is kept constant. Therefore, by moving the constriction forwards or backwards by a given amount one gets a concomitant shortening/lengthening of front and back cavities by the same amount. However, for the human speaker a resonance primarily associated with the back cavity retains a rather constant frequency, suggesting that the back cavity length may not shorten with tongue retraction.

A number of factors might conspire to maintaining a fairly constant back cavity length. Figure 9.11 overlays sketches of productions of two vowels [i], as solid lines, and [ɯ], as dashed lines, to help illustrate some of these factors.

Firstly, the larynx may move vertically with tongue fronting/backing, given the linkage between the tongue root, the hyoid bone and the larynx. Data for this was not available, but our subject, when asked to introspect on this possibility at a later time, felt it was reasonably likely.

Secondly, it could be argued that the contour of the tongue behind the constriction and the shape of the back cavity is different for [i] and [ɯ] in a way that may effect cavity length, which is measured at the midline between the tongue and the pharynx/roof of the mouth. Part of the difference in back cavity shape between [i] and [ɯ] is caused by the tongue body being further from the pharyngeal wall causing an increase in volume, but the difference in midline length of these cavities is not as great at the distance between the [i] and [ɯ] constriction locations.

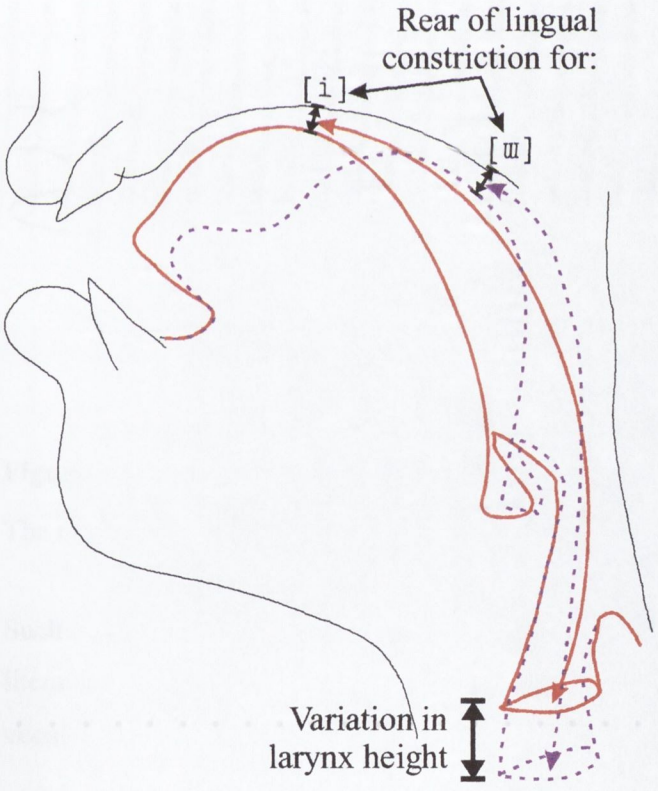



**Figure 9.11** Possible vocal tract shapes for [i] in solid curves and for [ɯ] in dashed curves. The midline of the back cavity for each vowel is marked with the arrow-headed lines.

Thirdly, with the shift in constriction from [i] to [ɯ], the length and shape of the lingual constriction changes as an inevitable consequence of the fact that the convex curve of the tongue is approximating an upper surface which is not flat. When the upper surface is concave, as for [i], one gets a relatively long constriction: with more retracted articulations the upper surface is less concave and constriction length gets shorter. (This can be seen, for example, in Figure 9.13.)

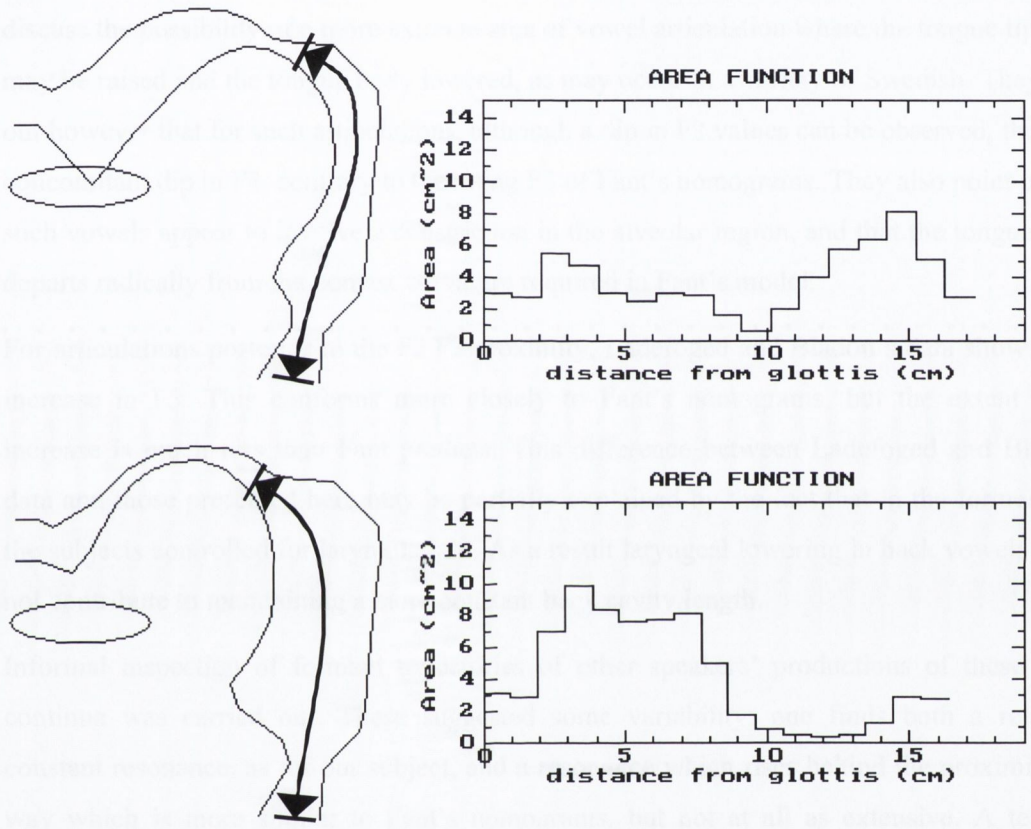



**Figure 9.12** vocal tract shapes from Maeda's model for [ɯ] (top) and for [i] (bottom). The midline of the back cavity is marked with arrow-headed lines.

Such variations in constriction length, larynx height, and back cavity shape are found in the literature. Maeda, 1990, presents an articulatory model derived from x-ray. Figure 9.12 shows vocal tract configurations and area functions from Maeda's model for approximations of [i] and [ɯ]. These were generated from settings for French vowels [i] and [u], and adjusted to provide a cross sectional area at the lingual constriction and at the lips that would be closer to the values of relevance here. The midline for the back cavity has been sketched freehand for the two articulations. When the tongue is bunched towards the front of the palate as for [i] the midline is straighter than when retracted for [ɯ], where it is more curved and closer

to the rear wall of the pharynx. The fact that the midline length is very similar for both articulations can also be inferred from the area functions.

It can therefore be concluded that the apparently constant back cavity resonance seen in the data can be reasonably accounted for. However, this constant resonance at about 2200 Hz, characteristic of the speaker in this study, is not necessarily found for all speakers. Ladefoged and Bladon, 1982, found, as was found here, that for articulations forward of the F2 F3 proximity, F2 did not dip below its peak value as predicted in Fant's nomogram. The authors discuss the possibility of a more extreme area of vowel articulation where the tongue tip/blade may be raised and the tongue body lowered, as may occur in a variety of Swedish. They point out however that for such articulations, although a dip in F2 values can be observed, there is a concomitant dip in F3, contrary to the rising F3 of Fant's nomograms. They also point out that such vowels appear to involve a constriction in the alveolar region, and that the tongue shape departs radically from the convex curvature required in Fant's model.

For articulations posterior to the F2 F3 proximity, Ladefoged and Bladon's data shows some increase in F3. This conforms more closely to Fant's nomograms, but the extent of the increase is much less than Fant predicts. This difference between Ladefoged and Bladon's data and those presented here may be partially explained by the fact that in the former study the subjects controlled for larynx height. As a result laryngeal lowering in back vowels would not contribute to maintaining a more constant back cavity length.

Informal inspection of formant trajectories of other speakers' productions of these vowel continua was carried out. These suggested some variability: one finds both a relatively constant resonance, as for our subject, and a resonance which rises behind the proximity in a way which is more similar to Fant's nomograms, but not at all as extensive. A tentative conclusion is therefore that the back cavity length does not shorten in a way that is linearly related with retraction of the vowel constriction for the reasons discussed above: whether it shortens, and the extent to which it shortens, is likely to vary insofar as the speaker may allow factors such as laryngeal height to vary. The steepness of the palatal dome also varies from speaker to speaker. This dome was very steep for the subject of this study. Such anatomical differences are likely to affect the shape and length of the lingual constriction for the most anterior articulations and might also be a source of some interspeaker variability.

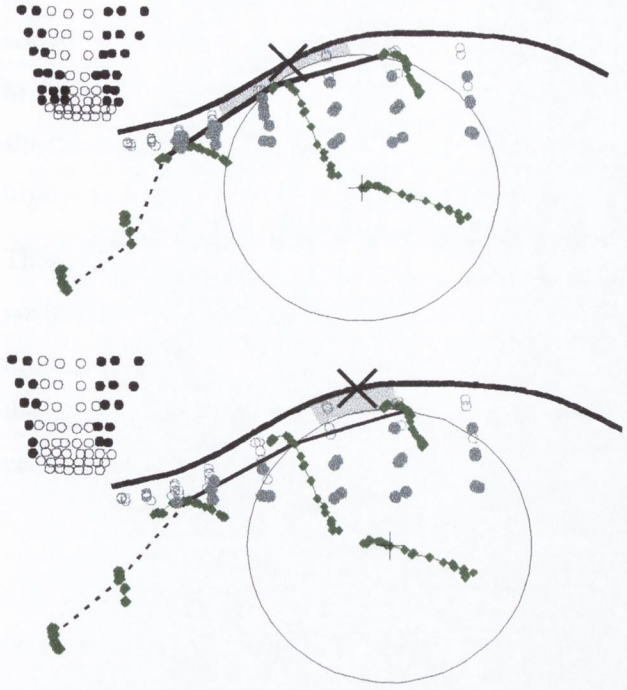


**Figure 9.13** Modelled data showing the narrowest point of constriction (X) and its approximate length (shaded in grey). Upper panel shows X at c. 3.5 cm from the teeth and the lower panel shows X at c. 4.5 cm. Also shown are the trajectories for the three tongue sensors, the lower jaw and lip sensors, and the centre of the circle model, for a single unrounded vowel continuum. EPG contact patterns are also shown.

The other divergence between the present data and Fant's nomograms concerned the very sharp fall in the front cavity resonance in the prepalatal to midpalatal region (3.5 cm to 5 cm from the teeth). The reason for this is likely to be very similar to the last mentioned reason for the back cavity's apparent invariant length, i.e. the fact that the length and shape of the constriction may alter with changing anterior articulations in a way that may have major implications for the front cavity length. Given the sharply curved surface of the prepalatal area, and the convex shape of the tongue, the length of the constriction increases as the tongue body moves from the midpalatal to the prepalatal region. Figure 9.13 above examines the lingual constriction that is estimated using the circle model. Two time-points in a repetition of [ɯ–i] are presented (using the viewer described in Figure 7.5). The predicted centre of the constriction is marked with a large, black X mark. Areas corresponding very approximately to the constrictions have been shaded in for both articulations. For a prepalatal lingual constriction as shown in the upper frame of Figure 9.13, a long narrow constriction is formed between the tongue and the curved surface of the prepalatal area. For a mid-palatal constriction, such as that in the lower frame, the constriction formed is shorter as the palate is less curved than the prepalatal region. Note that for a longer constriction length, much of the

constriction is forward of the narrowest point of constriction. In a progression from midpalatal to prepalatal, a small forward articulatory step in this region may result in a relatively sudden shortening of the front cavity. The prepalatal region, which forms part of the front cavity for high central vowels, may become part of the constriction for high front vowels.

This point raises a potential problem with this analysis. The articulatory measure that is correlated to acoustic measures is the distance between the narrowest point of constriction and the teeth. However this may not be a good predictor of front cavity length: partly because of the changing constriction length and partly because the narrowest point is not always the centre of the constriction.

# Chapter 10 Additional Information from EPG in the Study of 'Human Nomograms'

## *10.1 Introduction*

An important aim of this work was to develop visualisation and analysis tools to enable the parallel analysis of EMA and EPG data that would maximise the information yield from these techniques, individually and together. Such tools are discussed in Chapter 5, Chapter 6, and Chapter 7. This chapter looks at the results of analysis of simultaneously recorded EMA and EPG data using some of those tools.

The analysis presented in Chapter 9 was of data where EPG and EMA data were recorded simultaneously. That analysis was primarily based on EMA given its much higher spatial resolution. For the topic at hand, EPG suffers from limitations that it doesn't extend to the soft palate and that the distance between electrodes allows limited resolution. On the other hand trying to interpret tongue shape from the location of a small number of EMA sensors is difficult: in this experiment inferring the location of the lingual constriction required some level of modelling, for which a tongue circle model was developed. EPG data provides information on a greater number of points, albeit points on the roof of the mouth as opposed to on the tongue, which could in principle improve ones ability to locate the lingual constriction using this complementary data.

In §10.3 a number of methods are applied to the EPG data to investigate how information on the location of the lingual constriction could be extracted. In §10.3.1 the row on the EPG palate that exhibits the narrowest constriction is identified for a small number of EPG frames, and this is taken as the row at which the centre of the lingual constriction is located. The location estimated from EMA with the tongue circle model is overlaid to allow comparison of the two estimates. That these comparisons showed the EMA estimate to be within the region of the EPG row of greatest narrowing added to the confidence in the EMA data. Two methods are then used to see if a more precise estimate of the location of the lingual constriction could be made from EPG data. In §10.3.2 a method of drawing the upper edge of lingual contact with the EPG palate as seen from the side is used. These contours (called EPG tongue contours below) are an attempt to get a more detailed representation of the lingual contact through the use of curve-fitting through points based on the location of activated electrodes. EPG tongue contours are compared to corresponding EMA tongue circle representations to investigate the effectiveness of the contours at predicting the lingual contact and at facilitating insights into lingual activity. In §10.3.3 the centre of gravity index (as described in §5.3.2) is

investigated as a potential predictor of the lingual constriction's location. Centre of gravity values are compared to corresponding EMA circle model estimates of the location of the lingual constriction, and the relationship between the values is examined for general trends which would indicate that the index is working as a similar estimator to the circle model. The values that the index yields are also checked to see if they are interpretable for this experiment. Before comparing the date from the two devices though, §10.2 reviews the important differences between the measurements made with EMA and EPG.

The analyses illustrate the different aspects of lingual articulation the two recording methods measure and the strengths and weaknesses of analysis methods developed for this thesis to work with them. It is hoped that, along with validating the EMA analysis and adding qualitative information to the analysis, this work will help to guide how analysis methods could be further developed to effectively combine information from the two devices.

Part of this work was already presented in Fitzpatrick & Ní Chasaide, 2000.

## *10.2 Differences in measurements by EMA and EPG*

There are a number of differences between the data yielded by EPG and EMA. The differences listed here are some of the more important ones to be considered when analysing such data. Further discussion of the devices is found in Chapter 2.

EPG provides a relatively low resolution measure of the contact between the tongue and the EPG palate when compared to the resolution EMA provides for the location of sensors on the tongue. The distance between EPG electrodes on the speaker's palate varied between 2 mm and 8 mm. While the accuracy with which EMA measures the location of a sensor is dependent on factors such as the care with which the experiment is run and on lateral or twisting movements of the tongue, better than mm accuracy can be expected.

EPG does not track flesh points on the tongue but instead shows regions of the roof of the mouth with which some part of the tongue is in contact. So, for example, in the production of vowels the sides of the tongue may contact the sides of the EPG palate. For lingual stops contact typically shows across a full row of EPG electrodes, with contact involving the midline of the tongue as well as the sides of the tongue. The precise part of the tongue making contact with the EPG palate can only be inferred. EMA measures the absolute positions of flesh points on the midline of the tongue as the attached sensors move with the tongue during an utterance.

EPG does not extend to cover the soft palate so it yields less information for back vowels than for front vowels as there is less contact with the artificial palate. EMA is equally effective for front and back productions.

EPG has a large concentration of electrodes on the anterior part of the EPG palate so it can measure a lot of detail about tongue tip articulations where the tip makes contact with the EPG palate. Insofar as the most anterior EMA sensor on the tongue can be placed no closer than 0.5 cm from the tongue tip, and in practice is usually some mm further back, EMA cannot record tongue tip articulations. This means that EMA is less useful for apical and laminal stop articulations than EPG is.

EPG shows contact patterns for a large curved region of the roof of mouth whereas EMA tracks only a small number of flesh points on the midline. Thus EPG contains information about 3 dimensional aspects of lingual activity whereas the EMA information is only 2 dimensional.

These differences mean that although both instruments record aspects of lingual activity, the resulting data contains different information. This makes it difficult to find measures that make use of the potentially complementary information from both data sources.

## *10.3 Comparison of EMA Tongue Circle Model and EPG Data*

### 10.3.1 EPG Top views

In Chapter 9 the location of the centre of the lingual constriction was estimated from EMA data for sequences of high vowels. Those results are compared here with scaled representations of EPG contact patterns to see if the latter representation yields similar estimates of constriction location.

In the lower part of Figure 10.1, EMA based measurements are reproduced from Chapter 9. F2 and F3 values for all unrounded utterances are plotted as a function of the distance between the lingual constriction and the front of the upper incisors. That distance was calculated from EMA data. Immediately above this is shown the mid-sagittal outline of the roof of the mouth from the front of the artificial palate to a point approximately 2 cm behind the back of the artificial palate. EPG electrodes are shown as circles. Points on the palate at increments of 1 cm from the front of the teeth are joined by slanted lines to gridlines in the lower graph. This is a replica of Figure 9.6, and is discussed in detail in §9.3.2.

It is clear that access to the EPG palate gives us a way of relating the EMA sensors and our acoustic data to very precise landmarks on the roof of the mouth. For example, the F2/F3 crossover occurs just before row 7 on the artificial palate. This corresponds to the mid-palatal region for this speaker. This corresponds to slightly under 5 cm from the teeth.

Note that 5 cm from the teeth is a measurement of distance along the surface of the EPG palate. It is possible to measure distance from the teeth from EMA data alone, but distance

thus measured is the straight, point-to-point, distance. Having the EPG palate provides easy access to the anatomical reference points that phoneticians have traditionally worked with. Further to this, EPG rows can be visually inspected for their relationship to the contour of the plastercast of the speaker's palate. With EMA alone what the distance represents remains much more vague. For example, the junction between the hard and soft palate cannot easily be determined from EMA alone. Although it wasn't done in this study direct use of measurements of a plastercast could be used to provide the additional information needed to help interpret the EMA data.

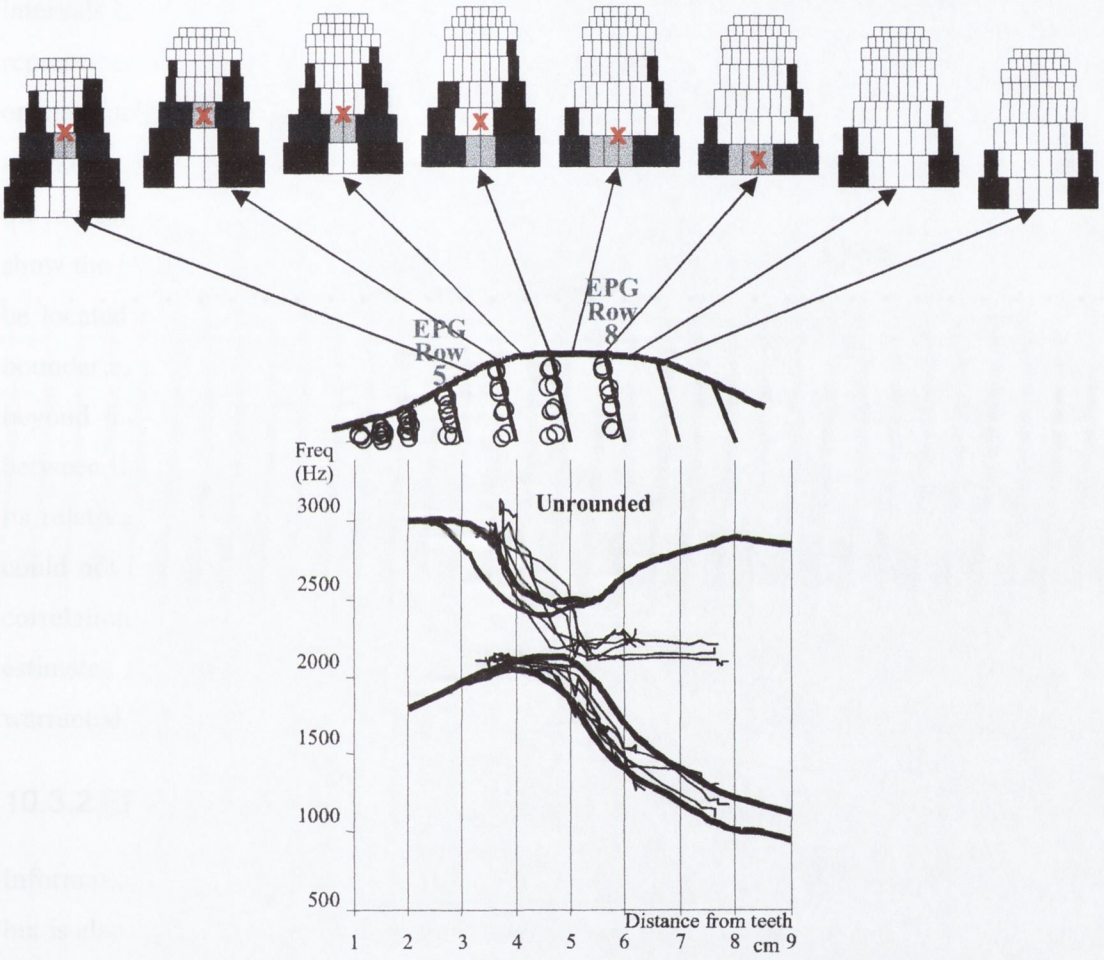



**Figure 10.1** A replica of Figure 9.6 is shown, with a mid-sagittal contour of roof of mouth from EPG and x-ray shown above. Slanted lines connect x-axis gridlines to points on this contour. EPG electrodes are shown as circles. Corresponding EPG top views are shown in the upper part of the figure from a single repetition of [i–ɯ]. On the top views the row with greatest EPG narrowing is coloured grey, and the location of the EMA constriction estimate is marked with an 'x'.

Returning now to the relationship between the formant trajectories and the location of the lingual constriction, the precise range of this speaker's high vowel articulations can be identified. The most anterior point is in the prepalatal region, midway between row 5 and row 6 of the artificial palate for this speaker. The most posterior vowel articulations are at a point in the velar region 1.5 cm behind row 8 on the artificial palate.

In the upper part of Figure 10.1 are shown EPG top views taken from a single repetition of [i-ɯ], at points where the corresponding EMA data indicated lingual constrictions at 0.5 cm intervals between 3.5 cm and 7 cm from the teeth. In the EPG top views an EPG electrode is represented as a rectangle in which it is centred. The dimensions of the rectangles are based on measurements of distances between electrodes and between rows of electrodes on the speaker's artificial palate. The black areas represent regions of tongue contact. In the row with the narrowest constriction the inactivated rectangles are shaded grey. An 'x' has been placed to show the location of the constriction as estimated from EMA data. One would expect the 'x' to be located in the grey region. In the top view frames the 'x' is located within or just at the boundaries of the grey regions in all frames except in the 4$^{th}$ frame, where the 'x' lies 0.5 cm beyond the boundary of the grey region. It was generally found that the correspondence between the two types of estimates was very good. Thus, although the EPG technique, given its relatively poor resolution for the hard palate and its lack of coverage of the soft palate could not on its own yield sufficiently accurate estimates of constriction location, the good correlation of the two types of information serves as reassurance that the EMA-based estimates are probably reasonably accurate and that the use of the tongue circle model is warranted.

### 10.3.2 EPG Derived Tongue Contours (Side Views)

Information about the location of the constriction is not only seen in the narrowest EPG row, but is also reflected in the contact on surrounding rows. Ideally this information could be used to improve the estimate of constriction location. In Chapter 9 the outline of the tongue body was approximated using a circle that was considered to be fixed to two EMA sensors. That approximation is discussed in §6.2.2 and was used to locate the lingual constriction. In this section a simple curve-fitting method was used to generate EPG contours which estimate the highest part of the tongue that is in contact with the EPG palate. The EPG tongue contour is discussed in §7.2.1, but to recap briefly, the EPG tongue contour is drawn on the basis of the innermost activated electrode on each row. A point halfway between each such electrode and the electrode inside of it is identified, and a smooth curve is fitted to pass through those points and ending at an estimated location of the root of the tongue. In this section EMA data with

the tongue circle model applied to them are overlaid on EPG data for which the EPG tongue contour is drawn.

In Figure 10.2 EPG contact patterns are drawn with EPG tongue contours overlaid. On top of these again are drawn corresponding EMA tongue circle model representations to facilitate comparisons between the representations. The basis of the displays is discussed in more detail in §7.2.1 for Figure 7.4. Three time-points are presented from a single repetition of [i–ɯ]. The time-points are in an order where the constriction location retracts progressively from (a) to (b) to (c).

In Figure 10.2, the tongue circle model suggests an interpretation of tongue dynamics that might not be apparent from perusal of the EMA sensors alone. Note that the horizontal movement of the 'pseudo-sensor' (the centre of the EMA tongue circle) is more extensive than would be inferred from either dorsal sensor, and that it correlates well with traditional descriptions (based on phoneticians' intuitions) of tongue movement for this utterance. The data further suggest that the reason for this is that tongue retraction in [i–ɯ] may involve different types of movement for different parts of the continuum. Stepping through Figure 10.2(a) to Figure 10.2(c), note large vertical movement in the anterior dorsal sensor while the posterior sensor barely moves. For articulations in first part of the continuum, approximately [i–ɨ], the lingual constriction retracts because the tongue body appears to be rotating in an anti-clock wise fashion at a point close to the posterior sensor. Further retraction from [ɨ–ɯ] appears to be achieved by simply retracting the tongue body. Further discussion of these types of movement patterns are presented in §9.3.1.

When the EMA data and the EPG tongue contours are compared a number of differences emerge. First of all, whereas for the EMA representation the highest point of the tongue retracts from (a) to (b) to (c), there is in the EPG representation no change between (b) and (c). This one assumes has to do with spatial resolution of EPG. A contributing factor might also be that the EPG contour, as currently drawn, masks asymmetry of contact, and this is a point that is returned to below. The second striking difference concerns the apparent disparity in tongue height for the two types of representations in (a), (b), and (c). The EPG dorsal contour is always somewhat lower, and this is most striking for time-point (c). Thirdly, the two contour lines are not parallel. An explanation for these last two differences probably lies in the fact that the EPG electrode activation is at the sides whereas the EMA sensors are in the mid-sagittal plane.

(a)

(b)

(c)

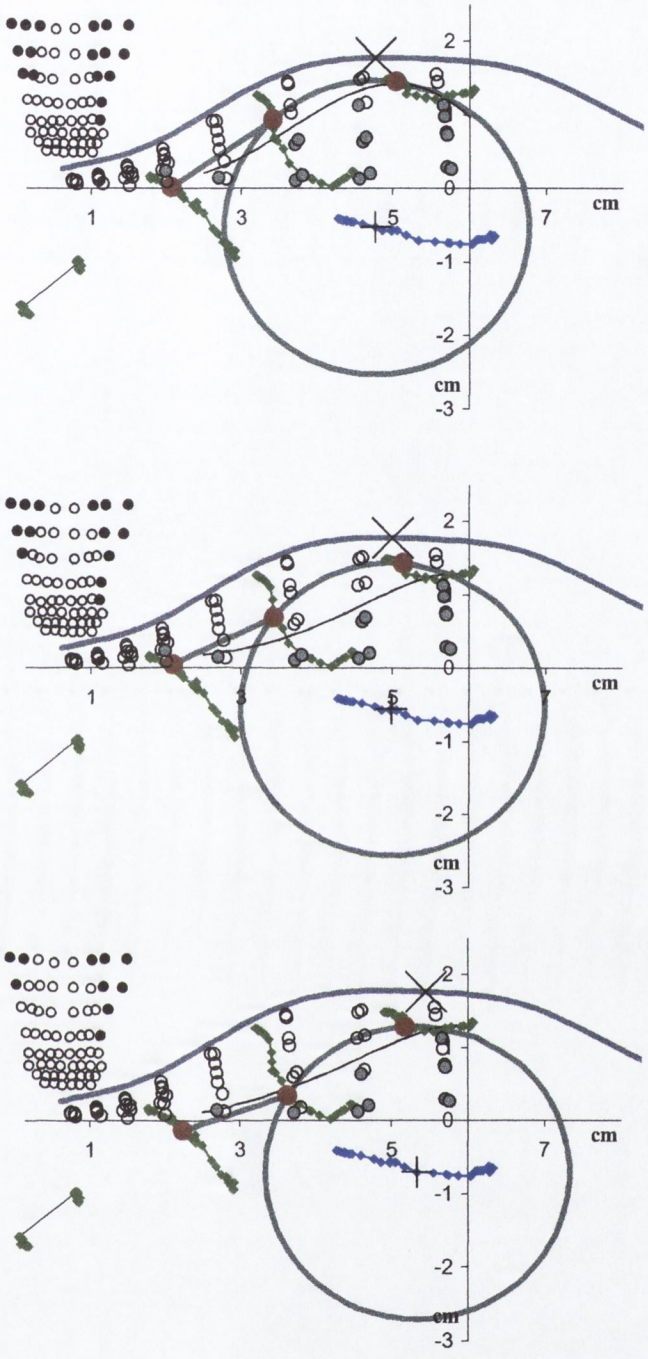



**Figure 10.2** Representation of tongue contours for EMA with tongue circle model and for the EPG derived tongue contours for three successive time-points from a single repetition of the [i–ɯ] continuum. The time-points are labelled (a), (b), and (c). EMA x-y movement of all sensors for the entire [i–ɯ] are indicated, along with that for "pseudo-sensor" for tongue circle centre. Rotated top views for EPG also shown.

It was hoped that the EPG contour might provide a way of showing the narrowest point of constriction in a way that might overcome some of the spatial limitations of EPG resolution as

well as its lack of coverage of the velar region. There is some indication that this might work. Looking closely at point (a) the EPG contour suggests a constriction located closer to row 7 than row 8. This mirrors the EMA estimate more closely than the traditional type of estimate based on the EPG top view, where the narrowest row is row 8. However, at time-points (b) and (c) there is a poor correspondence for constriction location. The EPG contour would indicate a constriction somewhere behind the artificial palate, and not between rows 7 and 8 as suggested by the EMA estimate. Not surprisingly, this correspondence was least good for more posterior constrictions, i.e., where EMA estimates place the constriction at about row 8 of the artificial palate or behind that point. Thus this EPG representation, as currently drawn, is not a reliable device for inferring the lingual contour behind the EPG palate.

There are a number of factors that limit the performance of this curve-fitting method of drawing contours. Firstly, it does not take into account the asymmetry of contact on the EPG palate. For example in Figure 10.2(c) there is an asymmetric contact in rows 7 and 8. In comparison, in Figure 10.2(b) there is no such asymmetry and the pattern suggests that the tongue must be relatively somewhat higher, and the constriction narrower. Yet, the EPG contours are identical for the two figures. This is because the innermost activated electrode was used in drawing the contour so this level of information is lost.

A second limitation of the EPG contour as a representation of tongue height arises out of the fact that the actual EPG electrodes activated are at the sides of the palate, not on the mid-sagittal plane, and there may be doming of the tongue towards the mid-sagittal plane. Such doming is illustrated in Figure 10.3 which is adapted from Fant, 1970. On the left of Figure 10.3 a mid-sagittal section of the vocal tract for the vowel [ɨ] is shown. On the section are lines labelled (1), (2), and (3). To the right are shown cross sections of the vocal tract at those points. Cross section (3) is located approximately at the narrowest point of the lingual constriction for the vowel. In the cross section the mid-point of the tongue is clearly higher than the points where the tongue comes into contact with the roof of the mouth. This might explain why the EPG contour is typically lower than the EMA data.

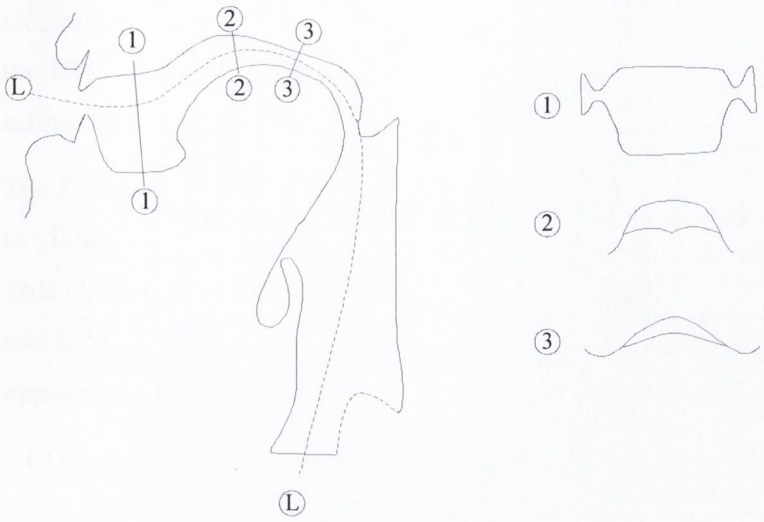



**Figure 10.3** The midline of the vocal tract for the vowel [i] on the left side, and three cross section outlines on the right side for the points labelled (1), (2) and (3). (Adapted from Fant, 1970, Fig 2.2-5).

Furthermore the vertical distance between the height of the tongue at the mid-sagittal plane and the height of the tongue where it is in contact with the EPG palate may vary along the length of the EPG palate for two reasons. Firstly, the lateral contact is at different distances from the mid-sagittal plane for different parts of the EPG contour. Where there is doming away from the mid-sagittal plane, the height at the point where the tongue touches the EPG palate will be lower the further that contact is from the mid-sagittal plane. Thus the extent to which mid-sagittal tongue height is underestimated by the EPG contour would be greater on rows where less central contact is seen. Secondly, there may be a difference in the cross section shape of the tongue along its length. Such a difference is seen in Figure 10.3 between cross-section (2) and cross-section (3). These differences may explain why the EPG tongue contour and the EMA data are not parallel in, for example, Figure 10.2(b).

Such observations draw attention to the three dimensional nature of both EPG data and of lingual activity. While the particular implementation of EPG tongue contours used here did not perform as well as might be hoped at making use of such information, it is clear that there is information here to be harnessed. Even as currently implemented it is useful in drawing attention to three dimensional aspects of the lingual contact.

### 10.3.3 EPG Centre of Gravity Index

There are a number of EPG indices discussed in §5.3.2 which can describe some aspect of EPG palate contact patterns as a single numerical value. One example is the centre of gravity (COG) index that was devised to specify the location of the main concentration of activated

electrodes across the whole palate (Hardcastle et al., 1991). As COG pinpoints a location on the EPG palate where the electrodes are concentrated it was considered a possible way of estimating the centre of the lingual constriction.

The formula used here to calculate the index is modified from that in Hardcastle et al., 1991 to yield a value which relates directly to EPG row numbers as used throughout the formula. This change does not effect the magnitude of the difference between the COG values for any two EPG contact patterns, but with this formula a higher COG indicates a more posterior, as opposed to a more anterior, concentration of electrodes. The formula used here is:

$$COG = 0.5 + ((½R1 + 1½R2 + 2½R3 + 3½R4 + 4½R5 + 5½R6 + 6½R7 + 7½R8) / total)$$

where R8 is the number of activated electrodes on row 8, etc., and total is the total number of activated electrodes. Thus a COG value of 6 indicates that contact is concentrated around row 6, and a COG of 6.5 indicates that contact is concentrated halfway between row 6 and row 7. (Note the original formula from Hardcastle et al., 1991, is shown in §5.3.2.)

In Figure 10.4, for all the utterances in the experiment, COG values are plotted against the EMA estimate of the distance between the teeth and the lingual constriction along the EPG palate. A thick black line shows the averaged trend-line of those plots.

If both COG and the EMA circle model act as predictors of the location of the oral constriction a simple relationship between the two calculations in this figure would be expected.

Considering only the thick black trend-line in Figure 10.4, it has a rather smooth slope in the region where the EMA estimated constriction location is between 3.7 cm and 5.6 cm from the teeth or, equivalently, where the COG is between 6 and 7 (i.e., from EPG row 6 to EPG row 7). This smooth slope is indicative of a simple relationship between the two methods. Where the lingual constriction is further from the teeth the relationship becomes increasingly non-linear, presumably due to the diminishing number of electrodes that the COG is calculated with. Further forward the relationship is also less linear. This may be caused by the rows of electrodes towards the front of the palate being closer together than those at the back. Contact closer to row 5 than row 6 is likely to have considerable contact spreading forward onto the more tightly packed row 4 and row 3 as well as backwards onto row 6 and row 7. As the standard COG index does not take account of the distance between rows, small forward shifts of the tongue cause relatively more electrode activation on the anterior rows. Consequently the index will change more quickly than the movement. (See Pompino-Marschall & Mooshammer, 1997, for a modified COG that takes distance between rows into account).

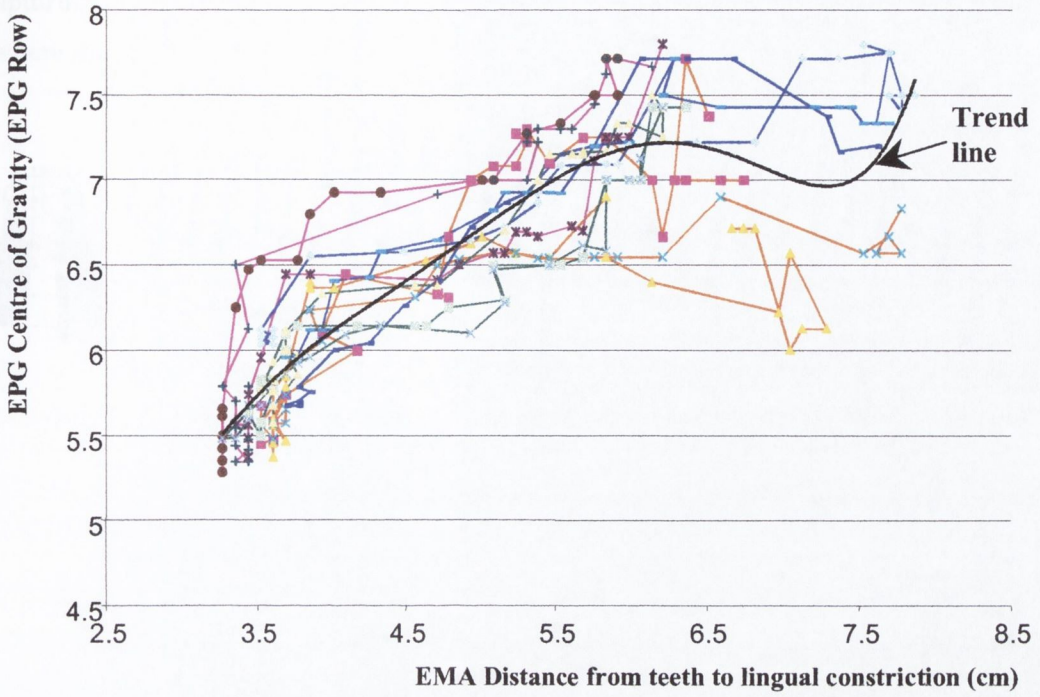



**Figure 10.4** Constriction location estimated on the basis of EPG Centre of Gravity (vertical axis) plotted against the EMA circle model estimate for each individual utterance. The thick black line is the overall trend line for all utterances.

The less linear relationship in the anterior region (i.e., forward of 3.7 cm) could also be due to estimation errors with the tongue circle model as discussed in §9.4. In §9.4 it was observed that that the circle model may estimate a more anterior location for the lingual constriction due in part to its failure to account for the raised tongue front that is immediately forward of the circle model in such articulations.

Even where the trend-line shows a convincingly linear relationship, looking at the plots for individual utterances in Figure 10.4 there is a lot of inter-utterance variation. A possible explanation is that the lower resolution of the EPG device yields values that change in larger steps than the EMA estimate does.

There may be a similar systematic reduction in the effectiveness of the COG for more posterior articulations where increasing amounts of lingual contact are behind the EPG palate. In Figure 10.5, Figure 10.1 is reproduced with the COG value shown for each EPG contact frame at the top of the figure. For the most anterior contact (leftmost EPG top view) the estimate from COG, shown as an 'O', is slightly posterior to the circle model estimate of lingual constriction, shown as an 'x'. But for successively more retracted vowels the COG index progresses to being anterior to the circle model estimated location. For the two most posterior articulations there is no change in COG value as the EPG device is no longer

capturing any information about the change in lingual position being detected with EMA (where the effect of such movement is showing in the acoustic signal).

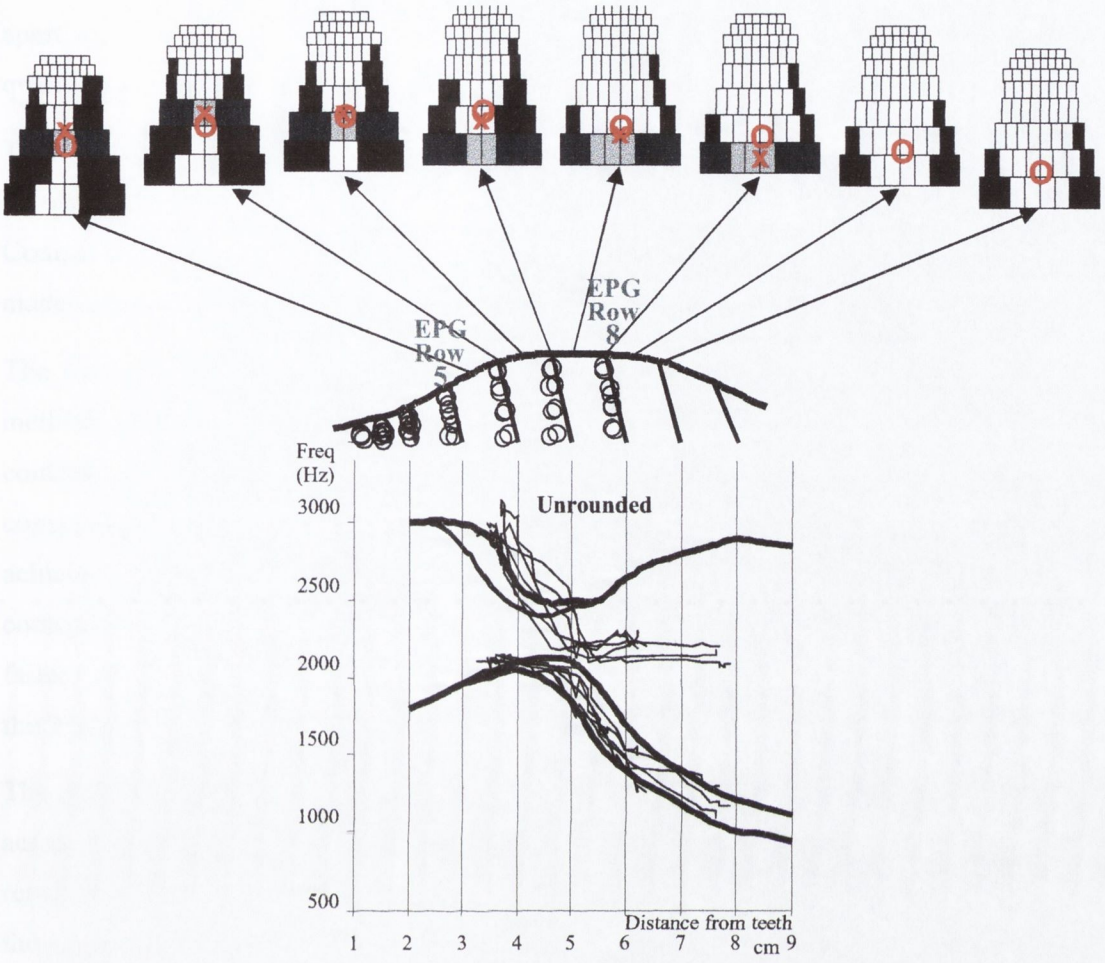



**Figure 10.5** A replica of Figure 10.1 is shown with an 'O' overlaid on the EPG top views to indicate the EPG centre of gravity value for the EPG contact patterns.

Though it is clear that the centre of gravity index in some way reflects the location of the lingual constriction, it is not clear how the index can be used to improve upon the estimate made with EMA and the circle model. While non-linear effects of differences in distance between EPG rows could be accounted for, the effects of contact being behind the palate on the index are more troublesome.

## *10.4 Discussion*

The location of lingual constriction that one would infer from EPG data is generally similar to that inferred from EMA data. In Figure 10.1 it is seen that using the narrowest EPG row to indicate the centre of the lingual constriction generally matched the corresponding estimate from EMA, notwithstanding the limits of EPG spatial resolution. The spatial resolution for

EPG is considerably lower than EMA on the palatal and post palatal region. With EMA, movements of a fraction of a mm are detected. With EPG, rows of electrodes are up to 8 mm apart on this speaker's artificial palate. This greatly limits the usefulness of EPG for making quantitative measures of place of lingual constriction.

There is however more information about the lingual constriction in an EPG pattern than can be obtained by simply considering the row on which the narrowest constriction is located. Contact forwards and backwards of that row also contributes information. Two attempts were made to utilise that information to make better estimates of the location of the constriction.

The first was the use of EPG tongue contours presented in Figure 10.2. The curve fitting method used to draw these curves makes use of the location of EPG electrodes, the EPG contact pattern, and the approximate location of the root of the tongue to produce a constrained curve. The intention was to estimate the contact at points between all rows and to achieve better results for posterior articulations. The particular method used produced contours that did not enhance the estimation of the lingual constriction due to factors such as a failure to consider asymmetrical contact patterns. They did however raise useful insights into the 3 dimensional nature of EPG data.

The second used the EPG centre of gravity (COG) index. When considering results averaged across all repetitions in Figure 10.4 the index appeared to produce results comparable with the result from EMA. The results were poorer for back vowels than front vowels, probably due to the smaller number of electrodes involved. Also in an individual utterance the results showed a jagged and often step-like changes, probably resulting from the resolution limits of EPG.

For the specific experiment in Chapter 9 EPG, used alone, has limitations such as relatively low resolution and a restriction to contact in the area between the top incisors and the back of the hard palate which would make it unsuitable as a principal method of locating the lingual constriction. But having simultaneous EPG data with EMA data enhanced the descriptive relevance of the study in a number of ways, and adds insights, not inferable from EMA alone. The EMA analysis involved a degree of abstraction through the use of a tongue circle model to permit interpretation of the sensor position data in a way that the centre of the lingual constriction could be calculated. The comparison of EPG data for some utterances served as a gross check on the accuracy of the results - even where the result of EPG analysis itself was insufficient as a basis for estimating the centre of the constriction.

A major benefit in employing simultaneous EMA and EPG was the possibility of locating EMA sensors very precisely relative to specific landmarks along the upper surface of the mouth, landmarks which are used in general phonetic descriptions of articulations. Thus it was possible here to locate with considerable precision where, in place of articulation terms,

the F2/F3 crossover, the extremes of the range of high vowel productions are located for the speaker, and for any location within the range of high vowels the corresponding resonance could be described.

In a typical EMA study of lingual articulations the point-to-point distance from an EMA sensor to the teeth can be calculated. However relating such measures to articulatory terms can be difficult, and generally requires additional information such as the precise shape of the speaker's palate and alveolar ridge. The EPG palate and its plastercast provide excellent access to this additional information. In this study the distance along the roof of the mouth of the location of the lingual constriction located from EMA data was related through the EPG palate and plastercast to a rather precise place of articulation.

A final point worth mentioning is that the EPG, EMA comparison was also useful in illustrating that the midline height of the tongue is not necessarily at the same height as where it comes in contact with the EPG palate. This yielded a richer understanding of the data, and pointed to ways that such information might be harnessed in future work.

# Part IV - Conclusions

# Chapter 11 Discussion and conclusions

## *11.1 Introduction*

In this chapter we review the motivations and achievements of the thesis, in addition we look at immediate and long term future work.

## *11.2 Motivation*

In experimental phonetics the instrumentation used for the analysis of speech production is usually applied to the acoustic signal or to measurements of movement of the speech organs. Analysis based on acoustics have traditionally received the greatest focus as the primary instrumentation is a relatively simple microphone. There is no equivalent piece of instrumentation to measure all the movement of speech organs. Instead many devices exist to measure different features of the movement of different organs. Each device has advantages and drawbacks in its operation.

The two devices that are the focus of this thesis are used to measure aspects of lingual articulation,. They are EPG (electropalatography) and EMA (electromagnetic mid-sagittal articulography). Either device provides data which contains only incomplete information on lingual activity. The data they provide is for multiple related points rather than, for example, a global image of the tongue. Perhaps for these reasons there are many different analysis methods used in research to analyse these data. With better analytical techniques more information can be extracted from the data from these devices, allowing for better inferences about the lingual activity they measure. This would allow questions which cannot be addressed today to be investigated. A better understanding of how the tongue performs during speech is of use in fields as diverse as speech therapy and man-machine interfaces.

## *11.3 Achievements*

The first major aim of this thesis was to develop enhanced analytical tools to help the experimental phonetician to exploit the measurement techniques. Such tools should make effective use of the data from those devices both individually and together. As two devices measure features of lingual articulation that are in some ways complementary one major challenge is to establish whether, and how, they might be used simultaneously to make best use of the complementary information. A variety of tools and techniques were developed to facilitate browsing, analysing, and displaying these types of data. Some of these tools are improvements on methods described in the literature, while others are more innovative. Examples of ways in which existing analytical methods were improved upon include a more

rigorous mapping to the anatomical locations of EPG electrodes on the palate when presenting EPG contact patterns, and 3 dimensional viewing of EPG and EMA data. Examples of more innovative methods include the use of the tongue-circle model for the analysis of EMA data, as seen in Chapter 9, and the differential contact pattern templates for EPG data, that were used in Chapter 8, which allow comparisons of similar EPG contact patterns.

The second major aim involved exploratory analytical studies. These studies addressed questions which are theoretically interesting in themselves, but were also interesting as test cases for the techniques developed in the thesis. Through the use of these techniques analysis was performed that would have been difficult or impossible otherwise.

The first study concerned the re-analysis of data for a stop assimilation experiment of assimilation in connected speech. The four Irish lingual stops were produced as paired clusters across a word boundary, and EPG and acoustic data were recorded. The non-homorganic clusters were classified in terms of anticipatory coarticulation, primarily on the basis of the similarity of the EPG frame of initial closure to equivalent frames for homorganic clusters through the use of differential contact pattern templates. The study was a re-analysis of the data of an experiment where the original experiment had encountered methodological limitations that restricted the degree to which the types of coarticulation could be categorised. Re-analysis using differential contact pattern templates allowed categorisation to be methodically and practically applied. From the results an ordering was observed for the propensity of the different stops to assimilate, and the way in which assimilations tended to occur could be described. Interesting differences between the assimilation of primary and secondary places of articulation were also observed.

Results were also considered in terms of their compatibility with the model of assimilation proposed by Browman & Goldstein, 1989. The main conclusion was that for most examples of assimilation observed an explanation in articulatory phonology terms could be found. However a small number of examples seemed not to yield to such explanations. In one example the required gestures for the primary and secondary articulators for a particular stop do not have the same temporal phasing pattern as is seen in the general case for that stop, whereas in articulatory phonology the gestures for a phonological unit are time-locked to each other. In another example the contact pattern observed suggests that the type of lingual contact (apical vs. laminar) and the place of the articulation of the tongue tip tract variable need to undergo separate reductions in magnitude, which articulatory phonology does not appear to support. As a more general point it is not clear to the author why gestures would vary in magnitude at a phonological level in connected speech.

The second study employed EMA and EPG which were simultaneously recorded along with the acoustic signal. The theoretical question addressed concerned articulatory/acoustic relationships, specifically whether this relationship as presented for a vocal tract model in Fant, 1970 is similar to that produced by a human speaker. In Fant, 1970, charts for formant frequency vs. location of the lingual constriction are presented for simple models of the vocal tract during non-nasal vowel-like productions. These charts are called nomograms. In this thesis nomograms are produced for a human speaker's production of high vowel continua. The location of the lingual constriction was estimated from EMA data. This was done by modelling the midline of the tongue body as a circle and determining the location of that circle from two EMA sensors attached to the tongue body. This meant that the location of the approximate lingual constriction could be calculated. Nomograms for the human speaker were then compared to equivalent nomograms from Fant, 1970. Broad similarities were observed but there were also some striking differences, particularly with back cavity resonances.

As the second study included EPG and EMA data it afforded the opportunity also to evaluate the advantages and potential problems of combining the techniques. As the utterances recorded were vowels, no hardware interference between the two devices was expected nor was any observed. The location of the lingual constriction, estimated from EMA, was compared to the narrowest row on corresponding EPG frames. Broad agreement was observed over a range of articulations where the location of the lingual constriction was neither close to nor behind the most posterior row of the EPG electrodes. A further comparison was made where the EMA estimate of constriction location was compared to a EPG based tongue contour estimate which was drawn on the basis of the locations of activated and inactivated electrodes. This was useful in enhancing the understanding of differences in the nature of the data from the two devices, and drawing attention to three dimensional aspects of articulation. However it proved disappointing as a method of estimating the location of the lingual constriction. A third comparison was made between the EMA estimate of constriction location and the EPG centre of gravity index. While the centre of gravity produced similar results over a range of articulations it was observed that the index was not effective for constrictions located close to the back edge of the palate

Both studies showed that the techniques developed for the thesis provide advantages in the study of lingual activity using EPG and EMA. They provide methodological advantages in the calculation of results and enhance the presentation of both basic data and of results. They facilitated the two studies in this thesis where interesting theoretical questions were addressed. They also provided indications as to how such work aimed at comprehensively addressing the simultaneous analysis of data from the two devices might proceed.

These techniques and tools should have a wider applicability than these studies. Techniques such as the tongue circle model should be of use in similar studies, while contact pattern templates, viewing tools and so on should have a more general applicability.

In summary, the primary questions underlying this work, as introduced in §0.2, were addressed as follows. Firstly, concerning how information might be more effectively extracted from EPG data, improvements and additions were made to the state of the art of tools and techniques for EPG analysis (see Chapter 5). The usefulness of these tools and techniques, and in particular of contact pattern templates, was shown in Chapter 8 where the types of coarticulation seen for abutting lingual stops of Irish could be more effectively classified than in a previous study (where such tools had not been available). Secondly, concerning how information might be more effectively extracted from EMA data, additions to the state of the art of analytical tools were made (see Chapter 6). In particular the technique of using a model of the tongue body as a circle (for high vowel productions) as a way of extracting information on the overall movement of the tongue body from two EMA sensors was shown to be a most useful analytical technique in a study of high vowels in Chapter 9. It facilitated locating the centre of the lingual constriction for high vowels - a location which was required to allow for the comparison of the articulatory/acoustic relationship seen for the production of high vowels by a human speaker to those for a vocal tract model. In the author's opinion, reliably determining such a location from EMA data would not have been possible without such a model. And thirdly, concerning how EMA and EPG can be used together as complementary data sources, enhanced methods of simultaneously presenting the data, through tools such as three dimensional viewers, were developed (see Chapter 7). Further, the compatibility and complementary nature of the data from the two sources was investigated (see Chapter 10).

## *11.4 Further work*

There are a number of avenues for further work leading from both the development of analytical tools for EMA and EPG and from the analytical studies that were carried out. But an important core of any such work is to find ways to make better use of data from the devices.

In the study of lingual stops, analysis of simultaneous EPG and EMA was hampered by the physical interaction between the devices where EMA sensors block some EPG contact. The study would very much benefit from the additional data, as it would provide more information on articulations that occur behind the EPG palate. It would also provide more information on whether traces of gestures which, on the basis of EPG alone, one might assume to have disappeared, are actually still present. Having EMA data without EPG data for this

experiment would mean that issues dealing with, for example, tongue tip articulations would be difficult to address. But as the EMA sensors can prevent lingual contact with the EPG palate for an area immediately surrounding the sensor, tools that reliably predict which EPG electrodes are inactivated because of the blocking effect of the EMA sensors, and that allow the experimenter to easily treat those electrodes as activated, are needed to make working with simultaneous data a practical undertaking for this type of experiment.

The study of high vowels would benefit from additional measurements of the cross section area of the lingual constriction. This could potentially be estimated from the EMA-based tongue midline, from the measured shape of the roof of the mouth, from the known contact between the tongue and the roof of the mouth, and from knowledge of the tongue's formation characteristics in such contexts. It would also benefit from a better predictor of tongue midline shape than that given by the tongue circle model. This again requires applying knowledge of what the tongue's shape could be expected to be, given the points on the tongue identified by the EMA sensors. Such knowledge could perhaps be drawn from work on the analysis of steady state articulations using MRI and x-ray based techniques. But progress could still also be made with curve-fitting techniques like that at the basis of the EPG tongue contour estimate in this thesis.

There are limits to the resolution and accuracy with which information about the position of the tongue can be determined from either, or both, of the devices. Those limits have not however been reached by tools developed in this thesis. But more adventurous methods risk capturing noise or information of little relevance to the actual experiment. This means that the need to check results with the basic data to see if they are feasible continues to be important.

Of course when EPG and EMA are superseded by devices which measure lingual activity more comprehensively in connected speech then experimental phonetics will be better served. For the analysis of static articulation devices are becoming available that perform better than EPG and EMA. However such devices are still some way from being suited to the analysis continuous speech, and further still from being generally available. Making best use of what is available today justifies, in the author's opinion, carrying out the additional work proposed here.

## References


Badin, P., Perrier, P., Boé, L., Abry, C. (1988), "Vocalic nomograms: acoustic & articulatory considerations upon formant convergences"; Bulletin du laboratoire de la communication parlée Grenoble, vol. 2, pp65-94.

Badin, P., Perrier, P., Boé, L.J., Abry, C. (1990), "Vocalic nomograms: acoustic and articulatory considerations upon formant convergences", Journal of the Acoustical Society of America, vol. 87, pp1290-1300.

Baer. T., Gore, J.C., Gracco, L.C., Nye, P.W. (1991), "Analysis of vocal tract shape and dimensions using magnetic resonance imagine: Vowels", JASA, Vol. 90, pp799-828.

Bailly, G. (1995), "Characterising formant trajectories by tracking vocal tract resonances"; check ref.

Bailly, G. (1995), "Caracterisation of formant trajectories by tracking vocal tract resonances", In Sorin, C., Mariani, J., Méloni, H., and Schoentgen, J. (eds.), Levels in speech communication : relations and interactions, Amsterdam: Elsevier, pp 91-102.

Barry, M. (1984), "Connected speech: processes, motivations, models", Cambridge papers in phonetics and experimental linguistics, vol. 3, pp1-16.

Barry, M. (1992), "Palatalisation, assimilation and gestural weakening in connected speech", Speech Communication, vol. 11, pp393-400.

Bateson, E.V. & Ostry, D.J. (1995), "An Analysis of the Dimensionality of Jaw Movement in Speech", Journal of Phonetics, vol. 23, pp101-117.

Branderud, P. (1985), "Movetrack - A movement tracking system", PERILUS, Stockholm University, Vol. 4, pp20-29.

Browman, C. and Goldstein, L. (1989), "Articulatory gestures as phonological units", Phonology 6, pp. 201-251.

Browman, C.P. (1995), "Assimilation as gestural overlap: comments on Holst and Nolan", in B. Connell and A. Arvaniti (eds.), Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV, Cambridge: Cambridge University Press, pp334-342.

Browman, C.P., Goldstein, L. (1992), "Articulatory phonology: an overview", Phonetica, vol. 49, pp155-180.

Byrd, D., Flemming, E., Mueller, C.A., Tan, C.C. (1995), "Using regions and indices in EPG data reduction", Journal of speech and hearing research, vol. 38, pp821-827.

Carter, J., Shadle, C., Davis, C. (1996), "On the use of structured light in speech research", Proc. 4th Speech Production Seminar, Autrans, pp. 229-232.

Catford, J.C. (1977), "Fundamental problems in phonetics", Bloomington: Indiana University Press.

Chiu, W.S.C, Shadle, CH., Carter, J.N. (1995), "Quantitative measures of the palate using enhanced electropalatography", European Journal of Disorders of Communication, vol. 30, pp149-160.

Chomsky, N. & Halle, M. (1968), "Sound pattern of English", New York: Harper and Row.

Clumeck, H. (1976), "Patterns of soft palate movement in six languages", Journal of Phonetics, Vol. 4, pp 337-351.

Coker, C. (1967). "Synthesis by rule from articulatory parameters", in J. Flanagan and L. Rabiner (eds), 1973, Speech Synthesis, pp396-399.

Coker, C. (1976). "A model of articulatory dynamics and control", Proc. IEEE vol. 64(4): pp452-460.

Cooper, J.W.A. (1994) " An EPG study of the consonants of the Irish dialect of Cois. Fhairrge", unpublished M. Phil. thesis, Trinity College Dublin.

De Bhaldraithe, T. (1975), "The Irish of Cois Fhairrge, Co. Galway", Dublin: The Institute for Advances Studies.

Engstrand, O. (1989), "Towards an electropalatographic specification of consonant articulation in Swedish", PERILUS, Stockholm University, vol. 10, pp115-156.

Faber, A. (1989), "Studies in sibilant articulation", unpublished manuscript, Haskins Laboratories, New Haven.

Fant, G. (1970), "Acoustic Theory of Speech Production (2nd ed.)", Mouton , The Hague. (First edition: 1960.)

Fant, G. (1982), "The voice source - acoustic modelling", STL-QPSR (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm), Vol. 4/1982, pp 28-48.

Fant, G. (1991), "What can basic research contribute to speech synthesis?", Journal of Phonetics, vol. 19, pp7-90.

Farnetani, E. (1988), "Asymmetry of lingual movements: EPG data on Italian", Quaderni del Centro di Studio per le Richerche di Fonetica del C.N.R., Vol. 7, pp211-228.

Farnetani, E. (1997), "Coarticulation and Connected Speech Processes", in W.J. Hardcastle & J. Laver, eds., The Handbook of Phonetic Sciences, Oxford: Blackwell Publishers Ltd., pp 371-404.

Farnetani, E., Hardcastle, W.J., Marchal A. (1989), "Cross-language investigation of lingual coarticulatory processes using EPG", in J.P. Tubach & J.J. Mariani, eds., Eurospeech '89, Paris, pp 429-432.

Fitzpatrick, L., Ní Chasaide, A. (1999), "Human speaker nomograms using EMA data", Proceedings of the XIVth International Congress of Phonetic Sciences, San Francisco, pp2021-2024.

Fitzpatrick, L., Ní Chasaide, A. (2000), " Inferring tongue articulation from simultaneous EMA & EPG: complementary and conflicting data", Proceedings of the 5th Speech Production Seminar, Kloster Seeon, pp201-204.

Fontdevila, J., Pallarès, M.D., Recasens, D. (1994), "The contact index method of electropalatographic data reduction", Journal of Phonetics, vol. 22, pp141-154.

Fowler, C. A. (1980), "Coarticulation and theories of extrinsic timing control", Journal of Phonetics, vol. 8, pp113-133

Fujimura, O., Tatsumi, I.F., Kagaya, R. (1972), "Computational processing of palatographic patterns", Journal of Phonetics, Vol. 1, pp47-54.

Gray, H. (1918), "Anatomy of the Human Body", Philadelphia: Lea & Febiger.

Hasegawa-Johnson, M. A. (1998), "Electromagnetic Exposure Safety of the Carstens Articulograph AG100," Journal of the Acoustical Society of America, vol. 104, pp. 2529-2532.

Hardcastle, W.J. (1976), "Physiology of speech production", London: Academic Press.

Hardcastle, W., Barry, W. (1989), "Articulatory and perceptual factors in /l/ vocalisation in English", Journal of the International Phonetic Association, vol. 15, pp3-17.

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

Hardcastle, W.J., Jones, W., Knight, C., Trugeon, A., Calder, G. (1989), "New developments in electropalatography: A state-of-the-art report", Clinical Linguistics and Phonetics, vol. 3, pp1-38.

Harshman, R., Ladefoged, P., Goldstein, L. (1977), " Factor analysis of tongue shapes"; Journal of the Acoustical Society of America, vol. 62, pp693-707.

Hawkins, S. and Stevens, K.N. (1985), "Acoustics and perceptual correlates of non-nasal/nasal distinction for vowels", Journal of the Acoustic Society of America, vol. 77, pp1560-1575.

Henke, W. (1966) "Dynamic articulatory models of speech production using computer simulation", PhD dissertation, MIT Press.

Holst T., Nolan, F. (1995), "The influence of syntactic structure on [s] and [ʃ] assimilation", in B. Connell and A. Arvaniti (eds.), Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV, Canbridge: Cambridge University Press, pp315-333.

Honda, M., Kaburagi, T. (1993), "Comparison of Electromagnetic and Ultrasonic Techniques for Monitoring Tongue Motion", FIPKM, Universität München, vol. 31, pp121-136.

Hong, G., Schönle, P.W., Conrad, B. (1987), "Electromagnetic Articulography: A new approach to the investigation of palatalization in Russian"; 11th International Congress of Phonetic Science, Tallin, USSR, vol. 1, pp27-30.

Hoole, P. (1993), "Methodological considerations in the use of electromagnetic articulography in phonetic research", FIPKM, Universität München, vol. 31, pp43-64.

Hoole, P. (1996), "Theoretische und methodische Grunglagen der Artikulationsanalyse in der experimentellen Phonetik", FIPKM, Universität München, vol. 24, pp3-173.

Hoole, P., Gfroerer, S., Tillmann, H.G. (1990), "Electromagnetic articulography as a tool in the study of lingual coarticulation", FIPKM, Universität München, vol. 28, pp107-122.

Hoole, P., & Nguyen, N. (1999), "Electromagnetic articulography in coarticulation research", in W.J. Hardcastle & N. Hewlett (eds.), Coarticulation: theory, data, and techniques (pp. 260-269), Cambridge: University Press.

Hoole, P., Nguyen-Trong, N., Hardcastle, W. (1993), "A comparative investigation of coarticulation in fricatives: electropalatographic, electromagnetic and acoustic data", Language and Speech, vol. 36, pp263-288.

Jones, W., Hardcastle, W.J. (1995), "New developments in EPG3 software", European Journal of disorders of communication, vol. 30, pp183-192.

Kaburagi, T, & Honda, M. (1994). "Determination of sagittal tongue shape from the positions of points on the tongue surface", Journal of the Acoustical Society of America, vol. 96, pp1356-1366.

Keating, P.A. (1985), "Universal phonetics and the organization of grammars", In V. Fromkin (ed.), Phonetics linguistics: essays in honor of Peter Ladefoged (pp. 115-132), Orlando: Academic Press.

Kroos, C., Hoole, P., Kühnert, B., Tillmann, H.G. (1997), " Phonetic evidence for the phonological status of the tense-lax distinction in German", FIPKM, Universität München, vol. 35, pp17-25.

Kühnert, B. & Nolan, F. (1999), "The origin of coarticulation" In W.J. Hardcastle & N. Hewlett (eds.), Coarticulation: theory, data, and techniques (pp. 7-30), Cambridge: University Press.

Ladefoged P., Antony J.F.K., Riley, C. (1971) "Direct measurement of the vocal tract", UCLA Working Papers in Phonetics, Vol. 19, pp. 4-13.

Ladefoged, P. & Bladon, A. (1982), "Attempts by human speakers to reproduce Fant's nomograms", Speech Communication, vol. 1, pp185-198.

Lee, W. (1999), "An articulatory and acoustical analysis of the syllable-initial sibilants and approximant in Beijing Mandarin", ICPHS 99, pp413-416.

Lindblom, B. (1983), "Economy of speech gestures", In P.F. MacNeilage (ed.), The Production of Speech (pp. 217-245), New York: Springer Verlag.

Ljungqvist, M. & Fujisaki, H. (1985), "A comparative study of glottal waveform models", Technical Report of the Institute of Electronics and Communications Engineers, Japan, EA85-58, pp23-29.

Löfqvist, A., Gracco, L., Nye, P.W. (1993), "Recording speech movements using magnetometry: one laboratory's experience", FIPKM, Universität München, vol. 31, pp143-162.

Maeda, S. (1988), "Improved articulatory model", Journal of the Acoustical Society of America, vol. 84, Sup. 1, S146.

Maeda, S. (1990), "Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model", in W.L. Hardcastle & A. Marchal (eds.), Speech production and speech modelling, Dordrecht: Kluwer Academic Publishers, pp 131-149.

Mermelstein, P. (1973), "Articulatory model for the study of speech production", Journal of the acoustical society of speech production, vol. 53, pp1070-1082.

Monahan, P. (1996), "Systems for Voice Source Analysis", Unpublished PhD thesis, Trinity College, Dublin.

Mooshammer, C., Hoole, P., and Kühnert, B. (1995), "On loops", Journal of Phonetics, 23, pp3-22.

Nguyen, N., Hoole, P. (1993), "Frequency variations of the lowest main spectral peak in sibilant clusters: an acoustic and EMA investigation", FIPKM, Universität München, vol. 31, 239-248.

Ní Chasaide, A., Fitzpatrick, L. (1995), "Assimilation of Irish velarised & palatalised stops", Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Vol. 2, pp334-337.

Nolan, F. (1992), "The descriptive role of segments: evidence from assimilation", in G.J. O'Docherty & D.R. Ladd (eds.), Papers in laboratory phonology II: Gesture, segment and prosody, Cambridge: Cambridge University Press, pp261-280.

Nolan, F. (1999), "The devil is in the detail", Proceedings of ICPhS 1999, vol. 1, pp1-8.

Nolan, F., Holst T., Kühnert, B. (1995), "Modelling [s] to [ʃ] accommodation in English", Journal of Phonetics, vol. 24, pp113-137.

Pandeli, H. (1993), "The articulation of lingual consonants", Unpublished Ph.D. thesis. Cambridge University.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I. & Jackson, M. (1992) "Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements", Journal of the Acoustic Society of America, vol. 92, pp3078-3096.

Pompino-Marschall, B., Mooshammer, C. (1997), "Factors of variation in the production of the German dorsal fricative", Proceedings of Eurospeech 1997, pp375-378.

Recasens, D. (1990), "The articulatory characteristics of palatal consonants", Journal of Phonetics, vol. 18, pp267-280.

Rouco, A. & Recasens, D. (1996), "Reliability of electromagnetic mid-sagittal articulometry and electropalatography data acquired simultaneously", Journal of the Acoustic Society of America, vol. 100, pp3384-3390.

Schönle, P.W. (1993), "The developmental genealogy of electromagnetic articulography (EMA)", FIPKM, Universität München, Vol. 31, pp83-90.

Shadle, CH., Carter, J.N. (1993), Presentation to a meeting of the Speech Maps project.

Shibata, S. (1968), "A study of dynamic palatography", Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo, No. 2 pp28-36.

Slattery, M. (1993) "Connected speech processes: an EPG study of assimilation in Irish stops", unpublished M. Phil. thesis, Trinity College Dublin.

Solé, M.J., Ohala, J.J. (1991), "Differentiating between phonetic and phonological processes: the case of nasalization", Proceedings of the XIIth International Congress of Phonetic Sciences, Aix-en-Provence, Vol. 2, pp110-113.

Stevens, K.N. (1989), "On the quantal nature of speech", Journal of Phonetics vol. 17, pp3-45.

Stone, M. (1997), "Laboratory techniques for investigating speech articulation", in W.J. Hardcastle & J. Laver, eds., The Handbook of Phonetic Sciences, Oxford: Blackwell Publishers Ltd., pp 11-32.

Wakumoto, M., Masaki, S., Honda, K. & Ohue, T. (1998), "A pressure sensitive palatography: application of new pressure sensitive sheet for measuring tongue-palatal contact pressure", Proceedings of the 5th International Conference on Spoken Language Processing, Vol. 7, pp 3151-3154.

Zierdt, A. (1993), "Problems of Electromagnetic Position Transduction for a Three-dimensional Articulographic Measurement System", FIPKM, Universität München, vol. 31, pp137-142.