Production and perception of lateral-final rimes in Australian English

Tünde Orsolya Szalay

Bachelor of Arts (English and American Studies)

Master of Arts (English and American Studies)

Master of Research (Linguistics)

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics

Centre for Language Sciences

Faculty of Human Sciences

Macquarie University

Sydney, Australia

Submitted: 31 December 2019

Contents

Thesis abstract

English /l/ is a multi-gestural segment produced with dorsal retraction and lowering and a central alveolar closure. In coda position, the tongue dorsum gesture precedes the tongue tip gesture, placing it temporally closer to the vowel gesture. In many varieties of English, including Australian English (AusE), the tongue dorsum gesture has been shown to influence the preceding vowel whereas the tongue tip gesture has been show to be lenited. This thesis examines the production and perception of lateral-final rimes in AusE, first focusing on the effect of coda /l/ on the preceding vowel, and then focusing on coda /l/ lenition. We begin with providing a systematic acoustic analysis of the AusE vowel space in prelateral position compared to pre-obstruent position. Confirming previous research, we find a reduced F1-F2 vowel plane and spectral contrast reduction between the members of the pairs /ʉː-ʊ, əʉ-ɔ, æɔ-æ/, and to a lesser extent /iː-ɪ/. We then examine how vowel-lateral coarticulation affects the perception of lateral final rimes, finding that coda /l/ increases the difficulty of vowel disambiguation compared to coda /d/; in particular, reduced perceptual contrast is found between the members of the pairs /ʉː-ʊ, əʉ-ɔ, æɔ-æ/. We examine listener-speakers' production and perception of duration contrast in /ʉː-ʊ, əʉ-ɔ, æɔ-æ, iː-ɪ/. We find that listener-speakers producing a longer durational contrast take longer to recognise /l/-final words and are only more accurate when the stimuli contains a larger durational contrast. We attribute spectral and perceptual contrast reduction in prelateral vowels to the coarticulatory influence of the tongue dorsum gesture of /l/. Having examined prelateral vowels, we turn our attention to the articulation and perception of coda /l/ to link our perceptual and articulatory understanding of /l/-vocalisation. Firstly, we contribute to the articulatory characterisation of AusE /l/ by showing that the tongue is elongated in /l/ compared to /d/ due to the simultaneous raising of the tongue tip and the lowering of the tongue dorsum. Secondly, we examine the effect of phonetic context on tongue tip lenition. We find that while speakers produce coda /l/ with a varying magnitude of lenition, they lenite less before a following alveolar and more before a following dorsal consonant due to their articulatory similarity to the coronal and the dorsal gesture of /l/. We link listeners' perception of /l/-vocalisation to articulatory characteristics and phonetic context of coda /l/ and find that while listeners' perception of /l/-vocalisation does not directly correlate with tongue tip lenition, they perceive coda /l/ as more vocalised in contexts in which /l/ is often lenited. This thesis thus provides an in-depth analysis of AusE rimes containing coda laterals, their production, their perception, and implications for a potential sound change. More generally, we also contribute to the understanding of the complex speech sound /l/ by providing further information on the influence of the dorsal gesture on the preceding vowel and lenition of the tongue tip gesture.

Declaration and statement of authorship

I hereby declare that the research presented in this thesis is original and has been submitted exclusively to Macquarie University for the degree of Doctor of Philosophy. This work has not been submitted for a Doctor of Philosophy degree at any other institution. All of the research contained in this thesis was undertaken while I was formally enrolled as a higher degree research candidate at Macquarie University. Experiment 1 in **Chapter 3** was conducted while I was enrolled as a Masters of Research student at Macquarie University; all other experiments were conducted while I was enrolled as a Ph.D. candidate. In accordance with Macquarie University guidelines, I include **Chapter 3** as a relevant paper for which at least half of the research has been undertaken during my Ph.D. candidature, and I mark Experiment 1 in **Section 3.2** as non-examinable. I have made every effort to indicate sources of information and to acknowledge where the work of others has been used.

The research in this thesis was conducted with ethics approval from the Macquarie University Human Sciences Research Ethics Subcommittee (Reference number: 5201700256).

Some of the material contained in this thesis has been published, accepted for publication, submitted for publication, or is currently in preparation for submission for publication. These items are listed below:

Chapter 2 is being prepared for submission to a specialist phonetics journal (such as Journal of the Acoustical Society of America or Journal of Phonetics) as:

Szalay, T., Benders, T., Cox, F., Palethorpe, S. & Proctor, M. (in preparation). Spectral contrast reduction in Australian English lateral-final rimes.

Chapter 3 is under revision as:

Szalay, T., Benders, T., Cox, F., & Proctor, M. (under revision). Limits of compensation for coarticulation: lateral-final rimes in Australian English. *Journal of the Association for Laboratory Phonology*

Chapter 4 has been published as:

Szalay, T., Benders, T., Cox, F., & Proctor, M. (2018). "Production and perception of length contrast in lateral-final rimes", in J. Epps, J. Wolfe, J. Smith, and C. Jones (Eds.) *Proceedings of the 17th Australasian International Conference on Speech Science and Technology*, (p. 127–132), Sydney, Australia.

Chapter 5 has been published as:

Szalay, T., Benders, T., Cox, F., & Proctor, M. (2019). "Lingual configuration of Australian English /l/" in Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren(Eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*, (p. 2816–2819), Melbourne Australia.

Chapter 6 is being prepared for submission to a specialist phonetics journal (such as Journal of Phonetics or Phonetica) as:

Szalay, T., Benders, T., Cox, F., & Proctor, M. (in preparation). Lenition of the coronal gesture in Australian English coda laterals.

Chapter 7 is being prepared for submission to a specialist phonetics journal (such as Journal of Phonetics or Phonetica) as:

Szalay, T., Benders, T., Cox, F., & Proctor, M. (in preparation). Perception of vocalised lateral and its phonetic correlates.

I state that I took leadership in conducting this research and was primarily responsible for all parts of this thesis, including the conception, development and design of all experiments, the bulk of data collection (see below), the preparation and annotation of data, stimuli creation, statistical analyses, and the writing of the thesis. My supervisors/co-authors, Dr. Michael Proctor, Prof. Felicity Cox and Dr. Titia Benders, provided advice on the research protocol, analyses, the interpretation of results, and the written component of the thesis. I additionally received detailed feedback from journal editors and reviewers on draft versions of some chapters, which has led to their improvement.

Data for **Chapter 2** were provided by Prof. Felicity Cox, and the data in **Section 3.2**, **Chapter 3** was collected during my Masters of Research candidature.

Tünde Orsolya Szalay

31 December 2019

Acknowledgements

First and foremost, I am extremely grateful for the support and guidance of my supervisors, Michael Proctor, Felicity Cox, and Titia Benders. I thank Michael Proctor for supporting my Ph.D. from the moment of my application to the very end of my candidature. I thank Felicity Cox, for her guidance and wisdom and for sharing her knowledge on anything pertaining to Australian English phonology and phonetics. I thank Titia Benders for spending time discussing statistics, interpreting results and appreciating the beauty of phonetic data.

I am very lucky to have been part of the MQ phonetics lab, whose expertise and feedback I profited from enormously. I thank all the lab members, especially Ivan Yuen and Louise Ratko, for their feedback and thought-provoking questions. I also thank Peter Humburg for his help in the statistical analyses. I thank the online R user community; they are people I have never met in person, but they have answered a lot of my questions. I also thank my participants without whom this dissertation would not have been possible.

I am thankful for my parents, Anikó and Tamás, and my partner Tibu, for their encouragement to finish my dissertation while maintaining a healthy-work life balance and attending as many live music events as possible.

1 Introduction

Contrast is a fundamental concept in every phonological theory. Members of the minimal pair *seed* and *seal* contrast in their coda consonants, making /d/ and /l/ phonemes of English (Trubetzkoy 1969). However, in several dialects of English, including Australian and General American English, the nuclear vowel /iː/ is also realised differently in these words: unlike the /iː/ in *seed*, the /iː/ in *seal* is realised with a schwa-like off-glide (Gick & Wilson 2006, Palethorpe & Cox 2003). This schwa-like offglide is due to the coarticulatory influence of /l/, caused by the opposing articulatory demands placed on the tongue by the high front place of articulation of /iː/ and the backed and lowered place of articulation of English /l/ (Gick & Wilson 2006).

Not all phonological theories define the smallest contrastive unit of language as the phoneme. Articulatory phonology (AP) proposes articulatory gestures as the smallest, contrastive, abstract, context-independent unit of language (Browman & Goldstein 1986). In the word *seal*, the nuclear vowel is specified for a high and fronted tongue body gesture, while the lateral /l/ is specified with three gestures: a tongue tip gesture, a tongue dorsum backing and lowering gesture, and a tongue lateralisation gesture (Browman & Goldstein 1995, Ladefoged & Maddieson 1996). Due to the antagonistic tongue dorsum specification of the vowel and the lateral, the tongue has to go through the canonical schwa-target to achieve an /l/ target after an /iː/ target (Gick & Wilson 2006).

To further complicate the phonological contrast between *seed* and *seal*, the tongue tip gesture of /l/ in *seal* may be lenited, as a result of which it may not achieve contact with the alveolar ridge (Giles & Moll 1975, Scobbie & Wrench 2003). The loss of the tongue tip contact, which can be accompanied by the delay in the tongue tip gesture and reduction of the lateral gesture, is one of the key defining feature of /l/-vocalisation (Strycharczuk et al. 2018, Strycharczuk & Scobbie 2019). /l/-vocalisation may be conditioned by various factors, including speech rate (American English), morphological complexity (American English, British, and Scottish English), lexical frequency (American English), and phonemic context (Australian English, British English) (Giles & Moll 1975, Scobbie & Wrench 2003, Lee-Kim et al. 2013, Lin et al. 2014). The effects of phonemic context can be understood as coarticulatory effects, because they have been explained in terms of articulatory similarity between /l/ and the adjacent segments: /l/-vocalisation is facilitated when the following consonant is articulatorily dissimilar to the tongue tip gesture of /l/ (e.g. in *bulk*) or when the preceding vowel is articulatorily dissimilar to the tongue dorsum gesture of /l/ (e.g. in *milk*) (Hardcastle & Barry 1989). As factors conditioning /l/-vocalisation have been examined in different dialects of English, little is known about how the effect of the aforementioned factors differs between dialects of English.

While the articulation of /l/ and its positional allophones is well understood, less is known about coarticulation between elements within /l/-final rimes. The aim of this Ph.D. dissertation is to address what coarticulatory relationships in /l/-final rimes tell us about phonological contrast, and its realisation and perception using articulatory, acoustic, and perceptual evidence.

1.1 Production of /l/-final rimes

1.1.1 Articulation and distribution of the allophones of /l/

Laterals are sounds produced with a mid-sagittal occlusion such that the airflow is not blocked along one or both sides of the tongue (Ladefoged & Maddieson 1996). The English lateral is produced with a central alveolar closure, lateral channel(s), and tongue dorsum retraction (Ladefoged & Maddieson 1996, Narayanan et al. 1997). Depending on the speaker, the alveolar closure can be laminal, i.e. produced with the tongue blade, or apical, i.e. produced with the tongue tip (Narayanan et al. 1997). The lateral channel allowing the airflow to escape at the side(s) of the tongue is created by a grooving behind the alveolar closure and inward lateral compression, i.e. narrowing of the tongue (Ladefoged & Maddieson 1996, Sproat & Fujimura 1993, Stone & Lundberg 1996, Narayanan et al. 1997). The channel's size increases behind the alveolar closure and decreases towards the back as tongue dorsum lowering and retraction creates an additional /ɔ/-like displacement in the velar and pharyngeal region (Stone & Lundberg 1996, Narayanan et al. 1997, Gick et al. 2002). The coordination of antagonistic coronal and dorsal gestures prototypically results in lingual elongation (Ladefoged & Maddieson 1996). Tongue grooving, lateral channel formation, and the tongue dorsum gesture show interspeaker and positional variation in their location and magnitude, to the extent that tongue grooving is not observed for all speakers (Giles & Moll 1975, Stone & Lundberg 1996, Narayanan et al. 1997).

Initial/onset and final/coda positions are associated with the clear and dark variants of /l/ (Giles & Moll 1975, Browman & Goldstein 1995). Variation between clear [l] and dark [ɫ] is caused by the different realisations and coordinations of the dorsal and the alveolar gestures (Giles & Moll 1975, Sproat & Fujimura 1993, Narayanan et al. 1997, Proctor et al. 2019). Dark, velarised [ɫ] differs from clear [l] in its less raised and fronted tongue tip gesture, in its more lowered and retracted tongue dorsum gesture and in their different gestural coordination, as the tongue tip gesture follows the tongue dorsum gesture in coda /l/, but precedes it in onset /l/ (Giles & Moll 1975, Sproat & Fujimura 1993, Browman & Goldstein 1995, Gick 2003, Proctor et al. 2019, Ying et al. 2017). Gick (2003) and Proctor et al. (2019) found that the lag between the tongue tip and tongue dorsum gestures is larger in the coda compared to the onset; however, Gick et al. (2006) found a larger lag in onset than in coda position. Results regarding the length of the lag also vary. Gick (2003) found that the tongue tip precedes the tongue dorsum gesture by 20 ms in onset and follows it by 30 ms in coda position (3 speakers of General American English, EMA). Proctor et al. (2019) found that the tongue tip gesture precedes the tongue dorsum gesture by 87 ms in onset [l] and follows the tongue dorsum gesture by 148 ms in coda [ɫ] (four speakers of General American English, real-time MRI). Gick et al. (2006) found that the tongue tip gesture precedes the tongue dorsum gesture in onset by 20 ms and follows it in coda by 20 ms (two speakers of Western Canadian English, ultrasound). Due to low number of participants in articulatory studies, it is impossible to know whether these discrepancies are due to interspeaker or dialectal variation, or to methodological differences.

Different timing of the coordination between the tongue tip and the tongue dorsum gesture in onset compared to coda position results in the tongue dorsum gesture being temporally closer to the vowel gesture than the tongue tip gesture in both positions (Sproat & Fujimura 1993). Because of the temporal proximity and the gestural similarity between the tongue dorsum gesture and back vowels, the tongue dorsum gesture is considered to be vocalic and the tongue tip gesture is considered to be consonantal (Giles & Moll 1975, Sproat & Fujimura 1993). In addition, clearer /l/ sounds were found to be more lateralised than darker /l/ sounds in New Zealand English (Strycharczuk et al. 2018), but not in Australian English (AusE) (Ying et al. 2017).). Again, it is impossible to know whether this discrepancy in dark [ɫ] lateralisation is due to dialectal or interspeaker variation.

Clear /l/ is associated with the utterance-initial and onset position while dark /l/ is associated with the utterance-final and the coda position (Giles & Moll 1975). Intervocalic and word final /l/ (e.g *kneeling, peel apples, peel bananas*) exhibit gradient variation on the clear-dark continuum: in these contexts, the height of the tongue tip gesture varies between being onset- and coda like, while the tongue dorsum and the coordination between the two gestures remain consistently coda-like (e.g. Gick 2003, Sproat & Fujimura 1993, Scobbie & Wrench 2003, Scobbie & Pouplier 2010). Gradient darkening is shown by more lowered and retracted tongue dorsum gesture, increasing duration of the tongue tip delay compared to the tongue dorsum, and increasing duration of the rime (Lee-Kim et al. 2013, Sproat & Fujimura 1993, Turton 2017)

Gradient /l/-darkening in intervocalic /l/ is conditioned by prosodic and morphological boundaries in American and British English (Sproat & Fujimura 1993, Lee-Kim et al. 2013, Turton 2017). In American English, darkening of word-final prevocalic /l/ correlates with the strength of the prosodic boundary following /l/: the stronger the boundary, the bigger the delay between maximal tongue dorsum displacement and maximum tongue tip displacement and the darker the /l/ (Sproat & Fujimura 1993). For example, intervocalic /l/ is darker before a major intonation boundary (e.g. *Neal, equate the actors!*) than before a verb-phrase boundary (e.g. *Neal equates the actors*), but clear before a word internal morpheme boundary (e.g *kneel-ing*) (Sproat & Fujimura 1993). Morpheme boundaries affect /l/ darkening similarly to prosodic boundaries: the stronger the boundary, the more lowered the tongue dorsum and thus darker the /l/ (Lee-Kim et al. 2013). That is, /l/ is the darkest word-finally (e.g. *tall hemlock*), less dark morpheme-finally before a vowel (e.g. *tall-est*) and the least dark morpheme-initially before a vowel (e.g. *flaw-less*). Manchester English utterance final /l/ exhibits a small but consistent increase in tongue dorsum backing compared to initial /l/, while morpheme- and word-final prevocalic /l/ falls between the two endpoints (Turton 2017). For instance, /l/-darkening increases from word initial (e.g. *leap*) to morpheme-final (e.g. *peeling*) to word-final prevocalic (e.g. *heal# V, heal# , V*) to absolute final (e.g. *peel*) in which /l/ is utterance final (Turton 2017).

Sproat & Fujimura (1993) argue that prosodic boundary effects can be reduced to acoustic duration: the stronger the boundary, the longer the acoustic duration and the darker the /l/, making acoustic duration of /l/ a good predictor of gradient /l/ darkening. However, Turton (2017) points out that the correlation between acoustic duration and /l/ darkness only holds for dark [ɫ], in which tongue tip delay increases with increased duration, whereas there is no correlation in clear /l/ in which the tongue tip delay does not increase with duration. In addition, /l/ seems to have the same acoustic duration before productive (e.g. *kneel-ing*) and non-productive morpheme boundaries (e.g. *tel-ic*) but appears to be systematically darker before a productive than a non-productive boundary (Lee-Kim et al. 2013).

1.1.2 Vowel-/l/ coarticulation

The temporal proximity of the dorsal gesture to the vowel gesture leads to an overlap between the two (Sproat & Fujimura 1993, Proctor et al. 2019). In American English, vowels have a stronger influence on the tongue dorsum gesture of onset /l/ than on that of coda /l/: the tongue dorsum gesture in coda /l/ is characterised by a stable position across all vowel contexts, while it is more likely to be displaced in onset position (Proctor & Walker 2012).

Coda laterals exhibit the opposite pattern from onsets, as coda laterals have a stronger coarticulatory influence on the prelateral vowel (e.g. in *peel*) than onset laterals have on the postlateral vowel (e.g. in *leap*), shown by the fact that prelateral vowels show bigger displacement from their prelabial counterpart than postlateral vowels (Proctor et al. 2019). As a result of vowel-lateral coarticulation, the vowel gesture is more retracted in *who'll* and *coal* than in *who'd* and *code* in AusE (Lin et al. 2011).

Coarticulation is reflected in acoustics, for instance the contrast between prelateral high vowels (e.g. *feel-fill*) is reduced in some Southern dialects of American English and in Standard Southern British English through the phonetic lowering of /iː/ (Altendorf & Watt 2008, Labov et al. 2008). The *pool-pull* contrast is reduced in Pennsylvanian and Southern British English due to the phonetic lowering of the vowel in *pool* (Altendorf & Watt 2008, Labov et al. 2008). The same contrast is also reduced in South Australian English, through a different mechanism: phonetic backing and lowering of the tense vowel /ʉː/ in the pre-lateral environment (Butcher 2006, Oasa 1989). The acoustic *pool-pull-pole* contrast is reduced in Ohio, as the vowels in *pool* and *pole* shift towards *pull* in the vowel space (Arnold 2015, Wade 2017). A perceptual merger between the mid and low front vowels /e/ and /æ/ has been observed in the pre-lateral environment (*hell-Hal*) in New Zealand English (Thomas & Hay 2005) and in Melbourne English (e.g. Loakes et al. 2012; 2014). Collectively, these findings suggest that different, dialect-specific mechanisms may be involved in vowel-lateral interactions in different varieties of English.

1.1.3 AusE vowels before coda /l/

AusE coda /l/ is typically realised as a dark [ɫ], therefore backing of prelateral vowels can be expected (Cox & Palethorpe 2007, Ying et al. 2017). Vowels show a gestural overlap with coda laterals, but most vowels with the exception of /ʉː, əʉ, oː/ and /æɪ/, resist vowel-/l/ coarticulation (Lin et al. 2011).

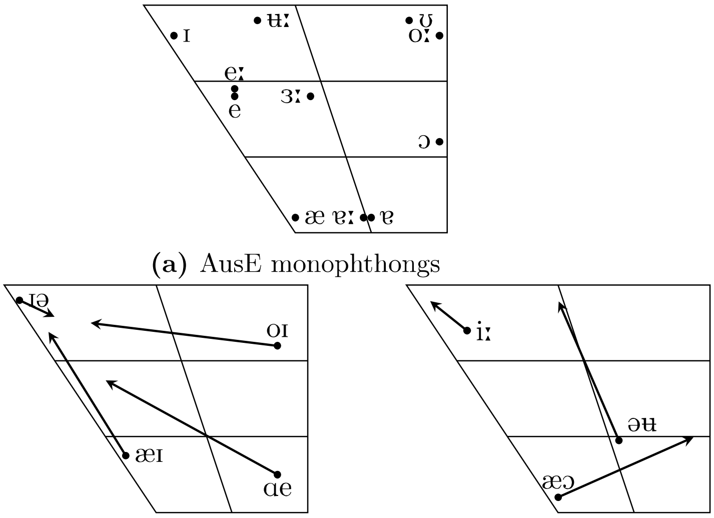
AusE vowel inventory

In AusE, coda /l/ has been shown to influence vowels in ways that can potentially reduce perceptual and acoustic vowel contrast, especially between the vowel pairs /ʉː-ʊ, æɔ-æ, əʉ-ɔ/ (Bradley 2004, Bernard 1985, Palethorpe & Cox 2003). AusE has a large vowel inventory consisting of 18 stressed vowels and schwa (**Figure 3.1**, **Table 1.1**), utilising both spectral and durational contrast (Cox & Fletcher 2017). Duration is contrastive for spectrally similar vowels, for instance, the vowel pairs /ɐː-ɐ, eː-e/ (e.g. *card-cud, shared-shed*) contrast mostly in duration, whereas the pair /ʉː-ʊ/ contrasts both in duration and in spectral quality (Cox & Fletcher 2017).

The AusE vowel system includes both diphthongs (/æɪ, ɑe, oɪ, æɔ, əʉ, ɪə/) and monophthongs (/iː, ɪ, e, eː, æ, ʉː, ɜː, ɐː, ɐ, ɔː, ɔ, ʊ/)[1] (Harrington et al. 1997, Watson & Harrington 1999). The vowels /iː/ and /ʉː/ are classified as monophthongs but are characterised by an onglide (Bernard 1970, Harrington et al. 1997, Cox & Palethorpe 2007, Cox et al. 2014). /ɪə/ is classified as a diphthong, but can be realised either as a monophthong /ɪː/ or as disyllabic /ɪːə/ (Harrington et al. 1997, Cox & Palethorpe 2007). Some of the diphthongs, having similar first target characteristics to a monophthong,

[1]The symbols used are those recommended by Cox & Palethorpe (2007) for Australian English

form pairs with monophthongs (Cox 1999). For example, /æɔ/ and /æ/ (e.g. *loud-lad*) share the first target of the diphthong, whereas /əʉ/ shares the location of the second target with the nucleus of /ʉː/ (e.g. *code-cooed*) (Cox 1999). These vowels are considered pairs as the members have moved in parallel in sound change: /æɔ/ lowering was accompanied by /æ/ lowering, and the fronting of /ʉː/ took place in parallel with the fronting of the second element of /əʉ/ (Cox 1999).



(a) AusE monophthongs

(b) AusE diphthongs 1 **(c)** AusE diphthongs 2

Figure 1.1: The AusE vowel inventory. Figures reproduced from Cox & Fletcher (2017).

Table 1.1: The AusE vowel inventory exemplified by Standard Lexical Sets

Keyword	IPA symbol recommended by Cox & Palethorpe (2007)
FLEECE	iː
KIT	ɪ
DRESS	e
SQUARE	eː
TRAP	æ
GOOSE	ʉː
NURSE	ɜː
BATH	ɐː
STRUT	ɐ
FOOT	ʊ
THOUGHT, NORTH	oː
LOT	ɔ
KIT	ɪ
NEAR	ɪə
FACE	æɪ
PRICE	ɑe
CHOICE	oɪ
MOUTH	æɔ
GOAT	əʉ

Effect of coda /l/ on monophthongs

Pre-lateral monophthongs differ from their non pre-lateral counterpart in many ways: front and central vowels are phonetically lowered and backed and some low and back vowels are phonetically raised (Bernard 1985, Cox & Palethorpe 2004, Palethorpe & Cox 2003). Front and central vowels exhibit phonetic lowering shown by increased $F1$ in /iː, ɪ, e,/ and /ʉː, ɜː/ (Bernard 1985, Cox & Palethorpe 2004, Palethorpe & Cox 2003). Front /iː, ɪ, e, æ/ and central /ʉː, ɜː/ are also characterised by lowered $F2$ representing phonetic retraction before /l/ (Bernard 1985, Cox & Palethorpe 2004, Palethorpe & Cox 2003). Among the low and back vowels only /ɐ/ and /oː/ are influenced by coda /l/, as the former exhibits phonetic raising and the latter phonetic backing (Bernard 1985, Palethorpe & Cox 2003, Cox & Palethorpe 2004).

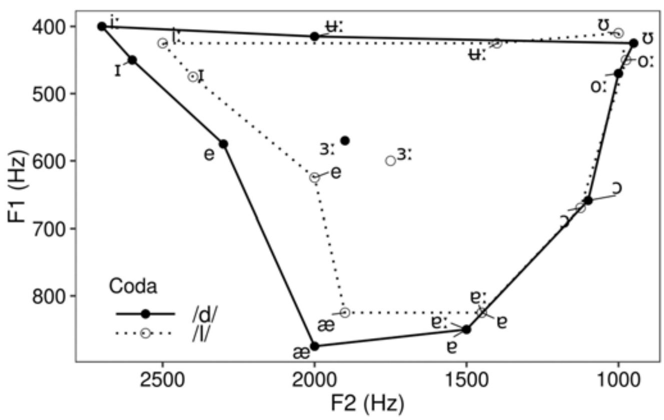


Figure 1.2: Monophthong targets before coda /d/ and coda /l/ in the F1-F2 vowel plane. Monophthongs before coda /l/ show reduced vowel contrast and smaller vowel dispersion compared to monophthongs before coda /d/. Figure reproduced from Palethorpe & Cox (2003).

Acoustic and auditory contrast reduction of some vowel pairs

The members of the long-short vowel pair /ʉːl-ʊl/ and the monophthong-diphthong pairs /əʉl-ɔl, æɔl-æl/ have been shown to undergo spectral contrast reduction (Bernard 1985, Bradley 2004, Palethorpe & Cox 2003). Spectral contrast between /ʉːl-ʊl/ and /əʉl-ɔl/ is reduced due to the $F2$ lowering in /ʉː/ and in the second target of the diphthong /əʉ/ (Palethorpe & Cox 2003). Bernard (1985) observed that the second target of the diphthongs /æɔ, əʉ/ is frequently lost before /l/ and commented on the lack observable change between the end of the vowels /æɔ, əʉ/ and /l/, which can potentially contribute to spectral contrast reduction between the members of the pairs /əʉl-ɔl/ and /æɔl-æl/. However, duration contrast between the members of these pairs is maintained (Palethorpe & Cox 2003).

Regional differences

AusE exhibits regional differences between some states in lateral-final rimes (e.g. Cox & Palethorpe 2004, Butcher 2006). The Victorian dialect of AusE shows contrast reduction between /el-æl/ and /ɐlC-ɔlC/ in production and in perception (Bernard 1985, Cox & Palethorpe 2004, Loakes et al. 2010a, Lewis & Loakes 2012). The $F1$ of /e/ is increased towards /æ/ (e.g. *hell, Hal* (Cox & Palethorpe 2004). The acoustic contrast reduction is reflected in a perceptual near-merger between /e/ and /æ/, as Victorian English listeners misperceive /e/ as /æ/ in the pre-lateral, but not in a pre-obstruent position when distinguishing minimal pairs (e.g. *telly-tally, pellet-palate* are confused but *head-had* are not) (Loakes et al. 2010a;b;c; 2011). Acoustic and perceptual contrast reduction has also been shown between /ɐlC-ɔlC/ *gulf-golf*) (Bernard 1985, Lewis & Loakes 2012).

South Australia shows a more prominent acoustic backing of /ʉː/ than New South Wales, despite /ʉː/ being backed in New South Wales as well (Butcher 2006, Oasa 1989). There is also some evidence for a potential merger between pre-lateral /iː/ and /ɪ/ in Adelaide and Hobart (Bradley 2004) but not in New South Wales (Palethorpe & Cox 2003).

Collectively, the findings presented in **Section 1.1.3** suggest that coda /l/ impacts the preceding vowel in various ways depending on vowel quality and speaker dialect and might lead to potential mergers. However, the impact of coda /l/ on preceding vowels and the potential for a loss of contrast has not been systematically examined as several of the observations on the effect of coda /l/ and on apparent contrast reduction were made on the basis of impressionistic observations or visual representations of formants (Bernard 1985, Palethorpe & Cox 2003).

1.1.4 /l/-vocalisation and its conditioning environments

While the magnitude and the coarticulatory impact of the tongue dorsum gesture increases in coda position compared to onset, the tongue tip gesture is lenited, to the point where dark /l/ can undergo /l/-vocalisation (Hardcastle & Barry 1989, Scobbie & Wrench 2003). /l/-vocalisation is commonly understood as a vowel-like realisation of coda /l/, as in the variant [mɪʊk] for /mɪlk/ (e.g. Giles & Moll 1975, Lin et al. 2014). /l/-lenition and /l/-vocalisation arises from the spatial reduction of the tongue-tip gesture (i.e. the reduction of the anterior closure), from the delay of the tongue tip gesture, and from the lenition of the tongue lateralisation (i.e. a flatter tongue shape) (e.g. Giles & Moll 1975, Browman & Goldstein 1995, Scobbie & Pouplier 2010, Strycharczuk et al. 2018, Strycharczuk & Scobbie 2019). The loss of the tongue tip gesture correlates with the loss of lateralisation (Strycharczuk et al. 2018) and the extended delay of the tongue tip gesture occurs with spatial reduction (Strycharczuk & Scobbie 2019).

The darker end of the continuum, consisting of word-final pre-pausal or pre-consonantal, and to a lesser extent word-final prevocalic /l/ can undergo /l/-vocalisation (Hardcastle & Barry 1989, Scobbie & Wrench 2003). Only coda /l/ can be vocalised, but /l/-vocalisation is optional and subject to prosodic, segmental, lexical, and stylistic factors, leading to a great amount of intraspeaker and interspeaker variation (Scobbie & Wrench 2003, Scobbie et al. 2007). The conditioning factors of /l/-vocalisation are similar to that of gradient /l/-darkening, for instance, word-final /l/ in connected speech is not only more likely to be darker in preconsonantal (e.g. *peel beavers*) than in pre-vocalic (e.g. *peel Eva*) context, but it is also more likely to be vocalised (Scobbie & Wrench 2003, Scobbie et al. 2007). In American English, /l/-vocalisation is also more likely in frequent words, compared to less frequent words (Lin et al. 2014).

The likelihood of /l/-vocalisation is also subject to segmental factors: articulatory similarity between the tongue tip gesture of /l/ and the following consonant inhibits /l/-vocalisation (Hardcastle & Barry 1989, Scobbie & Wrench 2003, Scobbie & Pouplier 2010). In coda clusters, /l/-vocalisation is more likely before a following velar (e.g. *milk*) than before an alveolar consonant (e.g. *tilt*), because the tongue tip contact of /l/ is more likely to be achieved when it is followed by a homorganic consonant (Hardcastle & Barry 1989). Vocalisation of word-final /l/ in a simple coda is more likely in pre-labial than in pre-pausal context, and similarly, it is more likely before labial /b/ than glottal /h/, despite the fact that /l/ occupies the coda position in all of these contexts (Scobbie & Wrench 2003, Scobbie & Pouplier 2010).

Articulatory similarity between the tongue dorsum gesture of /l/ and the preceding back vowel inhibits /l/-vocalisation (Hardcastle & Barry 1989).

For example, /l/ is less likely to be vocalised after back vowels (e.g. *halt* than after front vowels (e.g. *hilt*). The effect of vowel context is attributed to perceptual saliency rather than ease of articulation: the tongue dorsum gesture is similar to that of the preceding back vowel, therefore achieving tongue tip contact is necessary to create the percept of /l/, whereas lowering the tongue dorsum after a front vowel is enough to create the percept of /l/ without a tongue tip contrast (Hardcastle & Barry 1989).

1.1.5 /l/-vocalisation in AusE

AusE is known for having /l/-vocalisation (Wells 1982, Borowsky & Horvath 1997, Borowsky 2001, Horvath & Horvath 1997; 2001; 2002, Cox & Palethorpe 2007). /l/-vocalisation in AusE has been examined in a series of studies using impressionistic analysis of audio recordings collected in Adelaide, Brisbane, Melbourne, Sydney, Hobart, and Mount Gambier (Borowsky & Horvath 1997, Borowsky 2001, Horvath & Horvath 1997; 2001; 2002). The same dataset was used to examine dialectal and sociophonetic (Horvath & Horvath 1997; 2001; 2002), and prosodic and segmental factors (Borowsky & Horvath 1997, Borowsky 2001) that condition /l/-vocalisation.

In terms of sociophonetic factors, young (below 30) speakers were found to vocalise more than older speakers, female speakers more than male speakers, and working class speakers more than middle class speakers (Horvath & Horvath 1997; 2001; 2002). The fact that /l/-vocalisation was more common in the South Australian data collected in Mount Gambier and Adelaide indicates region-specific dialectal differences in AusE (Horvath & Horvath 1997; 2001; 2002).

In terms of prosodic factors, similarly to other varieties of English, the most important factor is the syllabic affiliation of /l/: /l/ can be vocalised in the nucleus, when it is syllabic, and in the coda, but not in onset position (cf. *metal, steel* but not *light*) (Borowsky & Horvath 1997, Borowsky 2001). In the coda, /l/ is most likely to be vocalised in a coda cluster (e.g. *milk, bolt*), closely followed by contexts containing word-final coda /l/ (e.g. *feel, full*) (Borowsky 2001). This contrasts with Horvath & Horvath's (2001) result, who found /l/-vocalisation to be more frequent in word-final coda position than in coda clusters. However, the data in both studies come from connected speech, as well as from individual words, and it is unclear whether the "word-final" category distinguishes between prevocalic and pre-consonantal environments (e.g. *feel angry* vs. *feel good*). When word-final /l/ was analysed separately, the vocalisation of word-final coda /l/ depended on whether it can be resyllabified in connected speech: when the following word begins with a vowel, resyllabification was possible (Borowsky 2001, Horvath & Horvath 2001). Therefore, similarly to other varieties of English, AusE /l/ is less likely to be vocalised in the l#V environment than in the prepausal or the pre-consonantal environment, and less likely before a pause than before a consonant (Borowsky 2001). The effect of syllabic affiliation on /l/-vocalisation is somewhat similar to its effect on /l/-darkening, as final /l/ is less dark and less likely to be vocalised in prevocalic than in preconsonantal position, (Sproat & Fujimura 1993).

Similarly to the articulatory study of Hardcastle & Barry (1989) on other dialects of English, in AusE too, articulatory similarity between the following consonant and the tongue tip gesture of /l/ inhibits /l/-vocalisation in coda clusters (Borowsky 2001). That is, impressionistic analysis of the effect of consonantal context on /l/-vocalisation supports articulatory results from other dialects of English, as /l/-vocalisation is the least likely before an alveolar, and more likely before a velar than a bilabial consonant (Hardcastle & Barry 1989, Borowsky 2001). However, in contrast with Hardcastle & Barry's (1989) result, articulatory similarity between the tongue dorsum gesture of /l/ and the preceding vowel facilitate /l/-vocalisation in AusE (Borowsky 2001). Impressionistic analysis provided by Borowsky (2001) contradicts articulatory analysis on the effect of vowel context, as articulatory methods showed that /l/-vocalisation is less likely after back than after front vowels (Hardcastle & Barry 1989). The impressionistic encoding used in Borowsky's (2001) study can explain the contradictory results. If indeed, the reason why front vowels facilitate /l/-vocalisation is that the tongue dorsum gesture without the tongue-tip gesture is sufficient to create the percept of /l/ after a front vowel (Hardcastle & Barry 1989), then the percentage of vocalised /l/ in the post-front vowel context may be underestimated in an impressionist analysis. The difference can also be attributed to speakers' dialects: participants were speakers of AusE in Borowsky's (2001) studies, while Hardcastle & Barry's (1989) study had four British English and one Australian English-speaking participant.

Vowel length effects on /l/-vocalisation are evident in the higher likelihood of /l/-vocalisation after a long monophthong or a diphthong than after a short monophthong. This might be explained by the differing syllable structure of /l/ final rimes differs with long and short vowels: /l/ tends to be syllabic after long vowels in AusE (Borowsky 2001).

These studies by Borowsky & Horvath (1997), Borowsky (2001) and Horvath & Horvath (1997; 2001; 2002) are invaluable in identifying numerous key factors in /l/-vocalisation. However, they are limited in scope, as the data were coded for /l/-vocalisation auditorily by a single researcher. /l/-vocalisation has been defined in articulatory rather than auditory terms (Hall-Lew & Fix 2012, Strycharczuk &

Scobbie 2019), but /l/-vocalisation in AusE has only been examined with auditory-impressionistic methods, which provides indirect measurements on /l/-vocalisation. Auditory-impressionistic coding method does not provide any information on the tongue tip contact on the alveolar ridge and might be prone to inconsistencies (Hall-Lew & Fix 2012).

1.2 Phonological contrast and coarticulation

The complex coarticulatory relationships between the elements within /l/-final rimes described in **Section 1.1** seem to affect vowel contrast in the prelateral environment and the realisation of coda laterals between vocalised /l/ and non-vocalised /l/. Both vowel-contrast reduction and /l/-vocalisation involve units smaller than a segment: prelateral vowel-contrast reduction is attributed to the influence of the tongue dorsum gesture of /l/ on the vowel, while /l/-vocalisation involves the lenition of the tongue tip gesture of /l/. Therefore these coarticulatory relationships can be better explained in the framework of Articulatory Phonology (AP) and in Task Dynamics, which use gestures as the smallest units of phonological contrast, than in a segmental framework (Browman & Goldstein 1986, Saltzman 1986). Gestures are defined as "a member of a family of functionally equivalent articulatory movement patterns that are actively controlled with reference to a given speech-relevant goal" (Saltzman & Munhall 1989) and they can be identified by observing the characteristic pattern of constriction formation and release of articulators (Browman & Goldstein 1989).

Gestures are context-independent, but as they are combined into larger units, such as syllables and words, the tongue has to travel with a continuous motion from one gestural target to the next (Lindblom 1963, Browman & Goldstein 1986, Iskarous et al. 2013). Therefore realisations of a gesture vary with the articulatory characteristics of the adjacent gestures, leading to gestural overlap or to gestural undershoot (Browman & Goldstein 1986). Variation in the realisation of gestures can be attributed either to the mechano-inertial constraints on the movement and the timing of the articulators (Lindblom 1963; 1983, Farnetani & Recasens 2012) or to the context-specific planning and co-production of articulatory gestures (Fowler 1980, Fowler & Saltzman 1993).

Temporal overlap of gestures, when the articulators realise more than one gesture at the same time, leads to coarticulation (Browman et al. 1990, Fowler & Saltzman 1993). When gestures activate the same articulator but place contradictory demands on it, the result of coproduction can be the blending of the observed output characteristics of two gestures. For instance, the constriction achieved by the tongue tip for /n/ in *ten things* will become more fronted due to the following labiodental fricative than the constriction of /n/ in *net* (Browman et al. 1990). Similarly, the articulatory backing of /ʉː/ before /l/ can be the result of gestural blending (Lin et al. 2011). Some segments, such as /k/ or /l/ are more likely to resist the coarticulatory influence of their neighbours and more likely to influence their neighbours than other segments (Bladon & Al-Bamerni 1976, Recasens 2002).

When two overlapping gestures are realised with two independent articulators, such as alveolar /t/ and bilabial /b/, they may overlap in time and can be realised independently from each other without perturbing each others' gestural trajectories (Browman et al. 1990). However, the gestures can overlap to such extent in the phrase *must be* that the acoustic correlates of /t/ are absent although the /t/ closure is fully realised (Browman et al. 1990). This kind of acoustic deletion is not to be confused with weakening of /t/ in casual speech, during which the amplitude and the magnitude of the tongue tip gesture of /t/ is reduced.

1.2.1 Perceptual effects of coarticulation

Gestural overlap can change or even obscure the acoustic cues for one segment without perturbing the articulatory trajectories of the gestures involved (Stevens & Keyser 2010). For example, gestural overlap between velum lowering and the vowel gesture causes acoustic vowel nasalisation, and overlap between the alveolar closure of /t/ and labial closure of /p/ in *top tag* leads to masking of the acoustic cues of /p/; however, the gestural trajectories are not perturbed (Krakow 1989, Browman et al. 1990, Stevens & Keyser 2010). Coarticulation affects the acoustic signal to varying degrees: some changes in articulation have smaller effects on the acoustics than others, due to the quantal relationship between articulation and acoustics (Stevens 1972, Stevens & Keyser 2010). Coarticulation can enhance acoustic contrast by providing additional and prolonged cues to the segment triggering coarticulation and it can also weaken or obscure acoustic cues to the segment undergoing coarticulation (Stevens & Keyser 2010). While acoustic contrast enhancement aids listeners' perception, weakened acoustic cues do not necessarily hinder listeners' perception (Mann 1980, Fowler 1984, Beddor et al. 2013, Zellou 2017).

Coarticulation enhances perceptual contrast

The fact that coarticulatory cues can aid the recognition of the influencing segment shows that coarticulation can enhance phonemic contrast (Fowler 1984, Beddor et al. 2013). Acoustic cues to a phone's identity are spread across several adjacent phones, therefore listeners can use information carried by an adjacent phone to correctly identify the source of the coarticulation (Fowler 1984, Beddor et al. 2013). For instance, in English, anticipatory vowel nasalisation aids the recognition of the following nasal consonant (Lahiri & Marslen-Wilson 1991, Beddor et al. 2013). When hearing a CV or a C$\tilde{V}$ sequence in a gating experiment, English listeners can correctly predict whether the following consonant is oral or nasal, based on the coarticulatory information carried by the vowel without hearing the consonant (Lahiri & Marslen-Wilson 1991, Ohala & Ohala 1995, Beddor et al. 2013). Listeners are faster to identify both oral and nasal consonants when the preceding vowel carries the appropriate coarticulatory cues, i.e. it is non-nasal before an oral and nasalised before a nasal consonant (Fowler & Brown 2000). This effect depends on the onset of nasalisation on the vowel, for example, listeners in an eye-tracking experiment looked at the CVN target earlier when the vowel nasalisation began early in the vowel compared to when it began late (Beddor et al. 2013).

Similarly to anticipatory vowel nasalisation, anticipatory vowel-lateral coarticulation allows British English listeners to reliably identify /l/. Listeners can identify /l/ when /l/ and the following /i/ are replaced by white noise in *belly* using coarticulatory cues on /e/ (West 1999). However, listeners cannot use regressive coarticulatory cues on /i/ to identify /l/, when /l/ and the preceding vowels are replaced by white noise (West 1999). The sound /ɹ/ can be identified on the basis of both anticipatory and regressive coarticulation: listeners can identify /ɹ/ both when the preceding as well as when the following vowel are replaced by white noise (West 1999).

The release burst of a consonant in a CV segment differs according to the place of articulation of the following vowel, and listeners can use this information to identify the V in a CV segment as /i/ or /u/ based on the consonant burst only, using coarticulatory information (Fowler 1984).

Coarticulatory cues can aid listeners across word boundaries too (Gow & McMurray 2007, Salverda et al. 2014). Eye-tracking captured that upon hearing the sentence *The... ladder is the target* listeners look at the word *ladder* sooner when the definite article contained coarticulatory cues for the upcoming /l/ compared to when the definite article did not carry any coarticulatory information (Salverda et al. 2014). Similarly, participants are quicker to identify the target *boat* in the phrase *green boat* when the /n/ in *green* contains coarticulatory cues to /b/ compared to when it does not or when the coarticulatory cues are misleading (Gow & McMurray 2007).

Perceptual compensation for coarticulation

Despite the acoustic changes in the influenced segment, coarticulation does not hinder its recognition due to listeners' ability to compensate for coarticulation. Perceptual compensation for coarticulation is understood as listeners' ability to factor out the influences of surrounding segments on a target segment and attribute them to the source segments, with the consequence that a single context-independent percept remains (Mann 1980, Zellou 2017). Listeners' ability to compensate for coarticulation has been demonstrated in studies showing that listeners interpret cues according to their contexts and thus perceive the same ambiguous signal as different segments in different contextual conditions (Mann & Repp 1980, Fowler 1984, Gaskell & Marslen-Wilson 1998, Kleber et al. 2012, Zellou 2017).

In the perception of consonants, when hearing a fricative that is ambiguous between /s/ and /ʃ/, English listeners reported perceiving /s/ when the fricative was followed by a rounded vowel, and reported perceiving /ʃ/

when it was followed by an unrounded vowel (Mann & Repp 1980). Listeners presented with a fricative+unrounded vowel sequence cannot attribute the low spectral frequency in the fricative to coarticulation with the vowel, which leads them to interpret the spectral frequency as an inherent part of the fricative, which therefore must be /ʃ/; however, in a fricative+rounded vowel sequence listeners attribute the same low frequencies to lip rounding and categorise the fricative as /s/ (Mann & Repp 1980, Smits 2001, Mitterer 2006). Similarly, a stop that is ambiguous between /d/ (with high $F3$ onset) and /g/ (with low $F3$ onset) is more likely to be perceived as /g/ in a /l+stop/ sequence than in /ɹ+stop/ (Mann 1980). Listeners interpret the ambiguos $F3$ to as the increased $F3$ of /g/ in the /l/+stop sequence, they interpret it as the lowered $F3$ of /d/ in the /ɹ/+stop sequence (Mann 1980). These effects might not be specific to speech, as under certain circumstances, a preceding low tone (corresponding to /ɹ/) or high tone (corresponding to /l/) have the same effect on the perception of contrast(Lotto & Kluender 1998, Fowler et al. 2000). Consonantal context has also been shown to affect vowel categorisation. For example, listeners accept a vowel with a relatively high $F2$ as /ʊ/ in the fronting /s_t/ context, whereas they categorise the same vowel as /ɪ/ in the non-fronting /w_l/ context despite the fact that prototypical /ʊ/ has a low $F2$ and prototypical /ɪ/ has a high $F2$ (Kleber et al. 2012).

These studies suggest that listeners attribute coarticulatory information to the influencing segment and factor coarticulatory effects out in the perception of the affected segment. That is, listeners can compensate for coarticulation by attributing acoustic cues resulting from coarticulation to their coarticulatory source. By compensating for coarticulation, listeners arrive at phoneme categories and category memberships despite contextual change to the signal. For instance, the release burst of /g/ differs acoustically between /gi/ and /gu/, but listeners perceive acoustically different /g/ bursts in the appropriate coarticulatory context as more similar to each other than two acoustically identical /g/ bursts when one of them is in a mismatching coarticulatory context (Fowler 1984). Similarly, English listeners perceived oral and nasal vowels as different when nasality cannot be attributed to context (e.g. both nasal and oral vowels presented in the context of oral consonants or in isolation) and as similar when nasality can be attributed to context (e.g. nasal vowels in the context of nasal consonants) (Beddor & Krakow 1999).

The examples provided so far show that the perception and identification of an ambiguous segment is affected by phonetic context. When contrastive cues of the target segment are affected, listeners only compensate for coarticulation in cases of ambiguous tokens. In a continuum of synthesised tokens between two unambiguous endpoints, listeners perceive the ambiguous tokens according to their context; however, the tokens at the endpoint corresponding to unambiguous phonemes as produced in natural speech tend to be perceived based on the features of the target sound, irrespective of context (Mann & Repp 1980, Mann 1980, Kleber et al. 2012). This shows that the boundary shift caused by segmental context and compensation for coarticulation does not impact the perception of unambiguous tokens.

In some phonetic contexts, acoustic cues to a segment are not ambiguous, but obscured due to assimiliation, for example /p/ in the phrase *top tag* can be assimilated to [t] (Stevens & Keyser 2010). Listeners can compensate for assimilative coarticulation and the lack of acoustic cues by integrating top-down lexical information when compensating for coarticulation (Gaskell & Marslen-Wilson 1998). For exmaple, listeners can identify /t/ *freigh<u>t</u> bearer*, even when it is realised with a final /p/ instead of a /t/ in *freight*, but cannot recognise a /t/ in nonwords, such as *preip bearer* (Gaskell & Marslen-Wilson 1998). As a result, even when coarticulation obscures certain acoustic cues, for instance the tongue tip gesture is reduced and the alveolar closure is not realised in the /nt/ cluster in *can't go*, nasalisation and glottalisation of /æ/ provide sufficient cues for the listeners to recover the intended word (Stevens & Keyser 2010).

1.2.2 Coarticulation and sound change

Coarticulation provides systematic and directional variation which may become the input for sound change, as sound change is often related to how coarticulation is produced by the speaker and perceived by the listener (Ohala 1993, Beddor 2009, Solé & Ohala 2010, Ohala 2012, Garrett & Johnson 2013, Harrington et al. 2018). Despite listeners being efficient at attributing the acoustic effect of coarticulation to its source, coarticulation can lead to sound change when listeners do not compensate for it (Ohala 1981) or the sound resulting from coarticulation is ambiguous between phonological categories (Harrington et al. 2018, Blevins 2007).

Ohala's (1981; 1993; 2012) model of sound change specifically identifies insufficient compensation for coarticulation by listeners as the phonetic process implicated in the initiation of sound change. Listeners, on hearing a coarticulated speech signal, can either compensate for coarticulation, retracing the acoustic signal to the speaker's intended form or they may not compensate for coarticulation, and take the utterance at face value (Ohala 1981). The scenario in which listeners compensate does not lead to sound change, whereas the scenario in which listeners take the signal at face value may impact listener-speakers' realisation and therefore result in a change in the pronunciation norm. A change in a pronunciation norm in some instances may provide favourable conditions for sound change to occur. For instance, the perceived equivalence between a nasal, the source of coarticulation, and a nasalised vowel, the effect of coarticulation, can lead to the emergence of nasalised vowels (Beddor 2009). Perceived equivalence between a nasal and a nasalised vowel was shown when listeners could not distinguish a sequence of a vowel with a long nasalised portion and short nasal from sequence of a vowel with a short nasalised portion followed by a long nasal. In contrast, listeners were able to tell the difference between tokens when the length of the nasal varied and vowel nasalisation was kept constant (Beddor 2009).

In contrast, in Harrington et al.'s (2018) interactive phonetic (IP) sound change model, insufficient compensation for coarticulation is not the cause, but the consequnce of and evidence for sound change. In the IP model of sound change, the prerequisite of sound change is that typical realisations of two phonemes are acoustically distinct, but highly coarticulated realisations of one phoneme become acoustically similar to the other phoneme (Harrington et al. 2018). As listeners and speakers interact, coarticulated realisations are incorporated to listeners' representation, shifting the representation closer to the second phoneme, and potentially leading to a merger (Harrington et al. 2018). This merger is signalled by failed compensation for coarticulation (Harrington et al. 2018).

Ohala's (1981; 1993; 2012) model and the IP model are not mutually exclusive; in fact, the IP model's aim was to bridge the gap between phonetically motivated and socially motivated models of sound change. For instance, evidence from /uː/-fronting in Standard Southern British English is consistent both with Ohala's (1981; 1993; 2012) and the IP model (Harrington et al. 2008; 2018). In Standard Southern British English, older speakers exhibit a contextual /uː/-fronting in words containing front /j/, (e.g. *feud*, /fjuːd/) compared to words comparing back /w/ (e.g. *swoop*, /swuːp/), indicating coarticulation, whereas younger speakers produce an equally fronted /uː/ in both words and compensate less for context effects in perception, indicating phonological fronting (Harrington et al. 2008). This result is consistent with Ohala's (1981; 1993; 2012) model of sound change, while Harrington et al.'s (2018) IP model correctly predicted the change in behaviour caused by the interaction between listener-speakers (Harrington & Schiel 2017, Harrington et al. 2018). However, not all allophonic variation in production leads to sound change (Ohala 1993) and failed compensation or miscategorisation of items does not always indicate sound change (Stevens & Harrington 2014, Harrington et al. 2018).

1.3 Research question

The research covered in **Section 1.1** shows that lateral-final rimes exhibit interesting patterns of gestural overlap and reduction in several dialects of English, including AusE. Studies in **Section 1.1** indicate that the tongue dorsum and the tongue tip gesture in the complex speech sound /l/ might behave differently from each other: while the tongue dorsum gesture overlaps with the preceding vowel, the tongue tip gesture is reduced compared to onset position. Therefore we further investigate how gestural complexity of /l/ is reflected in AusE phonetics.

The studies on coarticulation covered in **Section 1.2** suggests coarticulation in lateral-final rimes can potentially impact the production and perception of phonological contrast. However, the impact of coarticulation in lateral-final rimes on the production and perception of phonological contrast has not been systematically examined in AusE. Therefore the aim of this dissertation is to systematically characterise the production and perception of lateral-final rimes in this variety of English. We also aim to coalesce findings from different domains to advance our understanding of lateral-final rimes.

1.4 Organisation of thesis

Chapters 2-7 present six studies that address different aspects of the phonetics and the phonology of rimes containing coda /l/, focusing on the production as well as the perception of phonemic contrast. The first three studies examine prelateral vowel allophony conditioned by vowel-/l/ coarticulation and its effect on vowel contrast in speakers' production and listeners' perception. The last three studies examine the variation between consonantal and vocalised /l/ in speakers' articulation, with the last study linking speakers' articulation and listeners' perception.

1.4.1 Acoustic and perceptual vowel contrast reduction before coda /l/

The **first study** (**Chapter 2**) characterised the acoustic effect of coda /l/ on the AusE vowel inventory. Vowel contrasts may be reduced or neutralised before coda laterals in English (Bernard 1985, Palethorpe & Cox 2003, Labov et al. 2008), but the acoustic characteristics of vowel-lateral interaction in AusE rimes have not been systematically examined. Spectral and temporal properties of 16 prelateral and 16 preobstruent vowels produced by 29 speakers of AusE were compared. Acoustic vowel similarity in both environments was captured using random forest classification and hierarchical cluster analysis based on the first three DCT coefficients of

$F1$, $F2$, and $F3$, and duration values. Vowels preceding /l/ codas showed overall increased confusability compared to vowels preceding /d/ codas. In particular, reduced spectral contrast was found for the rime pairs /ʉːl-ʊl/ (*fool-full*), /əʉl-ɔl/ (*dole-doll*), and /æɔl-æl/ (*howl-Hal*). Potential articulatory explanations and implications for sound change are discussed.

Having established acoustic contrast reduction and the attenuation of cues contributing to phonological vowel contrast in lateral-final rimes, **Chapter 3** explores to what extent listeners compensate perceptually for coarticulation on the vowel or find lateral-final rimes ambiguous. To test the effect of vowel-/l/ coarticulation on vowel disambiguation, listeners categorised vowels in /hVd/ and /hVl/ contexts. Reduced accuracy in the /l/ context compared to the /d/ context showed that coda /l/ increases the difficulty of vowel disambiguation. In particular, reduced perceptual contrast was found for the rime pairs /ʉːl-ʊl, əʉl-ɔl/ and /æɔl-æl/ (e.g. *fool-full, dole-doll, howl-Hal*); however, duration contrast was maintained. A second experiment tested the effect of reduced perceptual contrast on word recognition. Listeners identified minimal pairs contrasting key vowel pairs in the /CVl/ and /CVd/ contexts. Reduced accuracy and increased response time in /l/ contexts showed the limits of listeners' ability to compensate for coarticulatory effects of final /l/. The relationship between limited compensation for coarticulation and sound change is discussed.

As we found that spectral contrast is reduced but durational contrast is maintained between the members of the pairs /iːl-ɪl, ʉːl-ʊl, æɔl-æl , əʉl-ɔl/, **Chapter 4** explores whether listeners can rely on durational cues to distinguish between the vowel pairs and whether listners' production affects their perception. We tested whether participants producing a larger durational contrast between word pairs containing the rimes /iːl-ɪl, ʉːl-ʊl, əʉl-ɔl, æɔl-æl/ were better at recognising minimal pairs contrasting the aforementioned rimes. 46 AusE speakers produced 24 /l/-final minimal pairs and identified the same minimal pairs spoken by two speakers. Participants producing a longer durational contrast took longer to respond and were only more accurate when the stimuli contained a bigger durational contrast, indicating a link between listeners' own production and perception. These studies combined indicate that spectral vowel contrast reduction in the prelateral environment contributes to perceptual vowel contrast reduction and therefore to potential language change either due to potential prelateral vowel mergers or to cue transfer from spectral to durational cues.

1.4.2 Coda /l/ lenition in articulation and perception

Having looked at variation and contrast in the acoustics and perception of pre-lateral vowels, we explores variation in coda /l/-lenition in different prosodic and phonemic contexts using articulatory measurements of coda /l/-lenition and impressionistic ratings of coda /l/-vocalisation to shed light on whether variation in coda /l/ lenition in production leads to variation in perception of vocalisation and therefore to potential ongoing language change. To capture the articulatory lenition of coda /l/, in **Chapter 5** we developed a method for measuring tongue elongation. In English /l/, the coordination of antagonistic coronal and dorsal gestures prototypically results in lingual elongation. Although intergestural coordination in laterals has been widely studied, less is known about articulatory configuration in AusE /l/ – a dialect characterised by /l/-lenition (Borowsky 2001, Borowsky & Horvath 1997). We explores tongue elongation as a potential metric of /l/-lenition. The timecourse of lingual elongation was examined in laterals produced by two AusE speakers using electromagnetic articulography.

Tongue elongation was greater in onsets and codas containing laterals compared to those with /d/, and greater in onset /l/ compared to coda /l/. Lingual elongation can potentially be used as a metric to differentiate onset /l/ from lenited or vocalised /l/ across a variety of vocalic and consonantal context in which the tongue tip and tongue dorsum gestures are often unmeasurable.

Chapter 6 explores coda /l/-lenition by examining the coronal constriction reduction in coda /l/ in different phonetic contexts. In auditory impressionistic studies, Borowsky (2001) identified numerous phonetic contexts which facilitate /l/-vocalisation in AusE in auditory-impressionistic studies: according to their analyses, a preceding back vowel, a preceding long vowel, and a following non-alveolar consonant all facilitate /l/-vocalisation. However, auditory-impressionistic rating of coda /l/ as vocalised does not provide any information about the tongue tip gesture. The goal of this study was to determine whether the phonetic contexts identified by Borowsky (2001) facilitate tongue tip lenition. To determine the lenition of the tongue tip gesture, we compared tongue tip position between onset and coda /l/ by systematically manipulating the place of articulation and length of the preceding vowel, the place of articulation of the following consonant, and coda complexity. We found that the magnitude of coda /l/-lenition varies between speakers, but in line with Borowsky's (2001) result, coda /l/ is less likely to be lenited before an alveolar consonant and more likely to be lenited before a labial or a dorsal consonant. A following tautosyllabic consonant has a larger effect than a heterosyllabic consonant. The effect of preceding vowel varies between speakers: a back vowel, similar to the dorsal gesture of /l/ may facilitate /l/-lenition for some speakers, while an articulatorily dissimilar front vowel might inhibit or facilitate /l/-lenition, depending on the speaker.

In **Chapter 7**, we compared auditory ratings of /l/-vocalisation to articulatory metrics, by measuring expert phoneticians' rating of /l/-vocalisation across several phonetic environments and testing if there was a correlation between tongue tip lenition and listeners' rating of /l/ as vocalised across several phonetic environments. In addition, we provide a detailed characterisation of the articulatory differences in /l/ associated with listeners' rating of /l/ as vocalised or consonantal. Listeners do not rely on coda /l/-lenition directly to identify coda /l/ as vocalised; however, listeners are more likely to perceive an /l/ as vocalised in phonetic contexts that facilitate /l/-lenition. These studies combined show that listeners tend to perceive /l/ as non-vocalised despite the prevalence of tongue tip lenition in the articulatory data which indicates that if /l/-vocalisation is a phonetically motivated language change in progress, it is below the consciousness of the listener.

1.4.3 General discussion

Chapters 8 and 9 provide a summary and general discussion of the studies that comprise this thesis. We coalesce findings on the acoustics and perception of lateral final-rimes, as well as on the articulation and perception of /l/-vocalisation to discuss their implications for sound change and to contribute to our understanding of the multigestural segment /l/. In addition, limitations of the dissertation and recommendations for future research are discussed.

1.5 References

Altendorf, U. & Watt, D. (2008). The dialects in the South of England: phonology. In B. Kortmann & C. Upton (Eds.), *Varieties of English* (Vol. 1, p. 194–222). Berlin, New York: Mouton de Gruyter.

Arnold, L. (2015). Multiple mergers: Production and perception of three pre-/l/ mergers in Youngstown, Ohio. *University of Pennsylvania Working Papers in Linguistics*.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 785–821.

Beddor, P. S. & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *The Journal of the Acoustical Society of America*, *106*(5), 2868–2887.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W. & Brasher, A. (2013, April). The time course of perception of coarticulation. *Journal of the Acoustical Society of America*, *133*(4), 2350–2366.

Bernard, J. (1985). Some local effects of post-vocalic (l). In J. E. Clark (Ed.), *The cultivated Australian: Festschrift in honour of Arthur Delbridge* (p. 319–332). Hamburg: Buske.

Bernard, J. R. (1970). Toward the acoustic specification of Australian English. *STUF-Language Typology and Universals*, *23*(1-6), 113–128.

Bladon, R. A. W. & Al-Bamerni, A. (1976). Coarticulation resistance in English /l/. *Journal of Phonetics*, *4*(2), 137–150.

Blevins, J. (2007). Interpreting misperception: beauty is in the ear of the beholder. In M.-J. Solé, P. S. Beddor & M. Ohala (Eds.), *Experimental approaches to phonology* (p. 144–154). Oxford: OUP.

Borowsky, T. (2001). The vocalisation of dark *l* in Australian English. In D. Blair & P. Collins (Eds.), *English in Australia* (p. 69–87). Philadelphia, Amsterdam: John Benjamins Publishing Company.

Borowsky, T. & Horvath, B. (1997). L-vocalisation in Australian English. In F. Hinskens, R. van Hout & W. L. Wetzels (Eds.), *Variation, change and phonological theory.* (p. 101–123). Amsterdam: Benjamins.

Bradley, D. (2004). Regional characteristics of Australian English: Phonology. In K. Burridge & B. Kortmann (Eds.), *Varieties of English: The Pacific and Australasia* (p. 645–655). Berlin: Mouton de Gruyter.

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

Browman, C. P., Goldstein, L. et al. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341–376.

Browman, C. P. & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*(1), 219–252.

Browman, C. P. & Goldstein, L. M. (1995). Gestural syllable position effects in American English. In F. Bell-Berti & J. L. Raphael (Eds.), *Producing speech: contemporary issues.* New York: AIP Press.

Butcher, A. (2006). Formant frequencies of /hVd/ vowels in the speech of South Australian females. In P. Warren & C. I. Watson (Eds.), *Proceed-*

ings of the 11th Australian international conference on speech science and technology (p. 449–453). Auckland, New Zealand.

Cox, F. (1999). Vowel change in Australian English. *Phonetica*, *56*, 1–27.

Cox, F. & Fletcher, J. (2017). *Australian English pronunciation and transcription.* Cambridge: Cambridge University Press.

Cox, F. & Palethorpe, S. (2004). The border effect: Vowel differences across the NSW-Victorian border. In C. Moskovsky (Ed.), *Proceedings of the conference of Australian Linguistics Society.* Newcastle, Australia.

Cox, F. & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association*, *37*(3), 341–350.

Cox, F., Palethorpe, S. & Bentink, S. (2014). Phonetic archaeology and 50 years of change to Australian English /iː/. *Australian Journal of Linguistics*, *34*(1), 50–75. Retrieved from `http://dx.doi.org/10.1080/07268602.2014.875455`

Farnetani, E. & Recasens, D. (2012). Coarticulation and connected speech processes. In W. J. Hardcastle, J. Laver & F. E. Gibbon (Eds.), *The handbook of phonetic sciences* (p. 316–352). Chichester, U.K.: Wiley-Blackwell.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, *8*(1), 113–133.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, *36*(4), 359–368.

Fowler, C. A. & Brown, J. M. (2000). Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics*, *62*(1), 21–32.

Fowler, C. A., Brown, J. M. & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human perception and performance*, *26*(3), 877.

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

Garrett, A. & Johnson, K. (2013). Phonetic bias in sound change. In A. C. L. Yu (Ed.), *Origins of sound change* (p. 51–97). Oxford, UK: Oxford University Press.

Gaskell, M. G. & Marslen-Wilson, W. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 380–396.

Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. *Phonetic interpretation: papers in laboratory phonology VI*, 222–236.

Gick, B., Campbell, F., Oh, S. & Tamburri-Watt, L. (2006). Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*, *34*(1), 49–72.

Gick, B., Kang, M. A. & Whalen, D. H. (2002). MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics*, *30*(3), 357–371.

Gick, B. & Wilson, I. (2006). Excrescent schwa and vowel-laxing: Corss-linguistic responses to conflicting articulatory targets. In L. M. Goldstein, D. H. Whalen & C. T. Best (Eds.), *Papers in laboratory phonology* (Vol. 8, p. 635–659). Berlin: Mouton de Gruyter.

Giles, S. & Moll, K. (1975). Cinefluorographic study of selected allophones of English /l/. *Phonetica*, *31*(3-4), 206–227.

Gow, D. W. & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. *Papers in laboratory phonology*, *9*(173-200).

Hall-Lew, L. & Fix, S. (2012). Perceptual coding reliability of (l)-vocalization in casual speech data. *Lingua*, *122*(7), 794–809.

Hardcastle, W. & Barry, W. (1989, 7). Articulatory and perceptual factors in /l/ vocalisations in English. *Journal of the International Phonetic Association*, *15*, 3–17.

Harrington, J., Cox, F. & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics*, *17*(155-184).

Harrington, J., Kleber, F. & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America*, *123*(5), 2825–2835.

Harrington, J., Kleber, F., Reubold, U., Schiel, F. & Stevens, M. (2018). Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in cognitive science*, *10*(4), 707–728.

Harrington, J. & Schiel, F. (2017). /u/-fronting and agent-based modeling: The relationship between the origin and spread of sound change. *Language*, *93*(2), 414–445.

Horvath, B. M. & Horvath, R. J. (1997). The geolinguistics of a sound change in progress: /l/ vocalization in Australia. *U. Penn Working Papers in Linguistics*, *4*(1), 109–124.

Horvath, B. M. & Horvath, R. J. (2001). A multilocality study of a sound change in progress: The case of /l/ vocalization in New Zealand and Australian English. *Language Variation and Change*, *13*, 37–57.

Horvath, B. M. & Horvath, R. J. (2002). The geolinguistics of /l/ vocalisation in Australia and New Zealand. *Journal of Sociolinguistics*, *6/3*, 319–346.

Iskarous, K., Mooshammer, C., Hoole, P., Recasens, D., Shadle, C. H., Saltzman, E. et al. (2013). The coarticulation/invariance scale: Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance. *The Journal of the acoustical society of America*, *134*(2), 1271–1282.

Kleber, F., Harrington, J. & Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, *55*(3), 383–405.

Krakow, R. A. (1989). *The articulatory organization of syllables: a kinematic analysis of labial and velar gestures.* Unpublished doctoral dissertation, Yale.

Labov, W., Ash, S. & Boberg, C. (2008). *The atlas of North American English, phonetics, phonology and sound change.* Berlin, Boston: Gruyter Mouton.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford, UK; Cambridge, Mass.: Blackwell.

Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition, 38*(3), 245–294.

Lee-Kim, S.-I., Davidson, L. & Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology, 4*(2), 475–511.

Lewis, E. & Loakes, D. (2012). The /ɐlC-ɔlC/ sound change in Australian English: preliminary resɔlts. In *Proceedings of the 14th australasian international conference on speech science and technology* (Vol. 14, p. 73–76). Macquarie University, Sydney.

Lin, S., Beddor, P. & Coetzee, A. (2014). Gestural reduction, lexical frequency, and sound change: a study of post-vocalic /l/. *Laboratory Phonology, 5*(1), 9–36.

Lin, S., Beddor, P. S. & Coetzee, A. W. (2011). Gestural reduction and sound change: an ultrasound study. In *Proceedings of the 17th international congress of phonetic sciences* (p. 1250–1253). Hong Kong.

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

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The production of speech* (p. 217–245). New York: Springer-Verlag.

Loakes, D., Clothier, J., Hajek, J. & Fletcher, J. (2014). An investigation of the /el/-/æl/ merger in Australian English: A pilot study on production and perception in South-West Victoria. *Australian Journal of Linguistics*, *34*(4), 436–452.

Loakes, D., Graetzer, N., Hajek, J. & Fletcher, J. (2012). Vowel perception in Victoria: Variability, confusability and listener expectation. In *Proceedings of the 14th australasian international conference on speech science and technology* (Vol. 14). Macquarie University, Sydney.

Loakes, D., Hajek, J. & Fletcher, J. (2010a). The /el/-/æl/ sound change in Australian English: A preliminary perception experiment. In Y. Treis & R. De Busser (Eds.), *Selected papers from the 2009 conference of the Australian Linguistic Society.* Melbourne, Australia.

Loakes, D., Hajek, J. & Fletcher, J. (2010b). Issues in the perception of the /el/-/æl/ contrast in Melbourne: Perception, production and lexical frequency effects. In *Proceedings of 13th australasian international conference on speech science and technology.* Melbourne, Australia.

Loakes, D., Hajek, J. & Fletcher, J. (2010c). (Mis)perceiving /el/-/æl/ in Melbourne English: A micro-analysis of sound perception and change. In *Proceedings of 13th australasian international conference on speech science and technology.* Melbourne, Australia.

Loakes, D., Hajek, J. & Fletcher, J. (2011). /æl/-/el/ transposition in Australian English: Hypercorrection or a competing sound change? In *Proceedings of the 17th international congress of phonetic sciences.* City University of Hong Kong.

Lotto, A. J. & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*(4), 602–619.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, *28*(5), 407–412.

Mann, V. A. & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*(3), 213–228.

Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics*, *68*(7), 1227–1240.

Narayanan, S. S., Alwan, A. A. & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. part I. The laterals. *The Journal of the Acoustical Society of America*, *101*(2), 1064–1077.

Oasa, H. (1989). Phonology of current Adelaide English. In P. Collins & D. Blair (Eds.), *Australian English: The language of a new society* (p. 271–287). St Lucia, Queensland: University of Queensland Press.

Ohala, J. J. (1981, May). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (Eds.), *Papers from the parasession on language and behaviour.* Chicago.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (p. 237–278). London: Longman.

Ohala, J. J. (2012). The listener as a source of sound change: An update. In M.-J. Solé & D. Recasens (Eds.), *The initiation of sound change*

(p. 21–36). Amserdam; Philadelphia: John Benjamins Publishing Company.

Ohala, J. J. & Ohala, M. (1995). Speech perception and lexical representation: the role of vowel nasalization in Hindi and English. In B. Connell & A. Arvaniti (Eds.), *Papers in laboratory phonology* (Vol. 4, p. 41–60). Cambridge: Cambridge University Press.

Palethorpe, S. & Cox, F. (2003). Vowel modification in pre-lateral environments. In *International seminar on speech production.* Sydney.

Proctor, M. & Walker, R. (2012). Articulatory bases of sonority in English liquids. In S. Parker (Ed.), *The sonority controversy* (p. 289–316). Berlin, New York: Mouton de Gruyter.

Proctor, M., Walker, R., Smith, C., Szalay, T., Narayanan, S. & Goldstein, L. (2019). Articulatory characterization of English liquid-final rimes. *Journal of Phonetics*, *77*, 100921.

Recasens, D. (2002). An EMA study of VCV coarticulatory direction. *The Journal of the Acoustical Society of America*, *111*(6), 2828–2841.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. *Experimental brain research series*, *15*, 129–144.

Saltzman, E. L. & Munhall, K. G. (1989). A dynamical approach to gestural pattern in speech production. *Ecological Psychology*, *1.*(4.), 333–382.

Salverda, A. P., Kleinschmidt, D. & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of memory and language*, *71*(1), 145–163.

Scobbie, J. M. & Pouplier, M. (2010). Syllable structure and external sandhi: an EPG study of vocalisation and retraction of word-final English /l/. *Journal of Phonetics*, *38*, 240–259.

Scobbie, J. M., Pouplier, M. & Wrench, A. A. (2007). Conditioning factors in external sandhi: an EPG study of English /l/vocalisation. In *Proceedings of the 16th international congress of phonetic sciences* (p. 441–444). Saarbrücken, Germany.

Scobbie, J. M. & Wrench, A. A. (2003). An articulatory investigation of word final /l/ and //l/-sandhi in three dialects of English. In *Proceedings of the 15th international congress of phonetic sciences.* Barcelona, Spain.

Smits, R. (2001). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology*, *27*(5), 1145–1162.

Solé, M.-J. & Ohala, J. J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. *Papers in laboratory phonology*, *10*, 607–655.

Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*(3), 291–311.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. *Human communication: A unified view.*

Stevens, K. N. & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*(1), 10–19.

Stevens, M. & Harrington, J. (2014). The individual and the actuation of sound change. *Loquens*, *1*(1), e003.

Stone, M. & Lundberg, A. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *The Journal of the Acoustical Society of America*, *99*(6), 3728–3737.

Strycharczuk, P., Derrick, D. & Shaw, J. (2018). How vocalic is vocalised /l/? Evidence from lateralisation. In *Conference of the association for laboratory phonology.* Lisbon, Portugal.

Strycharczuk, P. & Scobbie, J. M. (2019). Gestural delay and gestural reduction. Articulatory variation in /l/-vocalisation in Southern British English. In A. Prezwozny, C. Viollain & S. Navarro (Eds.), *Advances in the study of contemporary spoken English: Multifocal analyses of variation.* Edinburgh: Edinburgh University Press.

Thomas, B. & Hay, J. (2005). A pleasant malady: The Ellen/Allan merger in New Zealand English. *Te Reo*, *48*, 69–93.

Trubetzkoy, N. S. (1969). *Principles of phonology.* Berkeley: University of California Press.

Turton, D. (2017). Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalization in varieties of English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1).

Wade, L. (2017). The role of duration in the perception of vowel merger. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1).

Watson, C. I. & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America*, *106*(1), 458–468.

Wells, J. C. (1982). *Accents of English.* Cambridge: Cambridge University Press.

West, P. (1999). Perception of distributed coarticulatory properties of English /l/ and /r/. *Journal of Phonetics*, *27*(4), 405–426.

Ying, J., Carignan, C., Shaw, J. A., Proctor, M., Derrick, D. & Best, C. (2017). Temporal dynamics of lateral channel formation in /l/: 3D EMA data from Australian English. In *Proceedings of the 18th annual conference of the international speech communication association.* Stockholm, Sweden.

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, *61*, 13–29.

2 Spectral contrast reduction in Australian English lateral-final rimes

This chapter is based on the following paper, which is being prepared for submission to the Journal of the Acoustical Society of America:

Szalay, T., Benders, T., Cox, F., Palethorpe, S. & Proctor, M. (in preparation). Spectral contrast reduction in Australian English lateral-final rimes.

I certify that I was responsible for the development of the concept of this paper, in discussion with my supervisors/co-authors. The data was collected by Dr. Felicity Cox and Dr. Sallyanne Palethorpe for an unrelated study; the current subset of the data has not been published previously. I took leadership in conducting the research, and was responsible for the phonetic and statistical analysis, and the writing of all parts of the paper. My co-authors provided advice to improve the experimental design and methods, the analyses, the interpretation of the data, as well as the presentation of the written component.

Abstract

Vowel contrasts may be reduced or neutralised before coda laterals in English (Bernard 1985, Palethorpe & Cox 2003, Labov et al. 2008), but the acoustic characteristics of vowel-lateral interaction in Australian English rimes have not been systematically examined. Spectral and temporal properties of 16 pre-lateral and 16 preobstruent vowels produced by 29 speakers of Australian English were compared. Acoustic vowel similarity in both environments was captured using random forest classification and hierarchical cluster analysis of the first three DCT coefficients of $F1$, $F2$, and $F3$, and duration values. Vowels preceding /l/ codas showed overall increased confusability compared to vowels preceding /d/ codas. In particular, reduced spectral contrast was found for the rime pairs /ʉːl-ʊl/ (*fool-full*), /əʉl-ɔl/ (*dole-doll*), and /æɔl-æl/

(*howl-Hal*). Potential articulatory explanations and implications for sound change are discussed.

2.1 Introduction

Coarticulation, the influence of adjacent sounds on each other, causes predictable variation in speech with the potential to affect phonological contrast (Hyman 2013, Garrett & Johnson 2013). Vowel-lateral coarticulation in particular may reduce or neutralise phonemic vowel contrast in several varieties of English, including Australian English (AusE) (Altendorf & Watt 2008, Cox & Palethorpe 2004, Labov et al. 2008, Palethorpe & Cox 2003, Wade 2017). In AusE, vowel-lateral coarticulation has been shown to reduce the $F1$-$F2$ vowel space due the phonetic backing of front vowels in pre-lateral environment (Bernard 1985, Cox & Palethorpe 2004, Palethorpe & Cox 2003). For instance contrast reduction is regularly observed between *pool* and *pull* (Bernard 1985, Cox & Palethorpe 2004, Palethorpe & Cox 2003). However, carefully controlled and systematic analysis of AusE vowels is required to further our understanding of how coda laterals influence preceding vowels and reduce vowel contrast.

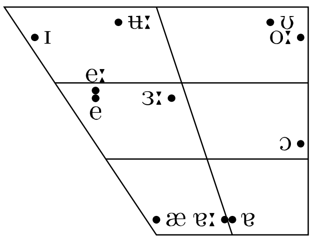
Both acoustic and perceptual vowel contrast reduction in pre-lateral environments have been reported in several dialects of English. The contrast between the high vowels in *feel-fill* is reduced in some Southern dialects of American English and in Standard Southern British English, through the phonetic lowering of the tense vowel /iː/ in the pre-lateral environment (Altendorf & Watt 2008, Harris 1994, Labov et al. 2008, Turton 2014). The *pool-pull* contrast is reduced in Pennsylvanian and Southern British due to the phonetic lowering of the vowel in *pool* (Altendorf & Watt 2008, Labov et al. 2008). The same contrast is also reduced in South Australian English, through a different mechanism: phonetic backing and lowering of the tense vowel /ʉː/ in the pre-lateral environment (Butcher 2006, Oasa 1989). The acoustic *pool-pull-pole* contrast is reduced in Ohio, as the vowels in *pool* and *pole* shift towards *pull* in the vowel space (Arnold 2015, Wade 2017). A perceptual merger between the mid and low front vowels /e/ and /æ/ has been observed in the pre-lateral environment (*hell-Hal*) in New Zealand English (Thomas & Hay 2005) and in Melbourne English (e.g. Loakes et al. 2012; 2014). Collectively, these findings suggest that different, dialect-specific mechanisms may be involved in vowel-lateral interactions in different varieties of English.

2.1.1 Pre-lateral vowels in Australian English

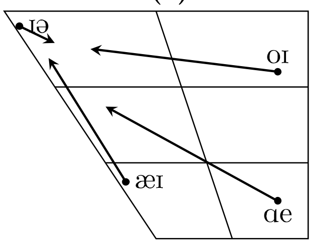
AusE vowel inventory

In Australian English, coda /l/ has been shown to influence vowels in ways that can potentially reduce perceptual and acoustic vowel contrast, especially between the vowel pairs /ʉː-ʊ, əʉ-ɔ, æɔ-æ/ (Bradley 2004, Bernard 1985, Szalay et al. 2018, Palethorpe & Cox 2003). AusE has a large vowel inventory consisting of 18 stressed vowels and schwa (**Figure 3.1**), utilising both spectral and durational contrast (Cox & Fletcher 2017). Duration is contrastive for spectrally similar vowels, for instance, the vowel pairs /ɐː-ɐ, eː-e/ (e.g. *card-cud, shared-shed*) contrast mostly in duration, whereas the pairs /ʉː-ʊ/ contrast both in duration and in spectral quality (Cox & Fletcher 2017).

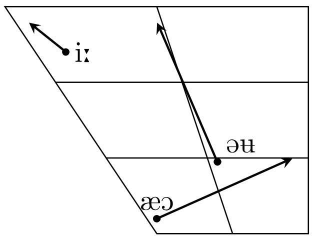
The AusE vowel system includes both diphthongs (/æɪ, ɑe, oɪ, æɔ, əʉ, ɪə/) and monophthongs (front: /iː, ɪ, e, eː, æ/, central: / ʉː, ɜː, ɐː, ɐ, and back: /oː, ɔ, ʊ/) (Harrington et al. 1997, Watson & Harrington 1999). The vowels /iː/ and /ʉː/ are classified as monophthongs but are characterised by an onglide (Bernard 1970, Harrington et al. 1997, Cox & Palethorpe 2007, Cox et al. 2014). /ɪə/ is classified as a diphthong, but can also be realised as a monophthong [ɪː] or as disyllabic [ɪːə] (Harrington et al. 1997, Cox & Palethorpe 2007). Some of the diphthongs, having similar first or second target characteristics to a monophthong, form pairs with monophthongs (Cox 1999). For example, /æɔ/ and /æ/ (e.g. *loud-lad*) share the first target of the diphthong, whereas /əʉ/ shares the location of the second target with the nucleus of /ʉː/ (e.g. *code-cooed*) (Cox 1999). These vowels are considered pairs as the members have moved in parallel in sound change: /æɔ/ lowering was accompanied by /æ/ lowering and the fronting of /ʉː/ took place in parallel with the fronting of the second element of /əʉː/ (Cox 1999).



(a) AusE monophthongs



(b) AusE diphthongs 1



(c) AusE diphthongs 2

Figure 2.1: The AusE vowel inventory. Figures reproduced from Cox & Fletcher (2017).

Effect of coda /l/ on monophthongs

AusE pre-lateral vowels differ from their non pre-lateral counterpart in many ways: front and central vowels are phonetically lowered and backed and some low and back vowels are phonetically raised (Bernard 1985, Cox & Palethorpe 2004, Palethorpe & Cox 2003). Front and central vowels exhibit phonetic lowering shown by increased $F1$ in /iː, ɪ, e,/ and /ʉː, ɜː/ (Bernard 1985, Cox & Palethorpe 2004, Palethorpe & Cox 2003). Front /iː, ɪ, e, æ/ and central /ʉː, ɜː/ are also characterised by lowered $F2$ representing phonetic retraction before /l/ (Bernard 1985, Cox & Palethorpe 2004, Palethorpe & Cox 2003). Among the low and back vowels only /ɔ, ʊ/ (e.g. *poll, pull*) and /ɐ/ (e.g. *hull*) are influenced by coda /l/, as the former two exhibit phonetic backing and the latter phonetic raising (Bernard 1985, Palethorpe & Cox 2003, Cox & Palethorpe 2004).

Acoustic and auditory contrast reduction of some vowel pairs

Spectral contrast reduction between the members of the pairs /ʉːl-ʊl/, /əʉl-ɔl/, and /æɔl-æl/ has been shown through auditory-impressionistic observations and visual representations of formant trajectories (Bernard 1985, Bradley 2004, Palethorpe & Cox 2003). Spectral contrast between /ʉːl-ʊl/ and /əʉl-ɔl/ is reduced due to the $F2$ lowering in /ʉː/ and in both the first and second target of the diphthong /əʉ/ (Palethorpe & Cox 2003). Bernard (1985) observed that the second target of the diphthongs /əʉ, æɔ/ is frequently lost before /l/ and commented on the lack of observable change between the end of the vowels /æɔ, əʉ/ and /l/, which can potentially contribute to spectral contrast reduction between the members of the pairs /əʉl-ɔl/ and /æɔl-æl/. However, duration contrast between the members of these pairs is maintained (Palethorpe & Cox 2003).

In line with the acoustic contrast reduction, perceptual contrast reduction between the members of the pairs /ʉːl-ʊl/, /əʉl-ɔl/, and /æɔl-æl/ has been noted (Loakes et al. 2012, Szalay et al. 2018). As spectral contrast is reduced between the members of these pairs, listeners rely on duration cues: listeners who maintain a larger duration contrast in their own production perceive the members of these pairs more accurately if the speaker maintains a larger duration contrast too (Szalay et al. 2018).

Regional differences

The Victorian dialect of AusE shows contrast reduction between /el-æl/ and /ɐlC-ɔlC/ in production and in perception (Bernard 1985, Cox & Palethorpe 2004, Loakes et al. 2010a, Lewis & Loakes 2012). The $F1$ of /e/ is increased towards /æ/ (e.g. *hell, Hal* (Cox & Palethorpe 2004). The acoustic contrast reduction is reflected in a perceptual near-merger between /e/ and /æ/, as Victorian English listeners misperceive /e/ as /æ/ in the pre-lateral, but not in a preobstruent position when distinguishing minimal pairs (e.g. *telly-tally, pellet-palate*) (Loakes et al. 2010a;b;c; 2011). Acoustic and perceptual contrast reduction has also been shown between /ɐlC-ɔlC/ *gulf-golf*) (Bernard 1985, Lewis & Loakes 2012).

Phonetic backing of pre-lateral /ʉː/ is more prominent in South Australia than in NSW, despite /ʉː/ being backed in New South Wales as well. There is also some evidence for a potential merger between pre-lateral /iː/ and /ɪ/ in Adelaide and Hobart (Bradley 2004) but not in New South Wales (Palethorpe & Cox 2003).

Collectively, these findings suggest that coda /l/ impacts the preceding vowel in various ways depending on vowel quality and speaker dialect, and might lead to potential mergers. However, the impact of coda /l/ on preceding vowels and the potential for a loss of contrast has not been systematically examined. Several of the observations on the effect of coda /l/ and on apparent contrast reduction were made only on the basis of impressionistic observations or visual representations of formants (Bernard 1985, Palethorpe & Cox 2003).

2.1.2 Aims and Hypotheses

The aim of the present study is to systematically characterise the spectral properties of AusE vowels produced in pre-lateral environments and determine the impact of vowel-lateral coarticulation on vowel contrast. We hypothesised that in the pre-lateral context 1) front vowels would have a higher $F1$; 2) front vowels would have a lower $F2$; and 3) spectral contrast would be reduced between /iː-ɪ, ʉːl-ʊl, əʉl-ɔl, æɔl-æl/.

To test Hypotheses 1) and 2), and to also systematically characterise the effect of coda /l/ on the spectral properties of non-front vowels, we examined the effect of /l/ on $F1$, $F2$, and $F3$ of monophthong targets in /l/-final rimes compared to monophthong targets in /d/-final rimes. To test Hypothesis 3) and to systematically characterise spectral contrast reduction in the pre-lateral vowel space, we modelled the dynamic properties of pre-/d/ monophthongs and diphthongs and each of the entire lateral-final rimes using discrete cosine transformation (DCT, see **Section 2.2.3**) of the first three formants. We quantified spectral contrast and similarity using random forest classification and agglomerative hierarchical cluster analysis of AusE vowels based on duration values and the first three DCT coefficients of $F1$, $F2$, and $F3$.

2.2 Methods

2.2.1 Participants

Data from twenty-nine female native monolingual speakers of AusE, born in NSW to Australian-born parents (age = 18–27, mean = 20.24) were analysed. None of the participants reported any speaking, hearing, or reading, difficulties.

2.2.2 Material and Procedure

16 stressed vowels of Australian English were elicited in two monosyllabic paradigms: hVd and hVl (**Table 2.1**). All phonotactically legal words and non-words were elicited in these two contexts. The vowels /ɪə/ and /eː/ were not elicited in the /l/-context as /ɪəl/ and /eːl/ are phonotactically illegal. The elicitation items include 14 non-words: *hal, hule, harl, hooll, holl, hile, hoil,* and *hude, hud, hod, hade, hoyd, howd, hode.* Non-words were included to provide a consistent phonetic frame of reference.

Speakers read each word as it was presented orthographically on a computer monitor. Non-words were accompanied by a rhyming helper word, e.g. *hule - sounds like tool.* Recordings were monitored by a phonetically trained native speaker of Australian English, and participants were asked to repeat erroneous items again with correct pronunciation using the rhyming prompt - no items were modelled by the researcher.

Items were presented in random order in three blocks. The task also included practice words at the beginning of the session, none of which contained coda /d/ or /l/, other contexts (hV, hVn, hVt), and each block of words was followed by 10 short sentences.

Table 2.1: Orthographic representation and IPA transcription of target words. Left columns: /l/-final targets. Right columns: /d/-final targets. Non-words are underlined.

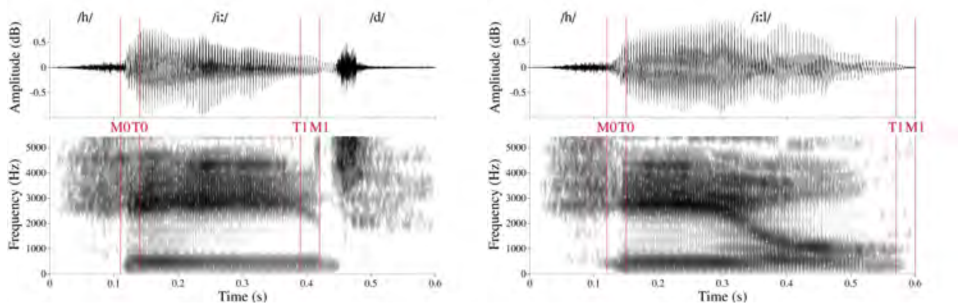
Coda /l/		Coda /d/	
Orthography	IPA	Orthography	IPA
heel	hiːl	*heed*	hiːd
hill	hɪl	*hid*	hɪd
hell	hel	*head*	hed
hal	hæl	*had*	hæd
hule	hʉːl	*who'd*	hʉːd
hurl	hɜːl	*herd*	hɜːd
harl	hɐːl	*hard*	hɐːd
hull	hɐl	*hud*	hɐd
hooll	hʊl	*hood*	hʊd
hall	hoːl	*horde*	hoːd
holl	hɔl	*hod*	hɔd
hail	hæɪl	*hade*	hæɪd
hile	hɑel	*hide*	hɑed
hoil	hɔɪl	*hoyd*	hɔɪd
howl	hæɔl	*howd*	hæɔd
hole	həʉl	*hode*	həʉd

Participants were recorded between 2004 and 2009 in a sound treated recording studio at Macquarie University, Sydney. Speech data were captured using an AKG C535 EB microphone, Cooledit 2000 audio recording software via M-Audio delta66 soundcard to a Pentium 4 PC at 44.1 kHz sampling rate.

2.2.3 Phonetic analysis

32 (targets) $\times$ 3 (repetitions) $\times$ 29 (participants)-1 = 2,783 tokens were analysed [1]. Segment boundaries were automatically located using MAUS forced aligner (Schiel 1999, Kisler et al. 2017) with the AusE grapheme-to-phoneme converter, and manually corrected. The vowel onset was determined on the basis of voicing onset and sudden increase in amplitude (M0;

[1] The second repetition of *who'd* by Speaker 134 is missing from the dataset.



(a) Acoustic landmarks defining the analysis window in the /d/-context, exemplified by *heed*

(b) Acoustic landmarks defining the analysis window in the /l/-context, exemplified by *heel*

Figure 2.2: Acoustic landmarks defining the analysis window, exemplified by *heed* (top) and *heel* (bottom). M0: vowel onset determined by MAUS. M1: vowel offset (pre-/d/ context) and rime offset (pre-/l/ context) determined by MAUS. T0 marks the beginning and T1 marks the end of the analysis window.

Figure 2.2). The vowel-/d/ boundary was determined on the basis of amplitude drop (M1; **Figure 2.2a**). The rime offset in /l/-final targets was not corrected (M1; **Figure 2.2b**). Because there is no discernible boundary between the vowel and the following /l/ in /hVl/ words, the entire /hVl/ rime was analysed instead of selecting an arbitrary boundary in the vowel-lateral transition (**Figure 2.2b**). Segmentation errors were corrected by a trained phonetician only when vowel onset or the vowel offset before coda /d/ was misplaced by more than 30 ms. To minimise potential imprecisions in formant measurements, the first and the last 30 ms of the vowel and the rime were discarded prior to extracting formant values (T0 and T1 in **Figure 2.2**). A boundary threshold larger than the customary 20 ms was chosen because pre-trained force aligners have been shown to be less accurate than train/align models, but are more appropriate for a relatively small dataset like the present one (Fromont & Watson 2016, González et al. 2018).

Formant frequencies were estimated at every 10 ms throughout the analysis window from a 5 ms Gaussian window with 75% overlap and 25 ms formant analysis window with 55 dB dynamic range and a pre-emphasis filter increasing spectral slope above 2700 Hz by 6 dB/octave in Praat (Boersma & Weenink 2013). To optimise formant settings for each speaker, four formants were tracked up to 4500 Hz ceiling for speakers who produced comparatively lower $F2$ and $F3$ or five formants were tracked up to a maximum frequency of 5000 Hz for speakers who produced a comparatively higher $F2$ or $F3$ trajectory. Formant trajectories were manually corrected by the first author using a Matlab-based interface that superimposed formant estimates over a broad band spectrogram with 5ms windows with 40% overlap allowing for corrections of estimates that did not align with the visible formants. After hand-correction, all formant values 1.5 times above or below the interquartile range for each formant in each vowel were rechecked.

Acoustic targets of monophthongs were located automatically in the corrected formant trajectories using the following criteria:

- $F1$ maximum
  - low vowels (/æ, ɐː, ɐ, ɔ/) before /d/ and /l/, as $F1$ maximum indicates the phonetically lowest point
- $F2$ minimum
  - high back vowels (/ʊ, oː/) before /d/, as $F2$ minimum indicates the phonetically backmost point
- $F2$ maximum
  - high front vowels (/iː, ɪ, e/) before /d/ and /l/, as $F2$ maximum indicates the phonetically frontmost point

  - /ʉː/ before /d/, as /ʉː/ is a fronted vowel, characterised by a high $F2$ in AusE
  - /ɜː/ before /l/, as $F2$ lowers considerably between an /ɜː/ target and an /l/ target

- temporal midpoint
  - /ɜː/ before /d/, as the formant trajectories of mid-central /ɜː/ do not show considerable formant change in the pre-/d/ context

- 25% of the rime
  - high back vowels (/ʊ, ʉː, oː/) before /l/, as there is no considerable formant change in these rimes

Acoustic targets were not located for diphthongs, as several diphthong tokens did not exhibit two targets in the pre-lateral context.

Discrete cosine transformations (DCT) were used to model the major dynamic properties of vowels in both types of rimes using emuR (Harrington & Cassidy 1994, Watson & Harrington 1999, Winkelmann et al. 2019). The first three DCT coefficients characterise formant change over time: the zeroth coefficient (k_0) represents the mean of a formant trajectory multiplied by $\sqrt{2}$; the 1st coefficient (k_1) represents the direction and magnitude of the curve of the trajectory: a greater negative k_1 corresponds to greater positive slope; the 2nd coefficient (k_2) represents the trajectory's curvature: a positive k_2 corresponds to an upward pointing curvature and a greater value corresponds to a narrow curvature (Harrington 2010). Each token was represented parametrically by a total of 9 DCT coefficients (3 formants $\times$ 3 coefficients).

2.2.4 Statistical analysis

Effect of coda /l/ on monophthong targets

The effect of coda consonants on the acoustic targets of the monophthongs was examined using Generalised Mixed-Effect Models (GLM) in the *lme4* package (Bates et al. 2015), followed by least square means tests in the *emmeans* package (Lenth 2019, Searle et al. 1980) to evaluate the effect of /l/ on the mean target of each vowel adjusted for the means of other levels of factors in the GLM. We constructed three GLMs with the dependent variables $F1$, $F2$, and $F3$, and the interacting independent variables Vowel (sum-coded) and Coda (treatment coded, comparing /l/ to the baseline /d/); we used the factor Vowel rather than vowel features to test whether all vowels pattern consistently according to their place of articulation (front vs. back, high vs. low). The model included a random by-participant intercept and a by-participant random slope for the effect of coda to account for interspeaker variation. $p-values$ were calculated with the *lmerTest* package (Kuznetsova et al. 2017) using Satterthwaite's degrees of freedom method. We constructed another three GLMs with the same structure, but without an interaction between Coda and Vowel to assess the effect of the Vowel-Coda interactions on model fit through model comparisons using a Chi-squared test. When adding Vowel-Coda interaction significantly improved model fit for $F1$, $F2$, and $F3$, least-square means analysis with Bonferroni correction was used to asses the effect of coda /l/ on the respective formant value of each vowel.

Spectral similarity

Spectral similarity across all diphthongs and monophthongs in the /d/- and /l/-context was tested by creating separate confusion matrices for pre-/d/ and pre-/l/ vowels using random forest classification in the randomForest package (Liaw & Wiener 2002). Random forest is a supervised classification algorithm that builds several decision trees and aggregates their result (Burger 2018). Each decision tree splits the dataset (e.g. formant values of vowels) into subsets (e.g. back versus front vowels) based on descriptor values (e.g. high or low $F2$) (Burger 2018). Building a random forest model consists of a training phase during which the algorithm learns the categories based on category labels (e.g. vowel labels) and descriptors (e.g. formant values, durational values) by building several binary decision trees (Burger 2018). Then, in the testing phase, the remaining data is classified into the previously learnt categories based on descriptors only (Burger 2018). Comparison of the original category labels and the category labels assigned by the random forest analysis provides a confusion matrix (Burger 2018).

During the training phase, random forest classification builds several decision trees to learn the categories present in the data. Each tree is based on a bootstrap sample from the training data (customarily and in this paper 75% of the data) and random selection of descriptors. As training uses several bootstrap samples and different selection of descriptors, cross-validation is not required (Breiman 2002). After a decision tree is built, the random forest classification makes a prediction, called out-of-bag prediction, about the data not in the bootstrap sample, based on the descriptors' values (Liaw & Wiener 2002). After a pre-set number of trees has been built, out-of-bag predictions are aggregated: a low out-of-bag error rate indicates that the algorithm made successful predictions about the data left out in the iterations, and learnt the categories successfully, whereas a high out-of-bag error rate indicates that the algorithm could not make accurate predictions about the data left out from the iteration and was less successful in learning the categories (Liaw & Wiener 2002).

Once the model is trained on a dataset, the second phase is the testing phase during which the model can be tested on the classification of novel data (customarily the remaining 25% of the original data), which are provided to the model without category labels. As a last step, the model's classification of the novel data is compared to the original category labels thus creating a confusion matrix between the original labels and the algorithm's labels, in which confusion rates indicate similarity between vowel categories.

To visualise the similarity between vowel categories and extract p-values, we ran a hierarchical cluster analysis on the confusion matrices output by the random forest analysis. Hierarchical cluster analysis takes the individual vowel categories as single-element clusters. At the first step, it merges two single-element clusters into a larger, binary-branching cluster. At each following step, it merges two clusters until it merges all the vowel categories into a single binary-branching cluster. Members within a cluster are maximally similar and the members of two separate clusters are maximally dissimilar; similarity was measured using Ward's method (Ward 1963). To attest the robustness of clusters made of two or more vowel categories, we extracted the Approximately Unbiased p-value for each multi-element cluster by repeating the hierarchical cluster analysis on the same confusion matrices using multiscale bootstrap sampling in the *pvclust* package (Suzuki & Shimodaira 2006, Efron et al. 1996). Approximately Unbiased p-value expresses the frequency with which a multi-element cluster appears in bootstrapping, and a multi-element cluster is considered to occur significantly frequently when it occurs in more than 95% of the resamples. The results of hierarchical cluster analysis is represented on a dendrogram: elements that are clustered together are similar to each other, and the lower the cluster is split from the other elements, the higher the spectral similarity between the members of the cluster. The location of nodes can be used for comparing between-cluster similarity across dendrograms.

To test spectral similarity in the /d/- and /l/-contexts, we first trained two random forest classification models to learn 16 vowel categories in the /d/-context and 16 vowel categories in the /l/-context based on the DCT coefficients, duration values, and vowel labels using 75% of the randomly selected /d/-final and 75% of the /l/-final tokens. The remaining 25% of the tokens were used to test the classifier, by grouping unlabelled values based just on DCT coefficients and duration values. Separate confusion matrices were created by coda-condition. Lastly, the confusion matrices were fed into an agglomerative hierarchical cluster analysis using Ward's (Ward 1963) to measure between-vowel similarity based on the confusability rates of the vowels. All statistical analyses were carried out in R (R Core Team 2018).

2.3 Results

2.3.1 Effect of /l/ on the monophthong targets

We compared model fits between GLMs with and without Vowel-Coda interactions and found that models including the interactions fit the data significantly better for $F1$, $F2$, and $F3$ ($p < 0.001$ for model comparisons). Therefore, we report the main effect of /l/ from the models containing the interaction. Coda /l/ overall increases $F1$ ($\beta = 33.32, t_{28.01} = 11.41, p < 0.001$), decreases $F2$ ($\beta = -249.88, t_{28.01} = -28.93, p < 0.001$), and increases $F3$

($\beta = 40.7, t_{28.01} = 4.77, p < 0.001$) (**Figures** 2.3a-2.4, **Tables** 2.2a-2.2b). Significant vowel-coda interactions are reported in **Table 2.3**.

Table 2.2: Mean formant values and durations

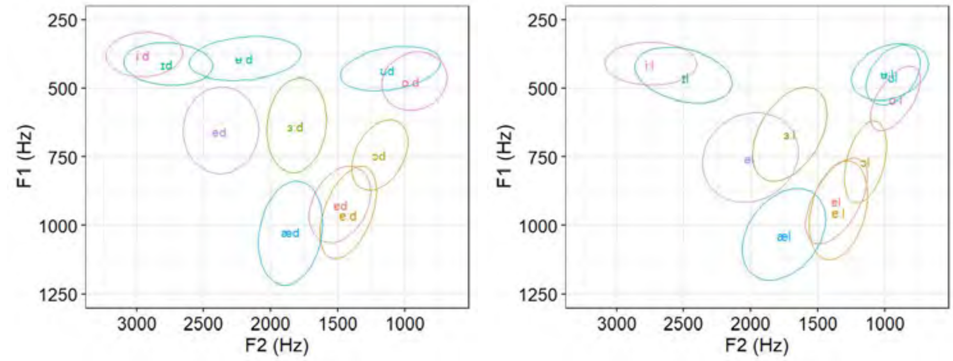
(a) Mean formant values (Hz) at vowel targets and mean vowel durations (ms) in hVd rimes

Vowel	$F1$	$F2$	$F3$	Vowel duration
iː	379	2954	3329	298
ɪ	413	2775	3255	177
e	658	2382	3149	171
æ	1023	1856	3005	209
ʉː	391	2197	2684	295
ɜː	638	1814	2886	307
ɐː	961	1419	3034	329
ɐ	927	1479	2995	158
ʊ	433	1132	2882	175
oː	475	953	3023	313
ɔ	743	1191	2984	169

(b) Mean formant values (Hz) at vowel targets and mean durations (ms) of hVl rimes

Vowel	$F1$	$F2$	$F3$	Rime duration
iː	413	2751	3204	424
ɪ	460	2489	3075	365
e	755	2011	3021	346
æ	1036	1750	2987	395
ʉː	446	983	3017	397
ɜː	668	1711	2865	431
ɐː	953	1347	3065	446
ɐ	910	1360	3079	362
ʊ	457	937	3131	375
oː	540	920	3186	428
ɔ	768	1146	3045	393

As the interactions significantly improved the model fit for all models, planned comparisons assessed the effect of coda /l/ on the $F1$, $F2$, and $F3$ of each vowel, using least-square means with Bonferroni correction (**Table 2.4**). Positive vowel-coda interactions (**Table 2.3**) show that, compared to the overall effect, $F1$ increases more in the /l/-context for /e, ʉː, oː/.



(a) Monophthong targets produced before /d/ codas.

(b) Monophthong targets produced before /l/ codas.

Figure 2.3: Acoustic monophthong targets produced before /d/ (right) and /l/ codas (left). IPA labels: mean $F1$ and $F2$ values (Hz). Ellipses: 95% confidence intervals.

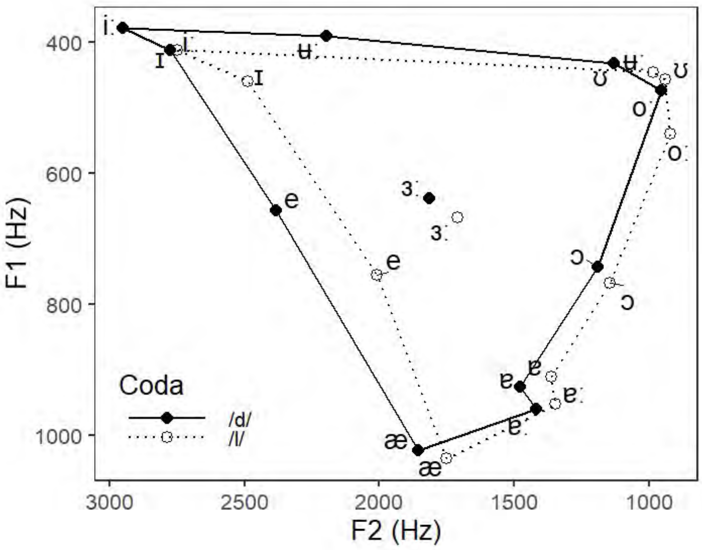


Figure 2.4: Mean acoustic monophthong targets produced before /d/ and /l/ codas.

Table 2.3: Significant vowel-coda interactions in modelling the effect of coda /l/ on pre-lateral vowel targets compared to pre-/d/ vowel targets.

Parameter	Vowel	β	*df*	*t-value*	*p-value*
$F1$	e	64.2	7.93	8.1	< 0.001
	ʉː	22.1	7.95	2.78	0.005
	ɐː	-40.9	7.93	-5.15	< 0.001
	ɐ	-49.5	7.93	-6.25	< 0.001
	oː	31.8	7.93	4.01	< 0.001
$F2$	iː	47.2	1835.01	2.71	0.006
	ɪ	-37.1	1835.01	-2.13	0.034
	e	-121.7	1835.01	-6.98	< 0.001
	ʉː	-964.1	1835.14	-55.12	< 0.001
	ɜː	146.3	1835.01	8.39	< 0.001
	ɐː	177.7	1835.01	10.19	< 0.001
	ɐ	130.4	1835.01	7.48	< 0.001
	ʊ	55.7	1835.01	3.19	0.001
	oː	216.6	1835.01	12.42	< 0.001
	ɔ	204.6	1835.01	11.73	< 0.001
$F3$	iː	-165.6	1835.01	-8.3	< 0.001
	ɪ	220.6	1835.01	-11.05	< 0.001
	e	-168.8	1835.01	-8.46	< 0.001
	ʉː	291.5	1835.18	14.56	< 0.001
	ɜː	-61.6	1835.01	-3.09	0.002
	ɐ	43.8	1835.01	2.20	0.028
	ʊ	207.4	1835.01	10.40	< 0.001
	oː	122.9	1835.01	6.16	< 0.001

Similarly, the negative interactions show that $F1$ increases less for /ɐː, ɐ/ than the overall effect (**Table 2.3**). In line with the negative interactions, least-squares mean test shows no significant difference for the already low /ɐː, ɐ/ vowels (**Table 2.4**). In addition, least-square means test found no significant effect of coda /l/ on low /æ/ (**Table 2.4**). All other vowels show a significantly higher $F1$ in the /l/-context (**Table 2.4**).

Vowel-coda interactions in the initial GLM show that $F2$ of /ɪ, e, ʉː/ is decreased before coda /l/ more than the overall effect, but $F2$ is lowered less than the overall effect for all other vowels. Least-square means test shows that even those vowels which showed a smaller effect for coda /l/ in the

GLM had a significantly lower $F2$ in the /l/-context, except for back /oː/ (**Table 2.4**).

Vowel-coda interactions show that the $F3$ of /ɪ, ʉː, ɐ, ʊ, oː/ is increased before coda /l/ more than the overall effect, but has a smaller than overall effect on /iː, e, ɜː/ (**Table 2.3**). Least-square means test shows that $F3$ in the /l/-context was significantly lower for front vowels /iː, ɪ, e/, and significantly higher for /ʉː, ɐ, ʊ, oː/ (**Table 2.4**). Coda /l/ did not have a significant effect on the $F3$ of /æ, ɜː, ɐː, ɔ/. Therefore least-mean square test does not show a consistent pattern on the effect of coda /l/ on $F3$.

The duration of short vowels was 57% of the long-vowel duration in the /d/ condition, and the duration of rimes containing short vowels was 88% of rimes containing long vowels in the /l/ condition.

Table 2.4: Effect of coda /l/ on $F1$, $F2$, and $F3$ values (Hz) at acoustic target compared to coda /d/. β shows the effect of coda /l/ compared to coda /d/ on the least-square mean of the vowel formant. *SE*, *t-ratio*, and *p-value* calculated from least square means.

Parameter	Vowel	β	*SE.*	*t-ratio*	*p-value*
	iː	33.8	8.45	3.999	0.0007
	ɪ	47.0	8.45	5.557	< 0.0001
	e	97.5	8.45	11.539	< 0.0001
	æ	12.8	8.45	1.518	1
	ʉː	55.4	8.48	6.538	< 0.0001
*F*1	ɜː	29.3	8.45	3.471	0.0059
	ɐː	−7.6	8.45	−0.895	1.0
	ɐ	−16.2	8.45	−1.920	0.6064
	ʊ	24.0	8.45	2.835	0.0553
	oː	65.1	8.45	7.704	< 0.0001
	ɔ	25.3	8.45	2.990	0.0343
	iː	−202.7	19.5	−10.412	< 0.0001
	ɪ	−286.9	19.5	−14.742	< 0.0001
	e	−371.6	19.5	−19.091	< 0.0001
	æ	−105.5	19.5	−5.419	< 0.0001
	ʉː	−1214.0	19.5	−62.207	< 0.0001
*F*2	ɜː	−103.6	19.5	−5.321	< 0.0001
	ɐː	−72.2	19.5	−3.709	0.0025
	ɐ	−119.4	19.5	−6.136	< 0.0001
	ʊ	−194.2	19.5	−9.978	< 0.0001
	oː	−33.3	19.5	−1.711	0.963
	ɔ	−45.3	19.5	−2.326	0.2239
	iː	−124.9	21.7	−5.755	< 0.0001
	ɪ	−179.8	21.7	−8.285	< 0.0001
	e	−128.0	21.7	−5.899	< 0.0001
	æ	−18.5	21.7	−0.852	1
	ʉː	332.2	21.8	15.261	< 0.0001
*F*3	ɜː	−20.9	21.7	0.962	1
	ɐː	30.9	21.7	1.422	1
	ɐ	84.6	21.7	3.895	0.0012
	ʊ	248.2	21.7	11.432	< 0.0001
	oː	163.7	21.7	7.539	< 0.0001
	ɔ	60.8	21.7	2.801	0.0574

2.3.2 Spectral similarity

Formant trajectories for all vowels were modelled using the first three DCT coefficients (**Table 2.6**). Two random forest classification models were trained on DCT coefficients, duration values, and vowel labels using 75% of the tokens to learn 16 vowel categories in each coda condition. Out-of-bag error rate in the testing phase was 3.55% in the /d/-context and 24.07% in the /l/-context, indicating that DCT coefficients and duration values can classify vowels more accurately in the /d/- than in the /l/-context. 25% of the tokens were used to test the classification algorithms; the output of the random forest classification algorithm was compared to the original vowel labels, resulting in two confusion matrices (**Figures** 2.5 and 2.6). In the /d/-context, seven vowels were classified with 100% accuracy (/ɪ, ʉː, ʊ, æɪ, ɔɪ, æɔ, əʉ/), whereas in the /l/-context only the vowel /e/ was classified perfectly. In the /d/-context error rates were small: the least accurately classified vowels were central /ɜː/ and back /ɔ/, identified with respectively 83% and 85% accuracy.

In the /l/-context, the vowel pairs whose members were hypothesised to undergo acoustic contrast reduction /uː-ʊ, əʉ-ɔ, æɔ-æ,/ have a high confusion rate (**Figure 2.6**): 26% of /ʉː/ tokens were classified as /ʊ/ and 28% of /ʊ/ tokens were classified as /ʉː/; 43% of /əʉ/ tokens were classified as /ɔ/ and 16% of /ɔ/ tokens were classified as /əʉ/; 30% of /æɔ/ tokens were classified as /æ/ and 30% of /æ/ tokens were classified as /æɔ/. In contrast, all of the /ʉː, ʊ, æɔ, əʉ/ tokens were identified correctly in the /d/-context, /æ/ was confused with /ɐ/ (9%), not with /æɔ/, and /ɔ/ was misidentified as /ʊ/ (12%) and not as /əʉ/. Members of the pair /iː-ɪ/ were also hypothesised to undergo spectral contrast reduction in the pre-lateral context and the confusion rate between /iː/ and /ɪ/ is higher in the /l/-context (19% of

/iː/ tokens misidentified as /ɪ/ while 5% of /ɪ/ tokens misidentifed as /iː/, without any confusion in the other direction) than in the /d/-context (5% of /iː/ tokens identified as /ɪ/). Despite the notable confusion between /iː/ and /ɪ/ in the /l/ context, it is smaller than for the other three vowel pairs that are confusable in this context

We used hierarchical cluster analysis to test whether the patterns of confusion correspond to statistically significant contrast reduction between AusE vowels. In the /d/-context, the only cluster that appears with significant frequency, namely in 100% of the bootstrap samples, was the cluster consisting of all diphthongs and monophthong vowels except the three back monophthongs /oː, ʊ, ɔ/ (**Figure 2.7a**). In the /d/-context no vowel pairs are confused with significant frequency; that is no two such vowels are found which are maximally similar to each other and maximally different from the rest **Figure 2.7a**). In the /l/-context, the cluster of /iː, ɪ, oː, oɪ, ɜː, e, æɪ, ɑe/ occurs with significant frequency. In addition, the pairs /iː-ɪ, ʉː-ʊ, əʉ-ɔ, æɔ-æ/ occur significantly frequently in a cluster, indicating that the members of these pairs are maximally similar to each other (**Figure 2.7b**).

Vertical location of the nodes of the dendrograms (**Figure 2.7**) indicates similarity between the clusters: the lower a node is located, the more similar the members of the cluster are. The close vertical alignment of the nodes in **Figure 2.7a** in the /d/ condition shows that members of the vowel dyads are only slightly more similar to each other than to other dyads. For instance, /ʊ/ and /ɔ/ merge into a cluster at approximately 1.4, and the /ʊ-ɔ/ cluster merges with /oː/ at approximately 1.5, indicating that /ʊ/ and /ɔ/ are only a little more similar to each other than the /ʊ-ɔ/cluster is to /oː/ (**Figure 2.7a**). In contrast, in the /l/ condition, members within the vowel pairs /ʉː-ʊ, əʉ-ɔ, æɔ-æ/ are maximally similar to each other, as the nodes of their respective dyads branch at 0.5 (/əʉ-ɔ/) or below (/ʉː-ʊ, æɔ-æ/) (**Figure 2.7b**).

Both random forest analysis and hierarchical cluster analysis indicate that spectral contrast is reduced between the members of the vowel pairs /iː-ɪ, ʉː-ʊ, əʉ-ɔ, æɔ-æ/. In the random forest analysis, the members of these pairs are systematically confused. In the hierarchical cluster analysis, these pairs form significantly frequently recurring dyads which are maximally similar to each other in the pre-lateral vowel space.

/iːl/ and /ɪl/ show a lower confusion rate in the random forest analysis compared to the other three key vowel pairs, and they are also merged later in hierarchical cluster analysis. For the pairs /ʉː-ʊ, əʉ-ɔ, æɔ-æ/, random forest provides more details than hierarchical cluster. Random forest analysis shows that /ʉː/ is primarily confused /ʊ/ and to a lesser extent with /oː/ (/ʉː/ and /ʊ/ are confused in almost 30% of the tokens for both vowels, and /ʊ/ and /oː/ are confused in 4% of the tokens for both vowels). The high confusion rate between /ʉː/and /ʊ/ leads to these vowels forming a dyad in hierarchical cluster analysis, while the smaller confusion rate between /ʉː, ʊ/ and /oː/ is not captured by hierarchical cluster analysis. Similarly, random forest misidentifies /ɔ/ as /ɐ/ (32%) and as /əʉ/ (16%) and misidentifies /æ/ as /ɐː, ɐ/ (6%-6%) and as /æɔ/ (30%). However, in the hierarchical cluster analysis /ɔ/ clusters with /əʉ/, not /ɐ/ due to 52% of /əʉ/-tokens being misidentified as /ɔ/, while /æ/ clusters with /æɔ/, not /ɐː, ɐ/ due to 30% of /æɔ/ tokens being misidentified as /æ/.

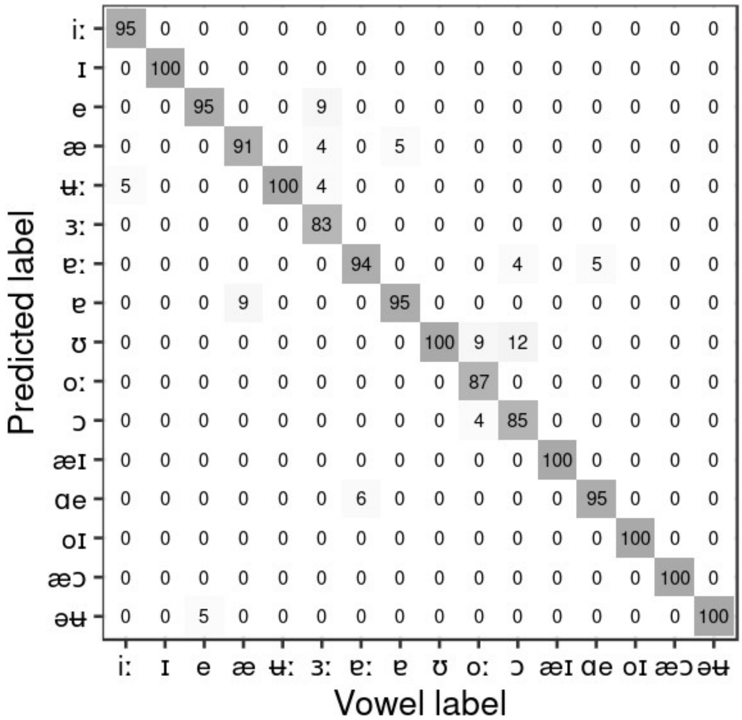


Figure 2.5: Confusion matrix of vowels produced before /d/ codas, based on DCT coefficients ($k0$, $k1$, $k2$) of formants ($F1$, $F2$, $F3$) and mean vowel duration. Columns show the percentage of tokens classified for each vowel target. Rows show the percentage of tokens classified by the random forest classification algorithm as a certain vowel.

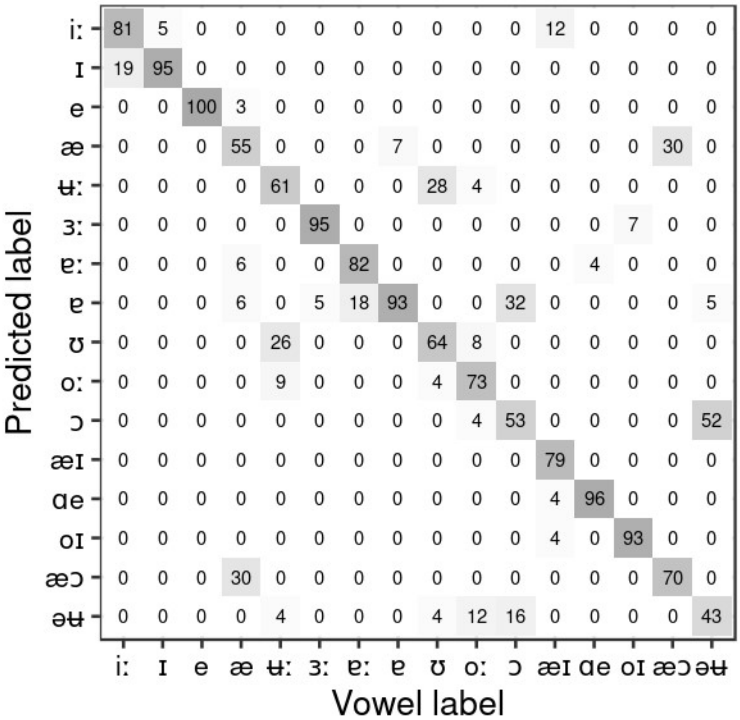
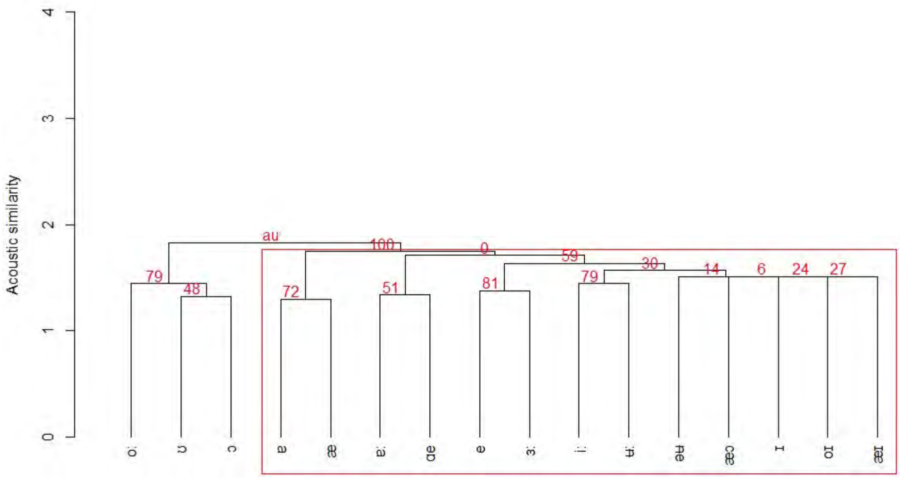
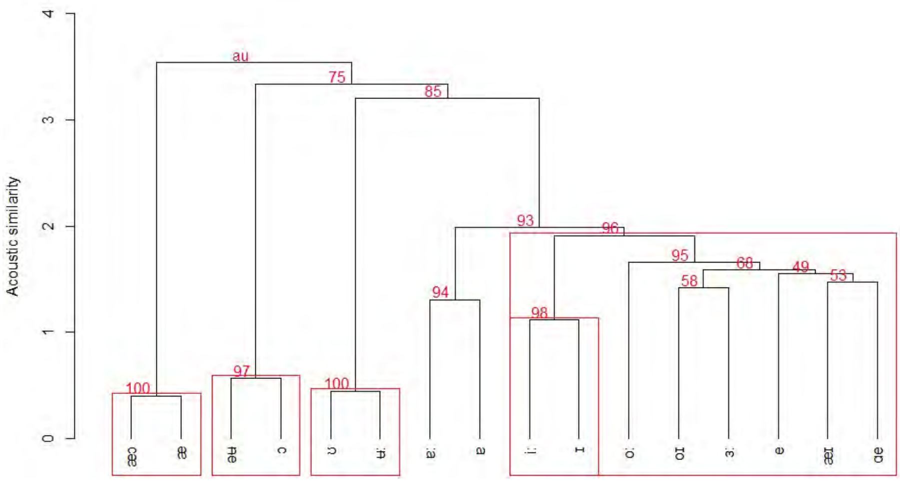


Figure 2.6: Confusion matrix of vowels produced before /l/ codas, based on DCT coefficients ($k0$, $k1$, $k2$) of formants ($F1$, $F2$, $F3$) and mean vowel duration. Columns show the percentage of tokens classified for each vowel target. Rows show the percentage of tokens classified by the random forest classification algorithm as a certain vowel.



(a) Acoustic vowel similarity before /d/ codas, based on vowel confusion.



(b) Acoustic rime similarity in /l/-final rimes based, on rime confusion.

Figure 2.7: Acoustic similarity in /d/- and /l/-final rimes: lower branching signals higher confusion rates. AU (red): Approximately unbiased p-value indicating the frequency with which a cluster appears in multiscale bootstrapping resampling. Red boxes highlights clusters that appear significantly frequently (in more than 95% of the samples) during resampling.

2.4 Summary of Results

1. Effect of coda /l/ on monophthong targets compared to coda /d/:

    (a) All vowels have a higher $F1$ except for /æ, ɐː, ɐ, ʊ/, indicating phonetic lowering before coda /l/;

    (b) All vowels have a lower $F2$ except for /oː/ and /ɔ/, indicating phonetic backing before coda /l/;

    (c) Front vowels /iː, ɪ, e/ have lower $F3$ before coda /l/, while /ʉː, ɐ, ʊ, oː/ have higher $F3$.

2. Spectral contrast reduction:

    (a) Increased out-of-bag error rate in random forest analysis indicates that a higher percentage of vowels were misidentified in the /l/-context than in the /d/-context;

    (b) Random forest analysis indicates that confusion of pre-/l/ vowels is pairwise and systematic; no such pattern is observed for pre-/d/ vowels;

    (c) Hierarchical cluster analysis shows that the members of the vowel-pairs /iː-ɪ, ʉː-ʊ, əʉ-ɔ, æɔ-æ/ are maximally similar to each other in the /l/-context; no such pairings were found among the pre-/d/ vowels.

2.5 Discussion

2.5.1 Acoustic patterns

F1 raising in monophthongs

Hypothesis 1) predicted that front vowels would have a higher $F1$, that is, they would be phonetically lowered in pre-lateral position compared to pre-/d/ position. Hypothesis 1) largely holds, as we found increased $F1$ for all front vowels (/iː, ɪ, e/) except for front /æ/. In addition, most back vowels were also found to be significantly lowered in pre-lateral contexts. The biggest lowering effect can be observed in /e/, whose target distribution shifts toward /æ/, similar to shifts observed in Melbourne/Victoria dialects (Cox & Palethorpe 2004, Loakes et al. 2010a). However, random forest and hierarchical cluster analysis did not classify /e/ as similar or confusable with /æ/ in the /l/-context, most probably due to the lack of overlap between /e/ and /æ/.

The only front vowel that did not lower before laterals was /æ/, which can potentially be explained by its already high $F1$ in the /d/ condition. The low vowels /ɐː/ and /ɐ/ did not lower either, similar to the observation of Bernard (1985) and Palethorpe & Cox (2003). The lack of phonetic lowering in /æ, ɐː, ɐ/ indicates that /æ/ might pattern with the phonologically low vowels due to its high $F1$. This pattern appears again as pre-/d/ /æ/ and /ɐ/ are classified as similar (**Figure 2.7a**).

F2 lowering in monophthongs

Hypothesis 2) predicted that front vowels would have a lower $F2$, that is, they would be phonetically backed in pre-lateral position compared to pre-

/d/ position. Hypothesis 2) holds, as we found decreased $F2$ for all front vowels before coda /l/, compared to coda /d/. In addition, back and low vowels were also phonetically backed except for /oː/ and /ɔ/.

The greatest backing effect was observed for /ʉː/, whose target $F2$ is on average 1214 Hz lower before coda /l/ than before coda /d/. As a result, /ʉː/ overlaps acoustically with /ʊ/ in the /l/-context, unlike in the pre-/d/ context, where it acoustically approaches /ɪ/ (**Figures** 2.3a and 2.3b). The backing influence of the lateral on /ʉː/ is further corroborated in the analysis of spectral similarity: in the /l/-context /ʉː/ shows similarity to /ʊ/ and to a lesser extent to long back /oː/. In contrast, in the /d/-context /ʉː/ shows some similarity to front /iː/ and central /ɜː/. The fact that /ʉː/ shows similarity to /iː/ and not to /ɪ/, even though the latter is acoustically closer to /ʉː/ in the $F1$-$F2$ vowel space, is due to the fact that the presented analysis of spectral similarity considers vowel length when classifying vowels. Therefore, in the /d/-context, long vowels are clustered with long vowels, but in the /l/-context long-short vowel pairs cluster together due to the reduction of the duration contrast.

In addition, we found that /e/ partially overlaps acoustically with /ɜː/ in the pre-lateral environment due to the lowering of its $F2$. However, we did not find spectral contrast reduction between /e/ and /ɜː/.

Acoustic contrast reduction

Hypothesis 3) predicted that acoustic contrast would be reduced between these vowel pairs. Analysis of spectral similarity shows that acoustic vowel contrast is reduced between the members of the vowel pairs /iː-ɪ, ʉː-ʊ, əʉ-ɔ, æɔ-æ/, as the members of each pair are maximally similar to each other.

/iːl/ and /ɪl/ are spectrally more similar to each other than to any other vowel; however, the extent of spectral similarity is smaller between the members of the pair /iː-ɪ/ than the members of the pairs /ʉː-ʊ, əʉ-ɔ, æɔ-æ/. These tentative results are best explained by the fact that both vowels are backed and lowered to a similar extent, with /iː/ remaining more peripheral than /ɪ/ (**Figures** 2.3 and 2.4). In addition, /iː/ and /ɪ/ might be differentiated by the presence of onglide in /iː/ (Cox et al. 2014). The high front target is followed by a steep $F2$ transition to /l/ (**Figure 2.8**).

Increased spectral similarity between /ʉːl/ and /ʊl/ is attributed to the $F2$ drop in /ʉː/ throughout the entire vowel, which makes high central /ʉː/ similar to high back /ʊ/ (**Figures** 2.4, 2.8). Not only is the vowel target backed, (see **Section 2.5.1**, **Figures** 2.3b and 2.4), but the entire $F2$ trajectory is low across the rime (**Figure 2.8**).

Increased spectral similarity between /əʉl/ and /ɔl/ is best explained by the diphthong's lowering and backing of the first target and the loss of the high central second target, shown by the overall lower $F2$ trajectory (**Figure 2.8**). As the high central second target of the diphthong is backed, it becomes similar to mid-back /ɔ/ (Figure 2.8). In contrast, $F2$ of /əʉ/ in the /d/-context shows a higher first target followed by a steep rise as it transitions from the schwa target to the [ʉː] target.

Increased spectral similarity between /æɔl-æl/ is best explained by the fact that the $F2$ trajectory of /æl/ becomes similar to that of /æɔl/ (**Figure 2.8**). /æɔ/ has a falling $F2$ both in the /d/- and in the /l/-context , as is expected in both conditions as the diphthong in the /d/-context and the rime in the /l/-context contains a transition from a high $F2$ to a low $F2$. In contrast, /æ/ has a rising $F2$ in the /d/-context due to the vowel-alveolar transition (Delattre et al. 1955), whereas /æ/ has a falling $F2$ in the /l/-

context due to the vowel-/l/ transition, making the *F*2 trajectory more similar to /æɔl/ (**Figure 2.8**).

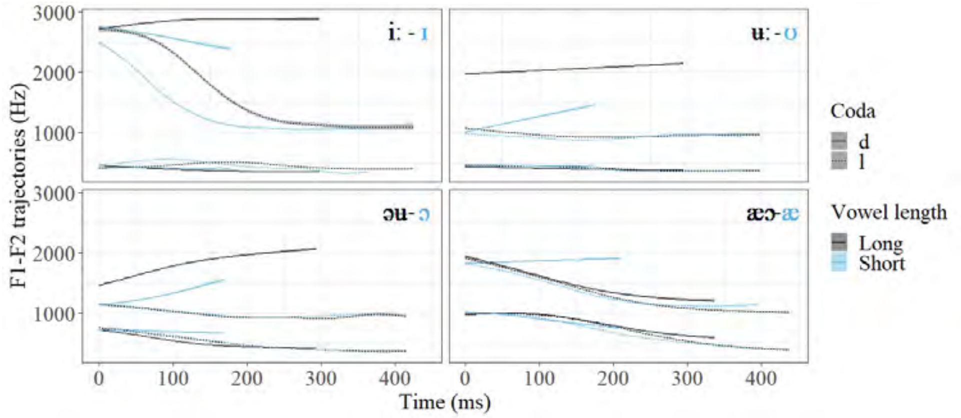


Figure 2.8: Mean *F*1 and *F*2 trajectories by coda context (solid: /d/, dashed: /l/) and vowel pair. Top: /iː-ɪ/ and /ʉː-ʊ/. Bottom: /əʉ-ɔ/ and /æɔ-æ/. Black: long vowel. Blue: short vowel.

The vowel pairs /iː-ɪ, ʉː-ʊ, əʉ-ɔ, æɔ-æ/ also contrast in terms of length. In the /d/-context, short vowel duration is 59% of long vowel duration, in line with Cox (2006), and mean duration of rimes with short vowels is 79% of the duration of rimes containing long vowels (**Tables** 2.5a-2.5b). In contrast, /l/-final rimes containing short vowels are 91% of /l/-final rimes containing long vowels (**Table 2.5c**). Reduced duration contrast in the /l/-context further increases similarity between key long-short vowel pairs, whereas the larger duration contrast in the /d/-context results in vowels being classified according to length (**Figure 2.7a**). However, duration contrast reduction between the /d/- and the /l/-context cannot be assessed without separating the vowel from the following liquid for which we have found no reliable method.

In contrast to the pairwise similarity of long-short vowels in the /l/-context, no two vowels show spectral similarity in the /d/-condition. The

Table 2.5: Duration contrast reduction from pre-/d/ long and short vowels to /l/-final rimes containing long and short vowels

(a) Duration (ms) and duration contrast between long and short vowels in the /d/-context

Vowel pair	long	short	Short:long
/iː-ɪ/	298	177	0.59
/ʉː-ʊ/	295	175	0.59
/əʉ-ɔ/	294	169	0.57
/æɔ-æ/	337	209	0.62
Mean	306	183	59

(b) Duration (ms) and duration contrast between /d/-final rimes containing long and short vowels

Vowel pair	Long	Short	Short:long
/iː-ɪ/	397	317	0.80
/ʉː-ʊ/	398	316	0.80
/æɔ-æ/	429	348	0.81
/əʉ-ɔ/	415	320	0.77
Mean	409	325	79

(c) Duration (ms) and duration contrast between /l/-final rimes containing long and short vowels

Vowel pair	Long	Short	Short:long
/iː-ɪ/	424	365	0.86
/ʉː-ʊ/	396	375	0.95
/æɔ-æ/	438	395	0.90
/əʉ-ɔ/	415	393	0.95
Mean	418	382	91

flat dendrogram in **Figure 2.7a** shows that vowel similarity within members of the cluster and between members of separate clusters is comparable.

Increased spectral similarity in the /l/-condition compared to the /d/ condition could be due to the fact that formant trajectories were measured in the rime, and thus all include /l/. However, if the overlap in the coda were the main cause of the increased confusion rates, all rimes would be confused to the same extent. That is, the dendrogram would be flat in the /l/-context, similar to that of the /d/-context, and it would not show the pairwise similarity of key vowel pairs.

The acoustic targets and the durations of pre-/d/ vowels in the current study are consistent with standard description of AusE (Cox 1999, Cox & Palethorpe 2001, Cox 2006). In addition, the pairing of /ʉː/ with /iː/ in the cluster analysis of the /d/-condition is in line with the fronting of the AusE /ʉː/ Harrington et al. (1997), Cox (1999), Cox & Palethorpe (2001), Elvin et al. (2016). Our results confirm the increased acoustic similarity between /iː-ɪ, ʉː-ʊ, əʉ-ɔ, æɔ-æ/ in the pre-lateral context noted by Palethorpe & Cox (2003).

2.5.2 Articulatory explanations

The phonetic backing and lowering of pre-lateral vowels can be attributed to the coarticulatory influence of the dorsal gesture of /l/ on the preceding vowel, as has been reported for American English (Giles & Moll 1975, Sproat & Fujimura 1993, Gick et al. 2002, Gick & Wilson 2006). In American English, tongue dorsum lowering and retraction typically precedes coronal articulation in coda laterals, and may overlap with the vowel Giles & Moll (1975), Sproat & Fujimura (1993), Proctor et al. (2019). The overall increase in $F1$ and overall decrease in $F2$ observed in AusE pre-lateral vowels is consistent with a pattern of production in which the lowered and retracted tongue dorsum gesture of coda /l/ coarticulates with the vowel gesture (Fant 1960). In particular, the phonetic backing of /ʉː/ observed here is consistent with the articulatory backing of this vowel observed in previous work for AusE (Lin et al. 2012). The backed tongue position in the production of pre-lateral /ʉː/ might make it articulatorily similar to /ʊ/.

The reduction in acoustic contrast between /æɔ-æ/ before laterals in the Australian English data is also consistent with the articulatory characterization of the dorsal gesture associated with American English laterals: an

MRI study of [ɫ] and /ɔ/ reported articulatory similarities between the dorsal gestures of [ɫ] and /ɔ/ (Gick et al. 2002). As a result, the monophthong /æ/ followed by an /ɔ/-like /l/ can be spectrally similar to the diphthong /æɔ/, whereas the second target of the diphthong /æɔ/ might be encroached upon by the following /ɔ/-like /l/.

Articulatory similarity between /ɔ/ and /l/ can potentially also play a role in the loss of the second target of /əʉː/, as the backed [ʉ] can be similar to /ɔ/ and therefore to /l/, leading to the loss of contrast between the second target of the diphthong and /l/. This account is consistent with the articulatory backing of the second target of /əʉ/ in the pre-/l/ context (Lin et al. 2012).

When coda /l/ is preceded by a high front vowel, the vowel and /l/ place competing demands on the tongue dorsum: the vowel target requires a raised and fronted tongue dorsum whereas the /l/ target requires it to be lowered and backed (Gick & Wilson 2006). These competing demands result in a long transition between the two segments during which the tongue passes through a schwa-like posture (Gick & Wilson 2006). Our acoustic data from Australian English are consistent with these articulatory accounts of American English, as /iː/ and /ɪ/ exhibited a relatively front target followed by a long steep $F2$ fall to reach the /l/ target.

2.5.3 Implications for sound change: pre-lateral vowel mergers?

A vowel merger is defined as the loss of contrast between two or more categories due to the loss of phonetic differentiation either across the board or in a particular phonological environment (Maguire et al. 2013). In Harrington et al.'s (2018) Interactive Phonetic model of sound change, the prerequisite of sound change is that typical realisations of two phonemes are acoustically distinct, but their highly coarticulated realisations become acoustically similar to each other. As listeners and speakers interact, atypical speaker realisations are incorporated to the listener's phoneme representation, shifting its boundary closer to the second phoneme until the categories overlap, potentially leading to a merger (Harrington et al. 2018).

Acoustic contrast reduction within the pairs /ʉː-ʊ, əʉ-ɔ, æɔ-æ/ in pre-lateral environments is consistent with the Interactive Phonetic model of sound change and with a contextual vowel merger conditioned by coda /l/. Vowel-lateral coarticulation creates atypical realisations for these vowels, shifting their boundaries closer to each other and leading to overlap. This is best exemplified by the vowel /ʉː/: /ʉː/ moves into the vowel plane of /ʊ/ (**Figure 2.3b**), making pre-lateral /ʉː/ a potential candidate for a vowel merger with pre-lateral /ʊ/ in the New South Wales dialect of AusE.

While our analysis of spectral similarity indicates that contrast is reduced even considering dynamic $F1$, $F2$, $F3$, and duration information, our methods cannot show whether the phonemes are differentiated: both Random Forest and Hierarchical Cluster Analysis classified the tokens into pre-defined 16 vowel categories. Increased similarity between categories is consistent both with a merger and with reduced acoustic contrast. To explore whether the phonemes undergo a conditional merger in the pre-lateral environment, an apparent time or a sociolinguistic study is needed to better understand the implications for the actuation of sound change in key pre-lateral vowels of Australian English.

2.6 Conclusion

In Australian English, $F1$ is increased and $F2$ is decreased in the acoustic target of prelateral vowels compared to coda /d/, indicating phonetic lowering and retraction of pre-lateral vowels. In addition, spectral and durational contrast is reduced within the pairs /iːl-ɪl, ʉːl-ʊl, əʉl-ɔl/ and /æɔl-æl/ (e.g. *fool-full, role-roll, howl-Hal*). Spectral contrast reduction is potentially caused by the coarticulatory effect of the dorsal gesture of /l/ reported in other varieties of English. The observed spectral contrast reduction may reflect necessary conditions for conditional vowel mergers in the pre-lateral environment.

2.7 Acknowledgement

This research was supported in part by iMQRTP 2015144, ARC DE150100318, and MQSIS 9201501719 grants. We thank Peter Humburg for his help in the statistical analysis, and the members of the Phonetics Lab, the Centre for Language Sciences and the ARC Centre of Excellence in Cognition and its Disorders at Macquarie University for their comments, feedback and support.

2.8 Appendix

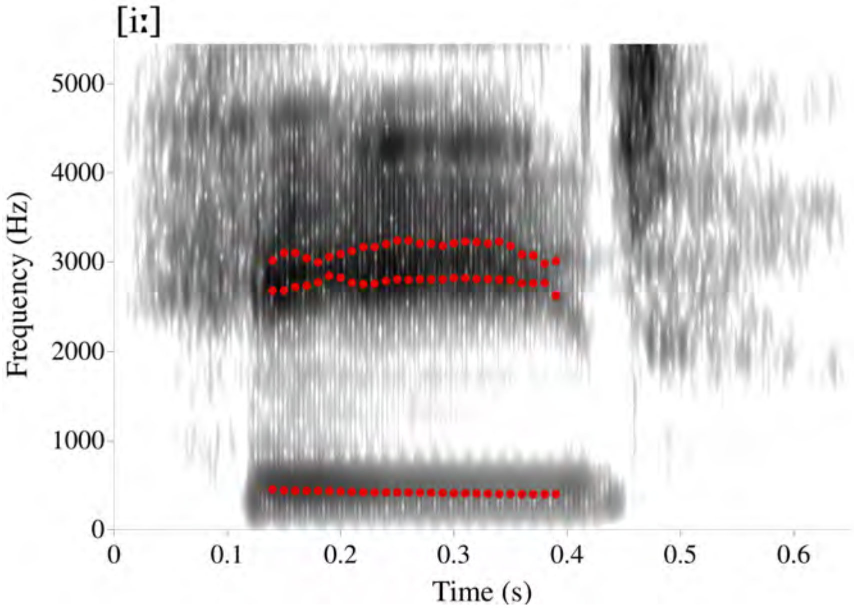
2.8.1 DCT coefficients characterising /d/- and /l/-final rimes

Table 2.6: DCT coefficients (k_0, k_1, k_2) of formants ($F1$, $F2$, $F3$) in vowels produced before coda /d/ and in /l/-final rimes.

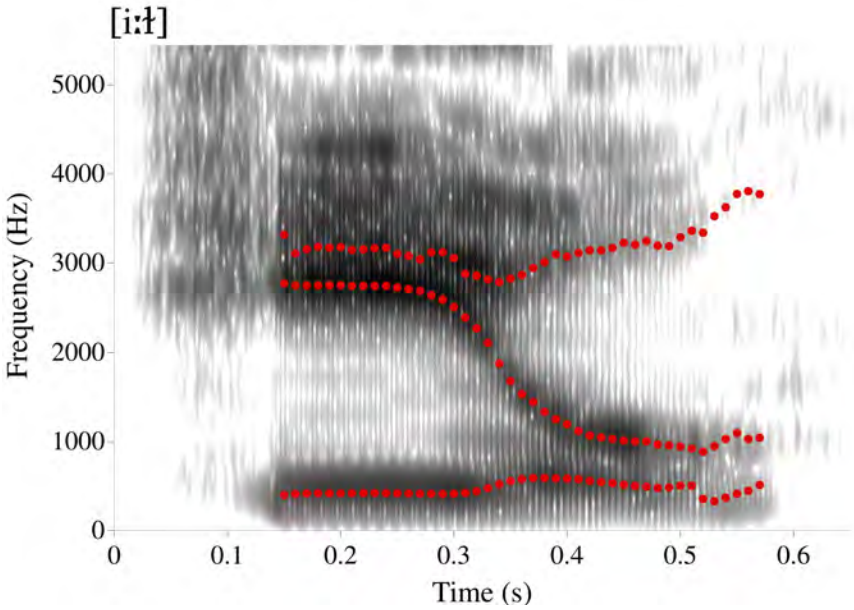
Coda	Vowel	$F1$			$F2$			$F3$		
		k_0	k_1	k_2	k_0	k_1	k_2	k_0	k_1	k_2
d	iː	563	41	10	4016	-59	-47	4634	-39	-20
	ɪ	593	-6	-3	3786	81	-25	4538	43	6
	e	915	31	-8	3274	30	-8	4440	-9	3
	æ	1330	88	-30	2629	-41	26	4287	-45	19
	ʉː	580	25	-1	2887	-59	3	3717	-16	32
	ɜː	889	21	-13	2574	-41	11	4096	-28	15
	ɐː	1265	42	-18	2035	-89	68	4302	-27	-11
	ɐ	1238	55	-18	2193	-104	27	4260	-26	2
	ʊ	649	-4	-4	1635	-166	53	4065	44	-14
	oː	730	-15	9	1374	-113	88	4280	34	-55
	ɔ	1002	16	-7	1723	-104	44	4201	27	-19
	æɪ	812	206	47	3579	-262	-65	4414	-54	-18
	ɑe	1251	95	-79	2398	-438	99	4224	-18	53
	oɪ	745	74	-41	2545	-867	-9	4155	-25	101
	æɔ	1257	155	-78	2192	341	15	4199	-56	31
	əʉ	780	136	21	2507	-243	-30	3743	52	28
	ɪ	677	73	-62	2104	624	267	4523	-236	32
	e	851	203	-32	1907	423	155	4622	-246	2
	æ	1087	295	-70	1980	366	36	4552	-276	22
	ʉː	590	48	-16	1370	43	55	4430	-186	10
	ɜː	835	130	-73	1925	332	14	4367	-308	80
	ɐː	1060	233	-76	1740	138	-25	4568	-226	37
	ɐ	897	252	6	1663	150	47	4683	-173	-51
	ʊ	587	63	-10	1316	5	52	4535	-126	-12
	oː	697	77	-31	1308	-4	32	4655	-163	19
	ɔ	778	194	5	1433	86	60	4634	-193	-24
	æɪ	826	191	-1	2393	698	-64	4387	-144	133
	ɑe	1075	221	-45	2124	147	-339	4304	-172	93
	oɪ	744	94	-30	2017	299	-289	4270	-244	141
	æɔ	1099	286	-81	1960	410	72	4451	-288	94
	əʉ	748	177	18	1398	86	64	4580	-201	3

2.8.2 Formant trajectories characterising /d/- and /l/-final rimes

/hVl/ and /hVd/ formant trajectories exemplified by Speaker 187. To select a typical speaker, nine GLMs with the dependent variables k_0, k_1, k_2 of $F1$, $F2$, and $F3$, and the independent variables Vowel (sum-coded) and Coda (treatment coded, comparing /l/ to the baseline /d/); the model included a Vowel-Coda interaction, a random by-participant and by-repetition effect on the intercept. By-participant random effects were extracted from the model, and a speaker who consistently showed a small random effect across all nine models was selected. Top: /hVd/ target words. Bottom: /hVl/ target words.

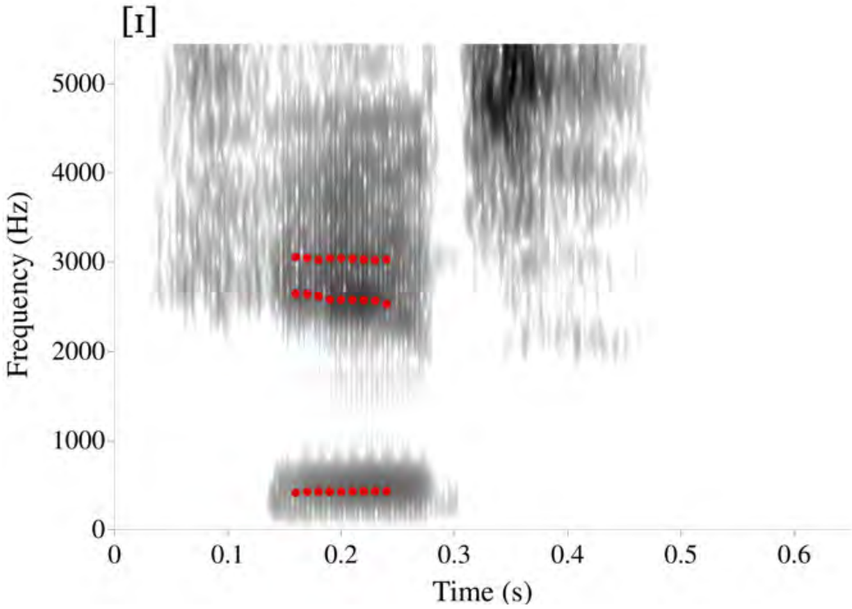


(a) Target: *heed*

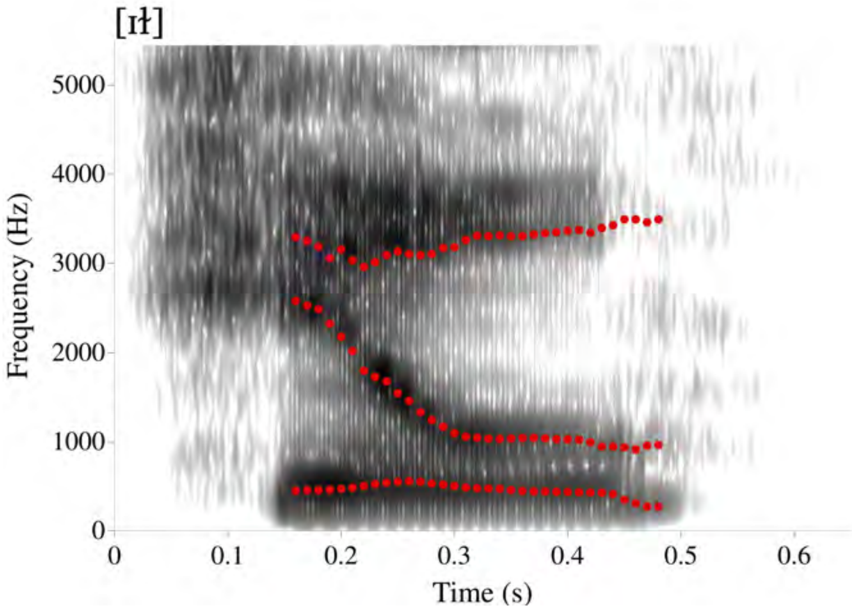


(b) Target: *heel*

Figure 2.9: Vowel /iː/

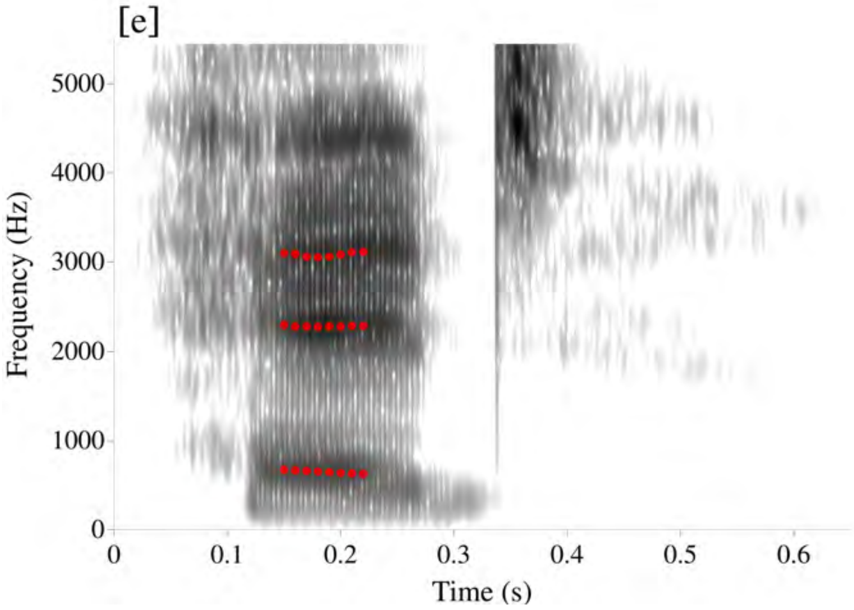


(a) Target: *hid*

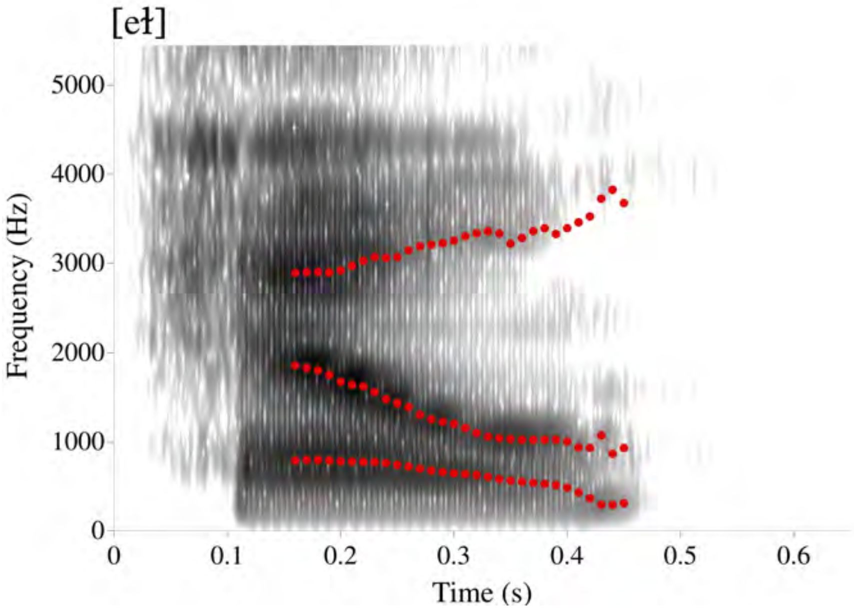


(b) Target: *hill*

Figure 2.10: Vowel /ɪ/

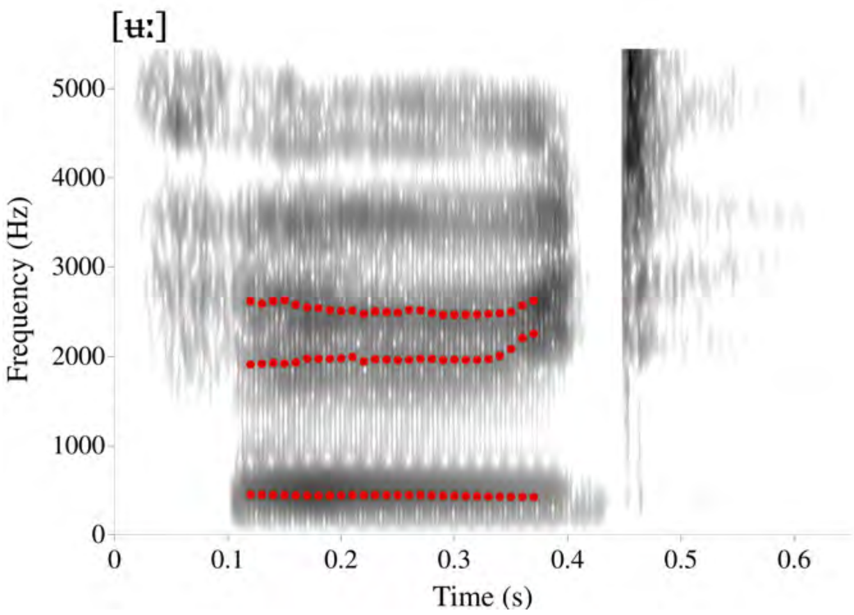


(a) Target: *head*

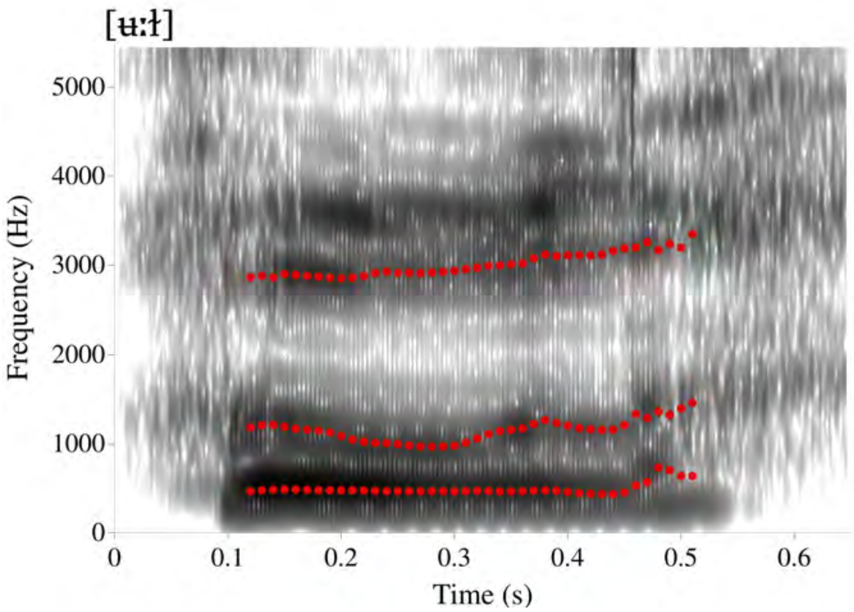


(b) Target: *hell*

Figure 2.11: Vowel /e/

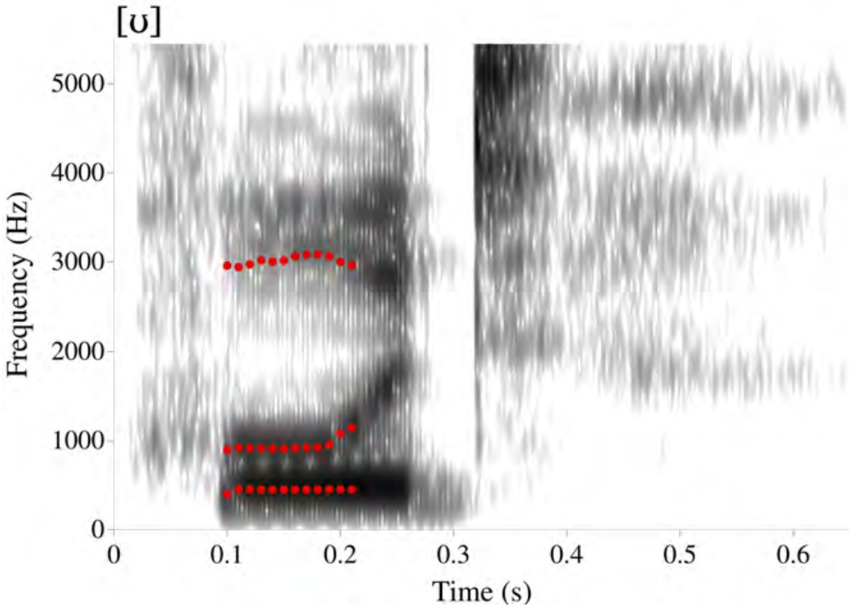


(a) Target: *who'll*

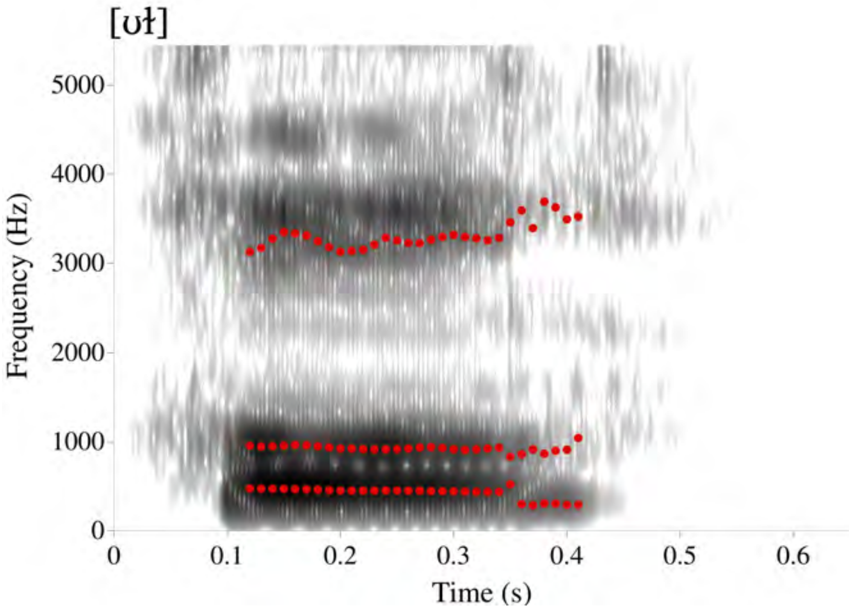


(b) Target: *hule*, Repetition: 1

Figure 2.12: Vowel /ʉː/

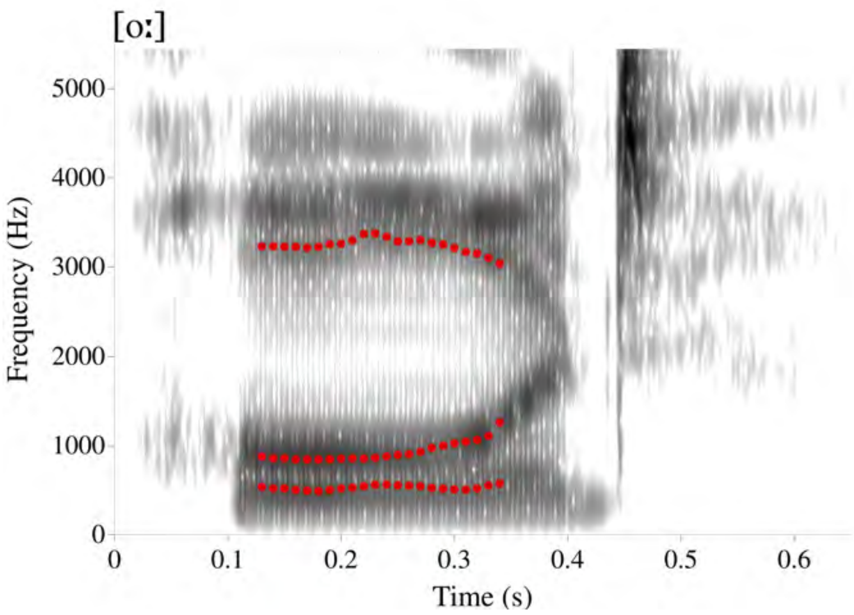


(a) Target: *hood*

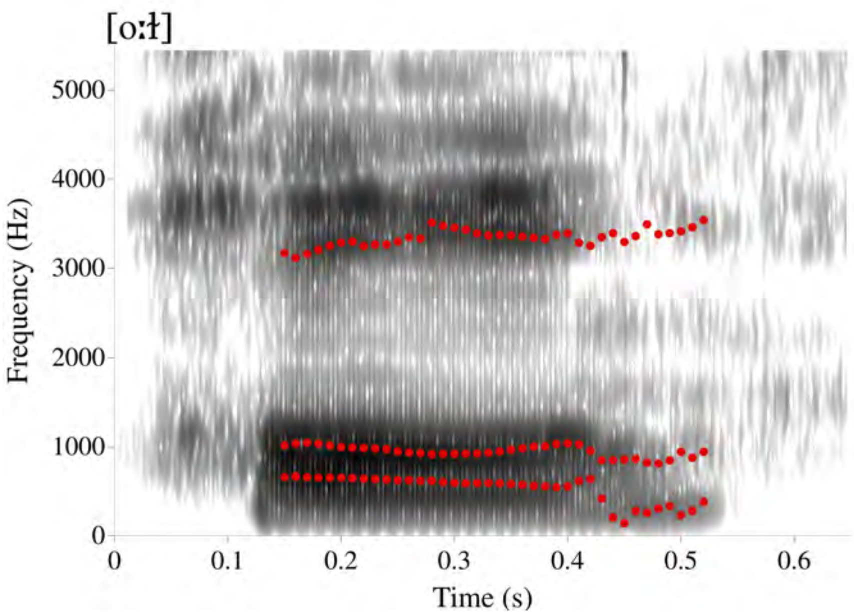


(b) Target: *hooll*

Figure 2.13: Vowel /ʊ/

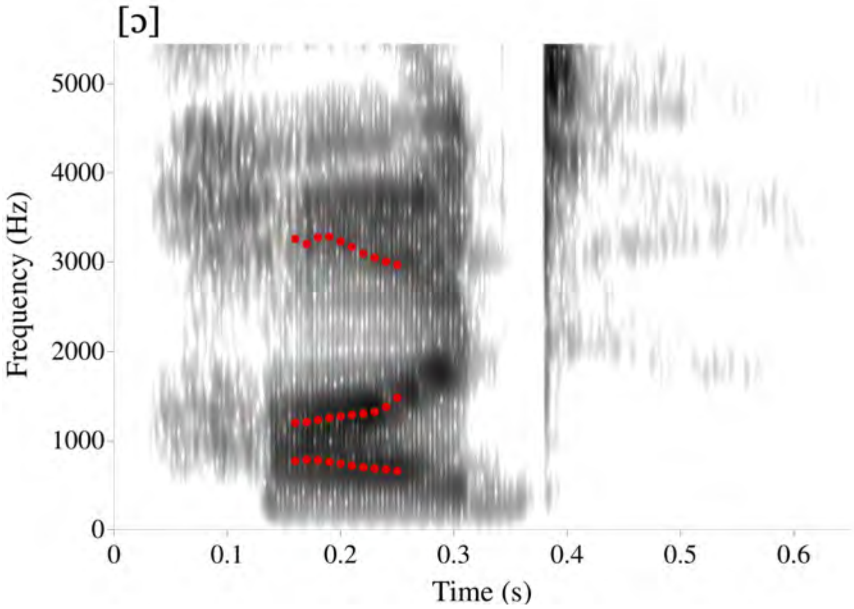


(a) Target: *horde*

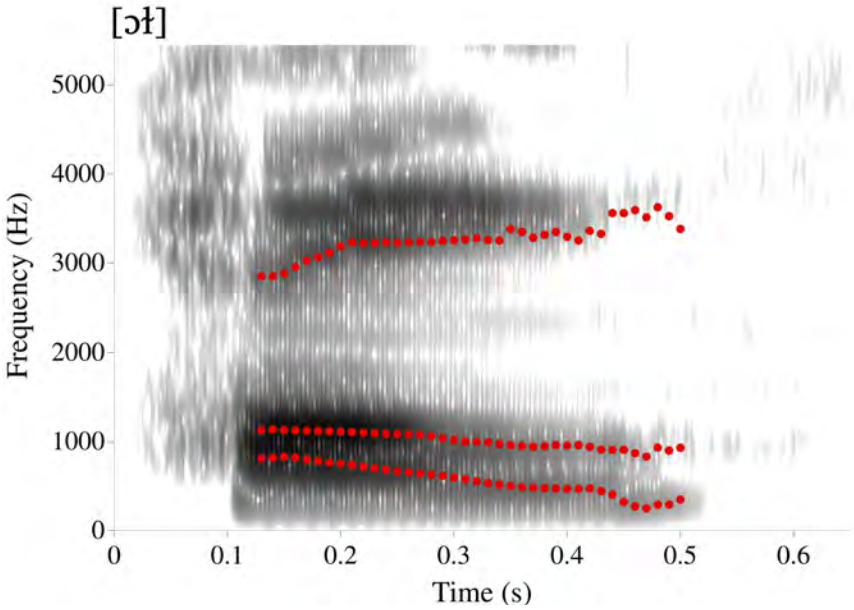


(b) Target: *hall*

Figure 2.14: Vowel /oː/

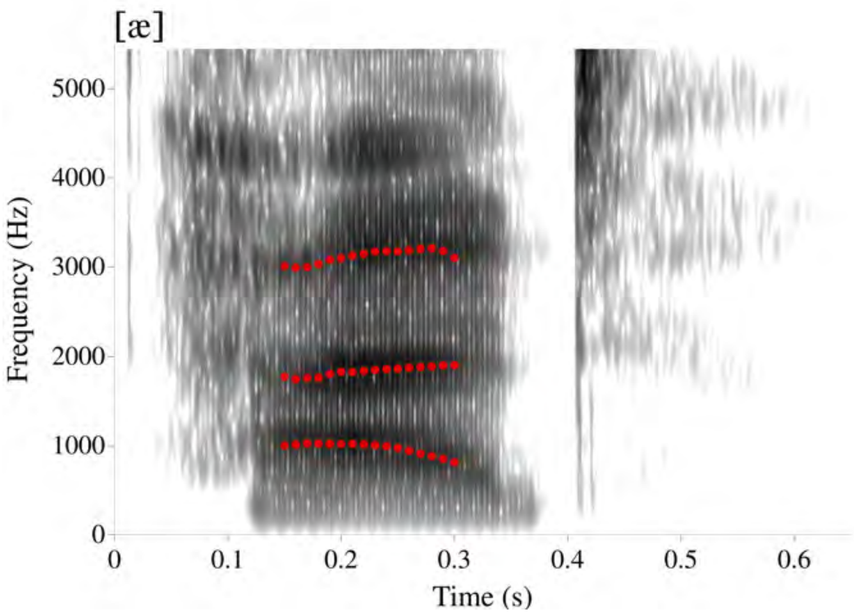


(a) Target: *hod*

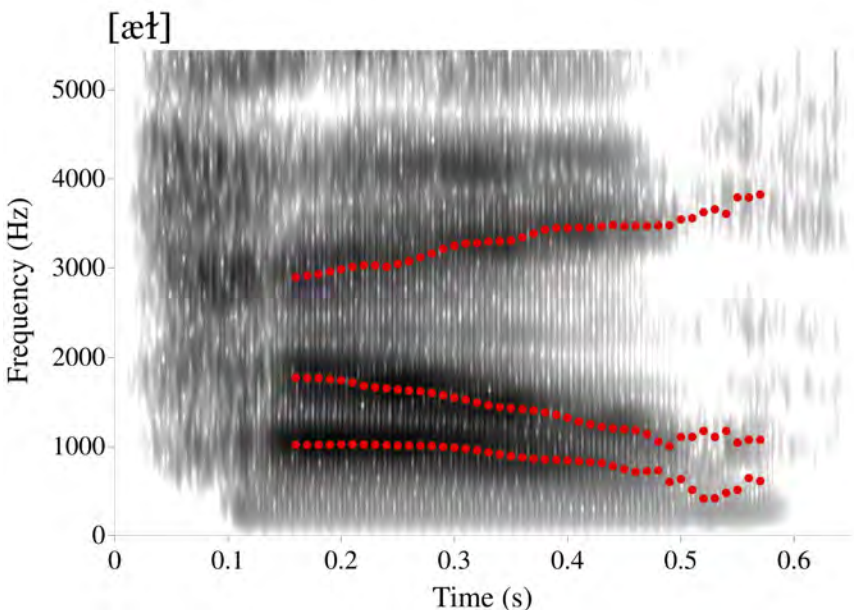


(b) Target: *holl*

Figure 2.15: Vowel /ɔ/

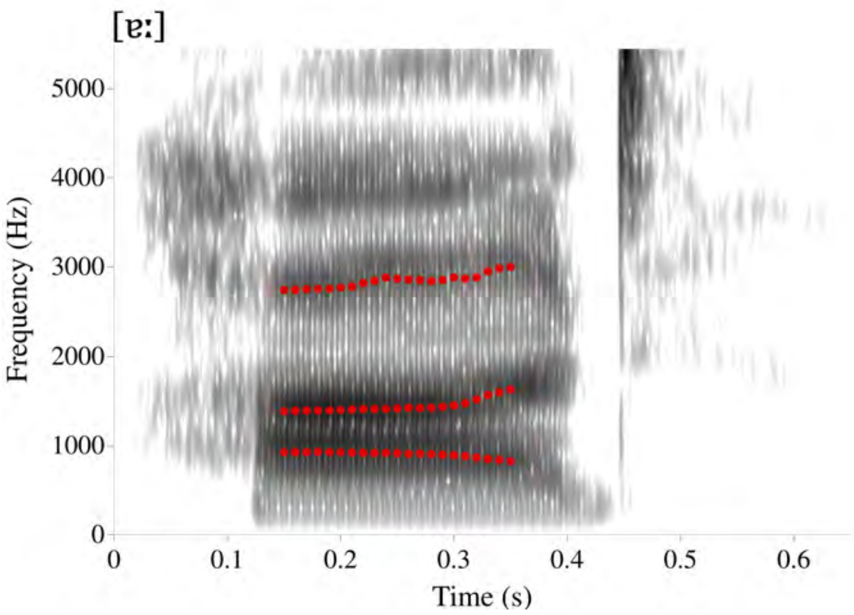


(a) Target: *had*

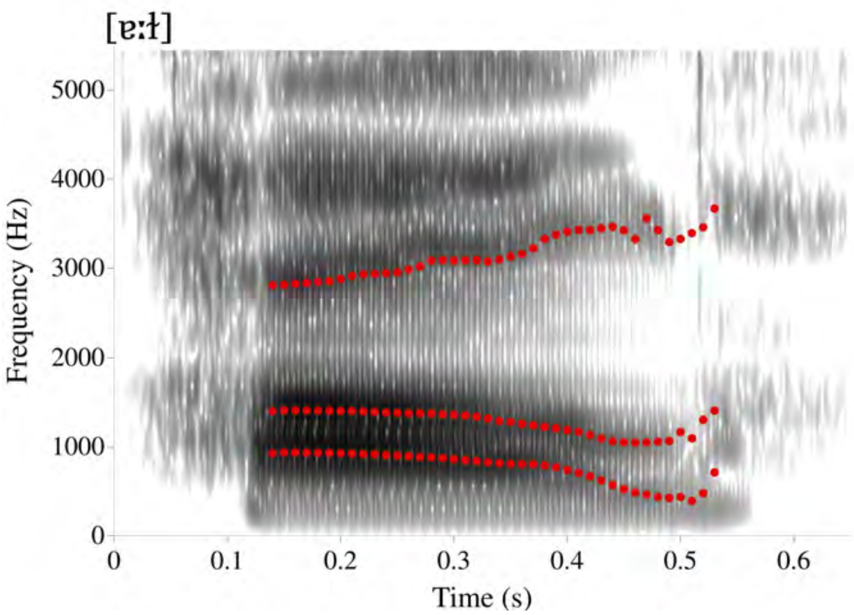


(b) Target: *Hal*

Figure 2.16: Vowel /æ/

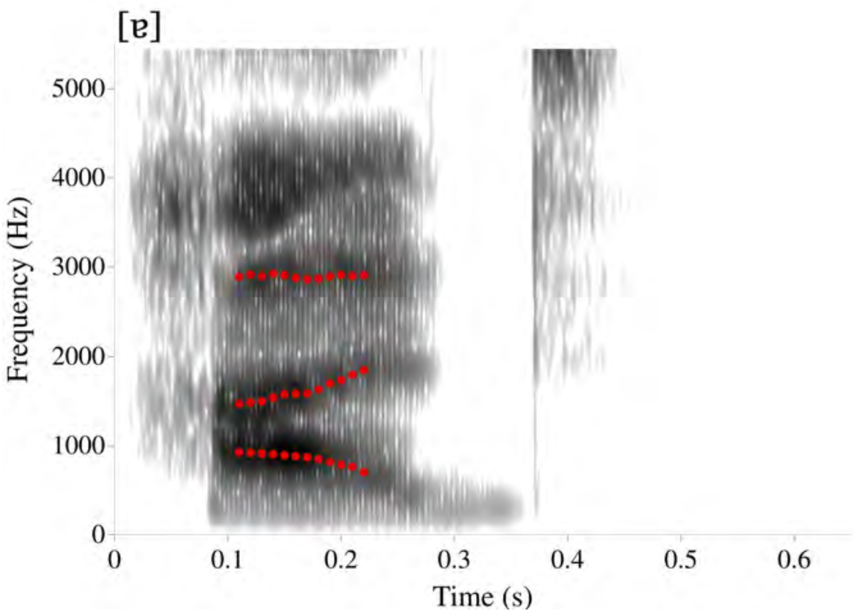


(a) Target: *hard*

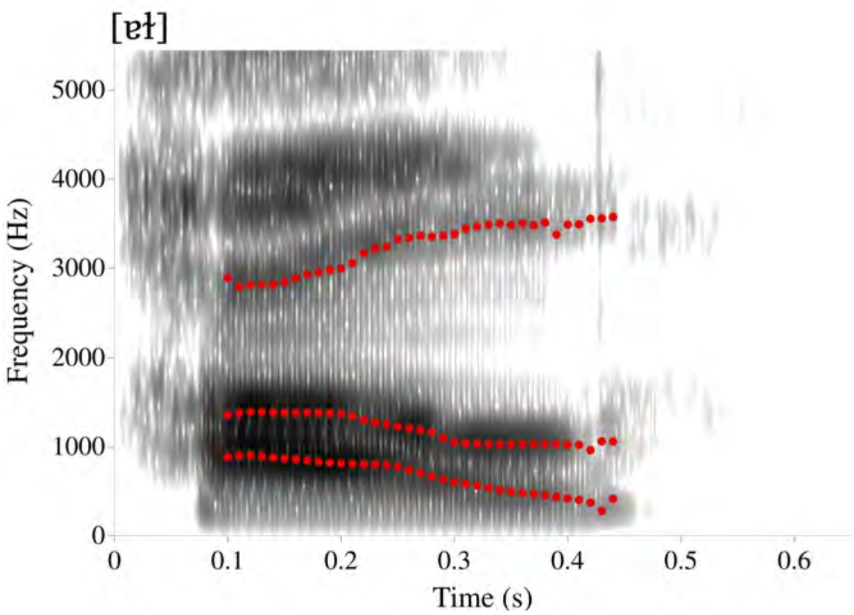


(b) Target: *harl*

Figure 2.17: Vowel /ɐː/

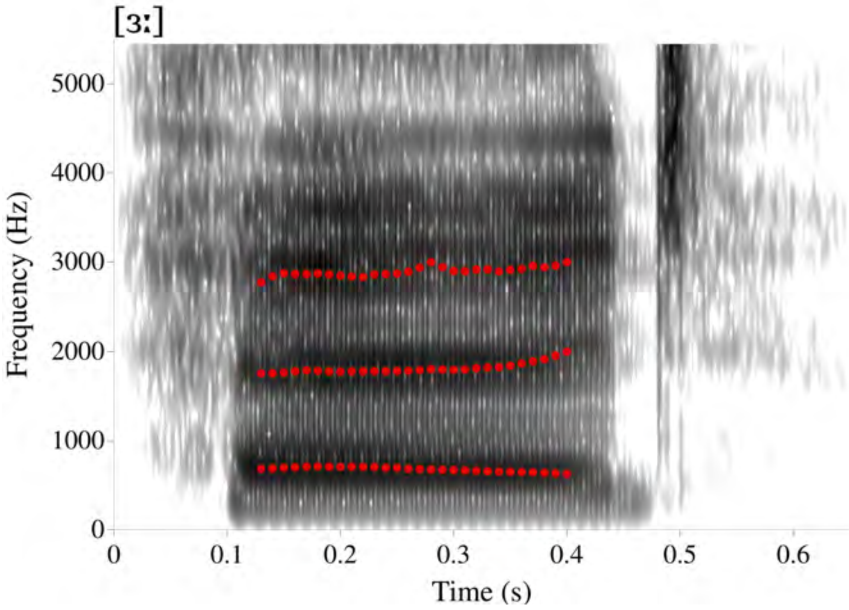


(a) Target: *hud*

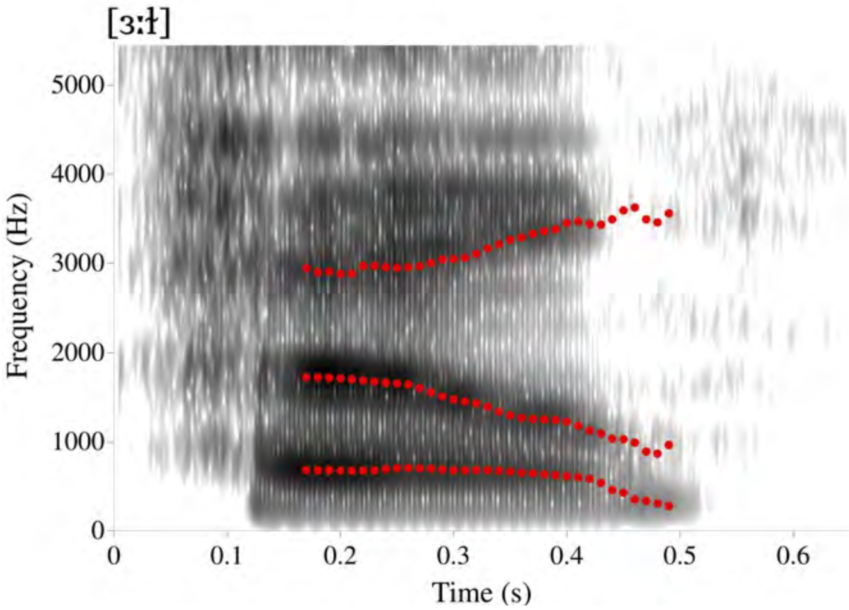


(b) Target: *hull*

Figure 2.18: Vowel /ɐ/

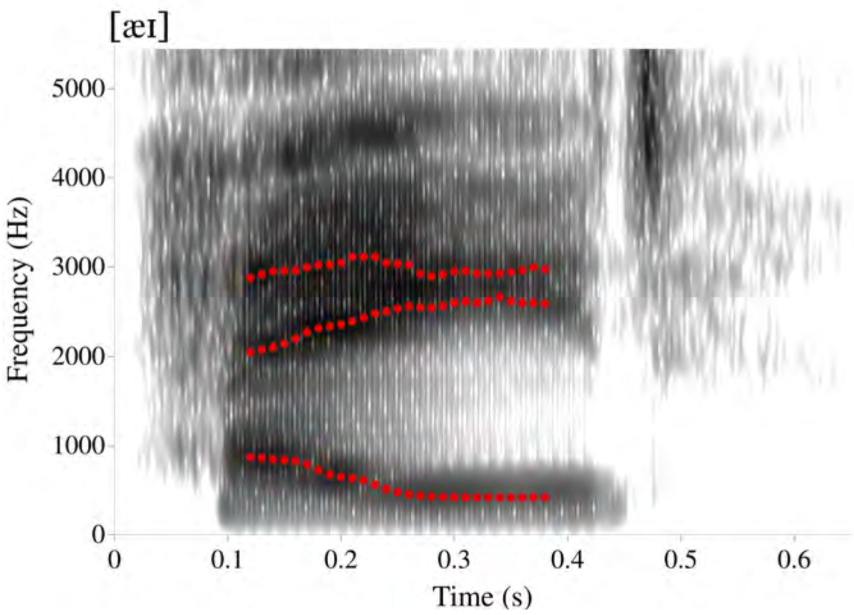


(a) Target: *herd*

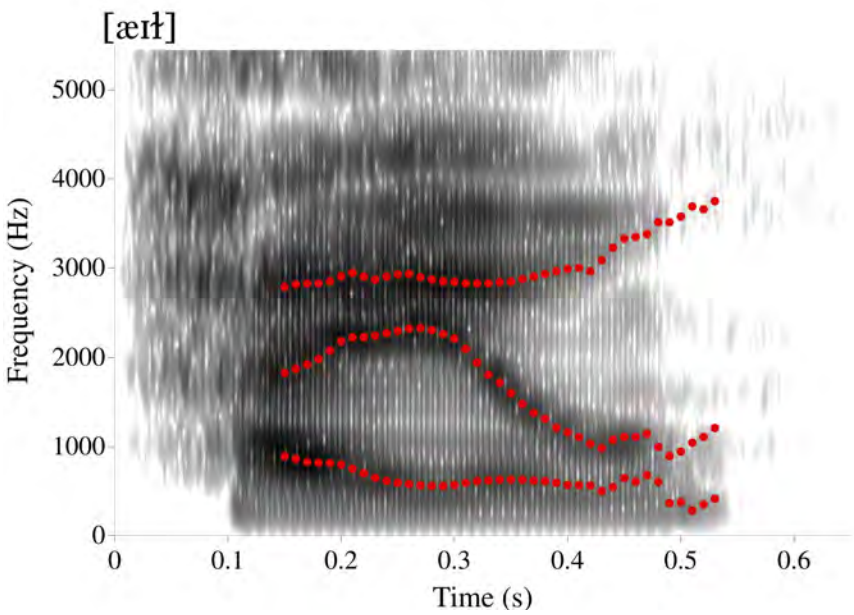


(b) Target: *herl*

Figure 2.19: Vowel /ɜː/

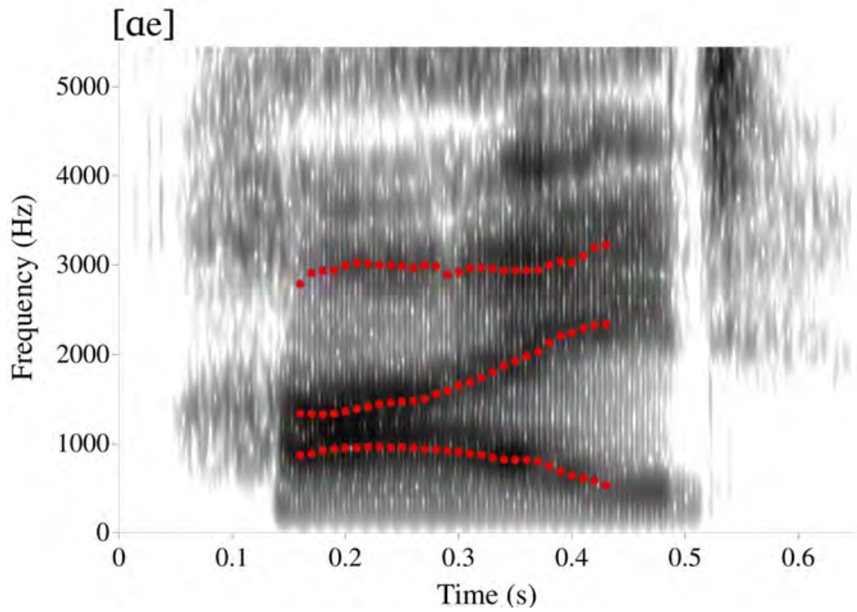


(a) Target: *hade*

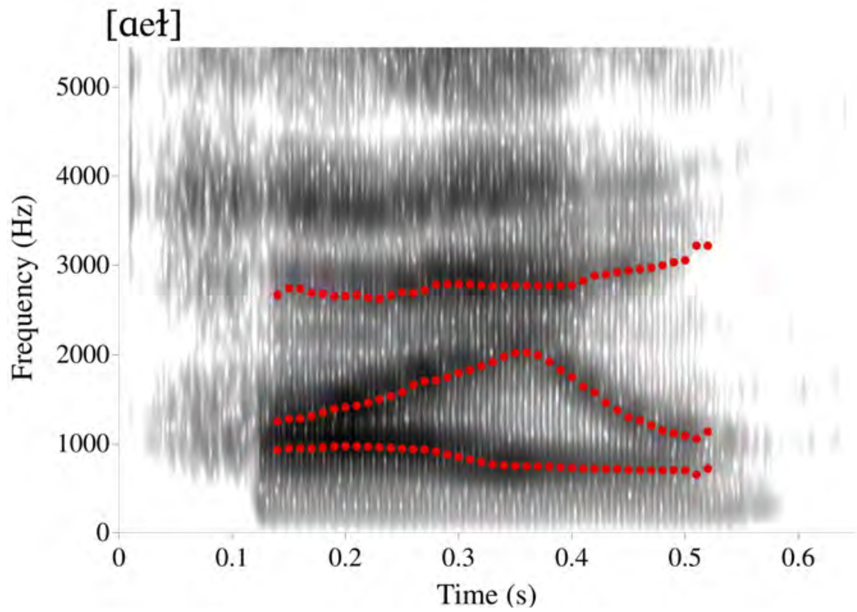


(b) Target: *hail*

Figure 2.20: Vowel /æɪ/

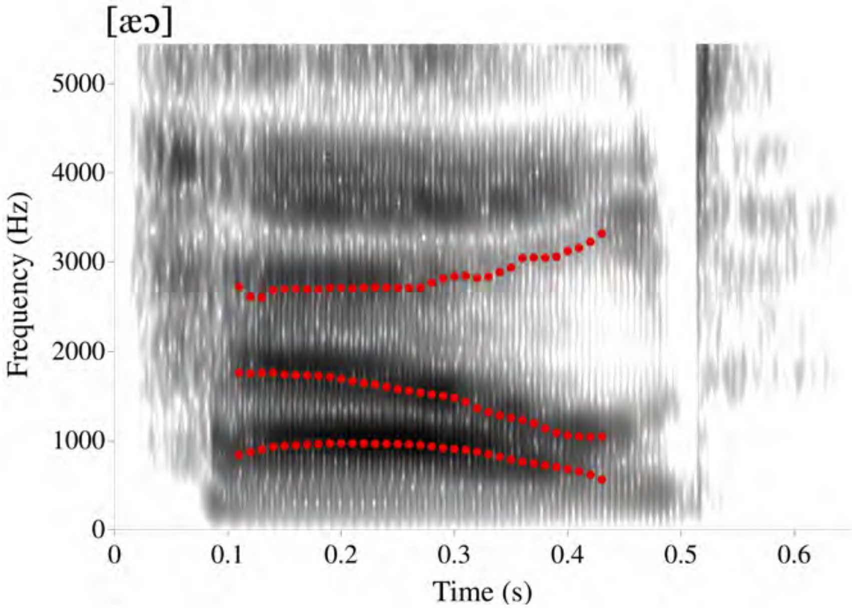


(a) Target: *hide*

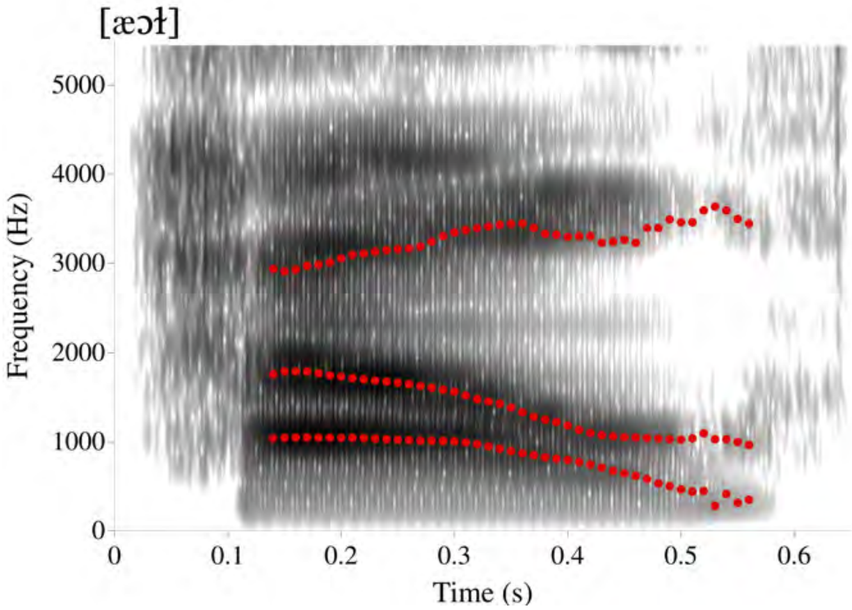


(b) Target: *hile*

Figure 2.21: Vowel /ɑe/

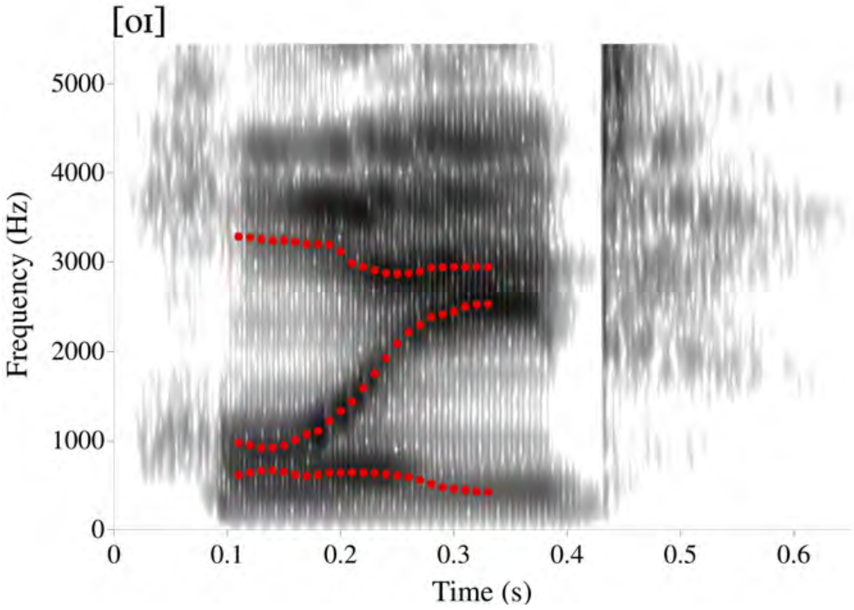


(a) Target: *how'd*

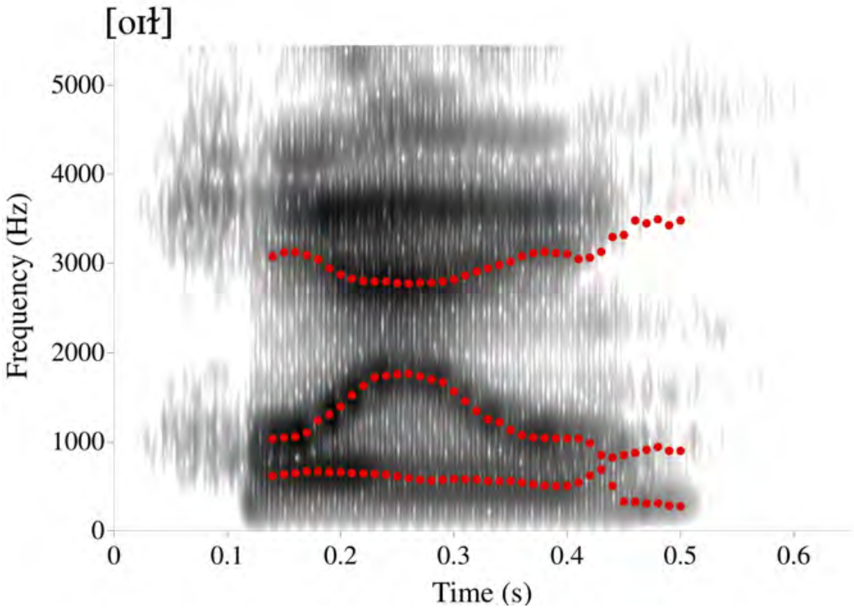


(b) Target: *howl*

Figure 2.22: Vowel /æɔ/

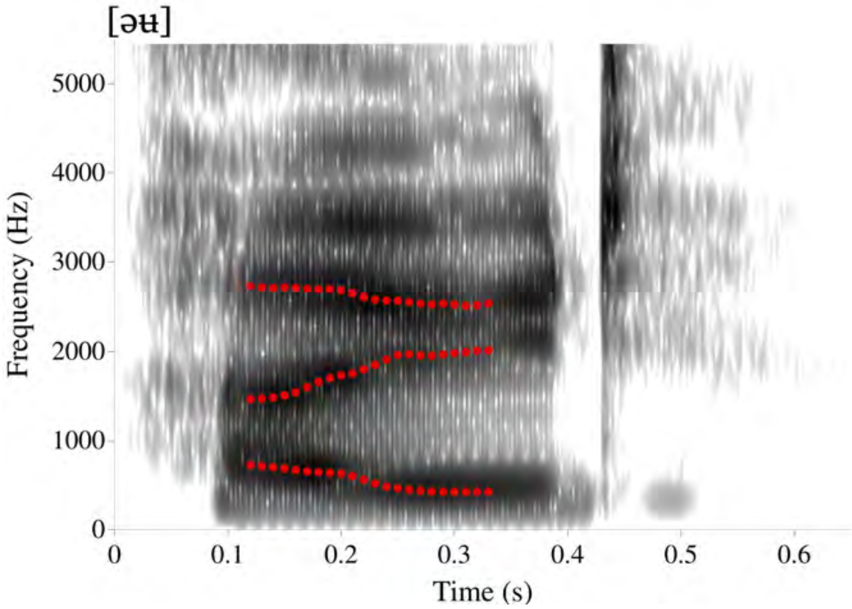


(a) Target: *hoyd*

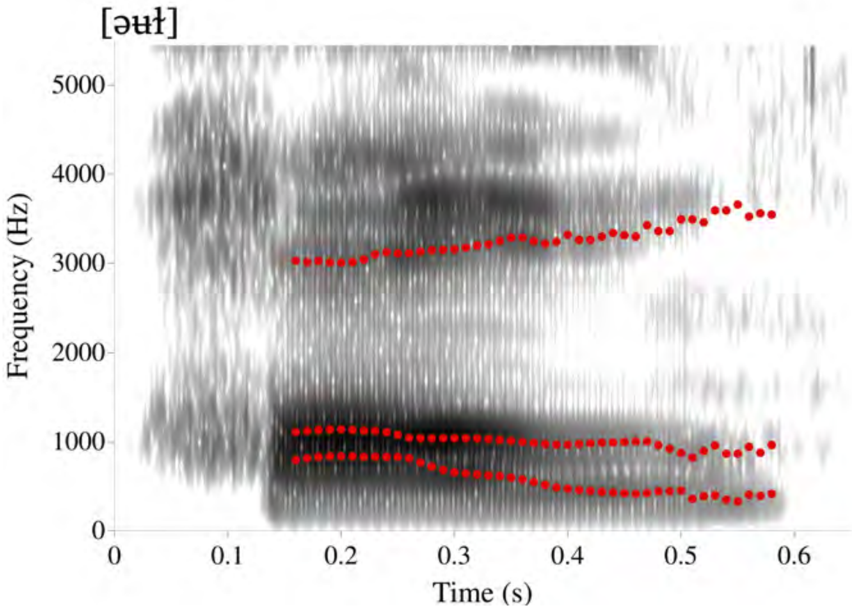


(b) Target: *hoil*

Figure 2.23: Vowel /oɪ/



(a) Target: *hode*



(b) Target: *hole*

Figure 2.24: Vowel /əʉ/

2.9 References

Altendorf, U. & Watt, D. (2008). The dialects in the South of England: phonology. In B. Kortmann & C. Upton (Eds.), *Varieties of English* (Vol. 1, p. 194–222). Berlin, New York: Mouton de Gruyter.

Arnold, L. (2015). Multiple mergers: Production and perception of three pre-/l/ mergers in Youngstown, Ohio. *University of Pennsylvania Working Papers in Linguistics*.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bernard, J. (1985). Some local effects of post-vocalic (l). In J. E. Clark (Ed.), *The cultivated Australian: Festschrift in honour of Arthur Delbridge* (p. 319–332). Hamburg: Buske.

Bernard, J. R. (1970). Toward the acoustic specification of Australian English. *STUF-Language Typology and Universals*, *23*(1-6), 113–128.

Boersma, P. & Weenink, D. (2013). *Praat 5.3.42.* Retrieved from `http://www.fon.hum.uva.nl/praat/`

Bradley, D. (2004). Regional characteristics of Australian English: Phonology. In K. Burridge & B. Kortmann (Eds.), *Varieties of English: The Pacific and Australasia* (p. 645–655). Berlin: Mouton de Gruyter.

Breiman, L. (2002). Manual on setting up, using, and understanding random forests v3. 1. *Statistics Department University of California Berkeley, CA, USA*, *1*, 58.

Burger, S. V. (2018). *Introduction to machine learning with r: Rigorous mathematical analysis*. Sebastopol, CA: O'Reilly Media, Inc.

Butcher, A. (2006). Formant frequencies of /hVd/ vowels in the speech of South Australian females. In P. Warren & C. I. Watson (Eds.), *Proceedings of the 11th Australian international conference on speech science and technology* (p. 449–453). Auckland, New Zealand.

Cox, F. (1999). Vowel change in Australian English. *Phonetica*, *56*, 1–27.

Cox, F. (2006). The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics*, *26*(2), 147–179.

Cox, F. & Fletcher, J. (2017). *Australian English pronunciation and transcription*. Cambridge: Cambridge University Press.

Cox, F. & Palethorpe, S. (2001). The changing face of Australian English vowels. In D. Blair & P. Collins (Eds.), *English in Australia* (p. 17–44). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Cox, F. & Palethorpe, S. (2004). The border effect: Vowel differences across the NSW-Victorian border. In C. Moskovsky (Ed.), *Proceedings of the conference of Australian Linguistics Society*. Newcastle, Australia.

Cox, F. & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association*, *37*(3), 341–350.

Cox, F., Palethorpe, S. & Bentink, S. (2014). Phonetic archaeology and 50 years of change to Australian English /i:/. *Australian Journal of Linguistics*, *34*(1), 50–75. Retrieved from `http://dx.doi.org/10.1080/07268602.2014.875455`

Delattre, P. C., Liberman, A. M. & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *The Journal of the Acoustical Society of America*, *27*(4), 769–773.

Efron, B., Halloran, E. & Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences*, *93*(23), 13429–13429.

Elvin, J., Williams, D. & Escudero, P. (2016). Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *The Journal of the Acoustical Society of America*, *140*(1), 576–581.

Fant, G. (1960). Acoustic theory of speech production. *The Hague: Mouton.*

Fromont, R. & Watson, K. (2016). Factors influencing automatic segmental alignment of sociophonetic corpora. *Corpora*, *11*(3), 401–431.

Garrett, A. & Johnson, K. (2013). Phonetic bias in sound change. In A. C. L. Yu (Ed.), *Origins of sound change* (p. 51–97). Oxford, UK: Oxford University Press.

Gick, B., Kang, M. A. & Whalen, D. H. (2002). MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics*, *30*(3), 357–371.

Gick, B. & Wilson, I. (2006). Excrescent schwa and vowel-laxing: Corss-linguistic responses to conflicting articulatory targets. In L. M. Goldstein, D. H. Whalen & C. T. Best (Eds.), *Papers in laboratory phonology* (Vol. 8, p. 635–659). Berlin: Mouton de Gruyter.

Giles, S. & Moll, K. (1975). Cinefluorographic study of selected allophones of English /l/. *Phonetica*, *31*(3-4), 206–227.

González, S., Grama, J. & Travis, C. (2018). Comparing the accuracy of forced-aligners for sociolinguistic research. In *Poster presented at coedl fest.* Melbourne.

Harrington, J. (2010). *Phonetic analysis of speech corpora.* Chichester, U.K.: Wiley-Blackwell.

Harrington, J. & Cassidy, S. (1994). Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English. *Language and Speech*, *37*(4), 357–373.

Harrington, J., Cox, F. & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics*, *17*(155-184).

Harrington, J., Kleber, F., Reubold, U., Schiel, F. & Stevens, M. (2018). Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in cognitive science*, *10*(4), 707–728.

Harris, J. (1994). *English sound structure.* Oxford, UK and Cambridge, USA: Blackwell.

Hyman, L. M. (2013). Enlarging the scope of phonologization. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (p. 3–28). Oxford, UK: Oxford University Press.

Kisler, T., Reichel, U. & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Labov, W., Ash, S. & Boberg, C. (2008). *The atlas of North American English, phonetics, phonology and sound change.* Berlin, Boston: Gruyter Mouton.

Lenth, R. (2019). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=emmeans (R package version 1.3.4)

Lewis, E. & Loakes, D. (2012). The /ɐlC-ɔlC/ sound change in Australian English: preliminary resɔlts. In *Proceedings of the 14th australasian international conference on speech science and technology* (Vol. 14, p. 73–76). Macquarie University, Sydney.

Liaw, A. & Wiener, M. (2002). Classification and regression by randomforest. *R News*, *2*(3), 18–22. Retrieved from https://CRAN.R-project.org/doc/Rnews/

Lin, S., Palethorpe, S. & Cox, F. (2012). An ultrasound exploration of Australian English /CVl/ words. In *14th Australasian International Conference on Speech Science and Technology* (p. 105–108). Sydney.

Loakes, D., Clothier, J., Hajek, J. & Fletcher, J. (2014). An investigation of the /el/-/æl/ merger in Australian English: A pilot study on production and perception in South-West Victoria. *Australian Journal of Linguistics*, *34*(4), 436–452.

Loakes, D., Graetzer, N., Hajek, J. & Fletcher, J. (2012). Vowel perception in Victoria: Variability, confusability and listener expectation. In *Proceedings of the 14th australasian international conference on speech science and technology* (Vol. 14). Macquarie University, Sydney.

Loakes, D., Hajek, J. & Fletcher, J. (2010a). The /el/-/æl/ sound change in Australian English: A preliminary perception experiment. In Y. Treis & R. De Busser (Eds.), *Selected papers from the 2009 conference of the Australian Linguistic Society.* Melbourne, Australia.

Loakes, D., Hajek, J. & Fletcher, J. (2010b). Issues in the perception of the /el/-/æl/ contrast in Melbourne: Perception, production and lexical frequency effects. In *Proceedings of 13th australasian international conference on speech science and technology.* Melbourne, Australia.

Loakes, D., Hajek, J. & Fletcher, J. (2010c). (Mis)perceiving /el/-/æl/ in Melbourne English: A micro-analysis of sound perception and change. In *Proceedings of 13th australasian international conference on speech science and technology.* Melbourne, Australia.

Loakes, D., Hajek, J. & Fletcher, J. (2011). /æl/-/el/ transposition in Australian English: Hypercorrection or a competing sound change? In *Proceedings of the 17th international congress of phonetic sciences.* City University of Hong Kong.

Maguire, W., Clark, L. & Watson, K. (2013). Introduction: what are mergers and can they be reversed? *English Language and Linguistics*, *17*(2), 229–239.

Oasa, H. (1989). Phonology of current Adelaide English. In P. Collins & D. Blair (Eds.), *Australian English: The language of a new society* (p. 271–287). St Lucia, Queensland: University of Queensland Press.

Palethorpe, S. & Cox, F. (2003). Vowel modification in pre-lateral environments. In *International seminar on speech production.* Sydney.

Proctor, M., Walker, R., Smith, C., Szalay, T., Narayanan, S. & Goldstein, L. (2019). Articulatory characterization of English liquid-final rimes. *Journal of Phonetics*, *77*, 100921.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Schiel, F. (1999, August). Automatic Phonetic Transcription of Non-Prompted Speech. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (Eds.), *Proceedings 14th international congress of phonetic sciences* (p. 607–610). San Francisco, CA, USA.

Searle, S. R., Speed, F. M. & Milliken, G. A. (1980). Population marginal means in the linear model: An alternative to least squares means. *The American Statistician*, *34*(4), 216–221. Retrieved from `https://www.tandfonline.com/doi/abs/10.1080/00031305.1980.10483031`

Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*(3), 291–311.

Suzuki, R. & Shimodaira, H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, *22*(12), 1540–1542.

Szalay, T., Benders, T., Cox, F. & Proctor, M. (2018). Production and perception of length contrast in lateral-final rimes. In J. Epps, J. Wolfe, J. Smith & C. Jones (Eds.), *Proceedings of the 17th australasian international conference on speech science and technology* (p. 127–132). Sydney, Australia.

Thomas, B. & Hay, J. (2005). A pleasant malady: The Ellen/Allan merger in New Zealand English. *Te Reo*, *48*, 69–93.

Turton, D. (2014). *Variation in English /l/: Synchronic reflections of the life cycle of phonological processes.* Unpublished doctoral dissertation, University of Manchester.

Wade, L. (2017). The role of duration in the perception of vowel merger. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1).

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244.

Watson, C. I. & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America*, *106*(1), 458–468.

Winkelmann, R., Jaensch, K., Cassidy, S. & Harrington, J. (2019). emur: Main package of the emu speech database management system [Computer software manual]. (R package version 2.0.3)

3 Limits of compensation for coarticulation: lateral-final rimes in Australian English

This chapter is based on the following paper, which has been submitted to the Journal of the Association for Laboratory Phonology:

Szalay, T., Benders, T., Cox, F., & Proctor, M. (under revision). Limits of compensation for coarticulation: lateral-final rimes in Australian English.

I certify that I was responsible for the development of the concept of this paper, in discussion with my supervisors/co-authors. For the first experiment, data collection was conducted during my Masters of Research at Macquarie University and preliminary analysis was submitted as a Masters of Research thesis. In accordance with Macquarie University guidelines, I include the present chapter as a relevant paper for which at least half of the research has been undertaken during my Ph.D. candidature, and I mark Experiment 1 in **Section 3.2** as non-examinable. The rest of the chapter, including conducting the entire second experiment and writing the current paper took place during my Ph.D. candidature. Data for the second experiment was collected as part of a larger project: detailed perceptual analyses are reported in the current chapter and perceptual-acoustic analyses are reported in **Chapter 4**. I took leadership in conducting the research, and was responsible for data collection and analysis, and the writing of all parts of the paper. My co-authors provided advice to improve the experimental design and methods, the analyses, the interpretation of the data, as well as the presentation of the written component. I have also received feedback from the reviewers of the Journal of the Association for Laboratory Phonology.

Abstract

Listeners show perceptual compensation for coarticulation by interpreting cues according to their phonetic contexts. In Australian English rimes, coar-

ticulation between coda /l/ and its preceding vowel attenuates cues that contribute to phonological vowel contrast. Therefore vowel-/l/ coarticulation may increase the potential ambiguity between pre-lateral vowels. We exploited this property of l-final rimes to explore the limits of listeners' ability to compensate for coarticulation using a vowel disambiguation and a word recognition task. To test the effect of vowel-/l/ coarticulation on vowel disambiguation, listeners categorised vowels in /hVd/ and /hVl/ contexts. Reduced accuracy in the /l/ context compared to the /d/ context shows that coda /l/ increases vowel disambiguation difficulty. In particular, reduced perceptual contrast was found for the rime pairs /ʉːl-ʊl, æɔl-æl/ and /əʉl-ɔl/ (e.g. *fool-full, Hal-howl, dole-doll*). A second experiment tested the effect of reduced perceptual contrast on word recognition. Listeners identified minimal pairs contrasting key vowel pairs in the /CVl/ and /CVd/ contexts. Reduced accuracy and increased response time in /l/ contexts shows that coda /l/ hinders listeners' ability to compensate for coarticulation. Results show the limits of listeners' ability to compensate for coarticulatory effects of final /l/. The relationship between limited compensation and sound change is discussed.
Keywords: vowel disambiguation, word recognition, coarticulation, lateral approximant, Australian English

3.1 Introduction

A fundamental issue in speech perception is how fine and varied phonetic details affect the identification and categorisation of speech into higher-level units. An intrinsic and pervasive source of variation in speech is coarticulation (Lindblom 1963, Iskarous et al. 2013). In order to map coarticulated speech to higher level units, listeners must compensate by attributing a coarticulatory effect to its source (Mann & Repp 1980, Gaskell & Marslen-Wilson 1998, Fowler 2005, Beddor et al. 2013, Harrington et al. 2016, Zellou 2017). That is, listeners effectively recognise which cues or properties are the result of coarticulation, and take these effects into account when perceiving speech.

Several studies have examined how the coarticulatory effects of nasal consonants on vowels are perceived in English (e.g. Beddor & Strange 1982, Beddor 2009, Beddor et al. 2013, Zellou 2017). These studies found that on the one hand, listeners can perceive fine-grained phonetic details, as they can differentiate between oral and nasal vowels, and between degrees of nasalisation (Beddor & Strange 1982, Beddor et al. 2013). On the other hand, listeners can compensate for the coarticulatory influence of nasals by attributing vowel nasalization to its consonantal source. This ensures that nasal coarticulation does not hamper vowel perception (Beddor 2009, Beddor et al. 2013, Zellou 2017). However, English does not have a phonemic contrast between oral and nasal vowels, and a nasalised vowel can and must only appear in a predictable pre-nasal or post-nasal environment. Therefore English listeners attribute vowel nasalisation as a cue to the following consonant instead of interpreting it as a cue to vowel identity.

Another segment that has been shown to have a strong coarticulatory influence on the preceding vowel is dark coda /l/ (Recasens 2002, Cox & Palethorpe 2007). Unlike nasals, coda /l/ affects cues that are contrastive in the English vowel inventory, as it reduces spectral cues to vowel contrast (Palethorpe & Cox 2003, Wade 2017). For instance, the backing effect of coda /l/ may reduce the contrast between central /ʉː/ and back /ʊ/ in *fool* and *full*. Therefore the quality of the nucleus in these words can be attributed to the coda but may also be interpreted as an intrinsic quality of the vowel. That is, coda /l/ potentially has the ability to mask acoustic cues used by listeners in vowel identification and word recognition.

The goal of this study is to investigate in new detail how coarticulation affects speech perception by examining listeners' ability to compensate for a coarticulatory process that affects phonologically contrastive cues. To exemplify such a coarticulatory process, we selected /l/-final words in Australian English (AusE). We hypothesised that if coarticulation with coda /l/ reduces perceptually contrastive vowel cues, listeners' ability to compensate for coarticulation would be hindered. Hindered compensation means that instead of attributing the coarticulatory influence of /l/ to its source, listeners may instead attribute this influence to an inherent property of the vowel. The effect would be evident through an increased difficulty in vowel disambiguation in the pre-/l/ context compared to a pre-/d/ context. We also expected that the most spectrally similar vowels would be the hardest to disambiguate before a lateral.

We tested these hypotheses in two experiments. In the first experiment we found that /l/-final rimes were disambiguated less easily than /d/-final rimes; in particular, the spectrally similar pairs /ʉːl-ʊl, æɔl-æl, əʉl-ɔl/ (e.g. *fool-full, howl-Hal, dole-doll*) were poorly discriminated compared to other /l/-final target-competitor pairs and to /d/-final minimal pairs contrasting the same vowels. The limitation of our first experiment was that it used a combination of real and non-words.

Because compensation for coarticulation is facilitated when it results in a lexical item rather than a non-word, (Gaskell & Marslen-Wilson 1998, Stevens & Keyser 2010), we conducted a second experiment using only real words to examine if the contrast-reducing influence of lateral codas also affects lexical access to /l/-final words. We hypothesised that if listeners cannot compensate for the coarticulatory influence of /l/, despite being presented with lexical information, they may map /l/-final words to incorrect lexical items (e.g. listeners might map the acoustic signal of *pool* to the lexical item *pull*). We found that listeners were less accurate and slower at accessing monosyllabic words within the pairs /iːl-ɪl, ʉːl-ʊl, æɔl-æl/ and /əʉl-ɔl/ compared to their /d/-final counterparts. These results suggest that some /l/-final word pairs may have an inherently ambiguous signal, which limits listeners' ability to compensate for coarticulation.

3.1.1 Compensating for coarticulation

Perceptual compensation for coarticulation is understood as listeners' ability to factor out the influences of surrounding segments on a target segment and attribute them to the source segments, with the consequence that a single context-independent percept remains (Mann 1980, Zellou 2017). Listeners' ability to compensate for coarticulation has been demonstrated in studies showing that listeners interpret cues according to their contexts and thus perceive the same ambiguous signal as different segments in different contextual conditions (Mann & Repp 1980, Fowler 1984, Gaskell & Marslen-Wilson 1998, Kleber et al. 2012, Zellou 2017).

In the perception of consonants, when hearing a fricative that is ambiguous between /s/ and /ʃ/, listeners reported perceiving /s/ when the fricative was followed by a rounded vowel, and reported perceiving /ʃ/ when it was followed by an unrounded vowel (Mann & Repp 1980). This is because in a fricative+unrounded vowel sequence listeners attribute the low frequencies to the fricative and categorise it as /ʃ/, whereas in a fricative+rounded vowel sequence listeners attribute the same low frequencies to lip rounding and categorise the fricative as /s/ (Mann & Repp 1980, Smits 2001, Mitterer

2006). Similarly, a segment that is ambiguous between /d/ (with high F3 onset) and /g/ (with low F3 onset) is more likely to be perceived as /g/ when it is preceded by /l/ than when it is preceded by /ɹ/ (Mann 1980). If listeners attribute the lowered F3 to the stop in the /l/+stop sequence they would categorise the stop as /g/, whereas a lowered F3 attributed to /ɹ/ in a /ɹ/+stop sequence may lead listeners to classify the stop as /d/ (Mann 1980). These effects might not be specific to speech, as under certain circumstances, a preceding low tone (corresponding to /ɹ/) or high tone (corresponding to /l/) have the same effect (Lotto & Kluender 1998, Fowler et al. 2000).

Consonantal context has also been shown to affect vowel categorisation. For example, listeners accept a vowel with a relatively high F2 as /ʊ/ in the fronting /s_t/ context, whereas they categorise the same vowel as /ɪ/ in the non-fronting /w_l/ context despite the fact that prototypical /ʊ/ has a low F2 and prototypical /ɪ/ has a high F2 (Kleber et al. 2012). These studies suggest that listeners attribute coarticulatory information to the influencing segment and factor coarticulatory effects out in the perception of the affected segment. That is, listeners can compensate for coarticulation by attributing acoustic cues resulting from coarticulation to their coarticulatory source.

There are instances of coarticulation that lead to assimilation, for example /p/ in the phrase *top tag* can be realised as [t] (Stevens & Keyser 2010). Listeners compensate for assimilative coarticulation in existing lexical items, such as *freight bearer*, realised with a final /p/ instead of a /t/ in *freight*, but not in nonwords, such as *preip bearer* (Gaskell & Marslen-Wilson 1998). These studies show that listeners integrate top-down lexical information when compensating for coarticulation.

By compensating for coarticulation, listeners can arrive at phoneme categories and category memberships despite contextual change to the signal. For instance, /g/ has an acoustically different release burst between /gi/ and /gu/, but listeners perceived acoustically different /g/ sounds in the appropriate coarticulatory context as more similar to each other than acoustically identical /g/ sounds when one of them was originally produced in a different phonetic context (Fowler 1984). Similarly, English listeners perceived oral and nasal vowels as different when nasality cannot be attributed to context (e.g. nasal vowels in the context of oral consonants or in isolation) and as similar when nasality can be attributed to context (e.g. nasal vowels in the context of nasal consonants) (Beddor & Krakow 1999).

When contrastive cues of the target segment are affected, listeners only compensate for coarticulation in cases of ambiguous tokens. In a continuum of synthesised tokens between two unambiguous endpoints, listeners perceive the ambiguous tokens according to their context; however, the tokens at the endpoint corresponding to unambiguous phonemes as produced in natural speech tend to be perceived based on the features of the target sound, irrespective of context (Mann & Repp 1980, Mann 1980, Kleber et al. 2012). This shows that segmental context shifts the category boundaries in ambiguous, but not in unambiguous tokens.

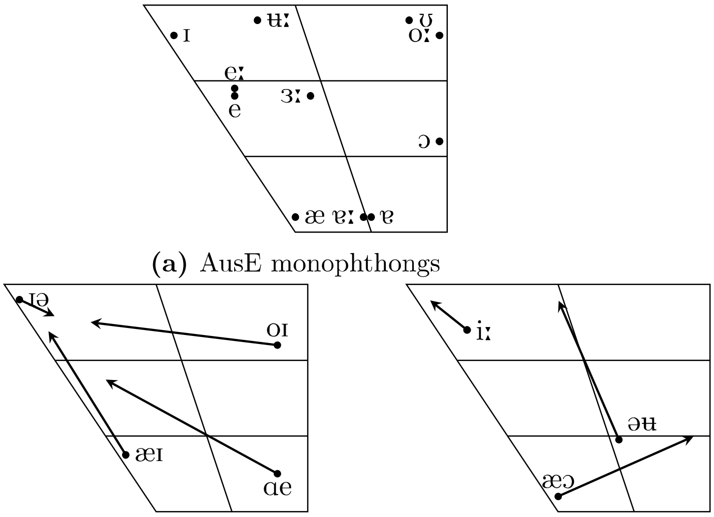
It is not clear from these studies whether listeners can compensate for coarticulation in unmanipulated speech when coarticulation reduces contrastive cues in such a way that acoustic cues might be attributed both to the segment undergoing coarticulation and to the segment causing it, potentially making the perceptual target inherently ambiguous. An environment where these interactions can be explored in more detail is lateral-final rimes, because vowel-lateral coarticulation affects contrastive vowel cues, making vowels potentially ambiguous in the pre-lateral context in natural speech.

3.1.2 The effect of coda /l/ on AusE vowels

General Australian English (AusE) uses a large vowel inventory consisting of 18 stressed vowels and schwa (**Figure 3.1**) (Cox & Fletcher 2017). The AusE vowel inventory utilises both spectral and durational contrasts, with phonemic vowel length contrast for spectrally similar pairs (Harrington et al. 1997, Cox & Palethorpe 2007). For instance, the vowel pairs /ɐː-ɐ, eː-e/ (e.g. *card-cud*, *shared-shed*) primarily contrast in length (Cox & Palethorpe 2007), and /iː-ɪ, ʉː-ʊ/ are realised with both durational and spectral contrast (Cox 2006). In addition, there are spectrally similar diphthong-monophthong pairs in which one of the diphthongal targets coincides with a monophthong, such as /æɔ-æ, æɪ-æ/ in *loud-lad, laid-lad*, and /əʉ-ʉː, æɔ-ɔ/ in *boat-boot, pout-pot* (Cox 1999). As a result, some AusE vowel pairs share spectral features.

English coda /l/ is typically realised as a dark [ɫ], articulated with a lowered and retracted tongue dorsum, and an alveolar tongue tip gesture (Sproat & Fujimura 1993). As the tongue dorsum gesture of [ɫ] may start during the vowel production, [ɫ] favours an anticipatory V-[ɫ] coarticulation, leading to the backing and the lowering of the vowel (Recasens 2002, Lin et al. 2012).

As **Chapter 2** showed, acoustically, a post-vocalic lateral results in an overall diminished vowel dispersion in the F1-F2 vowel plane and reduced vowel contrast between certain vowel pairs (**Figure 3.2**). Diminished vowel dispersion is the result of the backing of the front vowels: significantly lowered F2 was found for /iː, ɪ, e, ɐ, ɔ, ʉː, ɜː/ in pre-lateral environments



(a) AusE monophthongs

(b) AusE diphthongs 1

(c) AusE diphthongs 2

Figure 3.1: The AusE vowel inventory. Figures reproduced from Cox & Fletcher (2017).

(Palethorpe & Cox 2003, Cox & Palethorpe 2004, as well as **Chapter 2**). In addition to the overall reduced dispersion in the vowel space, the pairs /ʉː-ʊ, æɔ-æ/ and /əʉ-ɔ/ also showed reduced spectral contrast in pre-lateral position (Palethorpe & Cox 2003 and **Chapter 2**). Acoustic contrast between /ʉː-ʊ/ and /əʉ-ɔ/ is partially neutralised before a coda /l/, due to the lowering of the second formant of /ʉː/ (Palethorpe & Cox 2003, Cox & Palethorpe 2004 and **Chapter 2**). Contrast between /æɔ-æ/ is partially neutralised before coda laterals, as /l/ and the second element of the diphthong overlap substantially. In earlier analyses, the vowels /iː-ɪ/ not found to undergo acoustic contrast reduction at the steady state of the vowel, however the onglide of /iː/, which is one of the differentiating features between the two vowels (Cox 2006), is reduced and both vowels gain a schwa-like offglide (Palethorpe & Cox 2003). In contrast, in **Chapter 2**, the vowels /iː-ɪ/ were found to undergo acoustic contrast reduction, even if only to a lesser extent than the members of the vowel pairs /ʉː-ʊ, æɔ-æ/ and /əʉ-ɔ/.

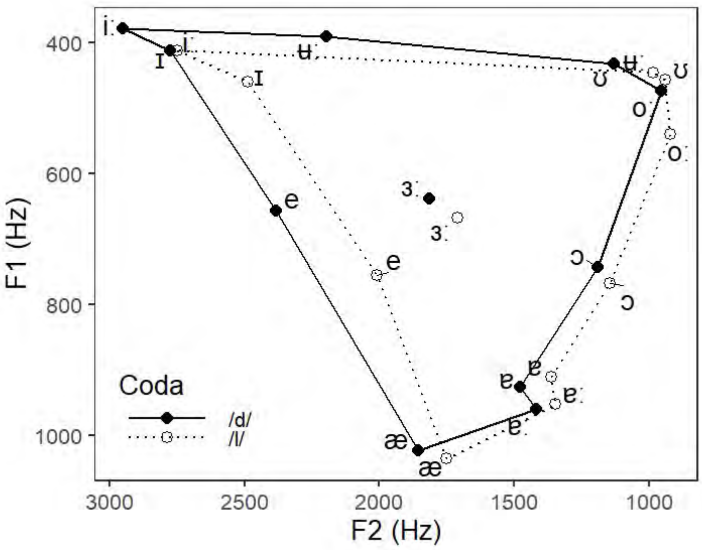


Figure 3.2: Monophthong targets before coda /d/ and coda /l/ in the F1-F2 vowel plane. Monophthongs before coda /l/ show reduced vowel contrast and smaller vowel dispersion compared to monophthongs before coda /d/. Figure reproduced from **Chapter 2**.

However, durational contrast was maintained between the vowel pairs in the pre-lateral context (Palethorpe & Cox 2003 and **Chapter 2**).

Reduced dispersion of vowels in the F1-F2 plane in the pre-lateral environment may hinder vowel perception, as a more dispersed F1-F2 vowel space has been demonstrated to facilitate intelligibility in clear speech (Bradlow et al. 1996, Ferguson & Kewley-Port 2007, Neel 2008; but see Krause & Braida 2004 for evidence to the contrary). Reduced dispersion may diminish spectral contrast and reduce intelligibility; for example, American English listeners confused the spectral neighbours /ɑ-ʌ/ and /ɛ-æ/ but never /i-ʌ/ or /ɛ-u/ (Neel 2008). Therefore reduced vowel dispersion caused by vowel-/l/ interactions might also increase the difficulty of vowel-, and thus potentially word identification.

Studies on the perception of English lateral-final rimes have shown that vowel-lateral coarticulation helps listeners identify /l/, but hinders identification of certain vowels. Anticipatory vowel-lateral coarticulation allowed British English listeners to reliably identify /l/ in *belly* when /l/ and the following sounds were replaced by white noise (West 1999). In contrast, listeners could not identify /l/, when /l/ and the preceding vowel were replaced by white noise: listeners could identify *belly* from [be##], but not from [b##i] (West 1999). Vowel identification has been examined in /l/-triggered vowel mergers in several dialects of English (Thomas & Hay 2005, Loakes et al. 2014b, Wade 2017). Listeners from Melbourne, Australia showed a limited ability to distinguish /el/ from /æl/ in a word identification task with minimal pairs (e.g. *Alan-Ellen*) (Loakes et al. 2010a;b;c; 2011; 2012; 2014a;b). Some speakers of New Zealand English were able to distinguish minimal pairs differing in /el/ and /æl/ despite merging /el-æl/ in production (Thomas & Hay 2005). In Ohio English, listeners could distinguish spectrally merged /oʊl-ul/ (e.g. *pole-pull*) and /ul-ʊl/ (e.g. *pool-pull*) using durational cues, but listeners from Vermont could not (Wade 2017).

Production and perception studies have demonstrated that vowel-lateral coarticulation reduces acoustic contrast between certain vowels. However, it is not clear if and how listeners can compensate for coarticulation when acoustic contrast is reduced. AusE lateral-final rimes offer the potential to gain insights into the issue of whether reduced acoustic contrast leads to a perceptually ambiguous vowel signal and whether final laterals limit listeners' ability to compensate for coarticulation. We address these questions by examining whether listeners attribute cues carried by the vowel to coda /l/ or if instead they interpret vowel quality as an inherent property of the vowel and cue to vowel identity.

3.2 Experiment 1: Disambiguation of /l/-final rimes

We tested listeners' ability to compensate for the effect of vowel-/l/ coarticulation using a rime disambiguation task. Participants were asked to identify an aurally-presented target by selecting one of two orthographic representations. Candidate pairs consisted of an exhaustive pairing of all 16 possible stressed /l/-final rimes in AusE and an exhaustive pairing of the same 16 stressed vowels in /d/-final rimes. Comparing accuracy and reaction time (RT) of responses to /d/- and /l/-final target words allowed us to test the extent to which vowel-lateral coarticulation affects vowel disambiguation. This task also allowed us to identify the most easily confused vowel pairs. We hypothesised that if vowel-lateral coarticulation masks cues that are vital to vowel disambiguation, listeners would perform worse on /l/-final rimes than on /d/-final rimes. We also predicted that/l/-final contexts would have a particularly strong negative effect on accuracy and reaction time compared to /d/-final contexts for vowel pairs that have been shown to exhibit reduced contrast in /l/-final contexts, namely /ʉː-ʊ, æɔ-æ, əʉ-ɔ/ (e.g. *fool-full, howl-hal, dole-doll*).

3.2.1 Methods

Participants

Thirty (F = 29, M = 1, bilingual = 19, age = 19–56, mean = 24.16) listeners of AusE (born in Australia or migrated to Australia before the age of 2) participated in the experiment. Participants were undergraduate students of linguistics at Macquarie University and received course credit for participation. All participants had linguistic training but were naive to the purpose of the experiment. None of the participants reported any current hearing, speaking, or reading difficulties.

Materials

The stimuli consisted of 16 AusE vowels embedded in /hVd/ and /hVl/ words. The vowels /ɪə/ and /eː/ were excluded as they never appear before final /l/. When a combination of /h/+V+/d/ or /h/+V+/l/ did not yield an existing word, the corresponding nonword was used. The two alternatives in the forced-choice task were the orthographic representations of the candidates spelled uniformly with an initial *h*. Nonwords were spelled according to English spelling and judged by native speakers of AusE for transparency (**Appendix 3.6.1**).

Stimulus materials were elicited from a 21-year old monolingual female university student born in Australia to Australian-born parents and recorded with an AKG C535 EB microphone at 44.1kHz sampling rate in a sound treated studio in 2006. The stimuli were amplitude-normalised, digitised as 16 bit WAV files, and truncated to have 1s silence before and after the word. Mean duration of target words in the /d/ condition was 486 ms (range = 320–650 ms), and 528 ms (range = 450–640 ms) in the /l/ condition.

Procedure

Participants familiarised themselves with the targets and they were introduced to the experiment with a short practice session, disambiguating the nonword targets. Feedback was provided after each trial. Familiarisation and practice were followed immediately by the experimental phase.

Participants were seated in front of a computer monitor located at eye height at a distance of 50 cm and wore Sennheiser 380 Pro headphones adjusted to their comfortable listening level. Participants were instructed to respond as quickly and accurately as possible. To begin each trial, a fixation cross was displayed in the centre of the screen. After 500 ms the two candidate items were displayed in lower case orthography, arranged horizontally, and presented in different coloured boxes. After 1500 ms the target word started playing, while the candidates remained on screen. Participants had 2000 ms from audio onset to select the candidate they heard (**Figure 3.3**). Selections were made with a Chronos button box whose input keys mapped to the colours on the screen. The experiment moved on to the next trial when participants responded. If participants did not answer within 2000 ms, a warning message let them know that they were too slow and they were instructed to press a button to continue. The experiment did not proceed to the next trial until the participants responded.

Each participant was tested either on 16 /d/-final targets and 15 competitor candidates or on 16 /l/-final targets and 15 competitor candidates, repeated in three blocks, once per block, with a 10s forced break between the blocks. Each participant was exposed to 240 (items) $\times$ 3 (repetitions) = 720 trials. In half of the trials, the target candidate was presented on the right, and in the other half on the left. Trials were randomised within the blocks. After the experiment, participants reported whether they found any of the words "unusal" or "difficult".

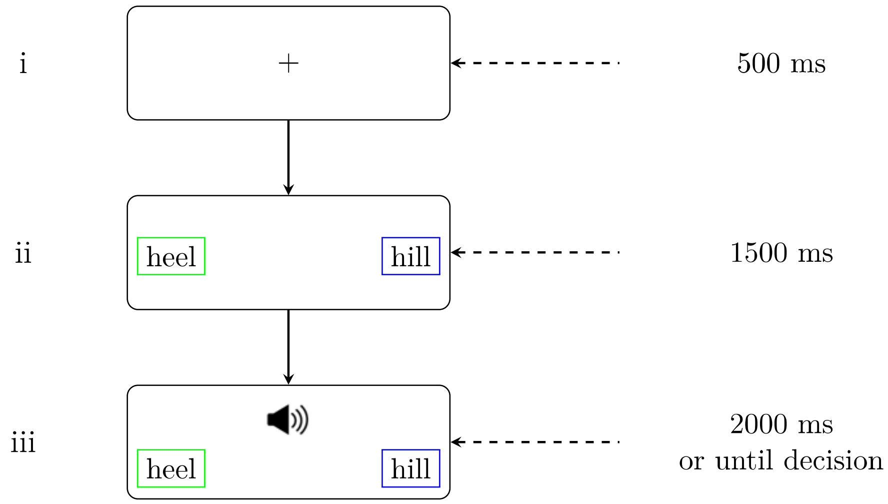


Figure 3.3: Structure of a trial i. Fixation; ii. Presentation of response alternatives; iii. Audio stimulus presentation

Analysis

Responses to 30 (participants) × 720 (trials) = 21,600 trials were collected. 63 observations, including all 45 trials with *hill* as target and *heel* as competitor, were excluded from the analysis due to errors in stimulus presentation. Trials in which response times were faster than 210 ms (Woods et al. 2015) or beyond mean ± 2 s.d. of the participant (Ratcliff 1993) were excluded from the analysis, leaving a total of 20,413 trials (94.8%) for the analysis.

Response accuracy was analysed using Generalised Linear Mixed-Effect Models (GLM) with the family binomial (Bates et al. 2015) and the BOBYQA (Bound Optimization BY Quadratic Approximation) optimiser (Powell 2009). $p - values$ were calculated with the *lmerTest* package (Kuznetsova et al. 2017) using Satterthwaite's degrees of freedom method. RT data was analysed using GLM models with the BOBYQA (Powell 2009) optimiser and the family gaussian with the logarithmic linking function because the distribution of RT was right-skewed and followed a log-normal distribution (**Figure 3.4**).

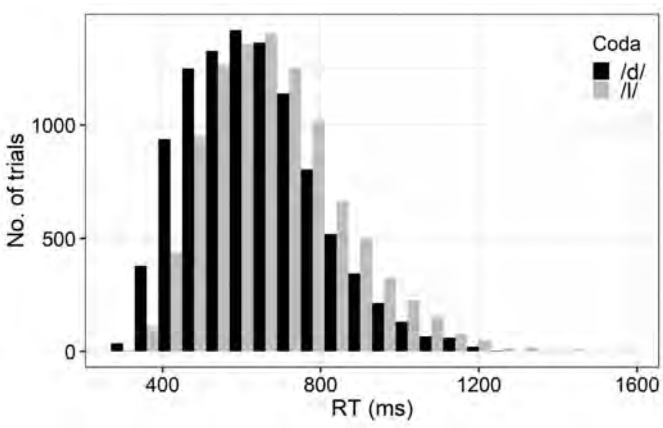


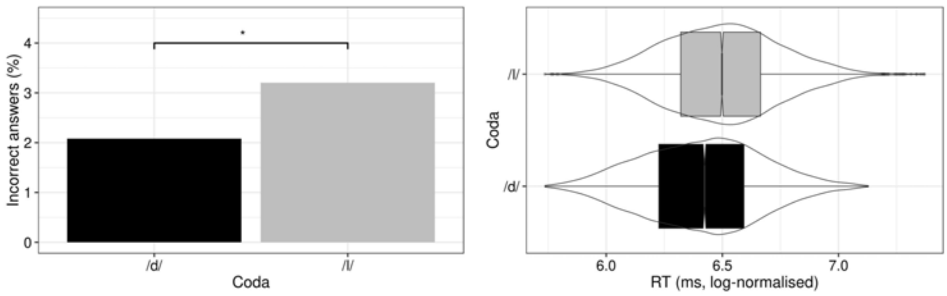
Figure 3.4: Distribution of reaction times (ms) for correct responses. Black bars: Coda /d/ condition. Grey bars: Coda /l/ condition.

To examine the effect of coda /l/ on accuracy and speed of rime disambiguation, we constructed two GLMs, one with the dependent variable Accuracy and another with RT of correct responses. The independent variables were Coda (treatment coded, comparing /l/ to the baseline /d/) and Lexical Status of Target (treatment coded, interacting); the model included a random by-subject effect on the intercept. To examine the speed-accuracy trade-off between the two coda conditions, we constructed a third GLM with the dependent variable RT, including RT of both correct and incorrect responses. The independent variables were Coda, Response Accuracy (interacting), and Lexical Status of Target (non-interacting); the model included a random by-subject effect on the intercept. Coda and Response Accuracy were treatment-coded so that the intercept was the RT of incorrect responses in the /d/ condition. To shed light on how the pairing of target and competitor vowels affects rime disambiguation, we used agglomerative hierarchical cluster analysis with Ward's method (Ward 1963). Hierarchical cluster analysis takes the individual vowels as single-element clusters and at each step merges two clusters into a group (a cluster) in such a way that the members of one cluster are maximally similar and the members of two separate clusters are maximally dissimilar.

3.2.2 Results

Effects of Coda

/l/-final rimes were disambiguated significantly less accurately ($\beta{=}-0.58, F(1,20,408){=}4.85$, $p{=}0.04$); however, there was no significant effect on the speed of disambiguation ($\beta{=}6.43, F(1,19,868){=}0.0001, p{=}0.1$) than /d/-final rimes (**Figure 3.5**).[1] Real words were disambiguated more accurately ($\beta{=}0.18, F(1,20,408){=}31.13$, $p < 0.001$), and quickly ($\beta{=}-0.01, F(1,19,868){=}132.12, p < 0.001$) than nonwords.



(a) Percentage of inaccurate responses. **(b)** RT (ms, log-normalised).

Figure 3.5: Effect of Coda /l/ (grey) compared to Coda /d/ (black) on response accuracy (left) and time (right).

The exploration of the speed-accuracy trade-off showed that RT of incorrect responses was slower in the /l/ condition than in the /d/ condition ($\beta{=}0.13, F(1,2,406){=}0.0001, p{=}0.02$). RT was slower for correct responses than for incorrect responses within the /d/ condition $\beta{=}0.04, F(1,2,406){=}0.27, p{=}0.038$). The difference between the RT of correct and incorrect responses was

[1]RT estimates are reported as log-normalised ms.

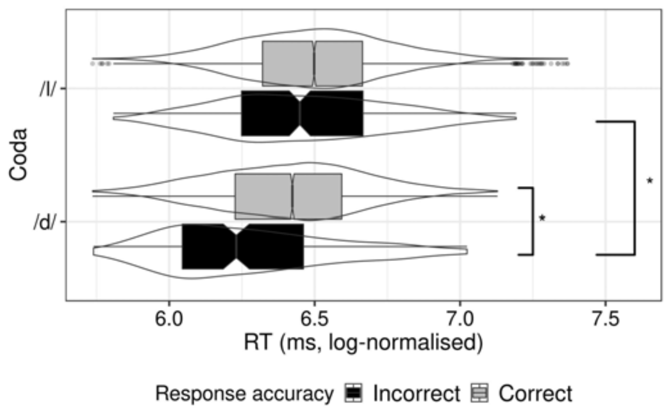


Figure 3.6: Speed-accuracy trade-off. Top panel: RT in the Coda /l/ condition. Bottom panel: RT in the Coda /d/ condition. Black: RT of incorrect answers. Grey: RT of correct answers.

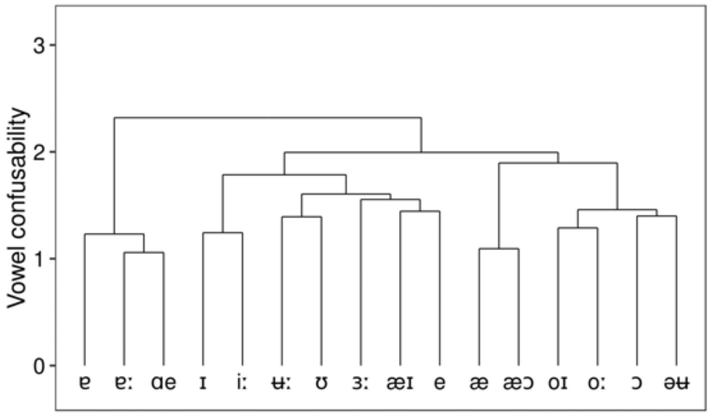
significantly smaller in the /l/ condition than in the /d/ condition ($\beta = -0.04, (F, 2, 406) = 3.91, p = 0.047$) (**Figure 3.6**). Real words were disambiguated more quickly ($\beta = -0.02, F(1, 2, 406) = 134.17, p < 0.001$) than nonwords.

3.2.3 Effect of Target- and Competitor vowels

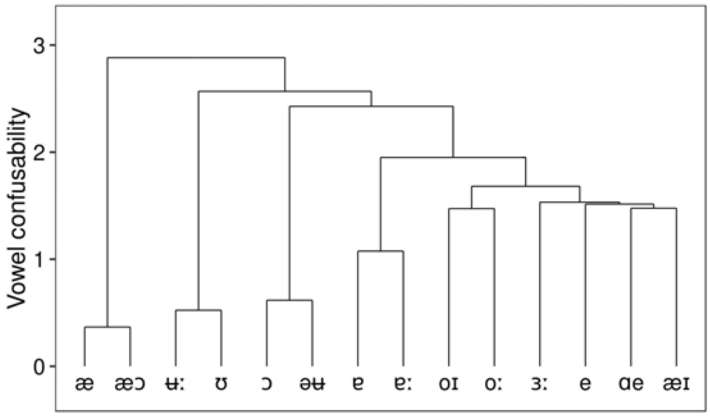
The effect of Target and Competitor vowels was examined using agglomerative hierarchical cluster analysis with Ward's method (Ward 1963), based on a confusion matrix of target- and competitor vowels (**Figure 3.7**). Vowels that form a dyad in **Figure 3.7** are vowels which were confused the most often when paired as target and competitor. The vertical location of the nodes indicates confusability: the lower a node is located, the higher the percentage of incorrect responses.

Target- and competitor vowels were most frequently confused when the two vowels shared a similar place of articulation (vowel frontness and height). Long-short vowel pairs were the hardest to disambiguate (e.g. /iː-ɪ/ and /ɐː-ɐ/) in the /d/ condition. In the /l/ condition, the vowel pairs /ʉː-

ʊ, æɔ-æ/ and /əʉ-ɔ/ were easily confused; however, this analysis does not establish whether articulatory similarity has a statistically significant effect on vowel disambiguation. Comparing the clusters between the /d/ and the /l/ condition shows that the rimes were harder to disambiguate in the /l/ condition, as two-member vowel clusters are separated earlier from other clusters, that is, the nodes are located lower in the /l/ condition.



(a) Coda /d/ condition



(b) Coda /l/ condition.

Figure 3.7: Perceptual vowel similarity based on rime confusion: closer clustering signals higher confusion rates.

3.2.4 Discussion

The aim of Experiment 1 was to examine the influence of coda lateral coarticulation on listeners' ability to disambiguate vowel contrasts. As predicted, these data revealed that vowel discrimination is significantly less accurate in pre-lateral than in pre-obstruent environments. Lower accuracy in the /l/ condition is consistent with the hypothesis that coda /l/ reduces perceptual vowel contrast. Listeners were not overall slower in disambiguating /l/ final rimes. However, incorrect responses were faster than correct responses in the /d/, but not in the /l/-condition, indicating a speed-accuracy trade-off for the former, but not for the latter. The presence of a speed-accuracy trade-off is consistent with a "fast-guess" model of RT that argues that decisions made quickly are guesses and therefore less likely to be accurate whereas decisions based on evidence are slow and highly accurate (Ollman 1966, Yellott Jr 1971). That is, incorrect answers are likely to be the result of fast guesses in the /d/ condition; however, when listeners allocated more time to make a decision they could disambiguate the vowel correctly. In contrast, we found no evidence for a speed-accuracy trade-off in the /l/-condition due to the increased RT of the incorrect responses, indicating that the incorrect answers were the result of processing difficulties, not of insufficient time taken to process the input. This suggests that when not opting for a fast-guess, listeners allocated the same amount of time to disambiguate the rime in both conditions; however, this time was not sufficient to make accurate decisions in the /l/ condition. That is, listeners' incorrect responses are the result of insufficient time in the /d/ condition, whereas in the /l/ condition they are the result of increased difficulties in vowel disambiguation.

We attribute the increased difficulty in vowel disambiguation in the pre-/l/ context to the coarticulatory influence of /l/ on the vowel. In the stimuli, vowel-/l/ coarticulation led to spectral contrast reduction, consistent with findings of Palethorpe & Cox (2003) (see **Appendix 3.6.2** for the formant trajectories of the most confused rimes). The overall negative effect of coda /l/ on vowel disambiguation indicates that coda /l/ masks some of the acoustic cues listeners rely on to the extent that listeners cannot compensate for it.

We also examined the effects of Target- and Competitor Vowel, expecting that spectrally similar vowels would be more likely to be confused with each other. This expectation was borne out both in the /d/ and the /l/ condition, as the most confused vowel pairs are similar to each other in place of articulation and formant trajectories, such as /ʉː-ʊ, æɔ-æ, əʉ-ɔ, oɪ-ɔ/. This is not surprising in the /d/ condition, as English listeners are only likely to confuse spectrally similar vowels (Neel 2008). However, English listeners have been shown to give more weight to length cues when spectral differences are inherently smaller (Bennett 1968) or not available any more due to a contextual merger (Wade 2017). This does not seem to be the case in our data: perceptual similarity between /ʉː-ʊ, æɔ-æ, əʉ-ɔ/ increased as spectral differences became smaller in the /l/ condition, even though the vowels within these pairs differed in length (**Appendix 3.6.3**). The high confusion rate of /ʉː-ʊ, æɔ-æ/ and /əʉ-ʉ/ shows that listeners interpret the coarticulatory effects of /l/ as an intrinsic property of the vowel and as a vowel cue, and not as a cue to the following consonant. This shows that vowel-/l/ coarticulation interferes with listeners' ability to map the signal to higher level units and disambiguate the rime.

A limitation of Experiment 1 was the nature of the task, which required listeners to map an auditory signal to a mixture of orthographically presented real words and non-words. Real-word status, word frequency, and familiarity all affect word recognition (Rubenstein et al. 1970, Forster & Chambers 1973, Segui et al. 1982, Meunier & Segui 1999). In addition, listeners were not exposed to variation in the coda, as listeners were assigned to either the /l/ or the /d/ condition. A lack of attention to the codas, which was predictable for all items, thus might have been partly responsible for the observed inefficient compensation. In Experiment 2, we used a word recognition paradigm to test whether vowel-lateral coarticulation affects how listeners compensate for context during lexical processing of words. Experiment 2 required the processing of the entire word and also presented words ending in /d/ and /l/ to all participants to draw participants' attention to the coda.

3.3 Experiment 2: Word recognition

We examined listeners' recognition of /l/-final words contrasting /iː-ɪ, ʉː-ʊ, æɔ-æ,/ and /əʉ-ɔ/ to assess whether listeners can compensate for vowel-/l/ coarticulation when required to process the information lexically. Participants listened to words contrasting the vowel pairs that had been identified as the most confusable in Experiment 1 (i.e. /ʉː-ʊ, æɔ-æ, ʉ-ɔ/ and in addition /iː-ɪ/, as their pre-/l/ allophones have acoustically similar offglides (Palethorpe & Cox 2003)) to determine how listeners map the acoustic signal of /CVl/ minimal pairs to lexical items.

3.3.1 Methods

Participants

Forty-six female native speakers of Australian English, born in Australia to Australian-born parents (monolingual = 33, age = 18–40 years, mean = 21.5) participated in the experiment. Participants received course credit or $15 for participation. None of the participants reported any current hearing, speaking, or reading difficulties.

Materials

The stimuli consisted of 32 unique CVC targets and 38 unique (C)V(C) fillers. For the 32 targets, 16 minimal pairs were chosen which contrasted the 4 vowel pairs (/iː-ɪ, ʉː-ʊ, æɔ-æ, əʉ-ɔ/), with two sets of minimal words per coda and per vowel pair. Due to the limited number of available minimal pairs, the target words varied in words class and lexical frequency. Frequency was measured in the AusE part of the GloWbe corpus (Davies 2013); mean frequency in the /d/ condition was 312.5 per million words (range = 0.3–2415), and 48.8 (range = 0.2–446) in the /l/ condition. Fillers were (C)V(C) words that did not contain /d/ or /l/ or the target vowels in any position. Fillers matched the candidates in part of speech and onset consonants and were chosen from the first 5000 most frequent words of the COCA database (Davies 2008). Mean frequency of fillers was 397 per million words (range = 10–2048).

Two sets of recordings of the stimulus materials were elicited, from a 57 and a 25 year-old female speaker of AusE. Stimuli were recorded with an AKG C535EB Condenser Microphone onto an iMac using Presonus Studio Live 16.2.4 AI Mixer at 44.1kHz sampling rate in a sound treated studio. The stimuli were amplitude-normalised and truncated to have a 1s silence before and after the end of the word. Formant change over time for the stimulus words is shown in **Appendix 3.6.4**. Mean duration of target words in the /d/ condition was 593 ms (range = 425–727 ms), and 644 ms (range = 474–841 ms) in the /l/ condition. Mean duration of the fillers was 662 ms (range = 474–844 ms).

Procedure

Prior to the experiment, participants familiarised themselves with the stimulus materials by reading them out loud as they were presented in random order on a computer monitor. Participants were introduced to the experiment with a short practice session, listening to audio recordings of ten words, and typing what they heard. Feedback was provided after each trial on spelling alternatives and acceptable responses. Familiarisation and practice were followed immediately by the experimental phase.

Participants were seated in front of a computer monitor located at eye height at a distance of 50 cm and wore Sennheiser 380 Pro headphones adjusted to their comfortable listening level. Participants were instructed to respond as quickly and accurately as possible. To begin each trial, a fixation cross was displayed in the centre of the screen. After 500 ms, the target word started playing and participants typed what they heard. Participants were allowed to use backspace but did not receive feedback on their responses.

Each participant was tested on 32 targets and 32 fillers, all repeated in four blocks, once per block, with a 30s forced break between the blocks. The first two blocks were spoken by the 57 year old informant and the last two by the 25 year old informant. The first and the third block were preceded by an additional six fillers at the beginning to habituate the listeners to the voice of the speaker. The 32 targets and the remaining 32 fillers were presented in a pseudo-random order. Each participant was exposed to 64 (items) $\times$ 2 (informants) $\times$ 2 (repetitions) + 12 (habituation) = 268 trials. The stimuli were presented with the software Expyriment (Krause & Lindemann 2014).

After the word recognition experiment, participants reported whether they found any of the words “unusal” or “difficult”.

Analysis

Responses to 46 (participants) × 268 (trials) = 12,328 trials were collected. 552 responses from the habituation trials and 5628 responses from fillers were excluded prior to any analysis. Nineteen tokens were excluded due to technical difficulties and coding errors.

The remaining 6,129 responses were rated for accuracy. Participants' responses were compared to the target and classified as Intended Answer, Phonetic Respelling, Typo, Minimal Pair Error and Other Error. Responses were classified as Intended Answer if spelled as the target or its homophone (e.g. both *would* and *wood* were classified as Intended Answer for /wʊd/). In addition, proper nouns spelled with lower case letters and contractions spelled without apostrophes were classified as Intended Answer. Unambiguous, phonetic, but nonstandard spellings of target words (e.g. *knowed* for *node*) were classified as Phonetic Respellings. Single letter deletions, additions, letter transpositions, and substitutions within one key distance of the target letter were classified as Typos (Luce & Pisoni 1998). Responses in which participants confused members of the minimal pairs (e.g. answered *fool* when the target was *full*) were classified as Minimal Pair Errors. Any other errors, such as misheard words errors, e.g. *cool* for *pool* were classified as Other Errors. Responses that were ambiguous between Typos and Other Errors, such as *how* for *howl* were also classified as Other Errors. 15 out of the 31 of Other Errors were ambiguous between Typos and Other Errors in the /d/ condition and 40 out of the 84 Other Errors were ambiguous in the /l/ condition. For the purposes of the analysis of accuracy, Intended An-

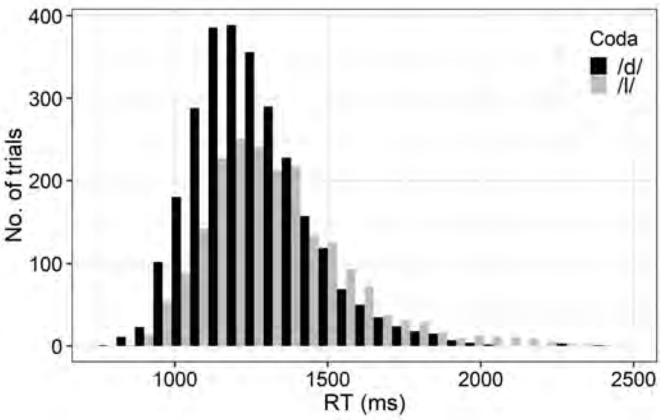


Figure 3.8: Distribution of RT (ms) for correct responses. Black bars: Coda /d/ condition. Grey bars: Coda /l/ condition.

swers, Phonetic Respellings and Typos were accepted as Correct; Minimal Pair Errors and Other Errors were rejected as Incorrect.

RT was measured from the onset of the stimulus to the first key-press. First, RT within 210 ms of stimulus onset (Woods et al. 2015) or above 5000 ms of stimulus onset (Baayen & Milin 2010) were excluded from further analysis (0.06% of responses), as were responses beyond mean ± 2 s.d. for each participant by coda condition (Ratcliff 1993), leaving a total of 5,591 trials (91%) for the analysis.

To measure the effect of coda /l/ on accuracy and speed of word recognition, we constructed two GLMs: one with the independent variable Accuracy and another with RT. The independent variables were Coda and Vowel (interacting) and Target Frequency (non-interacting); models included a random by-subject effect on the intercept. To explore whether the effect of Lexical Frequency on word recognition differs between coda conditions we created two models with the dependent variables Accuracy and RT and the independent variables Coda and Target Frequency (interacting), Vowel (non-interacting); models included a random by-subject effect on the intercept.[2]

For analysing the binary accuracy data, we used Generalised Linear Mixed-Effect Models (GLM) with the family binomial (Bates et al. 2015) and the BOBYQA optimiser (Powell 2009). $p - values$ were calculated with the *lmerTest* package (Kuznetsova et al. 2017) using Satterthwaite's degrees of freedom method. For analysing RT data, we used GLM models with the BOBYQA optimiser and the family gaussian with the logarithmic linking function because the distribution of RT was right-skewed and followed a log-normal distribution (**Figure 3.8**). In all four models Coda was treatment-coded, comparing /l/ to the baseline /d/. Vowel was deviation-coded, and the main effect of Vowel was investigated by comparing results for each vowel to the grand mean (instead of selecting one vowel as a baseline). Target Frequency was encoded as a continuous variable with the log-normalised per million words frequency of the target taken from the AusE section of GloWbE corpus (Davies 2013).

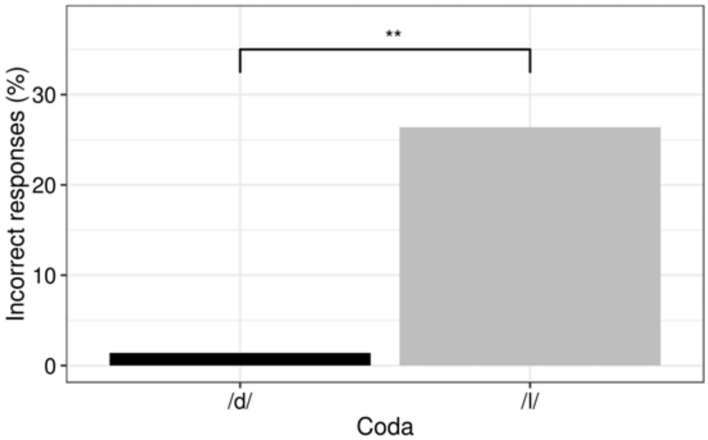
3.3.2 Results

/l/-final words were disambiguated less accurately ($\beta = -4.88$, $F(1, 5, 573)=306.26, p=0.01$) and more slowly ($\beta=0.07, F(1, 4, 797)=207.73$, $p < 0.001$) than /d/-final words (**Figure 3.9**).[3] To test that the accuracy results are due to confusion of minimal pairs, we repeated the analysis of accuracy data after removing responses classified as Other Errors and retaining only the responses classified as correct and Minimal Pair errors in a model with Coda, Vowel, and Lexical Frequency as

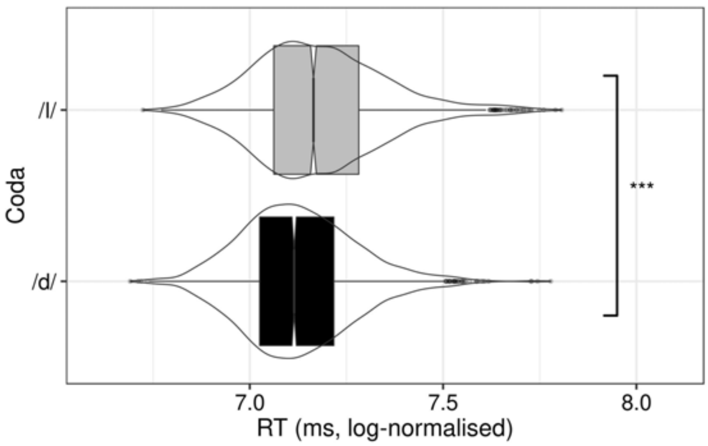
[2]The interaction effects between Coda and Target Frequency were tested in a separate model from the interaction effects between Coda and Vowel, as the stimuli were designed to test the latter, not the former.

[3]RT estimates are reported as log-normalised ms.

non-interacting factors. /l/-final words were disambiguated less accurately ($\beta = -4.99, F(1, 5, 469) = 147.43, p < 0.001$) with only Minimal Pair errors too (**Figure 3.10**).



(a) Percentage of inaccurate responses



(b) RT (ms, log-normalised)

Figure 3.9: Effect of Coda /l/ (grey) compared to Coda /d/ (black) on response accuracy and time.

Target vowels had no significant main effect on accuracy and did not show any significant interactions with Coda /l/. The lack of significant Vowel effects on accuracy is probably due to the fact that participans were at ceiling in the /d/ condition, therefore there was no variation between

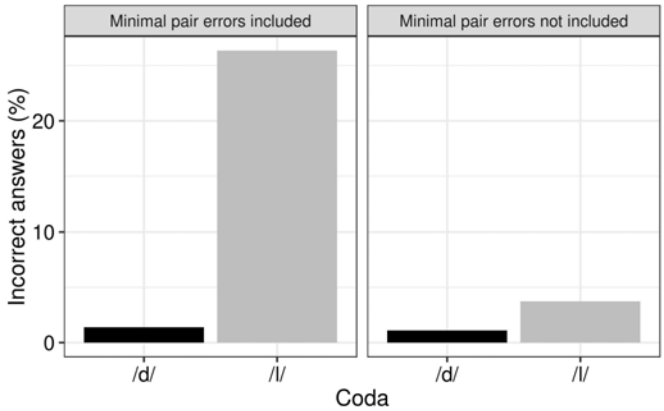


Figure 3.10: Percentage of inaccurate responses with and without minimal pair errors.

target Vowels in the /d/ condition. In addition, frequent and infrequent words were distributed unequally between the Coda and Vowel conditions therefore the different accuracy rates of different vowels (**Figure 3.11**) in the /l/ condition are better explained by the frequency of the carrier words than by vowel quality.

Target vowel significantly affected RT ($F(7, 4797)$=32.92). Response times for words containing the short target vowels /ɪ, ʊ, ɔ/ were significantly quicker than the grand mean, and response times to words containing long target vowels /iː, ʉː, æɔ/, but not /əʉ/, were slower than the grand mean (**Table 3.1**). Response times for word containing phonemically long vowels may have been slower because they were on average 132 ms longer than words containing short vowels, and RT was measured from acoustic stimulus onset.

Coda-Vowel interactions ($F(7, 4797)$=16.63) showed that the slowing effect of /l/ relative to /d/ was smaller on /iː, ɪ, ʉː, əʉ,/ and larger on /ɔ/ (**Table 3.2**, **Figure 3.11**).

Table 3.1: Vowel effects on RT. Top row: Estimate (β). Bottom row: p-value

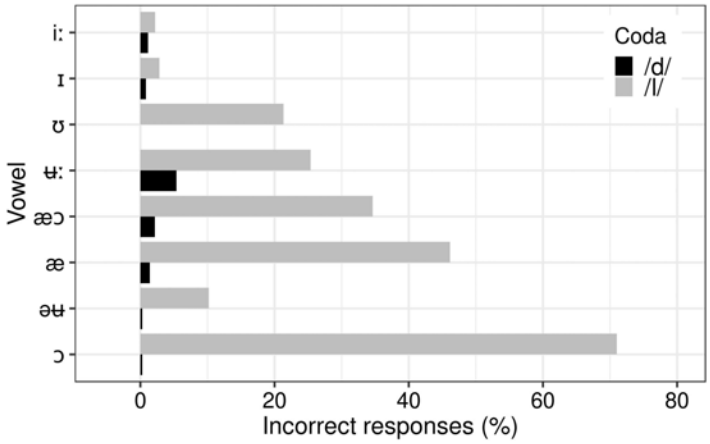
	iː	ɪ	ʉː	ʊ	æɔ	əʉ	ɔ
β	0.02	- 0.03	0.06	-0.04	0.08	-0.0001	-0.02
p	0.02	< 0.001	< 0.001	< 0.001	< 0.001	0.98	< 0.001

Table 3.2: Vowel-Coda /l/ interaction effects on RT. Vowel effects on RT. Top row: Estimate (β). Bottom row: $p - value$

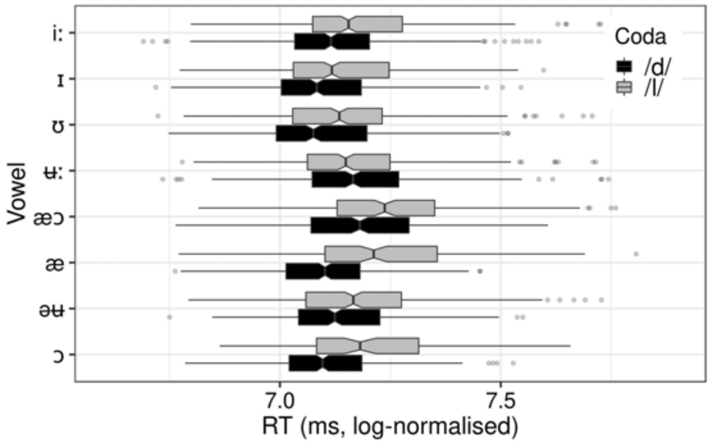
	iː	ɪ	ʉː	ʊ	æɔ	əʉ	ɔ
β	-0.05	-0.03	-0.09	-0.003	0.005	-0.02	0.07
p	< 0.001	0.01	< 0.001	0.74	0.61	0.04	< 0.001

Our first set of models contained an interaction between Coda and Vowel, but not between Coda and Lexical Frequency. These models suggested that more frequent words were disambiguated more accurately (β=0.18, $F(1, 5, 573)$=4.12, p=0.001) and more slowly (β=0.01, $F(1, 4, 797)$=8.74, $p < 0.001$), contrary to the established results on faster RT to more frequent words (Meunier & Segui 1999).

Our second set of models explored if the effect of Lexical Frequency differed between the coda /d/ and /l/ conditions, and therefore contained an interaction between Coda and Lexical Frequency, but not between Coda and Vowel. More frequent words were disambiguated more quickly ($\beta = -0.01$, $F(1, 4, 803)$=78.79, $p < 0.001$) and the effect was bigger in the Coda /l/ condition than in the /d/ condition (β=0.001, $F(1, 4, 803)$=7.97, p=0.004) (**Figure 3.12**). This result stands in apparent contrast to the result from our first set of models but it is in line with established frequency effects (Meunier & Segui 1999). The fact that the slowing effect of increased frequency disappears when frequency and coda interact, indicates that the lack of an interaction between Coda and Lexical Frequency may have resulted in the spurious result of longer RT to frequent words in the first set of models.



(a) Percentage of inaccurate responses by coda and vowel

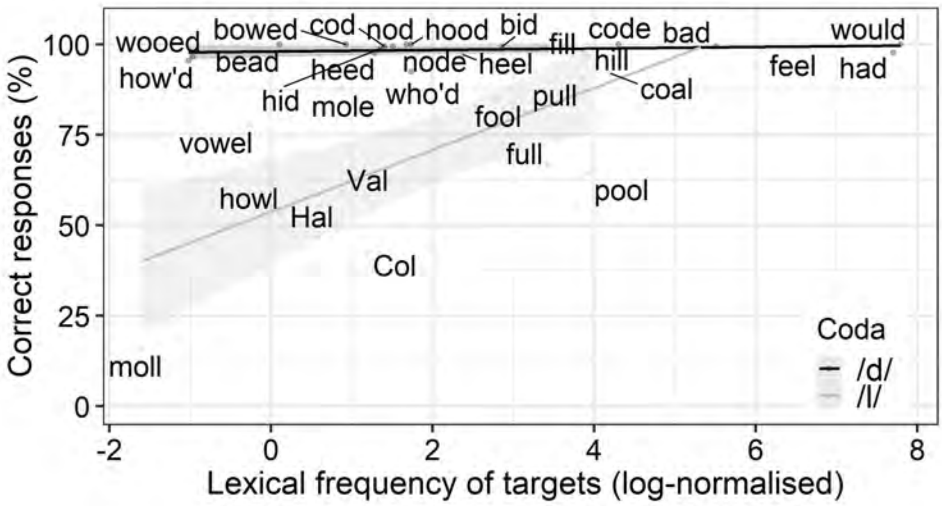


(b) RT of responses by coda and vowel

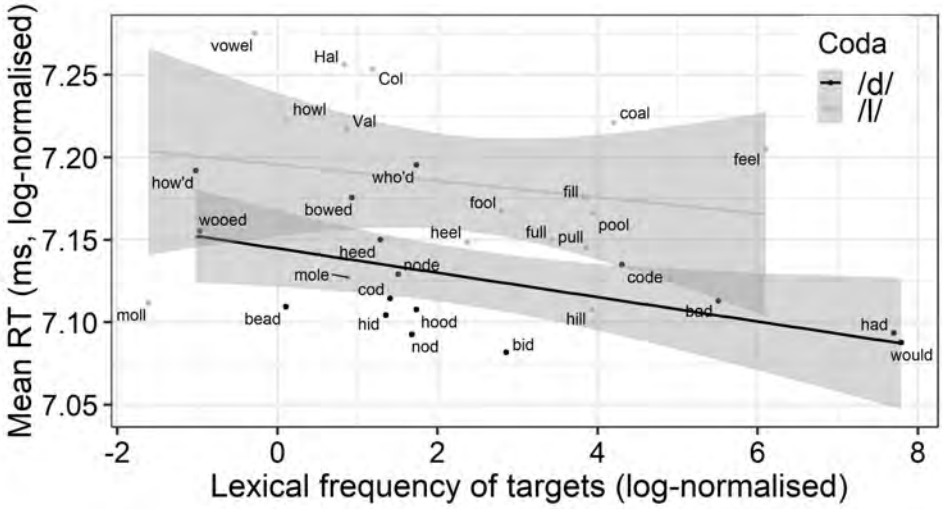
Figure 3.11: Effect of Coda and Vowel on the accuracy and RT of responses

The slower RT of more frequent words may also be caused by the fact that lexical frequency was not balanced between Coda and Vowel Conditions.

More frequent words were disambiguated more accurately (β=0.2, $F(1, 5, 579)$=227.73, p=0.01), as in the first set of models. There was no significant difference between the effect of frequency on the recognition of words ending in /l/ compared to words ending in /d/. Qualitative analysis of individual responses revealed whether participants prefer one member within the minimal pairs. Yet, the number of minimal pair errors exceed the number of correct responses only for *moll* and *Col*, indicating that participants responded with the more frequent words *mole* and *coal* to both members of these pairs (**Figure 3.13**). Other minimal pairs do not show a pattern that would indicate a default response (**Figure 3.13**).



(a) Percentage of accurate responses by lexical frequency and coda



(b) RT of responses by lexical frequency and coda

Figure 3.12: Effect of Coda and Vowel on the accuracy and RT of responses

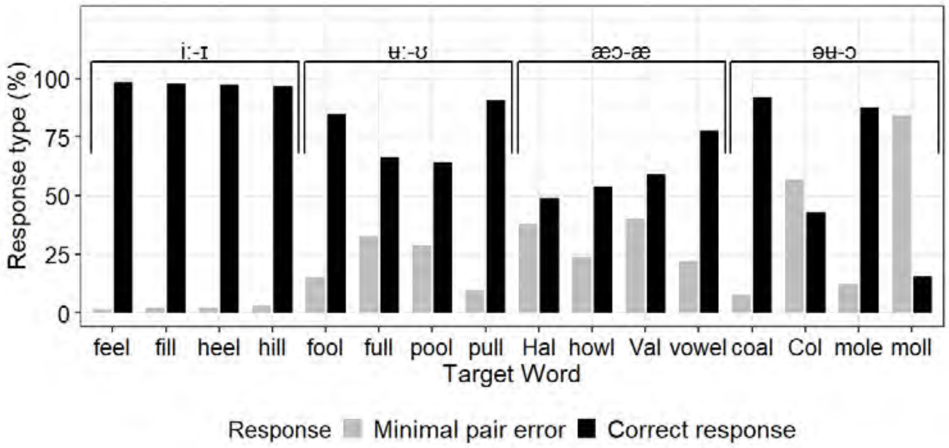


Figure 3.13: Correct responses and minimal pair errors by Target word in the /l/-condition

3.3.3 Discussion

The goal of this experiment was to gauge how listeners use word-level information when they compensate for coarticulation in /l/-final words. We found significantly less accurate and slower word recognition in /l/-final words compared to /d/-final words. The lower accuracy rates in the /l/ condition were driven by listeners tendency to confuse minimal pair competitors, a pattern that did not occur in the /d/ condition. We found that increased lexical frequency facilitated word recognition.

These findings that listeners sometimes map the acoustic signal inefficiently and even incorrectly indicate that listeners do not always compensate for the coarticulatory effect of /l/. That is, listeners cannot accommodate efficiently for vowel-lateral coarticulation in some minimal pairs. Instead of compensating for coarticulatory effects by attributing them to coda /l/, listeners may sometimes interpret coarticulatory effects as specific to the vowel instead of to the coda. This is despite the fact that typed responses showed that listeners identified the words as /l/-final. That is, listeners perceive the motivating environment for coarticulation but do not always compensate for its effect on the vowel.

The only two target words in which listeners sometimes missed the motivating environment were *howl* (/hæɔl/) and *Hal* (/hæl/), both of which were perceived as *how* (/hæɔ/) in respectively 22% and 13% of trials. Confusion of [æɔɫ#] and [æɫ#] with /æɔ#/ is not unexpected, given that the dorsal articulation of coda /l/ is inherently similar to that of a back vowel (Gick et al. 2002) and that acoustically, final /l/ can be absorbed in the preceding /æɔ/ (Palethorpe & Cox 2003). Furthermore, /l/-vocalisation is common after back vowels in AusE (Horvath & Horvath 1997, Borowsky 2001), which further increases the similarity between *howl* and *how*. In contrast, the low front /æ/ in *Hal* facilitates vocalisation to a lesser extent (Horvath & Horvath 1997, Borowsky 2001), but if listeners perceive final /l/ as a vowel (i.e. vocalised), it is very likely to be perceived as /ɔ/ due to the correspondence between /æɔ/ and /æl/ (Palethorpe & Cox 2003).

We did not find that the effect of /l/ on accuracy differed between words with different target vowels, despite an apparent difference in recognition accuracy (**Figure 3.11**). We detected a difference in the slowing effect of /l/ between words with different target vowels: the effect was smaller for /iː, ɪ, ʉː, əʉ/ and larger for /ɔ/, indicating increased difficulty for targets with /ɔ/. Overall vowel effects showed that words with short vowels were recognised more quickly compared to words with long vowels. The vowel effect could be the result of listeners waiting until stimulus offset, therefore taking longer to respond to stimuli with long vowels (mean length = 588 ms) compared to short vowels (mean length = 456 ms).

Frequent words were recognised more accurately but more slowly, when interactions between Coda and Target vowel were examined, partly consistent with previous findings (Morton 1969, Meunier & Segui 1999). The apparent slowing effect of increased lexical frequency might be the artefact of the stimuli not being balanced for lexical frequency. When interactions between Coda and Lexical Frequency were included in the model, frequent words were recognised more accurately and quickly, consistent with previous findings (Morton 1969, Meunier & Segui 1999). Effects of lexical frequency on RT were stronger in the /l/ condition compared to the /d/ condition, while no such differential effect of frequency was observed for accuracy. Nevertheless, an exploratory analysis revealed that only two minimal pairs in the /l/ condition were characterised by a listener preference for the frequent member of the minimal pair in the case of large frequency discrepancies: *mole-moll* (2.42 versus 0.2 occurences per million words (Davies 2013)) and *coal-Col* (66.96 versus 3.3 occurences per million words (Davies 2013)). That is, when the target was very infrequent, *Col* or *mol*, listeners defaulted to the more frequent minimal pair competitor, *coal* and *mole* instead of compensating for coarticulation. In addition, the slowing effect of /l/ was bigger for *Col* and *mol* than the overall slowing effect of /l/. This could be related to participants' unfamiliarity with the targets *Col* and *mol* (Connine et al. 1990), as some participants flagged the words *moll, Col, Hal, Val* as "unknown" or even "nonsense" words in the exit interview, but did not flag their minimally differing competitor. For the remaining targets, the percentage of correct responses exceeded minimal pair errors, showing that listeners somewhat compensated for coarticulation.

Lower accuracy and slower speed of recognition of lateral-final words indicate increased processing difficulty, which we attribute to the reduced acoustic contrast between the members of the minimal pairs. Reduced acoustic contrast can make word recognition harder by making the acoustic signal inherently ambiguous in perception. Furthermore, reduced acoustic contrast can also increase lexical activation of minimal pair competitors in the /l/-context compared to the /d/-context, which inhibits the recognition of the target (Luce & Pisoni 1998). That is, vowel-/l/ coarticulation does not only lead to increased processing difficulty, as shown in Experiment 1, but also hinders lexical access and limits listeners' ability to compensate. Listeners' minimal pair errors show that they mapped the acoustic signal to the competitor word instead of the target, indicating that CVl minimal pairs ending in /ʉːl-ʊl, əʉl-ɔl, æɔl-æl/ are inherently ambiguous between two lexical items.

3.4 Conclusion

The results of Experiment 1 and 2 combined show the limits of listeners' ability to compensate for the coarticulatory effects of /l/ on pre-/l/ vowels in /l/-final words. In Experiment 1, we found reduced perceptual contrast between the vowel-pairs /ʉː-ʊ, æɔ-æ, əʉ-ɔ/ which we attribute to the ambiguity of the acoustic signal. This is supported by the fact that vowels with similar place of articulation are confused with each other in the /d/ and increasingly so in the /l/ condition. Vowel cues are modified by the coarticulatory influence of the coda /l/ in such a way that contrastive cues are masked and the signal becomes ambiguous between two elements in the vowel inventory. In Experiment 2, we found that reduced perceptual vowel contrast and vowel ambiguity caused by the coarticulatory effects of /l/ also hinder lexical access and recognition of /l/-final minimal pairs contrasting /iː-ɪ, ʉː-ʊ, æɔ-æ/ and /əʉ-ɔ/. That is, listeners cannot always compensate for the effects of /l/, despite perceiving /l/ itself. Listeners' ability to compensate for coarticulation is limited by frequency of the word: for infrequent words, listeners map an ambiguous signal to a frequent competitor instead of compensating for coarticulation. The increase in vowel ambiguity and decrease in frequency makes the competing minimally different word more plausible and hinders listeners ability to compensate for coarticulation. We found hindered ability to compensate for coarticulation showing that listeners attribute the influence of /l/ on /l/-final rimes to intrinsic properties of the preceding vowel in the perception of /l/-final rimes. Thus they map the acoustic signal to a word that was not the speakers' intended target. The two experiments together show that vowel-lateral coarticulation limits listeners' ability to compensate for coarticulation both in vowel disambiguation and in word recognition.

Limited compensation for coarticulation has implications for theories of sound change, as sound change is often related to how coarticulation is produced by the speaker and perceived by the listener (Ohala 1993, Beddor 2009, Solé & Ohala 2010, Ohala 2012, Garrett & Johnson 2013, Harrington et al. 2018). Coarticulation provides systematic and directional variation which may become the input for sound change (Garrett & Johnson 2013). Ohala's (1981; 1993; 2012) model of sound change specifically identifies insufficient compensation for coarticulation, not its production, as a process implicated in the initiation of sound change. Listeners, on hearing a coarticulated speech signal, can either compensate for coarticulation, retracing the acoustic signal to the speaker's intended form or they may not compensate for coarticulation, and take the utterance at face value (Ohala 1981). The former scenario does not lead to sound change, whereas the latter may result in a change in the pronunciation norm, which in some instances may provide favourable conditions for sound change to occur. Viewed through this model, listeners' limited ability to compensate for /l/-influence on vowels in these data may be a precursor to a sound change, as listeners do not always retrace the acoustic signal to the speakers' intended form.

In contrast, in Harrington et al. (2018)'s interactive phonetic (IP) sound change model, insufficient compensation for coarticulation is not the cause, but the effect of and evidence for sound change. In the IP model, the prerequisite of sound change is that typical realisations of two phonemes are acoustically distinct, but highly coarticulated realisations of one phoneme become acoustically similar to the other phoneme (Harrington et al. 2018). As listeners and speakers interact, atypical realisations are incorporated to the phoneme's representation, shifting its boundary closer to the second phoneme, and potentially leading to a merger. This merger is signalled by failed compensation for coarticulation (Harrington et al. 2018). According to the IP model, the fact that the pre-lateral allophones of /ʉː, æɔ, əʉ/ are acoustically more similar to the pre-lateral allophones of /ʊ, æ, ɔ / respectively and less similar to their own pre-obstruent allophones (Palethorpe & Cox 2003) creates the necessary prerequisite of sound change. Failed compensation for coarticulation for these vowels indicates that the pre-lateral allophones of the vowel pairs /ʉː-ʊ, æɔ-æ, əʉ-ɔ/ might have merged, although we did not find the perception of pre-/l/ allophones of the vowel pairs to be skewed towards one phonemic category within the pair.

However, not all allophonic variation in production leads to sound change (Ohala 1993) and failed compensation or miscategorisation of items does not always indicate sound change (Stevens & Harrington 2014, Harrington et al. 2018). In order to explore this question, an apparent time or a sociolinguistic study is needed to better understand the implications for the actuation of sound change in the pre-lateral vowels of Australian English.

3.5 Acknowledgements

This research was supported in part by ARC DE150100318, iMQRTP 2015144, and MQSIS 9201501719 grants. We thank Peter Humburg for his help in the statistical analysis, the Phonetics Lab, the Centre for Language Sciences and the ARC Centre of Excellence in Cognition and its Disorders at Macquarie University for their comments, feedback and support. We would also like to thank our participants without whom this research would not have been possible.

3.6 Appendices

3.6.1 Target words in Experiment 1

Table 3.3: Carrier words for the 16 target vowels Left columns: orthographic representation and IPA transcription of carrier words in the /l/ condition. Right columns: orthographic representation and IPA transcription of carrier words in the /d/ condition.

Coda /d/		Coda /l/	
Candidate	Transcription	Candidate	Transcription
heel	hiːl	*heed*	hiːd
hill	hɪl	*hid*	hɪd
hell	hel	*head*	hed
hal	hæl	*had*	hæd
hule	hʉːl	*hude*	hʉːd
hurl	hɜːl	*herd*	hɜːd
harl	hɐːl	*hard*	hɐːd
hull	hɐl	*hud*	hɐd
hool	hʊl	*hood*	hʊd
hall	hoːl	*horde*	hoːd
holl	hɔl	*hod*	hɔd
hail	hæɪl	*hade*	hæɪd
hile	hɑel	*hide*	hɑed
hoil	hɔɪl	*hoyd*	hɔɪd
howl	hæɔl	*howd*	hæɔd
hole	həʉl	*hode*	həʉd

3.6.2 Target formant trajectories for /ʉː-ʊ, æɔ-æ, əʉ-ɔ/ in Experiment 1

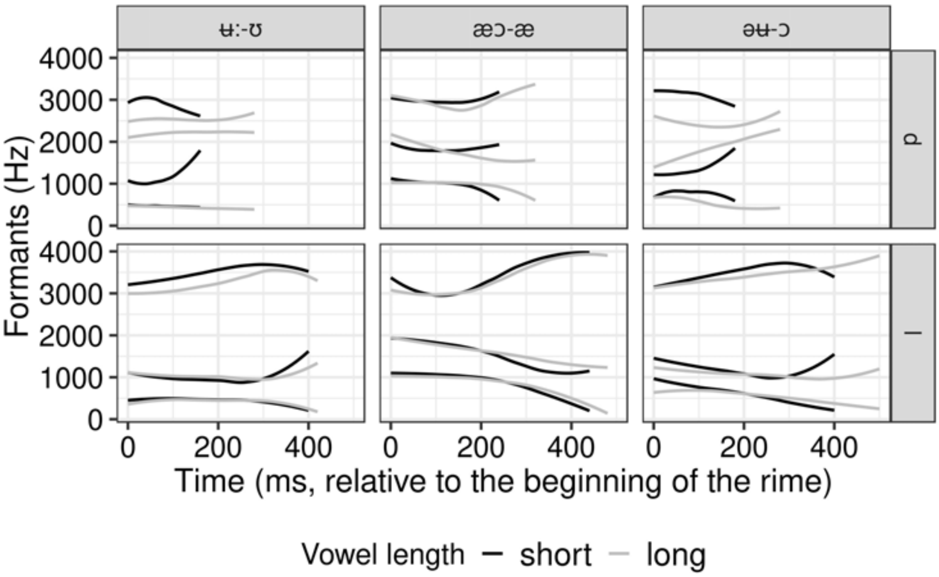


Figure 3.14: Smoothed formant trajectories (F1, F2, F3) measured at every 20 ms. Top row: /ʉː-ʊ, æɔ, əʉ- ɔ/ in pre-obstruent context. Bottom row: /ʉːl-ʊl, æɔl-æl, əʉl- ɔl/ rimes.

3.6.3 Acoustic duration for /ʉː-ʊ, æɔ-æ, əʉ-ɔ in Experiment 1

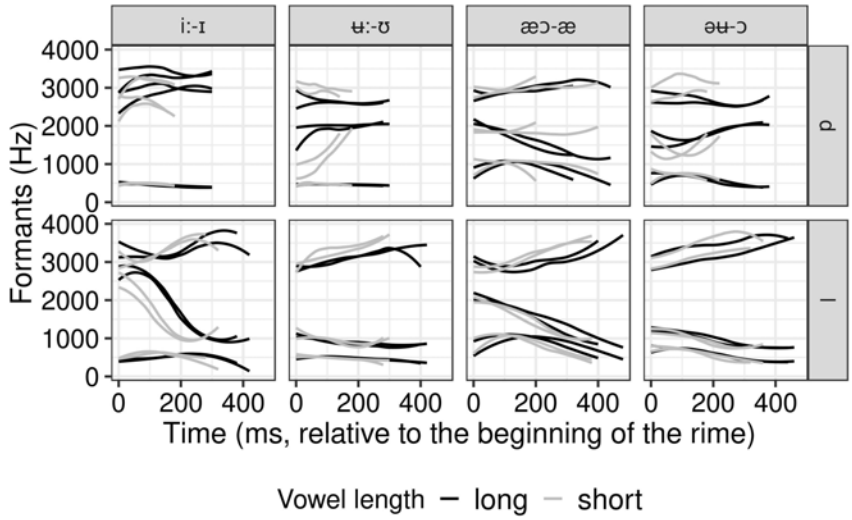
Table 3.4: Top row: vowel duration in /d/-final words (ms). Bottom row: rime duration in /l/-final words (ms).

	ʉː	ʊ	æɔ	æ	əʉ	ɔ
/d/-final	330	170	330	240	280	180
/l/-final	430	410	490	400	500	400

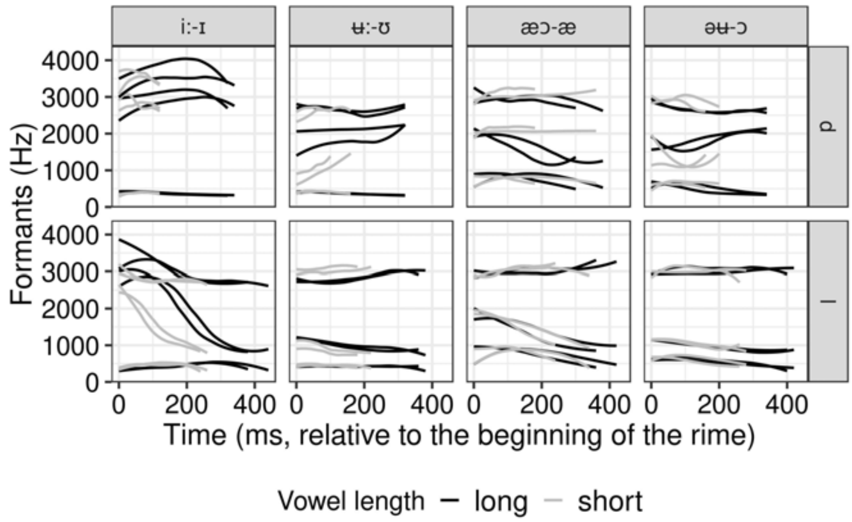
Table 3.5: Ratio of vowel duration (short/long). Top row: vowels in /d/-final words. Bottom row: /l/-final rimes

	ʉː-ʊ	æɔ-æ	əʉ-ɔ
/d/-final	0.52	0.73	0.64
/l/-final	0.95	0.91	0.80

3.6.4 Target formant trajectories in Experiment 2



(a) Speaker 1 (25 yrs old)



(b) Speaker 2 (57 yrs old)

Figure 3.15: Smoothed formant trajectories (F1, F2, F3) for stimulus items measured at every 20 ms. Top row: formant measurement in pre-obstruent vowels. Bottom row: formant measurements in /l/-final rimes.

3.7 References

Baayen, H. R. & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 785–821.

Beddor, P. S. & Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *The Journal of the Acoustical Society of America*, *106*(5), 2868–2887.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W. & Brasher, A. (2013, April). The time course of perception of coarticulation. *Journal of the Acoustical Society of America*, *133*(4), 2350–2366.

Beddor, P. S. & Strange, W. (1982). Cross-language study of perception of the oral–nasal distinction. *The Journal of the Acoustical Society of America*, *71*(6), 1551–1561.

Bennett, D. C. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech*, *11*(2), 65–85.

Borowsky, T. (2001). The vocalisation of dark *l* in Australian English. In D. Blair & P. Collins (Eds.), *English in Australia* (p. 69–87). Philadelphia, Amsterdam: John Benjamins Publishing Company.

Bradlow, A. R., Torretta, G. M. & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech communication*, *20*(3-4), 255–272.

Connine, C. M., Mullennix, J., Shernoff, E. & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1084–1096.

Cox, F. (1999). Vowel change in Australian English. *Phonetica*, *56*, 1–27.

Cox, F. (2006). The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics*, *26*(2), 147–179.

Cox, F. & Fletcher, J. (2017). *Australian English pronunciation and transcription.* Cambridge: Cambridge University Press.

Cox, F. & Palethorpe, S. (2004). The border effect: Vowel differences across the NSW-Victorian border. In C. Moskovsky (Ed.), *Proceedings of the conference of Australian Linguistics Society.* Newcastle, Australia.

Cox, F. & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association*, *37*(3), 341–350.

Davies, M. (2008). *The corpus of contemporary American English.* BYE, Brigham Young University.

Davies, M. (2013). *Corpus of global web-based English: 1.9 billion words from speakers in 20 countries.*

Ferguson, S. H. & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, *50*(5), 1241–1255.

Forster, K. I. & Chambers, M., Susan. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behaviour*, *12*, 627–635.

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, *36*(4), 359–368.

Fowler, C. A. (2005). Parsing coarticulated speech in perception: Effects of coarticulation resistance. *Journal of Phonetics*, *33*(2), 199–213.

Fowler, C. A., Brown, J. M. & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human perception and performance*, *26*(3), 877.

Garrett, A. & Johnson, K. (2013). Phonetic bias in sound change. In A. C. L. Yu (Ed.), *Origins of sound change* (p. 51–97). Oxford, UK: Oxford University Press.

Gaskell, M. G. & Marslen-Wilson, W. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 380–396.

Gick, B., Kang, M. A. & Whalen, D. H. (2002). MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics*, *30*(3), 357–371.

Harrington, J., Cox, F. & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics*, *17*(155-184).

Harrington, J., Kleber, F., Reubold, U., Schiel, F. & Stevens, M. (2018). Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in cognitive science*, *10*(4), 707–728.

Harrington, J., Kleber, F. & Stevens, M. (2016). The relationship between the (mis)-parsing of coarticulation in perception and sound change: Evidence from dissimilation and language acquisition. In *Recent advances in nonlinear speech processing* (p. 15–34). Cham: Springer International Publishing.

Horvath, B. M. & Horvath, R. J. (1997). The geolinguistics of a sound change in progress: /l/ vocalization in Australia. *U. Penn Working Papers in Linguistics*, *4*(1), 109–124.

Iskarous, K., Mooshammer, C., Hoole, P., Recasens, D., Shadle, C. H., Saltzman, E. et al. (2013). The coarticulation/invariance scale: Mutual information as a measure of coarticulation resistance, motor synergy, and articulatory invariance. *The Journal of the acoustical society of America*, *134*(2), 1271–1282.

Kleber, F., Harrington, J. & Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, *55*(3), 383–405.

Krause, F. & Lindemann, O. (2014, 1st Jun). Expyriment: A python library for cognitive and neuroscientific experiments. *Behavior Research Methods*, *46*(2), 416–428.

Krause, J. C. & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, *115*(1), 362–378.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Lin, S., Palethorpe, S. & Cox, F. (2012). An ultrasound exploration of Australian English /CVl/ words. In *14th Australasian International Conference on Speech Science and Technology* (p. 105–108). Sydney.

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

Loakes, D., Clothier, J., Hajek, J. & Fletcher, J. (2014a). Identifying /el/-/æl/: A comparison between two regional Australian towns. In J. Hay & E. Parnell (Eds.), *15th australasian international conference on speech science and technology.* Christchurch , New Zealand.

Loakes, D., Clothier, J., Hajek, J. & Fletcher, J. (2014b). An investigation of the /el/-/æl/ merger in Australian English: A pilot study on production and perception in South-West Victoria. *Australian Journal of Linguistics*, *34*(4), 436–452.

Loakes, D., Graetzer, N., Hajek, J. & Fletcher, J. (2012). Vowel perception in Victoria: Variability, confusability and listener expectation. In *Proceedings of the 14th australasian international conference on speech science and technology* (Vol. 14). Macquarie University, Sydney.

Loakes, D., Hajek, J. & Fletcher, J. (2010a). The /el/-/æl/ sound change in Australian English: A preliminary perception experiment. In Y. Treis & R. De Busser (Eds.), *Selected papers from the 2009 conference of the Australian Linguistic Society.* Melbourne, Australia.

Loakes, D., Hajek, J. & Fletcher, J. (2010b). Issues in the perception of the /el/-/æl/ contrast in Melbourne: Perception, production and lexical frequency effects. In *Proceedings of 13th australasian international conference on speech science and technology.* Melbourne, Australia.

Loakes, D., Hajek, J. & Fletcher, J. (2010c). (Mis)perceiving /el/-/æl/ in Melbourne English: A micro-analysis of sound perception and change. In *Proceedings of 13th australasian international conference on speech science and technology.* Melbourne, Australia.

Loakes, D., Hajek, J. & Fletcher, J. (2011). /æl/-/el/ transposition in Australian English: Hypercorrection or a competing sound change? In *Proceedings of the 17th international congress of phonetic sciences.* City University of Hong Kong.

Lotto, A. J. & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*(4), 602–619.

Luce, P. A. & Pisoni, D. B. (1998, Feb). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*, 1–36.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, *28*(5), 407–412.

Mann, V. A. & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*(3), 213–228.

Meunier, F. & Segui, J. (1999). Frequency effects in auditory word recognition: The case of suffixed words. *Journal of Memory and Language*, *41*, 327–344.

Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics*, *68*(7), 1227–1240.

Morton, J. (1969). Interaction of information in word recognition. *Psychological review*, *76*(2), 165.

Neel, A. T. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research*, *51*(3), 574–585.

Ohala, J. J. (1981, May). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (Eds.), *Papers from the parasession on language and behaviour.* Chicago.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (p. 237–278). London: Longman.

Ohala, J. J. (2012). The listener as a source of sound change: An update. In M.-J. Solé & D. Recasens (Eds.), *The initiation of sound change* (p. 21–36). Amserdam; Philadelphia: John Benjamins Publishing Company.

Ollman, R. (1966). Fast guesses in choice reaction time. *Psychonomic Science*, *6*(4), 155–156.

Palethorpe, S. & Cox, F. (2003). Vowel modification in pre-lateral environments. In *International seminar on speech production.* Sydney.

Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge*, 26–46.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532.

Recasens, D. (2002). An EMA study of VCV coarticulatory direction. *The Journal of the Acoustical Society of America*, *111*(6), 2828–2841.

Rubenstein, H., Garfield, L. & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 487–494.

Segui, J., Mehler, J., Frauenfelder, U. & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia*, *20*(6), 615–627.

Smits, R. (2001). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology*, *27*(5), 1145–1162.

Solé, M.-J. & Ohala, J. J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. *Papers in laboratory phonology*, *10*, 607–655.

Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*(3), 291–311.

Stevens, K. N. & Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *Journal of Phonetics*, *38*(1), 10–19.

Stevens, M. & Harrington, J. (2014). The individual and the actuation of sound change. *Loquens*, *1*(1), e003.

Thomas, B. & Hay, J. (2005). A pleasant malady: The Ellen/Allan merger in New Zealand English. *Te Reo*, *48*, 69–93.

Wade, L. (2017). The role of duration in the perception of vowel merger. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1).

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244.

West, P. (1999). Perception of distributed coarticulatory properties of English /l/ and /r/. *Journal of Phonetics*, *27*(4), 405–426.

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J. & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience*, *9*(131), 1–12.

Yellott Jr, J. I. (1971). Correction for fast guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology*, *8*(2), 159–199.

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, *61*, 13–29.

4 Perception and production of duration contrast in Australian English lateral-final rimes

This chapter is based on the following published paper:

Szalay, T., Benders, T., Cox, F., & Proctor, M. (2018). "Production and perception of length contrast in lateral-final rimes", in J. Epps, J. Wolfe, J. Smith, and C. Jones (Eds.) *Proceedings of the 17th Australasian International Conference on Speech Science and Technology*, (p. 127–132), Sydney, Australia.

I certify that I was responsible for the development of the concept of this paper, in discussion with my supervisors/co-authors. Data was collected as part of a larger project: detailed perceptual analyses are reported in **Chapter 3** and perceptual-acoustic analyses are reported in the current chapter. I took leadership in conducting the research, and was responsible for data collection and analysis, and the writing of all parts of the paper. My co-authors provided advice to improve the experimental design and methods, the analyses, the interpretation of the data, as well as the presentation of the written component. I have also received feedback from the reviewers and the audience of the 17th Australasian International Conference on Speech Science and Technology.

Abstract

Words containing /l/-final rimes challenge listeners as coda /l/ reduces certain vowel contrasts. Lateral-final rimes therefore allow us to gauge the link between individuals' word recognition and production. We tested whether participants producing a larger durational contrast between word pairs containing the rimes /iːl-ɪl, ʉːl-ʊl, æɔl-æl , əʉl-ɔl/ were better at recognising minimal pairs contrasting the aforementioned rimes. 46 Australian English speakers produced 24 /l/-final minimal pairs and identified the same minimal pairs spoken by two speakers. Participants producing a longer durational

contrast took longer to respond and were only more accurate when the stimuli contained a bigger durational contrast.

Keywords: durational vowel contrast, production, perception, lateral-final rimes, Australian English

4.1 Introduction

A growing body of experimental evidence shows that individuals' speech production and perception are linked (Newman 2003, Perkell, Matthies et al. 2004, Perkell, Guenther et al. 2004, Zellou 2017). Listeners who robustly produce a contrast are better able to perceive the same contrast than listeners with less robust contrast production (Newman 2003, Perkell, Matthies et al. 2004, Perkell, Guenther et al. 2004, Zellou 2017). For instance, English listeners who more accurately differentiate voiced from voiceless stops also produce longer voice onset time (Newman 2003). Listeners who are better at discriminating the /s-ʃ/ contrast in perception maintain a more consistent tongue-tip contrast in production (Perkell, Matthies et al. 2004). Listeners who are better at discriminating /ɑ-ʌ, u-ʊ/ produce greater spectral differentiation between members within these vowel pairs (Perkell, Guenther et al. 2004).

In perception, the phonological contrast between vowels is cued by several acoustic cues, i.e. formant values of vowel targets (Bennett 1968), vowel inherent formant change (Nearey & Assmann 1986), and duration (Bennett 1968). English, including Australian English, listeners rely more on spectral than durational contrast and use durational contrast only when spectral contrast is diminished or unavailable (Bennett 1968, Liu 2016). That is, in English, spectrally similar vowels are more likely to be confused (Neel 2008) even when they differ in length (Szalay et al. 2016).

Spectral contrast is weighted more heavily than durational contrast at an individual level (Wade 2017). Listeners from Burlington, Vermont, a speech community where spectral contrast is maintained between the vowels in PULL-POOL-POLE in the pre-/l/ context cannot discriminate these vowels in the speech of the Youngstown, Ohio, speech community, where only durational contrast is maintained (Wade 2017). However, speakers who reduce spectral difference but maintain durational difference in production can utilise durational cues in perception even when spectral cues are not available (Wade 2017). This indicates that listeners rely on the same cues in perception which they produce and perceive as phonologically contrastive. In contrast, in the production and perception of voiced and voiceless consonants, listeners were found to weight voice onset time and $f0$ differently (Shultz et al. 2012).

The aforementioned studies tested contrast perception on manipulated stimuli in a continuum, therefore little is known about if and how contrast production is associated with listeners' ability to cope with variation in unmanipulated speech. To further our understanding of the production-perception link, this study examined if and how contrast production is associated with word recognition and processing in Australian English (AusE) lateral-final rimes.

The AusE vowel inventory contrasts 18 stressed vowels, using both spectral and durational contrasts (Cox 2006). Some vowel pairs differentiated by duration exhibit smaller spectral differences (e.g. /ɐː-ɐ, iː-ɪ/, in *cart-cut*, *beat-bit*), others exhibit bigger spectral contrast (e.g. /ʉː-ʊ/, in *kook-cook*) (Cox 2006). There are diphthong-monophthong pairs in which the first or the second target of the diphthong coincides with a monophthong (Cox 2006). These inherent spectral similarities increase vowel confusion (Szalay et al. 2016). Coda /l/ further reduces the spectral contrast between /iː-ɪ, ʉː-ʊ, æɔ-æ, əʉ-ɔ/ (e.g. *feel-fill, fool-full, howl-Hal, dole-doll*); however contrastive duration may be maintained (Palethorpe & Cox 2003). It is not clear if listeners can use durational differences in /l/-final rimes.

This study examined perception of duration contrast in CVl minimal pairs contrasting /iː-ɪ, ʉː-ʊ, æɔ-æ, əʉ-ɔ/ in the speech of two Source Speakers, one of whom maintains a more robust duration contrast than the other. The association between participants' production of the same duration contrast and their perception was tested in three hypotheses:

1. if AusE listeners rely on durational cues in /l/-final rimes, increased duration contrast in the stimuli would aid word recognition for all listeners regardless of their contrast production

2. if production and perception are linked, listeners producing a consistent length contrast would have an overall advantage in recognising /l/-final words that differ in the duration of the rime in the speech of both Source Speakers

3. if listeners rely more on cues that they themselves produce, then listeners who produce a more robust duration contrast would only perform better when the Source Speaker does so too.

4.2 Method

4.2.1 Participants

Forty-six female [mean age = 21.5, range = 18 − 40] native speakers of AusE participated in the study[1]. All participants were born in Australia

[1]The participants and the stimuli in the present chapter are the same as in **Chapter 3**, Experiment 2.

to Australian-born parents. None of the participants reported any reading, hearing, or speaking disorders. Participants received course credit or $15 for participation.

4.2.2 Materials

The stimuli consisted of 32 CVC targets and 38 (C)V(C) fillers. 4 vowel pairs (/iː-ɪ, ʉː-ʊ, æɔ-æ, əʉ-ɔ/) were embedded in two sets of /l/-final and two sets of /d/-final minimal pairs to create 32 target words (see **Table 4.1** for the /l/-final words). Here we analyse only production and perception data of /l/-final words.

Table 4.1: *Target words ending in /l/*

Vowel pair			
/iː-ɪ/	/ʉː-u/	/æɔ-æ/	/əʉ-ɔ/
feel-fill,	*fool-full,*	*howl-Hal,*	*mole-moll,*
heel-hill	*pool-pull*	*vowel-Val*	*coal-Col*

To create the stimuli for the perception experiment, targets and fillers were read by two female native speakers (Source Speakers) of AusE upon orthographic random presentation on a computer monitor. Source Speaker 1 was 25, and Source Speaker 2 was 57 years old at the time of the recording. All stimuli were recorded with an AKG C535EB Condenser Microphone onto an iMac using Presonus Studio Live 16.2.4 AT Mixer in a sound treated studio. Stimuli were recorded at 44.1 KHz, amplitude-normalised, truncated to have 1 s silence before and after the word, and digitised as 16 bit WAV files.

Long:short rime duration ratios were calculated for the vowel-pairs /iː-ɪ, ʉː-ʊ, æɔ-æ, əʉ-ɔ/ from the experimental stimuli produced by the two Source Speakers (**Table 4.2**). Source Speaker 2 maintained a bigger long:short ratio, therefore maintained a bigger duration contrast for all vowel pairs except /æɔ-æ/.

Table 4.2: *Long:short rime duration ratios in the stimuli*

Informant	Vowel pair			
	/iː-ɪ/	/ʉː-u/	/æɔ-æ/	/əʉ-ɔ/
Source Speaker 1	1.27	1.3	1.23	1.23
Source Speaker 2	1.47	1.45	1.23	1.42

4.2.3 Procedure

The experiment consisted of a production task followed by a perception task, carried out in a one hour long session in a sound treated studio at Macquarie University, Sydney NSW. Participants were tested individually in the presence of the experimenter.

Firstly, participants read orthographically presented words aloud. Words were pseudo-randomised, presented one by one three times in three blocks and recorded with an AKG C535EB Condenser Microphone onto an iMac using Presonus Studio Live 16.2.4 AT Mixer. The production task helped participants familiarise with the stimuli for the perception task.

Next, participants carried out the perception task, consisting of a practice phase and a test phase. In the practice phase, 10 single words were individually presented auditorily. Participants were asked to type the word that they heard quickly and accurately and received immediate feedback on what the correct responses were. In the test phase, participants were presented with individual words auditorily and were asked to type the words as they perceived them as quickly and accurately as possible. First, participants heard the words spoken by Source Speaker 2, repeated twice in two blocks, and then by Source Speaker 1, repeated twice in two blocks; blocks were separated by 30 s long forced break. Items within a block were pseudo-randomised so that no /l/-final words followed each other. Stimuli were presented with Expyriment (Krause & Lindemann 2014) on an Asus X550JX laptop. Audio stimulus was presented via Sennheiser 380 Pro headphones at participants' preferred listening level. Participants' responses accuracy and response time (RT) of the first keypress were measured. After the word recognition task, participants were asked to fill out a self-evaluation questionnaire.

4.3 Data analysis

4.3.1 Production data

Recordings were segmented automatically (Schiel 1999); rime durations were extracted automatically (Boersma & Weenink 2013). Rime duration is a measure combining vowel and coda /l/ length. Duration values 1.5 times above or below the interquartile range for a given vowel were hand-checked and corrected for measurement errors.

Mean rime duration was calculated by participant and vowel. The ratio of long:short vowels for each vowel pair and for each participant was calculated; increased ratio indicates an increased duration contrast.

4.3.2 Perception data

Responses to 46 (participants) $\times$ 64 (/l/-final tokens) = 2944 trials were collected. Responses were rated for accuracy. Responses were classified as Intended Answer, Phonetic Respelling, Typo, Minimal Pair Error, and Other Error. Responses were classified as Intended Answer if spelled as the target. Unambiguous but nonstandard phonetic spellings (e.g. *cole* for *coal*) were classified as Phonetic Respellings. Single letter deletions, additions, letter transpositions, and substitutions within one key distance of the target letter were classified as Typos (Luce & Pisoni 1998), unless the result was an English lexical item. Confusion of members of minimal pairs (e.g. *fool* for *full*) was classified as Minimal Pair Error. Any other error (e.g. *cool* for *pool*, *howled* for *howl*) were classified as Other Error. For the purposes of the analysis of accuracy, Intended Answers, Phonetic Respellings and Typos were accepted as Correct; Minimal Pair Errors and Other Errors were classified as Incorrect.

RT of the first keypress was collected. RT within 210 ms (Woods et al. 2015) or above 5000 ms (Baayen & Milin 2010) of stimulus onset were excluded from analysis. Individual RT exceeding or less than mean+±2*sd for each participant were excluded from analysis (Ratcliff 1993). 5.1% of responses were excluded according to these criteria, leaving 2,794 tokens for analysis.

4.4 Results

4.4.1 Individual variation in production and perception

Participants produced /l/-final rimes with a mean long:short ratio of 1.34 and a range of 0.99-1.38.[2] Participants consistently produced a decreasing durational contrast from /iː-ɪ/ to /ʉː-ʊ/ to /æɔ-æ/ to /əʉ-ɔ/. In the perception data, participants were consistent across the vowel pairs.

4.4.2 Production-perception link

To measure the association between accuracy, RT, and duration ratio, we constructed two Generalised Linear Mixed-effect models (Bates et al. 2015)

[2] Mean long:short vowel ratio was 1.64 in /d/-final rimes, as in (Cox 2006).

with the dependent variables Accuracy and RT. The independent variables were Participant Duration Ratio (long:short, scaled), Vowel Pair (contrast coded and each compared against the grand mean), Source Speaker (contrast coded), and Lexical Frequency (from (Davies 2013), log-normalised); Participant and Block were random intercepts. All two-way interactions between Duration Ratio, Vowel Pair, and Source Speaker were included in the model, but three-way interactions were not; Lexical Frequency did not interact with the other independent variables. Effects on accuracy were tested using the binomial family and effects on RT with the gaussian family with log-normal link, as raw RT followed a log-normal distribution.

Participant Duration Ratio did not affect Accuracy significantly, but participants with larger Participant Duration Ratio had significantly slower RT (β=0.02, F(1, 4097)=9.53, p<0.001). Source Speaker did not affect Accuracy significantly, but participants responded more slowly to words produced by Source Speaker 2 (β=0.03, F(1, 4097)=0.0002, p=0.01). Participant Duration Ratio showed a significant positive interaction with Source Speaker 2 on accuracy (β=0.13, F(1, 5572)=9.74, p=0.002): participants with a larger long:short ratio recognised words more accurately when produced by Source Speaker 2, who produced larger duration contrast. Participant Duration Ratio and Source Speaker did not show significant interaction on RT.

Vowel Pair effects showed that /iː-ɪ/ was disambiguated more accurately (β=1.43, F(3, 5572)=105.95, p<0.001) and more quickly (β=-0.64, F(3, 4097)=99.11, p<0.001) than other Vowel Pairs. /ʉː-ʊ/ was disambiguated less accurately (β=-0.92, F(3, 5572)=105.92, p<0.001) but more quickly (β=-0.05, F(3, 4097)=99.11, p<0.001) than other Vowel Pairs. /ʉː-ʊ/ and Source Speaker 2 showed a negative interaction on RT (β=-0.02, F(3, 4097)=8.25, p<0.001): the RT difference between responses to Source

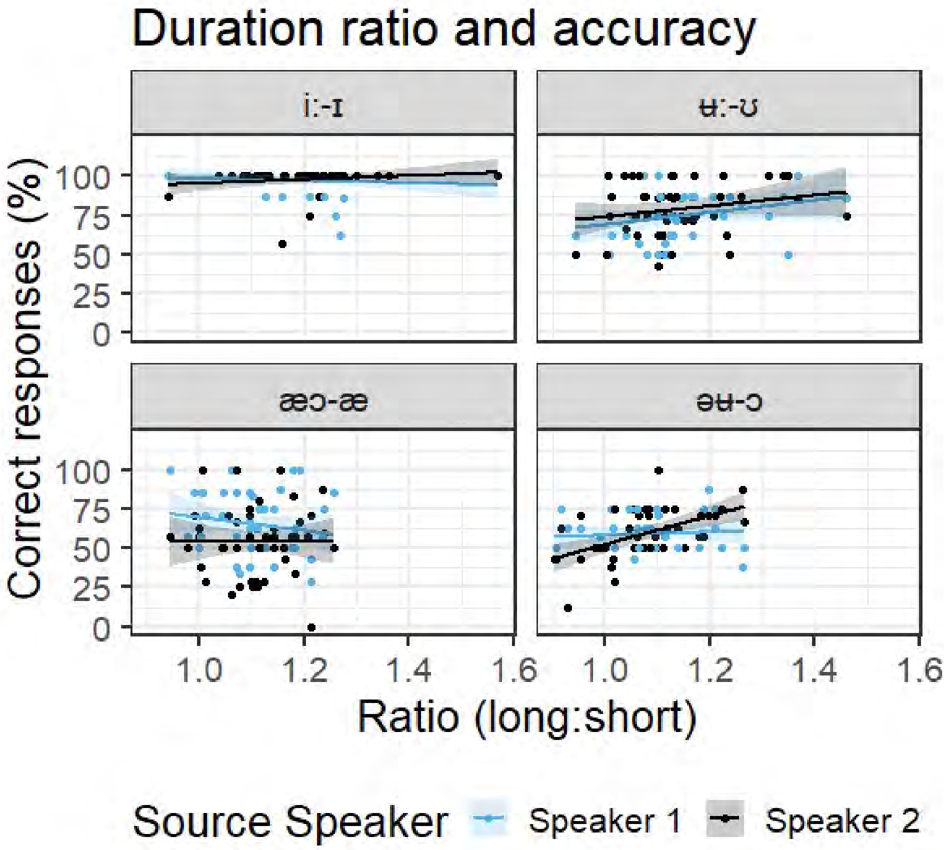


Figure 4.1: Correlation of participants' duration ratio (x-axis) and recognition accuracy (y-axis) by Source Speaker (blue: Speaker 1, black: Speaker 2) and Vowel Pair (panels). Top: /iː-ɪ/ and /ʉː-ʊ/ contrast. Bottom: /æɔ-æ/ and /əʉ-ɔ/ contrast.

Speaker 1 and 2 was smaller for /ʉː-ʊ/ than for other Vowel Pairs. /æɔ-æ/ was disambiguated less accurately (β=-0.18, $F(3, 5572)$=105.92, p=0.048) and more slowly (β=0.11, $F(3, 4097)$=99.11, $p<0.001$) than other pairs with 59% response accuracy and log-normalised 7.23 ms RT, in contrast with the overall response accuracy of 73% and log-normalised RT of 7.18 ms. Interactions between Participant Duration Contrast and Vowel Pair /æɔ-ɔ/ showed that participants with larger long:short ratio disambiguated /æɔ-ɔ/ less accurately (β=-0.22, $F(3, 5572)$=3.08, p=0.012) and more slowly (β=0.1, F(3,

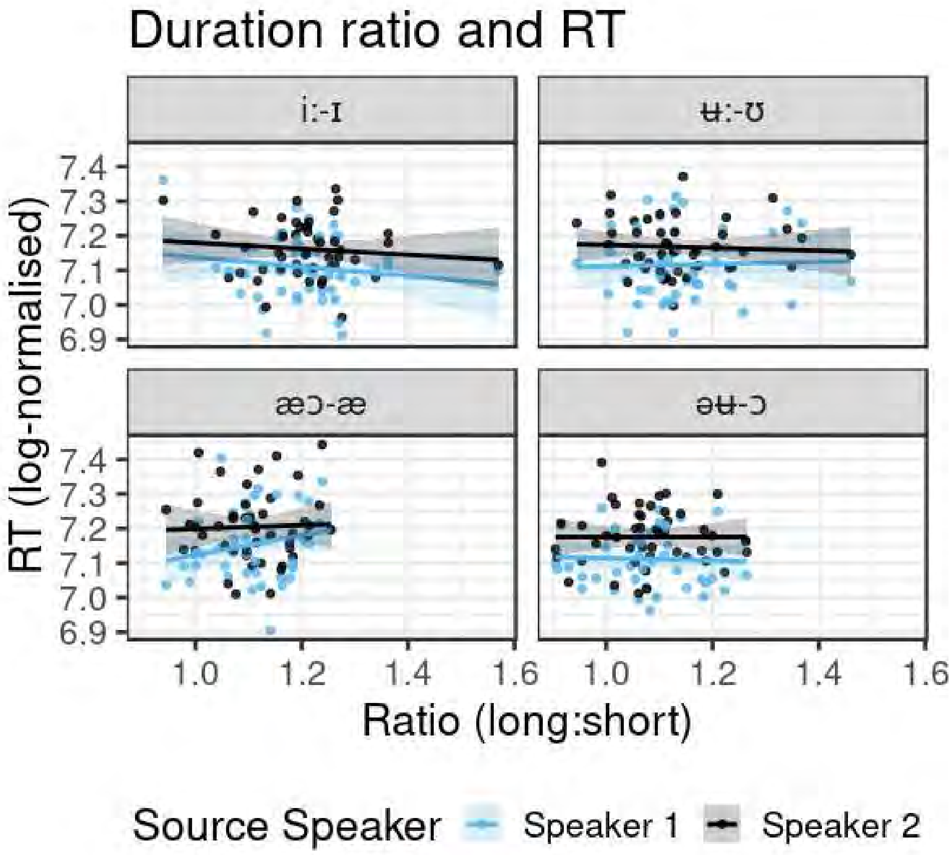


Figure 4.2: t Correlation of participants' duration ratio (x-axis) and perceptual RT (y-axis) by Source Speaker (blue: Speaker 1, black: Speaker 2) and Vowel Pair (panels). Top: /iː-ɪ/ and /ʉː-ʊ/ contrast. Bottom: /æɔ-æ/ and /əʉ-ɔ/ contrast.

4097)=1.69, p=0.04). Interaction between Source Speaker 2 and Vowel Pair /æɔ-æ/ showed that /æɔ-æ/ was disambiguated less accurately when produced by Source Speaker 2 (β=-0.25, F(3, 5572)=7.23, p=0.001).

Increased Lexical Frequency lead to increased accuracy (β=0.52, F(1, 5572)=256.3, p<0.001) and to increased RT (β=0.02, F(1, 4097)=63.23, p<0.001). Increase in RT with the increase in Lexical Frequency was probably due to the fact that there were more high frequency words among the targets with long acoustic duration.

4.4.3 Summary of findings

1. Contra to hypothesis 1, increased durational contrast in the speech of Source Speaker 2 did not assist word recognition, suggesting that not all listeners rely on durational cues.

2. Contra to hypothesis 2, participants who produced an increased durational contrast were not overall better at word recognition but they were overall slower.

3. In accordance with hypothesis 3, participants producing a larger duration contrast were more accurate on the contrast produced by Source Speaker 2, who, like them, maintained a larger durational contrast.

4.5 Discussion

Accuracy data showed that increased duration contrast in the stimuli aided word recognition only when participants also produced a more robust durational contrast. This indicates that perception is aided by cues that speakers themselves produce, but speaker-listeners without a robust durational cue production could not gain perceptual benefits. We found no evidence for overall better perception by participants with more robust duration contrast, contrary to (Perkell, Matthies et al. 2004, Perkell, Guenther et al. 2004). These discrepancies may be attributed to the differing methods, as we used an open-ended word recognition task, not contrast discrimination.

RT data showed that participants' increased rime duration contrast was associated with overall longer RT, indicating that these participants might consistently monitor for durational contrast. Durational contrast might take longer to process than spectral cues, as spectral cues may be available earlier in the vowel, whereas the whole rime needs to be processed for the perception of durational cues [25, 26, c.f. 27]. The overall increase in RT with the increase in durational contrast in production indicates that speaker-listeners who rely on durational contrast in perception always monitor for it. However, the fact that these speaker-listeners are not overall more accurate indicates that they cannot always find durational contrast.

All participants responded more slowly to Source Speaker 2, despite Source Speaker 2 producing overall shorter target words than Source Speaker 1. The reason might lie in the potentially different spectral quality of the Source Speakers' vowels, in Source Speaker 2 always being presented first, or in the fact that Source Speaker 1 was closer in age to the participants.

Words contrasting the four vowel pairs were recognised differently and showed complex interactions with participants' production. Words contrasting /iː-ɪ/ were recognised more efficiently, potentially due to the F2 differences between /iː/ and /ɪ/ at vowel onset in the stimuli. Minimal pairs contrasting /æɔ-æ/ were poorly recognised, probably because neither of the Source Speakers produced a robust durational contrast for this vowel pair. All participants performed less accurately on Source Speaker 2's production of the /æɔ-æ/ contrast. Moreover, participants with a bigger durational contrast performed *worse* on the overall recognition of the /æɔ-æ/ contrast. That is, participants with bigger durational contrast did not perform better on Source Speaker 2, contrary to their performance with other vowel contrasts, as they may have been looking for a durational contrast that was not present. Patterns of minimal pair recognition contrasting /æɔ-æ/ are consistent with hypothesis 3, in which listeners' perception is aided by cues that they themselves produce.

These findings suggest that listeners can only benefit from durational cues in vowel perception when they themselves produce it. Similarly, in Wade's (2017) study listeners who could not use durational contrast were members of a different speech community and maintained spectral contrasts (and presumably a non-phonological durational contrast as well), whereas participants in our study were members of a single speech community. These results do not allow us to determine the cues that listeners without a durational contrast use to identify /l/-final words. Future work will analyse listeners' spectral contrast production and link it to their perception of /l/-final minimal pairs.

4.6 Conclusion

Slower discrimination of /l/-final rimes by individuals who produce larger durational contrast implies that these speaker-listeners may monitor for durational contrast. This makes word identification slower, but only leads to increased accuracy when the speaker produces a sufficient durational contrast too. This implies that robust durational contrast production may come at a price and with limited benefits in word recognition.

4.7 Acknowledgements

We thank the Phonetics Lab at Macquarie University. This research was supported in part by iMQRTP 2015144, ARC DE150100318, and MQSIS 9201501719 grants.

4.8 References

Baayen, H. R. & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bennett, D. C. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech*, *11*(2), 65–85.

Boersma, P. & Weenink, D. (2013). *Praat 5.3.42.* Retrieved from `http://www.fon.hum.uva.nl/praat/`

Cox, F. (2006). The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics*, *26*(2), 147–179.

Davies, M. (2013). *Corpus of global web-based English: 1.9 billion words from speakers in 20 countries.*

Krause, F. & Lindemann, O. (2014, 1st Jun). Expyriment: A python library for cognitive and neuroscientific experiments. *Behavior Research Methods*, *46*(2), 416–428.

Liu, S. (2016). *The effect of vowel duration on native Mandarin listeners' perception of Australian English vowel contrasts in voiced and voiceless coda contexts.* Unpublished master's thesis.

Luce, P. A. & Pisoni, D. B. (1998, Feb). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*, 1–36.

Nearey, M., Terrance & Assmann, P. F. (1986). Modelling the role of intherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, *5*(80), 1297–1308.

Neel, A. T. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research*, *51*(3), 574–585.

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America*, *113*(5), 2850–2860.

Palethorpe, S. & Cox, F. (2003). Vowel modification in pre-lateral environments. In *International seminar on speech production.* Sydney.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M. et al. (2004). Cross-subject correlations between measures of vowel production and perception. *Journal of the Acoustical Society of America*, *116*(4), 2338–2344.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N. et al. (2004). The distinctness of speakers' /s/—/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of speech, language, and hearing research*, *47*(6), 1259–1269.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532.

Schiel, F. (1999, August). Automatic Phonetic Transcription of Non-Prompted Speech. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville &

A. C. Bailey (Eds.), *Proceedings 14th international congress of phonetic sciences* (p. 607–610). San Francisco, CA, USA.

Shultz, A. A., Francis, A. L. & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, *132*(2), EL95–EL101.

Szalay, T., Benders, T., Cox, F. & Proctor, M. (2016). Disambiguation of Australian English vowels. In C. Carignan & M. D. Tyler (Eds.), *Proceedings of the 16th australasian international conference on speech science and technology* (p. 73–76). Sydney: Australasian Speech Science and Technology Association.

Wade, L. (2017). The role of duration in the perception of vowel merger. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1).

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J. & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience*, *9*(131), 1–12.

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, *61*, 13–29.

5 Lingual configuration of Australian English /l/

This chapter is based on the following published paper:

Szalay, T., Benders, T., Cox, F., & Proctor, M. (2019). "Lingual configuration of Australian English /l/" in Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren(Eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*, (p. 2816–2819), Melbourne Australia.

I certify that I was responsible for the development of the concept of this paper, in discussion with my supervisors/co-authors. I took leadership in conducting the research, and was responsible for the construction of the stimuli, all data collection, the majority of the articulatory and all of the statistical analyses, and the writing of all parts of the paper. Data was collected as a part of a larger project: methodological innovations are reported in the current chapter, detailed articulatory analyses are reported in **Chapter 6** and perceptual-articulatory analyses are reported **Chapter 7**. My co-authors provided advice to improve the experimental design and methods, the analyses, the interpretation of the data, as well as the presentation of the written component. I have also received feedback from the reviewers and the audience of the 19$^{\text{th}}$ International Congress of Phonetic Sciences.

Abstract

English /l/ is a multi-gestural segment produced with dorsal retraction and lowering and a central alveolar closure. The coordination of antagonistic coronal and dorsal gestures prototypically results in lingual elongation. Although intergestural coordination in laterals has been widely studied, less is known about articulatory configuration in Australian English /l/ —a dialect characterised by coda /l/-lenition (Borowsky 2001, Borowsky & Horvath 1997). We explored tongue elongation as a potential metric of /l/-lenition.

The timecourse of lingual elongation was examined in laterals produced by two Australian English speakers using electromagnetic articulography. Tongue elongation was greater in onsets and codas containing laterals compared to onsets and codas containing /d/. Coda laterals showed less elongation than onset laterals. Quantifying lingual elongation can potentially differentiate onset /l/ from lenited or vocalised /l/ across a variety of vocalic and consonantal contexts by capturing a key characteristic of /l/ in environments where coronal and dorsal gestures are often unmeasurable.
Keywords: laterals, goals of /l/ articulation, /l/-vocalisation, Australian English

5.1 Introduction

The English lateral approximant and its allophonic variation between clear, dark, and vocalised /l/ has been studied widely both because of its social salience and implications for syllable structure (Borowsky 2001, Browman & Goldstein 1995, Gick 2003, Gick et al. 2002, Giles & Moll 1975, Horvath & Horvath 1997; 2001; 2002, Proctor & Walker 2012, Scobbie & Pouplier 2010, Sproat & Fujimura 1993). /l/-vocalisation —the realisation of /l/ with no alveolar closure —has been studied with articulatory (Hardcastle & Barry 1989, Lee-Kim et al. 2013, Lin et al. 2014, Scobbie & Wrench 2003) and acoustic impressionistic (Borowsky 2001, Horvath & Horvath 1997; 2001; 2002) methods, showing that the likelihood of /l/-vocalisation depends on the place of articulation of adjacent segments.

However, /l/-articulation is hard to measure, as articulatory analysis requires contrasting /l/ gestures with the gestures of surrounding segments. As the coronal gesture of /l/ does not contrast with following homorganic alveolar consonants, some studies avoided a following alveolar (e.g. Scobbie

& Pouplier 2010, Wrench & Scobbie 2003) or used it as a baseline to elicit unvocalised /l/ (Lin et al. 2014; 2011). As the tongue dorsum gesture is similar to that of back vowels (Gick et al. 2002), some studies have focused on lateral production only in front vowel contexts (e.g. Scobbie & Pouplier 2010, Sproat & Fujimura 1993, Ying et al. 2017). /l/-vocalisation is characterised by lenition of tongue tip contact; therefore, it is difficult to capture before a coronal consonant because a tongue tip gesture can be attributed both to the alveolar consonant and to /l/. Auditory impressionistic classification can distinguish vocalised and non-vocalised /l/ (e.g. Borowsky 2001); however, that is an indirect measurement.

We aimed to develop a technique that has the potential to quantify and characterise /l/ lenition and vocalisation by tracking change in tongue elongation during /l/ production. Tongue elongation results from /l/ having complex articulation: /l/ involves a temporal overlap between the raising and/or fronting of the tongue apex and the retraction of the tongue dorsum (Giles & Moll 1975, Ladefoged & Maddieson 1996, Sproat & Fujimura 1993), resulting in lingual elongation along the midline of the vocal tract (e.g. Ladefoged & Maddieson 1996, Proctor & Walker 2012). In contrast, coronal stops and non-front vowels would not be expected to show tongue elongation as coronal stops do not require tongue retraction and non-front vowels do not require tongue tip fronting. Vocalised /l/ may be expected to show reduced tongue elongation, as it is articulated without a tongue tip contact with the alveolar ridge (Giles & Moll 1975, Hardcastle & Barry 1989). This suggests that tongue elongation might be a metric that can distinguish non-vocalised /l/ from vocalised /l/. To capture tongue elongation, we computed the distance between the tongue tip and the tongue dorsum during /l/ production in front, back, and low vowel contexts. We also compared tongue tip and tongue dorsum trajectories of /l/ to /d/ to determine how they contribute to tongue elongation. We hypothesised that (1) in accordance with previous research, the tongue would be more elongated in /l/ than in /d/ in all vowel contexts; and (2) onset /l/ might be more elongated than coda /l/, due to potential lenition or vocalisation of coda /l/.

5.2 Methods

5.2.1 Participants

Two female native speakers of AusE participated in the study. Participants were students of linguistics, naive to the purpose of the experiment, who did not report any hearing, speaking, or reading difficulties. Participants received $80 for their time.

5.2.2 Material

Twenty-four unique monosyllabic words containing /iː, ɪ, ɐː, ɐ, oː, ɔ/ were selected from an experimental corpus (5.1). Target words combined real words of varying frequency and non-words. Although /l/-vocalisation is sensitive to lexical frequency (Lin et al. 2014), we did not find a difference in tongue tip position and elongation between real words and nonwords in a pilot with one participant. Target words were elicited in a carrier phrase with antagonistic vowel contexts: "*far;* __ *HARP*" and "*fee;* __ *HEAP*" for front and non-front vowels respectively. Non-target consonants were /p/, /f/, or /h/ to minimise lingual coarticulation. A semicolon was introduced after the first word to minimise resyallbification between target and carrier phrase. The last word was set in capitals to maintain consistent prosody across trials.

Table 5.1: Target words without carrier phrase.

Vowel Context		/d/		/l/	
		Onset	Coda	Onset	Coda
Front	/iː/	*deep*	*peed*	*leap*	*peel*
	/ɪ/	*dip*	*pid*	*lip*	*pill*
Back	/oː/	*dorp*	*poured*	*lorp*	*Paul*
	/ɔ/	*dop*	*pod*	*lop*	*pol*
Low	/ɐː/	*darp*	*pard*	*larp*	*parl*
	/ɐ/	*dup*	*pud*	*lup*	*puhl*

5.2.3 Procedure

Participants were seated approximately 50 cm from a computer screen and were introduced to the task and the experimental materials with a short practice block. Participants read the phrases aloud. Each trial began with a blank screen for 500 ms, followed by the stimulus for 2000 ms. After 2000 ms, the experiment automatically moved on to the next trial. Items were presented once per block in a random order. The block was repeated 8 times, providing 192 target words per participant.

5.2.4 Data acquisition

Articulatory data were acquired using an NDI Wave system sampling each sensor at a rate of 100 Hz. Eleven sensors were attached to the participant. Five sensors were attached to the tongue to track lingual articulation: three midsagittal (tongue tip (TT), tongue body, tongue dorsum (TD)) and two parasagittal sensors (right and left) (**Figure 5.1**). One sensor was attached to the lower and one to the upper lip to track lip aperture and rounding. A sensor was attached to the lower gumline to track jaw movement. Three

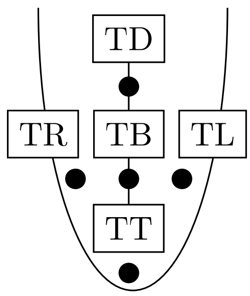


Figure 5.1: Tongue sensor placements viewed from top.

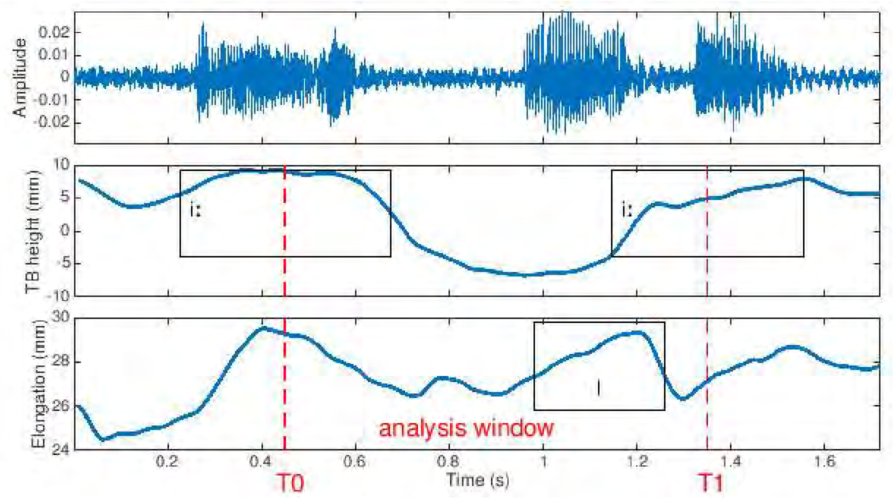
reference sensors (nasion, left and right mastoid) were used to correct for head movement. The occlusal plane was located with a bite trial and the palate was traced with a palate probe.

5.2.5 Data analysis

24 (targets) × 8 (repetitions) × 2 (participants) = 384 tokens were recorded. 14 tokens (4 from W1, 10 from W2) were excluded from analysis due to being misread, leaving 370 tokens for analysis. A maximal analysis window was defined from the midpoint of the vowel gesture in the first word (T0) to the midpoint of the vowel gesture in the last word (T1) of the carrier phrase (**Figure 5.2**). That is, for the phrase *fee; Paul HEAP*, the gestural midpoint of /iː/ in *fee* was selected as T0, and the gestural midpoint of /iː/ in *HEAP* was selected as T1 (**Figure 5.2**). For each token, the gestures defining the analysis window were determined visually using MView (Tiede 2005). From this window, unfiltered trajectories of TT and TD movement were extracted (Wieling 2018). We calculated a tongue elongation trajectory (TE) as the Euclidean distance between the TT and TD sensors (horizontal and vertical positions) at each point in time.

We analysed tongue movement trajectories in the selected window using generalised additive modeling (GAM) (Wieling 2018). GAM is a non-linear

Figure 5.2: Analysis window exemplified by *fee Paul HEAP*. Top panel: waveform. Middle panel: vertical location of tongue body. Bottom panel: tongue elongation. Boxes mark gestures. T0 marks the start of the analysis window at the gestural midpoint of the first vowel and T1 marks the end at the gestural midpoint of the last vowel.



regression model which can be used to analyse change in articulatory trajectories over time by computing the best-fitting non-linear basis function for a trajectory (Wieling 2018).

TE, and horizontal and vertical TT and TD trajectories were time-normalised to account for differing length of the trajectories and modeled separately for both speakers and both positions; as a result 5 (trajectories) × 2 (participants) × 2 (onset and coda) = 16 models were built. We modeled TE, TT, TD trajectories as the function of consonant segment (/l/ compared to baseline /d/) and vowel context (front and back vowels compared to baseline low) using GAM with thin plate regression splines as basis functions. Random effects were not added as speakers were modelled separately.

5.3 Results

5.3.1 Tongue elongation

Tongue elongation was greater in /l/ than in /d/ for W1 in both syllable onset ($\beta = 0.88, F(1, 8714) = 7.48, p = 0.006$) and coda ($\beta = 0.74, F(1, 9050) = 6.18, p = 0.01$). For W2, tongue was more elongated in /l/ than in /d/ in the onset ($\beta = 1.96, F(1, 5492) = 11.85, p < 0.001$), but less elongated in the coda ($\beta = -0.47, F(1, 5146) = 7.79, p = 0.005$) (**Figure 5.3**). Tongue elongation occurs in the first half of the analysis window in onset /l/, and in 50%–80% of the analysis window in coda /l/ for W1 (**Figure 5.4**). Greater tongue elongation in W2's coda /d/ might be an artifact of a too-large analysis window as the tongue seems to be more elongated in /l/ than in /d/ in 50%–75% of the analysis window, and less elongated elsewhere (**Figure 5.4**). Tongue elongation in onset and coda /l/ was not compared in the same model; however, comparing estimates across models indicates greater tongue elongation in onset than in coda /l/.

Figure 5.3: Change in tongue elongation over normalised time (T0 to T1). Left: /l/ vs. /d/ in onset. Right: /l/ vs. /d/ in coda. Top: W1. Bottom: W2. Shaded bands show 95% confidence intervals. Red vertical bars mark greater tongue elongation associated with /l/ as in **Figure 5.4**.

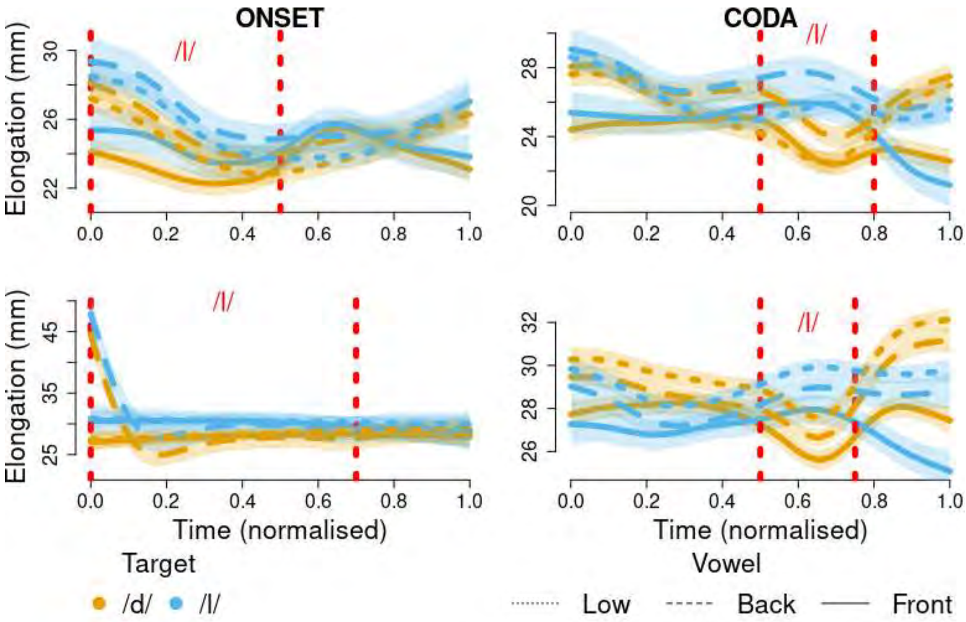
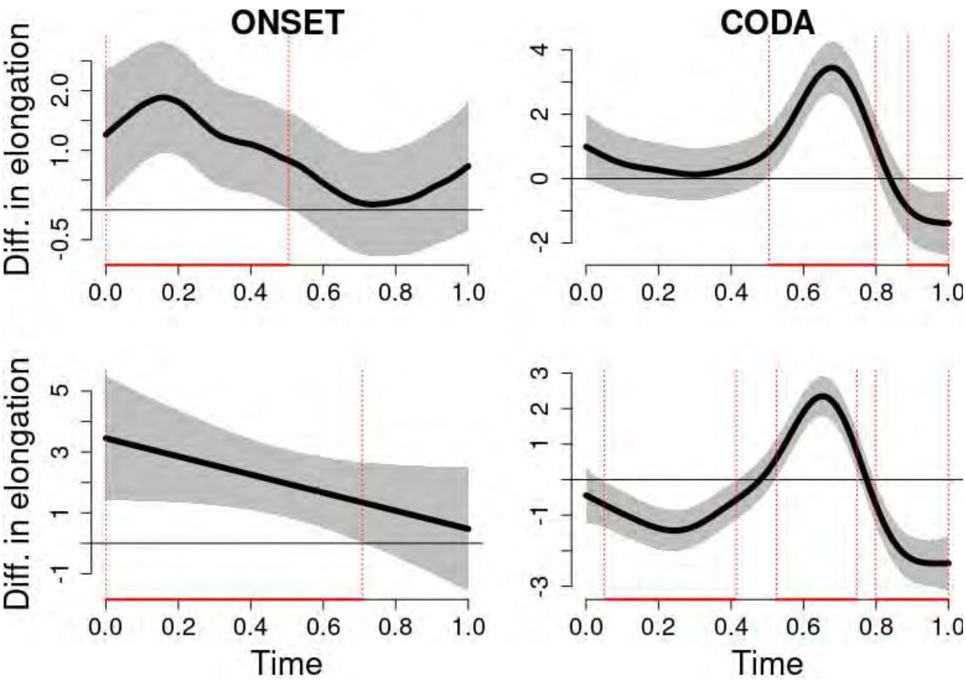


Figure 5.4: Difference in tongue elongation over normalised time (T0 to T1) comparing /l/ to /d/. Left: /l/ vs. /d/ in onset. Right: /l/ vs. /d/ in coda. Top: W1. Bottom: W2. Shaded bands show 95% confidence intervals. Red lines on the X-axis and red vertical bars indicate areas of significant difference.



5.3.2 TT and TD trajectories

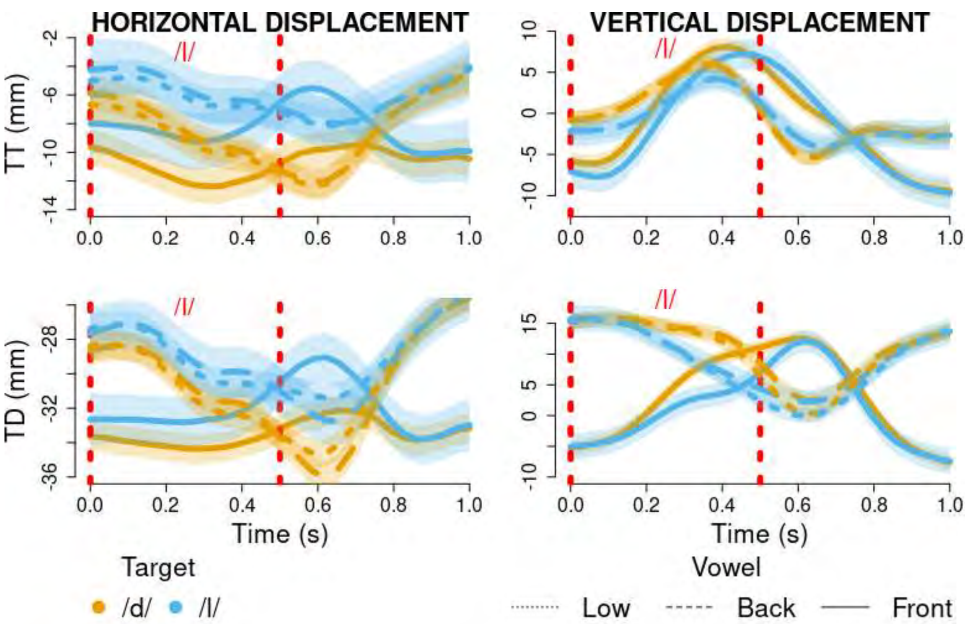
Tongue tip and tongue dorsum were fronted during the production of /l/ compared to /d/ in onset and coda position for both speakers (**Table 5.2**, TTx and TDx trajectories). Tongue dorsum was lowered during the production of /l/ compared to /d/ in coda position for both speakers, and in W1's onset (**Table 5.2**, TDz trajectories). Tongue tip gesture of /l/ was only lowered compared to /d/ in W1's coda. TT fronting was always greater than TD fronting, except for W2's coda (**Table 5.2**, β). TT and TD trajectories are illustrated by W1's production, as W2 produced a similar pattern (**Figure 5.5**).

Table 5.2: Effect of /l/ on TT and TD. TTx and TDx: horizontal movement. TTz and TDz: vertical movement.

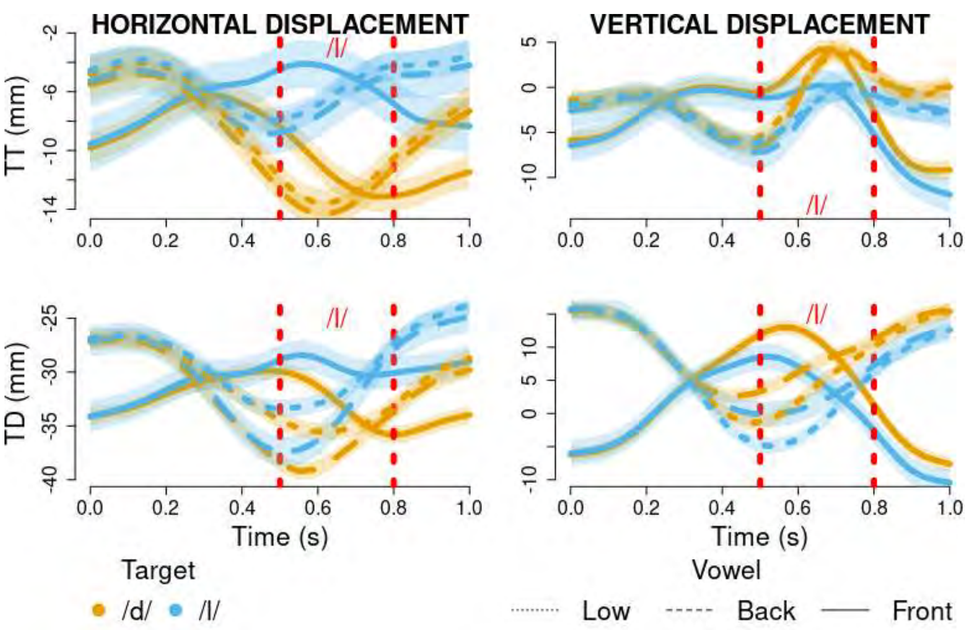
Position	Speaker	Trajectory	β	p
Onset	W1	TTx	2.49	< 0.001
		TDx	1.51	0.001
		TTz	-0.68	0.2
		TDz	-1.88	0.01
	W2	TTx	3.83	< 0.001
		TDx	2.35	< 0.001
		TTz	0.02	0.95
		TDz	0.47	0.63
Coda	W1	TTx	3.26	< 0.001
		TDx	2.25	< 0.001
		TTz	-1.40	0.002
		TDz	-2.43	< 0.001
	W2	TTx	0.98	< 0.001
		TDx	1.33	0.002
		TTz	0.56	0.11
		TDz	-3.17	0.001

Figure 5.5: Change in TT and TD movement over normalised time (T0 to T1) in W1's speech. Left: horizontal displacement. Right: veritcal displacement. Top row: TT. Bottom row: TD. Shaded bands indicate 95% confidence intervals. Red vertical bars mark greater tongue elongation associated with /l/ as in **Figure 5.4**.

(a) Analysis window contains /d/ and /l/ in onset position.



(b) Analysis window contains /d/ and /l/ in coda position.



5.4 Discussion

The aim of this study was to develop a metric of tongue elongation that can potentially quantify /l/-vocalisation. In accordance with our first hypothesis and previous research (Browman & Goldstein 1995, Ladefoged & Maddieson 1996), the tongue was more elongated in /l/ than in /d/, except for W2's codas. Tongue elongation may distinguish /l/ from surrounding segments, whereas the tongue tip gesture of /l/ is similar to coronal stops and the tongue dorsum gesture is similar to non-front vowels. Thus, the tongue elongation metric could be used to automatically identify the point in time at which lingual elongation is maximised in different environments.

Tongue elongation may occur because of the fronting of the tongue tip and the lowering of the tongue dorsum gestures. Both tongue tip and dorsum are fronted in /l/ compared to /d/, but greater fronting in the tongue tip compared to the dorsum fronting leads to elongation. The more extensive tongue tip fronting may indicate that only the tongue tip gesture of /l/ has a fronted target, whereas tongue dorsum fronting might result from being coarticulated with the tongue tip.

In W1's speech, the magnitude of tongue elongation of /l/ compared to /d/ was greater in syllable onset than in coda, which is consistent with our second hypothesis, showing lenition in coda /l/. Although the tongue tip was more fronted in coda compared in onset position, it was also lowered, indicating lenition. This finding indicates that reduced tongue elongation may provide a consistent measurement of coda /l/ lenition in a variety of segmental contexts. In contrast with tongue elongation, tongue tip position is likely to be conflated with a following alveolar consonant.

W2's coda /l/ production shows a different pattern: the overall estimate showed the tongue dorsum to be more fronted in coda /l/ compared to coda /d/ relative to the tongue tip difference between these two consonants. Consequently, tongue elongation was smaller in coda /l/ than in coda /d/. This discrepancy in tongue elongation between W1 and W2 can be attributed to individual differences in lateral production or to the method of data analysis. In contrast to the overall estimate, the tongue seemed to be more elongated in /l/ compared to /d/ in the part of the analysis window associated with the coda. In the rest of the analysis window, corresponding to the vowels in the carrier phrase and the target word, tongue was more elongated when the target word contained coda /d/ compared to coda /l/. That is, the results are inconclusive as the overall effects might indicate /l/-vocalisation, whereas a more detailed temporal analysis suggests that the analysis window needs to be smaller.

5.5 Conclusion

These data demonstrate the utility of tracking change in tongue elongation as a metric for lateral production. The tongue might be less elongated in coda /l/ compared to onset /l/, consistent with the lenition of the tongue tip gesture in coda /l/ observed in AusE. Future research on the articulatory characterisation of /l/ may include direct comparison of tongue elongation in onset and coda position to quantify /l/ vocalisation.

5.6 Acknowledgements

This research was supported in part by iMQRTP 2015144, and ARC DE150100318 grants. We thank Martijn Wieling for his help with GAM, and Louise Ratko for her help in data collection.

5.7 References

Borowsky, T. (2001). The vocalisation of dark *l* in Australian English. In D. Blair & P. Collins (Eds.), *English in Australia* (p. 69–87). Philadelphia, Amsterdam: John Benjamins Publishing Company.

Borowsky, T. & Horvath, B. (1997). L-vocalisation in Australian English. In F. Hinskens, R. van Hout & W. L. Wetzels (Eds.), *Variation, change and phonological theory.* (p. 101–123). Amsterdam: Benjamins.

Browman, C. P. & Goldstein, L. M. (1995). Gestural syllable position effects in American English. In F. Bell-Berti & J. L. Raphael (Eds.), *Producing speech: contemporary issues.* New York: AIP Press.

Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. *Phonetic interpretation: papers in laboratory phonology VI*, 222–236.

Gick, B., Kang, M. A. & Whalen, D. H. (2002). MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics*, *30*(3), 357–371.

Giles, S. & Moll, K. (1975). Cinefluorographic study of selected allophones of English /l/. *Phonetica*, *31*(3-4), 206–227.

Hardcastle, W. & Barry, W. (1989, 7). Articulatory and perceptual factors in /l/ vocalisations in English. *Journal of the International Phonetic Association*, *15*, 3–17.

Horvath, B. M. & Horvath, R. J. (1997). The geolinguistics of a sound change in progress: /l/ vocalization in Australia. *U. Penn Working Papers in Linguistics*, *4*(1), 109–124.

Horvath, B. M. & Horvath, R. J. (2001). A multilocality study of a sound change in progress: The case of /l/ vocalization in New Zealand and Australian English. *Language Variation and Change*, *13*, 37–57.

Horvath, B. M. & Horvath, R. J. (2002). The geolinguistics of /l/ vocalisation in Australia and New Zealand. *Journal of Sociolinguistics*, *6/3*, 319–346.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford, UK; Cambridge, Mass.: Blackwell.

Lee-Kim, S.-I., Davidson, L. & Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology*, *4*(2), 475–511.

Lin, S., Beddor, P. & Coetzee, A. (2014). Gestural reduction, lexical frequency, and sound change: a study of post-vocalic /l/. *Laboratory Phonology*, *5*(1), 9–36.

Lin, S., Beddor, P. S. & Coetzee, A. W. (2011). Gestural reduction and sound change: an ultrasound study. In *Proceedings of the 17th international congress of phonetic sciences* (p. 1250–1253). Hong Kong.

Proctor, M. & Walker, R. (2012). Articulatory bases of sonority in English liquids. In S. Parker (Ed.), *The sonority controversy* (p. 289–316). Berlin, New York: Mouton de Gruyter.

Scobbie, J. M. & Pouplier, M. (2010). Syllable structure and external sandhi: an EPG study of vocalisation and retraction of word-final English /l/. *Journal of Phonetics*, *38*, 240–259.

Scobbie, J. M. & Wrench, A. A. (2003). An articulatory investigation of word final /l/ and //l/-sandhi in three dialects of English. In *Proceedings of the 15th international congress of phonetic sciences.* Barcelona, Spain.

Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*(3), 291–311.

Tiede, M. (2005). Mview: software for visualization and analysis of concurrently recorded movement data. *New Haven, CT: Haskins Laboratories*.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116.

Wrench, A. A. & Scobbie, J. M. (2003). Categorising vocalisation of English /l/ using EPG, EMA and ultrasound. In *Proceedings of the 6th international seminar on speech production* (p. 314–319).

Ying, J., Carignan, C., Shaw, J. A., Proctor, M., Derrick, D. & Best, C. (2017). Temporal dynamics of lateral channel formation in /l/: 3D EMA data from Australian English. In *Proceedings of the 18th annual conference of the international speech communication association.* Stockholm, Sweden.

6 Coronal lenition in Australian English coda laterals

This chapter is based on the following paper, which is being prepared for submission:

Szalay, T., Benders, T., Cox, F., & Proctor, M. (in preparation). Coronal lenition in Australian English coda laterals.

I certify that I was responsible for the development of the concept of this paper, in discussion with my supervisors/co-authors. I took leadership in conducting the research, and was responsible for the construction of the stimuli, all data collection, the majority of the articulatory and all of the statistical analyses, and the writing of all parts of the paper. Data was collected as a part of a larger project: methodological innovations are reported in **Chapter 5**, detailed articulatory analyses are reported in the current chapter and perceptual-articulatory analyses are reported in **Chapter 7**. My co-authors provided advice to improve the experimental design and methods, the analyses, the interpretation of the data, as well as the presentation of the written component.

Abstract

The lateral approximant /l/ has been observed to undergo vocalisation in coda position in Australian English (Horvath & Horvath 1997, Borowsky 2001). /l/-vocalisation is frequent, but shows inter- intraspeaker variation: one of the factors conditioning /l/-vocalisation is the place of articulation of segments adjacent to /l/ (Horvath & Horvath 1997, Borowsky 2001). Although /l/-vocalisation is attributed to the lenition of the coronal gesture of /l/, /l/-vocalisation in Australian English has only been examined using auditory-impressionistic methods; articulation of coda /l/ has not been systematically explored. We examined TT aperture trajectories and targets of coda /l/ in the speech of six native speakers of Australian English using elec-

tromagnetic articulography. Speakers differed from each other in the degrees of lenition with which they produced coda /l/. Speakers produced coda /l/ with varying degree of lenition across phonetic contexts. Articulatory similarity between /l/ and the following consonant has a strong effect on /l/ lenition: coronal lenition is less frequently observed when the following consonant is coronal and more frequently observed when it is dorsal. The effect of the following consonant is stronger when /t/ or /k/ are tautosyllabic with /l/.

6.1 Introduction

English /l/ is a multi-gestural segment, canonically produced with a central alveolar closure, lateral channel(s), and tongue dorsum retraction (Giles & Moll 1975, Sproat & Fujimura 1993, Ladefoged & Maddieson 1996, Narayanan et al. 1997). Variation in the production of these gestures leads to a gradient variation between clear, dark, and vocalised /l/ (Giles & Moll 1975, Browman & Goldstein 1995, Sproat & Fujimura 1993). Dark [ɫ] is articulated with a delayed, less raised and fronted tongue tip gesture, and a more lowered and retracted tongue dorsum gesture compared to clear [l] (Giles & Moll 1975, Sproat & Fujimura 1993). Vocalised /l/ is articulated with a lenited tongue-tip gesture (i.e. reduced anterior constriction degree), increased delay of the tongue tip gesture compared to dark [ɫ], and lenited tongue lateralisation (i.e. flatter tongue shape) (Giles & Moll 1975, Browman & Goldstein 1995, Scobbie & Pouplier 2010, Strycharczuk et al. 2018, Strycharczuk & Scobbie 2019). Vocalisation of /l/ is associated with prosodic, lexical, segmental and sociolinguistic factors in several dialects of English, including Australian English (AusE) (Giles & Moll 1975, Borowsky 2001, Lin et al. 2014).

Many of the prosodic, segmental, and sociolinguistic factors conditioning /l/-vocalisation in AusE have been identified using auditory-impressionistic ratings of /l/ as vocalised or consonantal (Borowsky & Horvath 1997, Horvath & Horvath 1997, Borowsky 2001, Horvath & Horvath 2001; 2002). Our aim is to provide greater clarity on variation in coronal lenition in /l/ by measuring the tongue tip gesture to characterise coronal lenition in some key phonetic and prosodic contexts that have been associated with variation in /l/-vocalisation.

6.1.1 Articulatory characteristics of clear and dark /l/

Articulatory variation in the production of American, Standard Southern British, Scottish English and AusE /l/ leads to gradient variation on the continuum from clear [l], associated with onset, to dark [ɫ], associated with coda position (Giles & Moll 1975, Sproat & Fujimura 1993, Browman & Goldstein 1995, Turton 2017, Ying et al. 2017). Dark [ɫ] differs from clear [l] in three respects: its less raised and less fronted tongue tip gesture; its more lowered and retracted tongue dorsum gesture; and the delayed tongue tip gesture, as the tongue tip gesture follows the tongue dorsum gesture in dark [ɫ], but precedes it in clear [l] (Giles & Moll 1975, Sproat & Fujimura 1993, Browman & Goldstein 1995, Gick 2003, Proctor et al. 2019, Ying et al. 2017). As a result of the different coordination between the tongue dorsum and the tongue tip gestures, the tongue dorsum gesture is temporally closer to the vowel, while the tongue tip gesture is farther both in clear and dark [l] (Sproat & Fujimura 1993). In addition, it is unclear whether clear and dark /l/ differ in their lateralisation gesture. In New Zealand English, clearer /l/ sounds were found to be more lateralised than darker /l/ sounds (Strycharczuk et al. 2018). In AusE, the timing and degree of lateralisation were not significantly different between clear [l] and dark [ɫ] (Ying et al. 2017).

Gradiency along the clear/onset-dark/coda continuum is observed in intervocalic word-final /l/: the tongue tip gesture exhibits variation between onset-like and coda-like realisations, while the realisation of the tongue dorsum gesture and the presence of the delay between the tongue tip and tongue dorsum gestures remain coda-like in the /Vl#V/ context (e.g. Gick 2003, Scobbie & Wrench 2003). The magnitude of the tongue tip delay corresponds to gradient darkening: the larger the delay between the tongue dorsum and the tongue tip gestures, the darker the [ɫ]. Gradient darkening in intervocalic /l/ is conditioned by prosodic and morphological boundaries in American and British English (Sproat & Fujimura 1993, Lee-Kim et al. 2013, Turton 2017). In American English, darkening of word-final prevocalic /l/ correlates with the strength of the prosodic boundary following /l/: the stronger the boundary, the bigger the delay between maximal tongue dorsum displacement and maximum tongue tip displacement and the darker the /l/ (Sproat & Fujimura 1993). Morpheme boundaries affect /l/ darkening similarly to prosodic boundaries: the stronger the boundary, the more lowered the tongue dorsum and thus darker the /l/ (Lee-Kim et al. 2013). That is, /l/ is darkest word-finally (e.g. *tall*), less dark morpheme-finally before a vowel (e.g. *tall-est*), and least dark morpheme-initially before a vowel (e.g. *flaw-less*).

Sproat & Fujimura (1993) argue that boundary effects can be reduced to acoustic rime duration: the stronger the boundary, the longer the acoustic duration and the darker the /l/, making acoustic duration of /l/ a good predictor of gradient /l/ darkening. However, Turton (2017) points out that the correlation between acoustic duration and /l/ darkness only holds for dark /l/, in which the tongue tip delay increases as duration increases, whereas there is no correlation in clear /l/, in which the tongue tip delay does not increase as duration increases. In addition, /l/ is systematically darker before productive than non-productive boundaries, but has the same acoustic duration before both (Lee-Kim et al. 2013). The analyses by Turton (2017) and the data by Lee-Kim et al. (2013) indicate that acoustic rime duration cannot fully account for /l/-darkening.

6.1.2 Conditioning factors in /l/-vocalisation

The darker end of the continuum – word-final pre-pausal or pre-consonantal, and to a lesser extent word-final prevocalic /l/ – can undergo /l/-vocalisation (Giles & Moll 1975, Hardcastle & Barry 1989, Scobbie & Pouplier 2010, Lin et al. 2014). /l/-vocalisation is commonly understood as a vowel-like realisation of coda /l/, as in the variant [mɪʊk] or [mɪɤk] for /mɪlk/ (e.g. Giles & Moll 1975, Lin et al. 2014). /l/-vocalisation arises from gestural reduction and delay in the tongue tip gesture and from lenition of the tongue lateralisation (i.e. a flatter tongue shape) (Giles & Moll 1975, Browman & Goldstein 1995, Strycharczuk et al. 2018, Strycharczuk & Scobbie 2019). The lenition of the tongue tip gesture and lateralisation are correlated but not directly causally linked (Strycharczuk et al. 2018). The lenition of the tongue tip and tongue-lateralisation gestures can be understood in terms of gestural lenition in coda position. Coda consonants, especially coda stops, often show reduction in the magnitude and duration of closure, leading to coda lenition (Byrd 1996, Keating et al. 2004, Byrd et al. 2005). However, during /l/-vocalisation, the tongue tip and the tongue lateralisation gestures are lenited whereas the tongue dorsum gesture causes a /w/- or /ʊ/-like percept. The exact realisation of a vocalised /l/, such as the extent of coronal constriction reduction, temporal delay of the coronal gesture, or the deletion of the coronal gesture varies between speakers (Strycharczuk & Scobbie 2019).

/l/-vocalisation, much like gradient /l/ darkening, is subject to prosodic, segmental, lexical, and stylistic factors, leading to a great amount of intraspeaker variation (Hardcastle & Barry 1989, Scobbie & Pouplier 2010, Lin et al. 2014). In connected speech, /l/-vocalisation is the most likely in preconsonantal context, and more likely in pre-pausal than in pre-vocalic context (Scobbie & Wrench 2003, Scobbie et al. 2007).

/l/-vocalisation is inhibited by articulatory similarity between the gestures of /l/ and adjacent segments (Hardcastle & Barry 1989). Vocalisation of preconsonantal /l/ in coda clusters is inhibited by a following alveolar consonant that shares the coronal gesture of /l/ and by a preceding back vowel that shares the dorsal gesture of /l/ (Hardcastle & Barry 1989). The effect of alveolar consonant is attributed to alveolars being homoorganic with the tongue tip gesture of /l (Hardcastle & Barry 1989). In contrast, the effect of a back vowel is attributed to perceptual, rather than articulatory reasons: the tongue dorsum gesture of /l/ is similar to that of the preceding back vowel which makes achieving tongue tip contact necessary for creating different percepts for the back vowel and the /l/ (Hardcastle & Barry 1989). In contrast, lowering the tongue dorsum after a front vowel is enough to create different vowel and /l/ percepts and the tongue tip contact contributes less to the percept of /l/ (Hardcastle & Barry 1989). All these factors affect the likelihood of /l/-vocalisation within speakers; in addition, interspeaker variation in the overall frequency of /l/-vocalisation has also been observed (Hardcastle & Barry 1989, Strycharczuk & Scobbie 2019).

6.1.3 /l/-vocalisation in Australian English

AusE is known for having /l/-vocalisation, but most of the data on its conditioning factors come from impressionistic analyses of audio recordings (Wells 1982, Borowsky & Horvath 1997, Borowsky 2001, Horvath & Horvath 1997; 2001; 2002, Cox & Palethorpe 2007 and **Chapter 7**). Similarly to other varieties of English, the most important phonological factor in /l/-vocalisation is the syllabic affiliation of /l/: dark, syllabic or coda [ɫ] can be vocalised, but clear onset [l] cannot (cf. *metal, steel* but not in *light*) (Borowsky 2001). Word-final /l/ is thus the least likely to be vocalised in the l#V environment, where it can be resyllabified. Word-final /l/ is less likely to be vocalised before a pause than before a consonant (Borowsky 2001, Horvath & Horvath 2001). It is unclear whether /l/ in a coda cluster is more likely to be vocalised than word-final /l/, as Borowsky (2001) found more vocalisation in clusters while Horvath & Horvath (2001) found more vocalisation in word-final position. However, these studies analysed both connected speech and individual words, and it is unclear whether their "word-final" categories distinguish between prevocalic, preconsonantal, and pre-pausal environments (e.g. *feel angry* vs. *feel good*).

Another factor conditioning /l/-vocalisation in AusE is articulatory similarity between /l/ and segments adjacent to /l/. In AusE, as in other dialects of English, articulatory similarity between the following consonant and the tongue tip gesture of /l/ inhibits /l/-vocalisation in coda clusters: /l/-vocalisation is least likely before an alveolar, and more likely before a velar than a bilabial consonant (Borowsky 2001, Hardcastle & Barry 1989). Borowsky's (2001) impressionistic analyses of the effect of consonantal context on /l/-vocalisation in AusE provide the same results as articulatory analyses on British English (Hardcastle & Barry 1989). However, in AusE, articulatory similarity between the tongue dorsum gesture of /l/ and the preceding vowel facilitates /l/ vocalisation (Borowsky 2001), while Hardcastle & Barry (1989), despite two out of six speakers being AusE, found that a preceding back vowel inhibits /l/-vocalisation. These contradictory results can be explained by methodological differences or by differences in speakers' dialect. Hardcastle & Barry (1989) measured tongue tip contact using EPG, while (Borowsky 2001) relied on acoustic-impressionistic encoding of tokens of /l/. If front vowels facilitate /l/-vocalisation because the tongue dorsum gesture without the tongue-tip gesture is sufficient to create the percept of /l/ after a front vowel (Hardcastle & Barry 1989), then the percentage of vocalised /l/ following front vowels may be underestimated in impressionistic analyses. The difference can also be attributed to speakers' dialects: perhaps, speakers of AusE (as tested by Borowsky (2001)) are more likely to vocalise in the context of back vowels, whereas speakers of British English (which constituted four out of five speakers in Hardcastle & Barry's (1989) study) are more likely to vocalise in the context of front vowels.

The third factor to be considered in /l/-vocalisation in AusE is the length of the preceding vowel: a preceding long monophthong or a diphthong facilitates /l/-vocalisation compared to a preceding short monophthong. This might be explained by the differing syllable structure of /l/ final rimes differs with long and short vowels: /l/ tends to be syllabic after long vowels in AusE (Borowsky 2001).

6.1.4 Aims and hypotheses

These studies by Borowsky & Horvath (1997), Borowsky (2001) and Horvath & Horvath (1997; 2001; 2002) are invaluable in identifying numerous key factors in /l/ vocalisation. However, they are limited in scope, as the data was coded for /l/ vocalisation by a single person, using auditory-impressionistic methods. This coding method does not provide any information on the presence or the absence of the tongue tip contact on the alveolar ridge, which is one of the key differentiating features between vocalised and non-vocalised /l/.

The aim of the present study, therefore, was to provide an articulatory characterisation of coronal contact and coronal constriction reduction[1] in coda /l/ in AusE. Firstly, we aimed to establish whether AusE coda /l/ is characterised by a larger magnitude of spatial reduction than other coda coronals, as the descriptions of AusE as a dialect characterised by /l/ vocalisation suggest. Secondly, we aimed to explore the impact of different phonetic environments on the degree of constriction in coda /l/ by selecting two segmental and two prosodic environments previously identified as affecting vocalisation of coda /l/: place of articulation of preceding vowel and following consonant, and preceding vowel length and coda complexity. We hypothesised that: 1) the coronal constriction would be more reduced in coda compared to onset for /l/ than for coronal obstruents; 2) the loss of coronal contact would be inhibited by similarity in place of articulation between the tongue tip gesture of /l/ and the following consonant, while it would be facilitated by similarity between the tongue dorsum gesture of /l/ and the preceding vowel; and 3) syllables containing either a long vowel or a coda cluster would facilitate the loss of coronal contact compared to syllables containing a short vowel and a simple coda.

[1]In Articulatory Phonology's terminology constriction degree decreases from wide, characterising low vowels, to closed, characterising obstruents (Browman & Goldstein 1989). In this dissertation, we use the term "constriction reduction" to refer to a gradient increase in TT aperture.

6.2 Methods

6.2.1 Participants

Ten native speakers of AusE participated in an electromagnetic articulography (EMA) study. Two were excluded from analysis due to producing retroflex /d/ potentially due to the sensors placed on the tongue tip (W1 and W9), one was excluded due to reporting hearing disorders after the experiment (W6), and one was excluded due to difficulties during data collection (W8). The remaining six female native speakers of AusE (mean age = 23.4, range = 20–27) were analysed in this corpus. All participants were born and raised in NSW. All but one participant had two NSW-born parents; W2 had one NSW-born and one Victoria-born parent. Participants received course credit and/or $40/hour for participation. None of the participants whose speech was analysed reported any current or past reading, hearing, or speaking disorders.

6.2.2 Material

The stimuli consisted of 51 3-word long phrases, the second of which contained the target consonant gesture: 24 phrases in the baseline context and 27 phrases in the vocalising context (**Table 6.1**). The baseline context contrasted the target alveolars, /d/ and /l/, in onset and coda position (12 /dVp, lVp/ onsets and 12 /pVd, pVl/ codas). The nuclear vowel was counterbalanced for place of articulation (8 front, 8 back, and 8 low vowels) and length (12 short and 12 long vowels). In the baseline context, the word containing the alveolar target was always followed by a word beginning with /h/ (**Table 6.1a**).

The vocalising context contrasted coda /l/ in 3 vowel contexts, 3 consonantal contexts, and 3 syllable types, giving a total of 27 words (3 vowel contexts × 3 consonantal contexts × 3 syllable types) (**Table 6.1b**). To vary vowel context, coda /l/ was preceded by a high, low, or back vowel. To manipulate consonant context, coda /l/ was followed by a labial, alveolar, or dorsal consonant. Pre-pausal /l/ was not elicited. The syllable type manipulation combined vowel length and coda complexity: 9 words contained a short vowel and a simple coda, 9 words contained a long vowel and a simple coda, and 9 words contained a short vowel and a complex coda. Words containing a long vowel and a complex coda would be phonotactically illicit and were not tested. The consonant following /l/ was the onset consonant of the following word when /l/ was in a simple coda and it was the following segment of the cluster when /l/ was in a coda cluster. To provide a consistent phonetic frame of reference, non-words were used when necessary (**Table 6.1**).

Target words were placed in a carrier phrase with antagonistic vowel contexts; that is words with front nucleus were placed in the carrier phrase "far ____ harp" and words with low or back nucleus were placed in the carrier phrase "fee ____ heap". To minimise lingual coarticulation, all non-target consonants were /f, p, h/. Target words and the carrier phrase were spelled according to the rules of English orthography and judged by native speakers of AusE for transparency.

Table 6.1: Target words containing alveolar /l, d/ followed by the last word of the carrier phrase.

(a) Baseline context

Vowel Context	/d/ Onset	/d/ Coda	/l/ Onset	/l/ Coda
Front	*dip hVp* *deep hVp*	*pid hVp* *peed hVp*	*lip hVp* *leap hVp*	*pill hVp,* *peel hVp*
Back	*dop hVp* *dorp hVp*	*pod hVp* *pord hVp*	*lop hVp* *lorp hVp*	*Paul hVp,* *pol hVp*
Low	*dup hVp* *darp hVp*	*pud hVp* *pard hVp*	*lup hVp* *larp hVp*	*parl hVp,* *puhl hVp*

(b) Vocalising context

Vowel Context	Word-final Bilabial	Word-final Alveolar	Word-final Dorsal	Cluster Bilabial	Cluster Alveolar	Cluster Dorsal
Front	*pill pVp,* *peel pVp*	*pill tVp,* *peel tVp*	*pill kVp,* *peel kVp*	*pilp hVp*	*pilt hVp*	*pilk hVp*
Back	*Paul pVp,* *pol pVp*	*Paul tVp,* *Pol tVp*	*Paul kVp,* *Pol kVp*	*polp hVp*	*polt hVp*	*polk hVp*
Low	*parl pVp,* *puhl pVp*	*parl tVp,* *puhl tVp*	*parl kVp* *puhl kVp*	*pulp hVp*	*pult hVp*	*pulk hVp*

6.2.3 Procedure

Participants read each word as it was presented orthographically on a computer monitor. Participants were seated approximately 150 cm from a computer screen and were introduced to the task and the experimental materials with a short practice block. Each trial began with a blank screen for 500 ms, followed by the written stimulus for 2000 ms. After 2000 ms, the experiment automatically moved on to the next trial. Target words were divided into two blocks, one for baseline targets and one for targets intended to elicit vocalised /l/. Targets were randomised within blocks and the order of the blocks was counterbalanced between participants. Each block was repeated 8 times, providing 408 target words per participant.

6.2.4 Data acquisition

Articulatory data were acquired using electromagnetic articulography (EMA). EMA records the movement of articulators over time in an electromagnetic field by tracking sensors attached to the participant. Eleven sensors were used. Five sensors were attached to the tongue to track lingual articulation: the tongue tip (TT), tongue body (TB), tongue dorsum (TD), and the left lateral and right lateral sensors; data from the parasagittal sensors were not analysed because they were not reliable. Two sensors were attached to the lips (upper- and lower lips) to track lip aperture and lip rounding. One sensor was attached to the gumline below the lower incisor to measure jaw movement. There were three reference sensors to correct for head movement (nasion, left mastoid, right mastoid). The occlusal plane was located with a bite trial and the palate was traced with a palate probe. The intersection of the occlusal plane and the incisors was defined as the origin for all sensor measurements: vertical displacement is expressed relative to the occlusal plane and horizontal displacement relative to the upper incisors.

Audio was acquired using two microphones located 150 cm from the participants' lips and offset by 15°. A Røde NTG-1 was connected through a Focusrite OctoPre MkII preamplifier to the NDI Wave system, recording synchronised acoustic data simultaneously with the spatial data from the sensor coils. A second microphone (Røde NT1-A) was connected through a separate Focusrite OctoPre MkII preamplifier to the computer presenting the experimental stimuli, capturing the same responses as a series of WAV files using SpeechRecorder (Draxler & Jansch 2017).

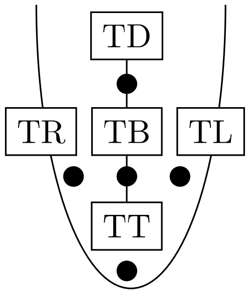


Figure 6.1: Tongue sensor placements viewed from top.

6.2.5 Data analysis

Exclusion criteria

408 (tokens) × 6 (participants) = 2,448 tokens were collected. Tokens were excluded if they were misread (84 tokens), if the sensors were tracked incorrectly (171) or when a sensor fell off and reattaching it resulted in a somewhat different, therefore incomparable sensor configuration (291 tokens). In total, 1,902 tokens were analysed, 909 in the baseline context and 993 in the vocalisation context.

Articulatory data analysis

The shape of the palate was recorded by tracing the midline of the roof of the oral cavity from the soft palate towards the upper incisors, using a 6D probe sensor. The location of the palate was also estimated from the complete lingual trajectories of the three lingual sensors (TD, TB, TT). Because the tongue comes into contact with the roof of the mouth during obstruent production and when resting against the palate, a convex hull defined over this set of points describes the upper limit of lingual excursion, onto which the probe-defined palate trace can be mapped. This latter method of palate estimation was used to define the palate trace for each subject so as to

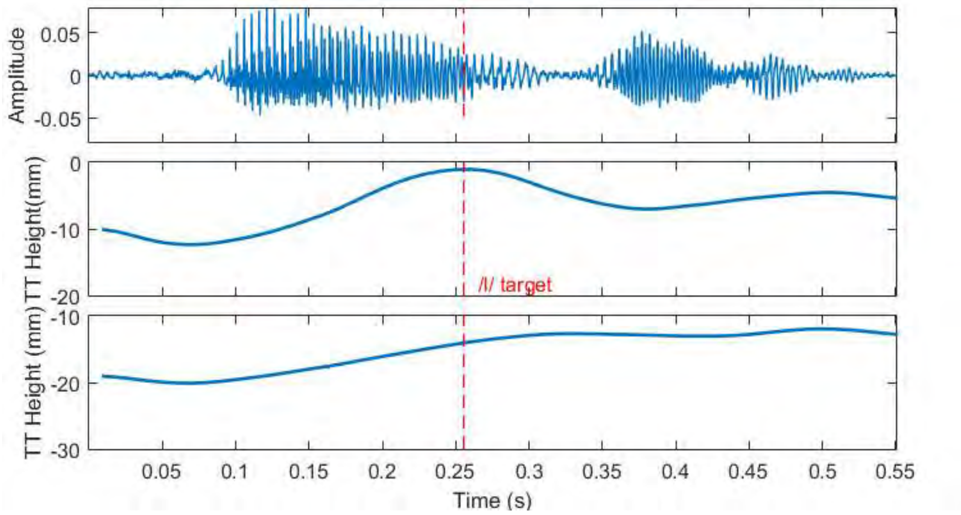


Figure 6.2: Identifying coronal targets in coda laterals: *pol heap* (W2, 8^{th} repetition). Top panel: acoustic speech waveform. Middle panel: vertical tongue tip trajectory. Bottom panel: horizontal tongue tip trajectory. Red line: coronal target of coda /l/ identified at maximum TT height and fronting.

correct for any minor alignment errors arising from rotation of partially and variably parasagittal EMA data into a common mid-sagittal reference system. Individual differences between participants' palates can be observed in **6.4** and **6.5**.

Targets of the tongue tip gesture of /l/ and /d/ in the baseline target phrases (**Table 6.1a**) were identified on the basis of velocity profiles using a semi-automatic labelling procedure *lp_findgest*, implemented in MView, a MATLAB software package (Tiede 2005). The coronal gestural target for /d/ and /l/ was identified when when the tongue tip achieved maximum height and fronting (**Figure 6.2**).

In some cases – especially vocalised laterals – coronal gestures could not be identified reliably from velocity profiles, either due to tongue tip lenition or to gestural overlap between adjacent segments. **Figure 6.3** illustrates a case of the token *pol keep* in which the tongue tip raising gesture of /l/ is

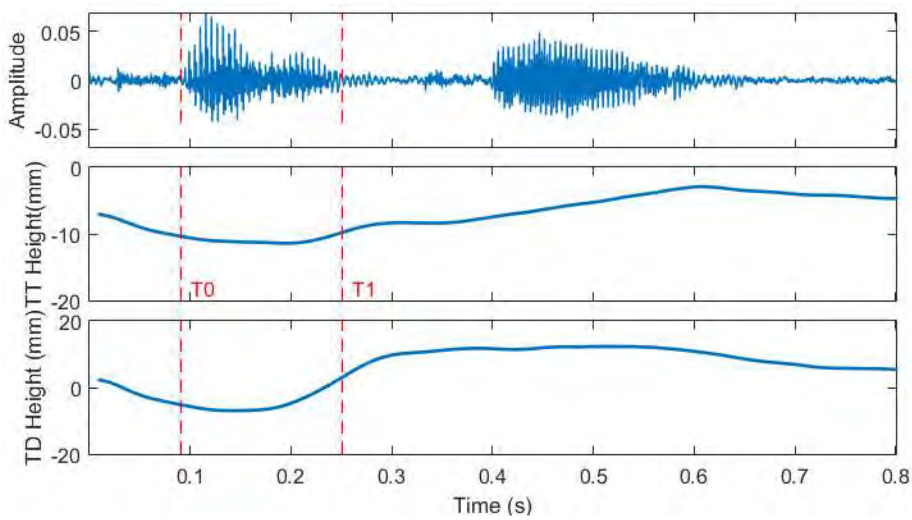


Figure 6.3: Identifying analysis window in coda laterals without clear articulatory targets: *pol keep* (W2, 2nd repetition). Top panel: acoustic speech waveform. Middle panel: vertical tongue tip trajectory. Bottom panel: vertical tongue dorsum trajectory. T0: start of the analysis window identified at the acoustic vowel onset. T1: end of the analysis window identified at the acoustic offset of /l/.

delayed and overlaps with the tongue dorsum raising gesture of the following /k/. Because the beginning and the end of /l/ could not be identified reliably in the acoustic data due to the lack of discernible boundary between vowels and /l/, the analysis window was defined from the acoustic beginning of the vowel (T0 in **Figure 6.3**) to the end of /l/ (T1 in **Figure 6.3**) using the forced-aligner MAUS (Schiel 1999). From this window, unfiltered articulatory trajectories were extracted for all tokens designed to elicit vocalised /l/ (**Table 6.1b**) (Wieling 2018).

TT aperture was calculated as the Euclidean distance of the TT sensor to the closest point on the palate: 0 aperture indicates a full closure, while large TT aperture indicates lenition. In the baseline context, TT aperture was calculated at the coronal target (**Figure 6.2**). **Figure 6.4** illustrates individual coronal targets for each speaker in onset and coda /d/, and in

Table 6.2: Tokens visualised in **Figure 6.4**

Segment	Position	Target	W2	W3	W4	W5	W7	W10
/d/	Onset	*dorp heap*	Rep. 5	Rep. 2	Rep. 2	Rep. 2	Rep. 2	Rep. 1
	Coda	*poured heap*	Rep. 2	Rep. 3	Rep. 3	Rep. 2	Rep. 2	Rep. 8
/l/	Onset	*lorp heap*	Rep. 3	Rep. 1	Rep. 2	Rep. 2	Rep. 4	Rep. 1
	Coda	*Paul heap*	Rep. 3	Rep. 5	Rep. 5	Rep. 1	Rep. 1	Rep. 1

onset and coda /l/ (**Table 6.2**). **Figure 6.5** illustrates the mean of coronal targets in the baseline context averaged over all vowels (low, front, and back; long and short); discrepancies in the coronal target between **Figures 6.4** and **6.5** might arise from the fact that individual tokens represent tongue shapes at the coronal target of /d/ and /l/ in the context of long back vowel, while mean targets were calculated across all vowel contexts. In the vocalisation context, TT aperture was calculated at every 10 ms in the analysis window (**Figure 6.3**: T0 and T1), and trajectories were time-normalised across speakers and tokens to account for differing length of the trajectories. Having observed the TT aperture trajectories, alveolar targets of laterals were located automatically in the TT aperture trajectories at the TT aperture minima in the last 40% of the analysis window.

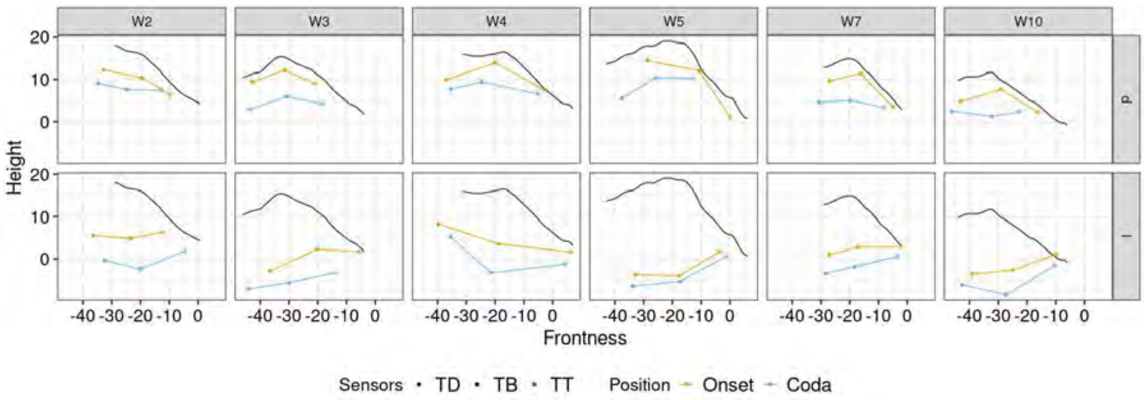


Figure 6.4: Tongue shape at coronal target for /d/ and /l/ in individual tokens. Top panel: tongue shape at the coronal target of /d/ in *dorp heap* and *poured heap*. Bottom panel: tongue shape at the coronal target of /l/ in *lorp heap* and *Paul heap*. Orange: onset position in .dorp heap and *lorp heap* . Blue: coda position in *poured heap* and *Paul heap*. Star: TT sensor. Square: TB sensor. Circle: TD sensor.

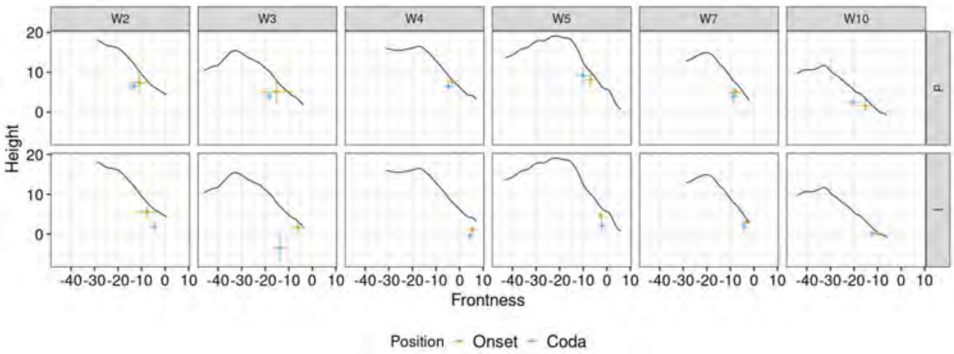


Figure 6.5: Coronal target of /d/ and /l/ in the baseline context by participant. Top panel: mean±s.d. coronal target of /d/. Bottom panel: mean±s.d. coronal target of /l/. Orange: onset position. Blue: coda position.

Statistical analysis

We analysed the baseline context and the vocalisation context separately. In the baseline context, tongue tip lenition in coda /l/ compared to coda /d/ was examined with Linear Mixed Effect Models (LM) using the *lme4* in R (Bates et al. 2015, R Core Team 2018). *p*-values were calculated with the *lmerTest* package (Kuznetsova et al. 2017) using Satterthwaite's degrees of freedom method. To establish that tongue tip lenition is larger in coda /l/ than in coda /d/, we built two LMs with the dependent variable TT aperture. The independent variables were Segment (treatment coded, comparing /l/ to the baseline /d/) and Position (treatment coded, comparing Coda to the baseline Onset). The models contained a two-way interaction between Segment and Position expressing whether tongue tip lenition differs between coda /l/ and coda /d/. In the first model, a random by-participant intercept and slope for Position was used. To test whether coda /l/ lenition varies between participants, a random by-participant slope for the Segment-Position interaction was added in the second model[2]. Model comparison showed that a random slope significantly improves model fit for TT aperture($\chi^2 = 134.43, p < 0.001$), therefore we report results from the model containing the random slope. In addition, the improvement indicates that tongue tip lenition in coda /l/ varies between participants. Therefore, to explore individual variation, we built a linear model with the fixed factors Segment, Coda, and Participant (all interacting), followed by least square means test in the *emmeans* package. Least square means tests with Bonferroni correction tested for each participant whether the difference between onset /l/ and coda /l/ is significantly larger than the difference between coda /d/ and onset /d/.

In the vocalisation context, TT aperture trajectories were modelled for each speaker separately due to the low number of participants and the inter-speaker variation observed in the baseline context. TT aperture trajectories were modelled using generalised additive modelling (GAM) in the *mgcv* and

[2]The model failed to converge with random slopes for Segment, Position, and the Segment-Position interaction. Random effects with the largest by-subject variance were selected.

itsadug libraries of R (Wieling 2018, Woods et al. 2015, van Rij et al. 2017). GAM is a non-linear regression model which can be used to analyse change in articulatory trajectories over time by computing the best fitting non-linear basis function on a trajectory (Wieling 2018). We modelled trajectories as the function of Consonant Place of Articulation (treatment coded, comparing alveolar and dorsal to the baseline labial), Vowel Place of Articulation (treatment coded, comparing front and back to the baseline low), and Syllable Type (treatment coded, comparing Long Nucleus-Simple Coda and Short Nucleus-Complex Coda to Short Nucleus-simple Coda). The models included all three-way interactions. Random effects were not added as speakers were modelled separately. Autocorrelations between measurements at adjacent time points were accounted for in the model.

As GAMs modelling trajectories test the effect of phonetic context on the vocalic portion of the rime as well as on the lateral, we provide a more focused analysis by testing the effect of phonetic context on the lateral targets. The effect of phonetic context on lateral targets in the vocalisation context were examined using linear models in the *lme4* package (Bates et al. 2015). We constructed six models, one for each participant, with the dependent variables tongue tip aperture, and the interacting independent variables Consonant Place of Articulation (sum coded), Vowel Place of Articulation (sum coded), and Syllable Type (sum coded). The three-way interactions were added to all models to improve the reliability of main effects and two-way interactions, but are not discussed. Random effects were not added as speakers were modelled separately.

6.3 Results

6.3.1 Tongue tip lenition in coda /l/ in the baseline data

TT aperture in the baseline data of all participants (**Table 6.1a**) were analysed using GLM; estimates (β), t-values with degrees of freedom and p-values calculated with Satterthwaite's degrees of freedom method are reported. Analysis of all participants showed that TT aperture at onset /d/ target did not differ significantly from TT aperture at onset /l/ target ($\beta = -0.07, t_6 = 4.27, p = 0.56$), indicating no evidence for lenition in onset /l/. TT aperture was 1.6 mm larger at coda /d/ target than at onset /d/ target ($\beta = 1.62, t_5 = 3.76, p = 0.01$), indicating coda lenition in /d/. Contra Hypothesis 1, the increase in TT aperture at coda /l/ target was not significantly larger than the increase in TT aperture at coda /d/ target ($\beta = 0.69, t_5 = 0.9, p = 0.41$), indicating no evidence for different coda /l/ lenition compared to coda /d/ lenition.

Interaction quantifying the onset-coda difference in /l/ compared to /d/ was associated with significant by-subject variance. Analysis of individual variation in the baseline data, using planned comparisons on the model including Participant as a fixed factor, showed that the onset-coda difference in TT aperture was significantly larger in /l/ compared to /d/ for participants W3 and W5, and smaller for W10 (**Table 6.3**, **Figure 6.7**). Significantly larger coda lenition in /l/ compared to /d/ in the speech of W3 and W5 partially supports Hypothesis 1.

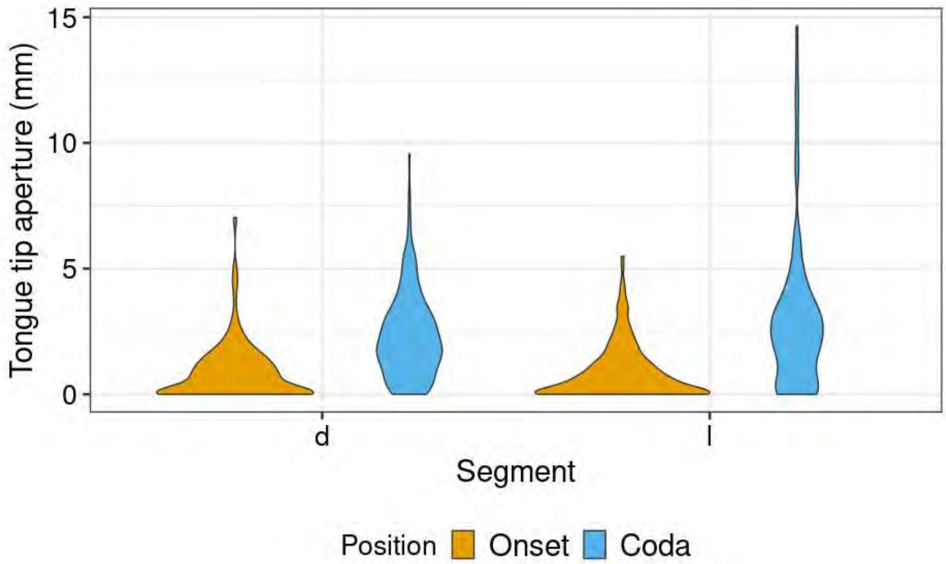


Figure 6.6: Distribution of TT aperture at coronal target across all participants. X-axis: Segment. Orange: onset. Blue: coda. Larger tongue tip aperture in coda position indicates coda lenition.

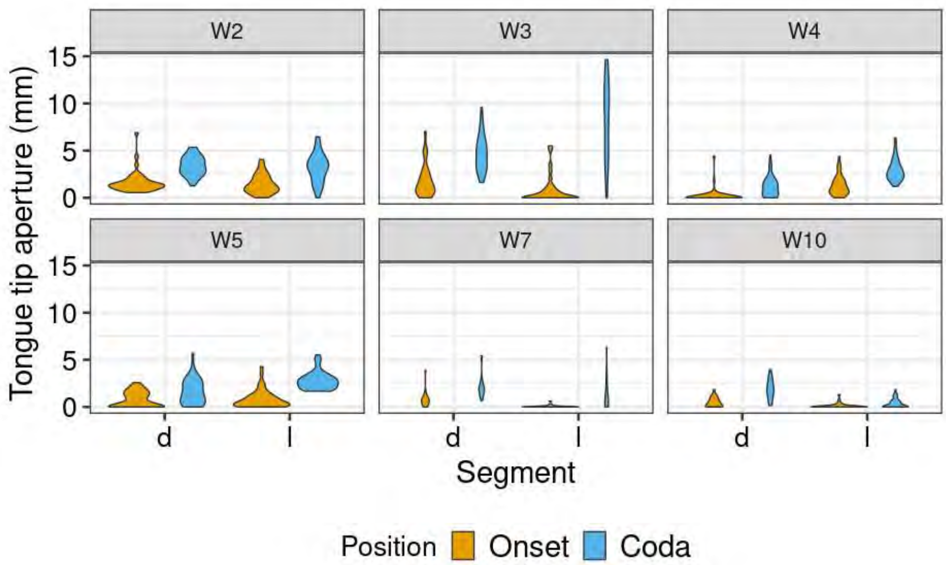


Figure 6.7: Distribution of TT aperture at coronal target by participant. X-axis: Segment. Orange: onset. Blue: coda. Larger TT aperture in coda position indicates coda lenition.

Table 6.3: Coda /l/ lenition compared to coda /d/ lenition. β shows the onset /l/ - coda /l/ difference in tongue tip aperture compared to the onset /d/ - coda /d/ difference. SE, t-ratio, and p-value calculated from least square means.

Participant	β	Standard Error	t-ratio	p-value
W2	-0.04	0.41	-0.09	1
W3	4.48	0.48	9.3	< 0.0001
W4	0.63	0.37	1.7	0.55
W5	1.05	0.38	2.74	0.04
W7	-0.1	0.53	0.19	1
W10	-1.74	0.39	-2.98	0.02

6.3.2 Coda /l/ trajectories across phonemic contexts

TT aperture trajectories across phonemic contexts were analysed separately for each participant, using GAM. A TT aperture at 0 indicates a full closure, while large TT aperture indicates lenition. Data are presented for each participant, first contrasting W10, who almost always achieved tongue tip contact in coda /l/ in both the baseline and in the vocalising context, with W3 who rarely achieved tongue tip contact in coda /l/ in either the baseline or the vocalising context, followed by W2, W7, W5, and W4 who produced coda /l/ with a varying degree of coronal constriction reduction (**Figures 6.8–6.13**). Model outputs are reported in **Appendix 6.7.1**; however estimates need to be interpreted carefully, because estimates in a GAM model indicate the average difference between two trajectories and therefore might reflect differences in the vocalic part of the analysis window, and not differences associated with the lateral. Conclusions are based on mean trajectories and confidence intervals in **Figures 6.8–6.13**.

W10 consistently reached full tongue tip closure in coda /l/ across most phonemic contexts, except in target words containing a back vowel and a coda cluster with /k/ (e.g. *polk*), in which she produced larger TT aperture

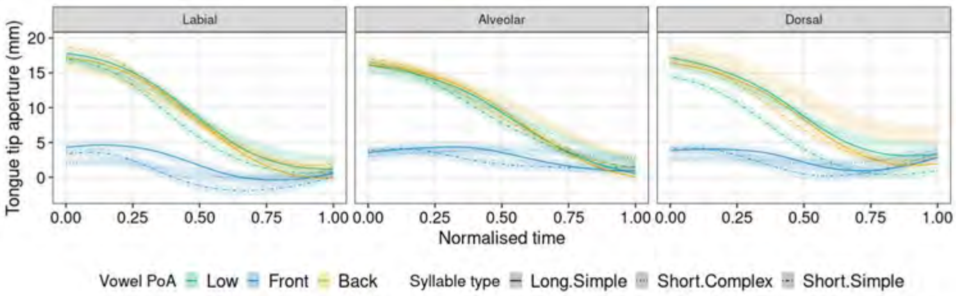
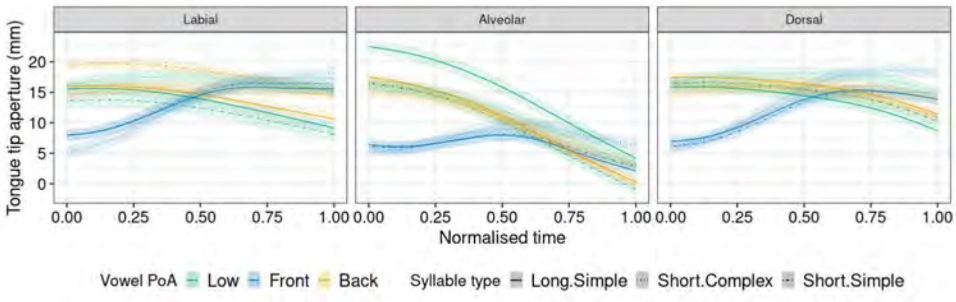


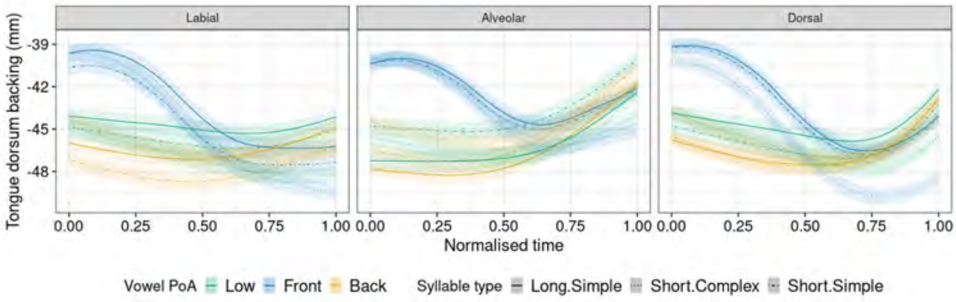
Figure 6.8: W10's production of TT aperture trajectories. X axis: normalised time in the analysis window identified by the acoustic onset of the vowel and the acoustic offset of /l/. Left panel: labial consonant context. Middle panel: alveolar consonant context. Right panel: dorsal consonant context. Green: low vowel context. Blue: front vowel context. Yellow: low vowel context. Solid line: long nucleus, simple coda context. Dotted line: short nucleus, complex coda context. Dot-dashed line: short nucleus, simple coda context.

(**Figure 6.8**). The coronal target of /l/ seems to be achieved earlier in the front vowel context compared to other vowel contexts (**Figure 6.8**). The negative TT aperture in the front vowel context is due to the GAM model oversmoothing the trajectories; the original dataset did not contain a TT aperture value below zero.

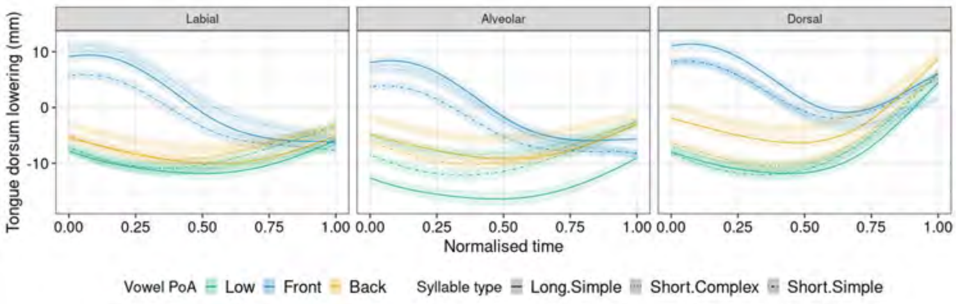
In contrast to W10, W3 only produced full alveolar closure within the analysis window in two target words in the alveolar context: *Paul teep* (/poːl tiːp/) and *puhl teep* (/pɐl tiːp/). TT aperture trajectories in other target words in the alveolar context showed a steep downward trajectory approaching but not reaching zero. As a result, the range of TT aperture was large (0–10mm) in the alveolar context. TT aperture decreased gradually in the dorsal and labial context when the vowel was not front, resulting in a 10–18 mm minimum TT aperture. When the vowel was front and the following consonant was labial or dorsal, TT aperture increased in /l/ compared to the front vowel, and a coronal target can be observed 15–18 mm away from the alveolar ridge. To investigate the increasing tongue tip aperture in tar-



(a) TT aperture



(b) TD backing

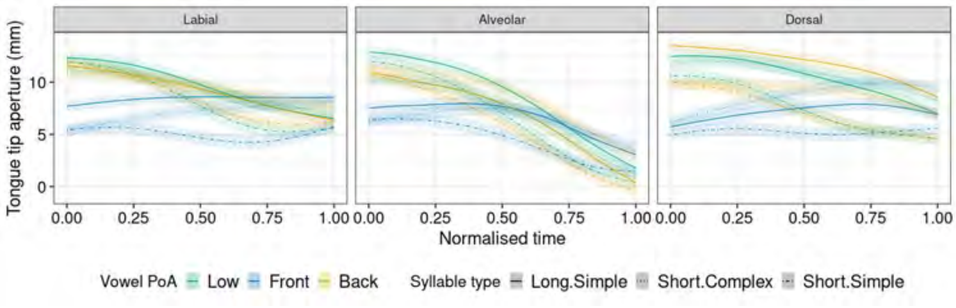


(c) TD lowering

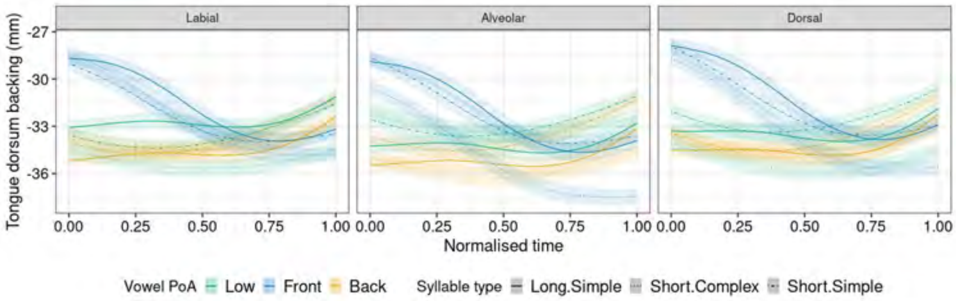
Figure 6.9: W3's production of TT aperture, TD backing, and TD lowering trajectories. Top panel: TT aperture. Middle panel: TD backing. Bottom panel: TD lowering. X axis: normalised time in the analysis window identified by the acoustic onset of the vowel and the acoustic offset of /l/. Left panel: labial consonant context. Middle panel: alveolar consonant context. Right panel: dorsal consonant context. Green: low vowel context. Blue: front vowel context. Yellow: low vowel context. Solid line: long nucleus, simple coda context. Dotted line: short nucleus, complex coda context. Dot-dashed line: short nucleus, simple coda context.

get words containing a front vowel and a non-alveolar coda, tongue dorsum fronting and raising trajectories were modelled (**Figures 6.9b 6.9c**). The increasing TT aperture trajectories in the front vowel-labial consonant context can be attributed to the increasing tongue dorsum backing and lowering trajectories in the same context (**Figures 6.9b, 6.9c**). In the front vowel-labial consonant context the tongue dorsum was backed and lowered in /l/ compared to the preceding vowel, and remained in a low-back position until the end off the analysis window. In the non-front-labial context, the tongue dorsum started to raise and move to the front at the end of the analysis window. This difference can be attributed to the carrier phrase in which target words were followed by a word with an antagonistic vowel (e.g *peel parp* and *pol peep*). In contrast, the increasing TT aperture in the front vowel - dorsal consonant context was independent from the tongue dorsum movement, as tongue dorsum trajectories indicate that the tongue dorsum was raised at the end of the analysis window due to the following dorsal consonant (**Figures 6.9b, 6.9c**).

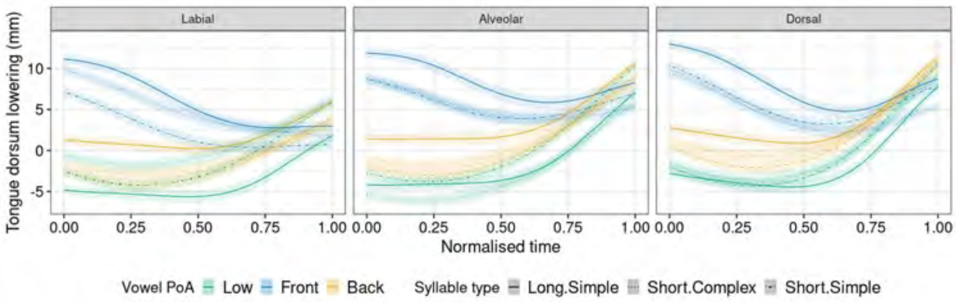
W2 only achieved full tongue tip closure in coda /l/ when /l/ was followed by an alveolar consonant: the range of tongue aperture was between 0–5 mm in the alveolar context (**Figure 6.10**). TT aperture within the alveolar context was smaller when the nucleus was short and the coda was simple. TT aperture appeared to be larger in the dorsal context (5–10 mm) than in the labial context (5–7.5 mm). Similarly to W3, TT aperture decreased gradually in the dorsal and labial context for most contexts, but increased when the vowel is front and the coda was complex or the nucleus was long (*pilp, pilk, peel parp, peel karp*). In the target words *pill parp, pill karp* a coronal target was achieved early (at 0.6 in the analysis window), similarly to the early coronal targets produced by W10 across all



(a) TT aperture



(b) TD backing

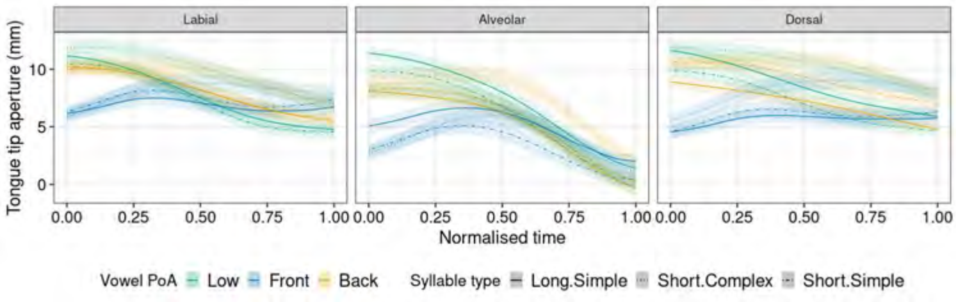


(c) TD lowering

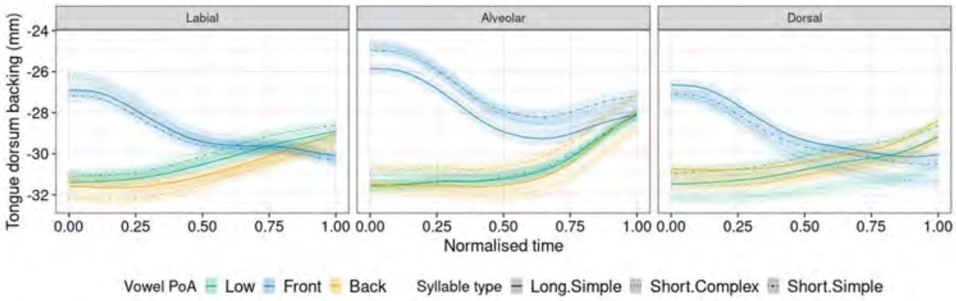
Figure 6.10: W2's production of TT aperture, TD backing, and TD lowering trajectories. Top panel: TT aperture. Middle panel: TD backing. Bottom panel: TD lowering. X axis: normalised time in the analysis window identified by the acoustic onset of the vowel and the acoustic offset of /l/. Left panel: labial consonant context. Middle panel: alveolar consonant context. Right panel: dorsal consonant context. Green: low vowel context. Blue: front vowel context. Yellow: low vowel context. Solid line: long nucleus, simple coda context. Dotted line: short nucleus, complex coda context. Dot-dashed line: short nucleus, simple coda context.

front vowel contexts. To investigate the increasing tongue tip aperture in the *pilp, pilk, peel parp, peel karp* target words, tongue dorsum fronting and raising trajectories were modelled (**Figures 6.10b 6.10c**). The increasing TT aperture trajectories in the *pilp, pilk, peel parp, peel karp* target words cannot be attributed to an increasing tongue dorsum backing or lowering. Although the tongue dorsum remained back and low until the end of the analysis window in *pilp, peel parp* due to following low vowel in the carrier phrase, tongue dorsum backing and lowering were not greater in *pilp, peel parp* than in *pill parp* (**Figures 6.10b, 6.10c**). In fact *pill parp* shows both larger tongue dorsum lowering and smaller TT aperture than *pilp, peel parp* (**Figure 6.10c**). In the dorsal context, *pilk, peel karp* did not show larger tongue dorsum lowering or backing than in *pill karp* (**Figures 6.10b, 6.10c**), therefore tongue dorsum movement cannot explain the increased TT aperture.

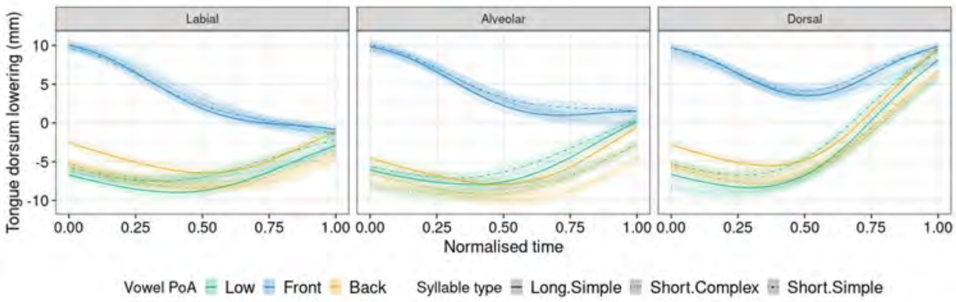
W7 only achieved full tongue tip closure in coda /l/ when /l/ was followed by an alveolar consonant: the range of tongue aperture was between 0–2.5 mm in the alveolar context (**Figure 6.11**). TT aperture appears to be larger in the dorsal context than in the labial context, as the trajectories were steeper in the labial context compared to the dorsal context. Similarly to what has been observed for W3, TT aperture decreased gradually in the dorsal and labial context for most vowel- and syllable contexts, but increased when the vowel was front and the coda was complex (*pilp, pilk*). When the vowel was front and the coda was not complex (*pill parp, pill karp, peel parp, peel karp*), a local minima indicating a coronal target was achieved (at 0.75 in the analysis window), similarly to the early coronal targets produced by W10 across all front vowel contexts. To investigate the increasing tongue tip aperture in the *pilk* target words, tongue dorsum fronting and



(a) TT aperture



(b) TD backing



(c) TD lowering

Figure 6.11: W7's production of TT aperture, TD backing, and TD lowering trajectories. Top panel: TT aperture. Middle panel: TD backing. Bottom panel: TD lowering. X axis: normalised time in the analysis window identified by the acoustic onset of the vowel and the acoustic offset of /l/. Left panel: labial consonant context. Middle panel: alveolar consonant context. Right panel: dorsal consonant context. Green: low vowel context. Blue: front vowel context. Yellow: low vowel context. Solid line: long nucleus, simple coda context. Dotted line: short nucleus, complex coda context. Dot-dashed line: short nucleus, simple coda context.

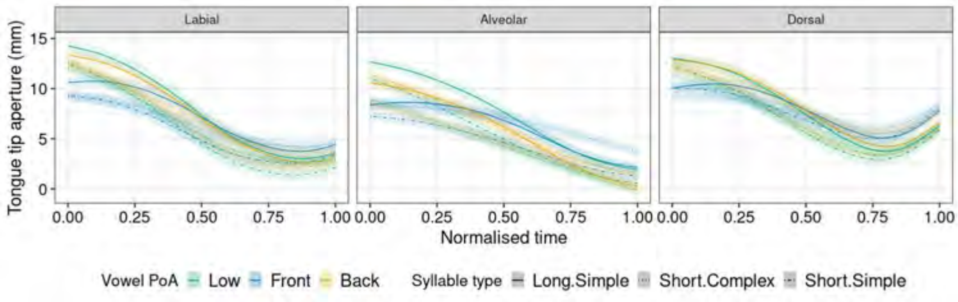


Figure 6.12: W5's production of TT aperture trajectories. X axis: normalised time in the analysis window identified by the acoustic onset of the vowel and the acoustic offset of /l/. Left panel: labial consonant context. Middle panel: alveolar consonant context. Right panel: dorsal consonant context. Green: low vowel context. Blue: front vowel context. Yellow: low vowel context. Solid line: long nucleus, simple coda context. Dotted line: short nucleus, complex coda context. Dot-dashed line: short nucleus, simple coda context.

raising trajectories were modelled(**Figures 6.11b 6.11c**). The increasing TT aperture in the *pilk* target words can be attributed to a larger tongue dorsum backing in *pilk* compared to *pill karp* and *peel karp*, but not in the *pilp* target words, as *pilp, pill parp* and *peel parp* did not differ in tongue dorsum backing and lowering.

W5 only achieved full tongue tip closure in /l/ when /l/ was followed by an alveolar consonant: the range of tongue aperture was between 0–2.5 mm in the alveolar context (**Figure 6.12**). In the labial context, local TT aperture minima indicating the alveolar target appears to be smaller (1–5 mm), longer, and to come later (0.77–0.87) than in the dorsal context (2.5 – 6 mm at 0.75–0.77).

W4 only achieved full tongue tip closure in /l/ when /l/ was followed by an alveolar consonant: the range of tongue aperture was between 0–2.5 mm in the alveolar context (**Figure 6.13**). TT aperture appears to be larger in the dorsal context than in the labial context, as the trajectories were steeper in the labial context compared to the dorsal context. All other

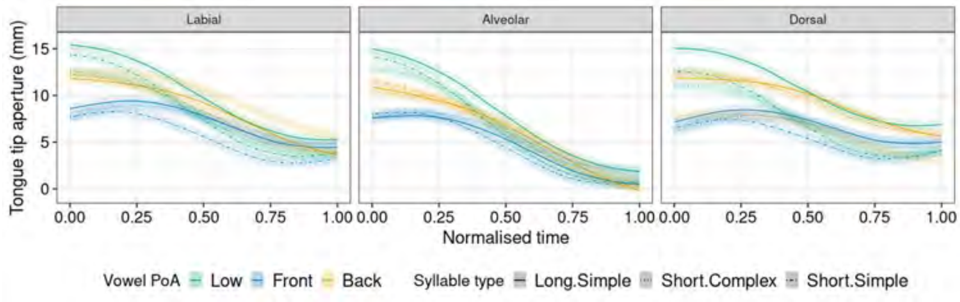


Figure 6.13: W4's production of TT aperture trajectories. X axis: normalised time in the analysis window identified by the acoustic onset of the vowel and the acoustic offset of /l/. Left panel: labial consonant context. Middle panel: alveolar consonant context. Right panel: dorsal consonant context. Green: low vowel context. Blue: front vowel context. Yellow: low vowel context. Solid line: long nucleus, simple coda context. Dotted line: short nucleus, complex coda context. Dot-dashed line: short nucleus, simple coda context.

contexts being equal, local TT aperture minima indicates that the alveolar target seem to appear earlier in the short nucleus - simple coda context. In the labial and dorsal contexts, there was a smaller TT aperture following a front vowel when the coda was not complex compared to other vowels.

With respect to the effect of segmental context on /l/ lenition, we hypothesised that a following alveolar consonant would inhibit /l/-vocalisation due to it being homorganic with the alveolar gesture of /l/ (Hypothesis 2. In line with Hypothesis 2, we found that the pre-alveolar context was the only context in which all participants achieve full closure. W2 and W5 seem to have larger TT aperture in the dorsal than in the labial context, while W4 and W7 produced a steeper decrease in TT aperture in the labial than in the dorsal context. TT aperture trajectories do not indicate a consistent pattern between speakers in the dorsal and labial contexts. Although W4, W2, W3, and W7 did not achieve coronal closure before a labial or a dorsal consonant within the analysis window, TT aperture trajectories indicate decreasing aperture until the end of the analysis window in most vowel contexts (**Figures 6.9–6.11** and **6.13**). Therefore, it is possible that the TT constriction is not reduced before a labial or dorsal consonant, but delayed and falls outside of the analysis window. In contrast, W5 achieves a local minimum TT aperture within the analysis window and TT aperture starts increasing before the end of the analysis window (**Figure 6.12**).

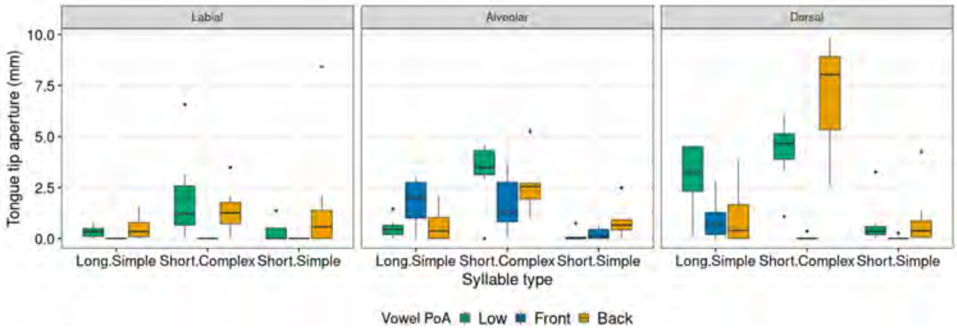
We also hypothesised that a preceding back vowel would facilitate /l/-vocalisation due it its articulatory similarity to the tongue dorsum gesture of /l/ (Hypothesis 2. We found no observable differences between the TT aperture trajectories in the low and the back vowel contexts; however, front vowels seem to differ from non-front vowels. W10 produced an early alveolar target associated with /l/ in the front vowel context, while W4 produced a smaller TT aperture target in *pill parp, peel parp, pill karp, peel karp* compared to other vowels. W7, W2, and W3 all produced an increasing TT aperture trajectory from the front vowel towards the /l/ in the labial and dorsal contexts. W7 produced a local maxima when the preceding vowel was front and coda was complex prior to the decrease in TT aperture at the end of the analysis window. W2 produced a gradual increase till the end of the analysis window after a front vowel when the coda was complex or the nucleus was long. W3 also produced a gradual increase till the end of the analysis window after a front vowel in all pre-labial and pre-dorsal contexts, irrespective of syllable type.

Lastly, regarding the effect of syllable type on /l/ lenition, we hypothesised that syllables containing a complex coda or a long nucleus would facilitate /l/-vocalisation compared to syllables containing a short nucleus with a simple coda (Hypothesis 3). When the following consonant is labial or dorsal, trajectories indicate a larger TT aperture in /l/ in the complex coda and long nucleus contexts for W3, and in the complex coda context for W7. We observed no effects consistent with this hypothesis for W2, W4, W5, and W10.

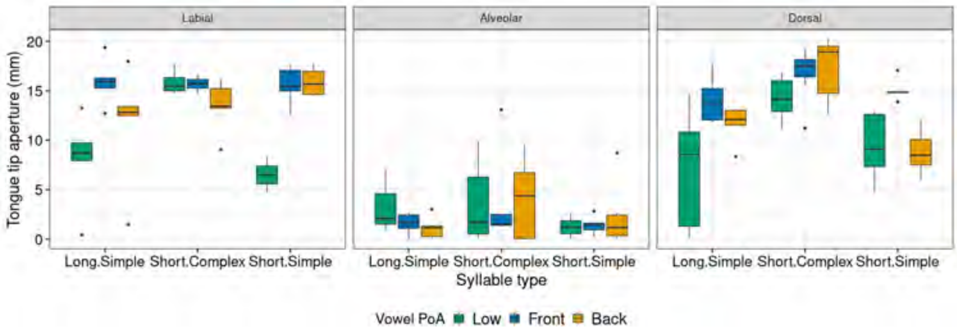
6.3.3 Coda /l/ targets across phonetic contexts

TT aperture targets were analysed separately for each participant, using linear regression. A positive estimate for a main effect indicates that a context facilitates coronal constriction reduction, while a negative estimate indicates that a context inhibits vocalisation. A significant interaction has three possible interpretations. An interaction with the same sign as as both main effects indicates that the two main effects strengthened each other. An interaction with a sign opposite to both main effects indicates that the main effects did not strengthen each other. Lastly, when two main effects oppose each other, a significant interaction shows that one main effect mitigated the effect of the other. Main effects rarely varied between participants, indicating consistent patterns in the effect of phonetic context on /l/-vocalisation; however, interactions varied greatly between participants, indicating interspeaker variability.

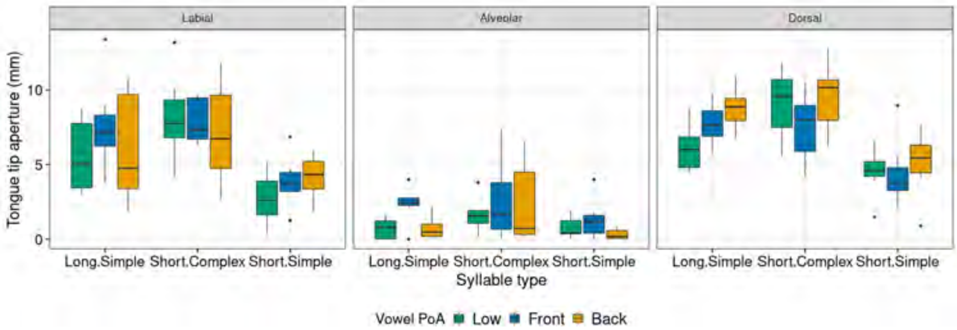
TT aperture was significantly increased in the dorsal consonant context for all participants and in the context of a complex coda for all participants except W4 (**Table 6.4**). A further increase in TT aperture when the dorsal was part of an /lk/ cluster suggests that these effects strengthened each other for W3, W5, W7, and W10. TT aperture was significantly increased in the back vowel context for W4 and W10. Increase in TT aperture when the target contained a back vowel and a dorsal consonant indicates that these effects strengthened each other for W4, W10, and for W2, although she did not show a significant back vowel effect. The back vowel and the complex coda effects only strengthened each other for W10 (**Table 6.4**).


Labial
Alveolar
Dorsal
Tongue tip aperture (mm)
10.0
7.5
5.0
2.5
0.0
Long.Simple Short.Complex Short.Simple
Syllable type
Vowel PoA Low Front Back

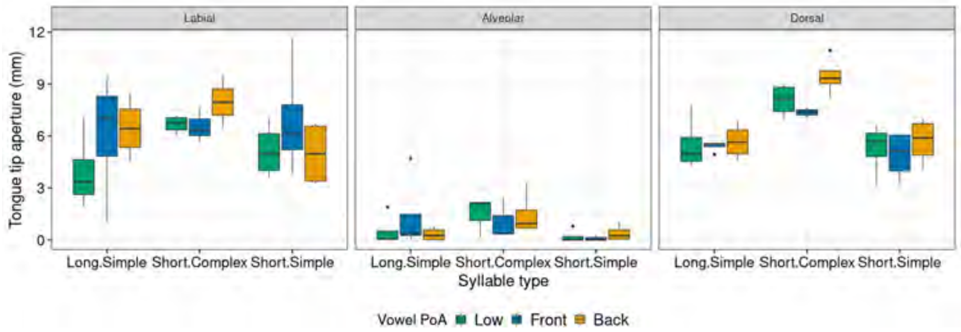
(a) W10


Labial
Alveolar
Dorsal
Tongue tip aperture (mm)
20
15
10
5
0
Long.Simple Short.Complex Short.Simple
Syllable type
Vowel PoA Low Front Back

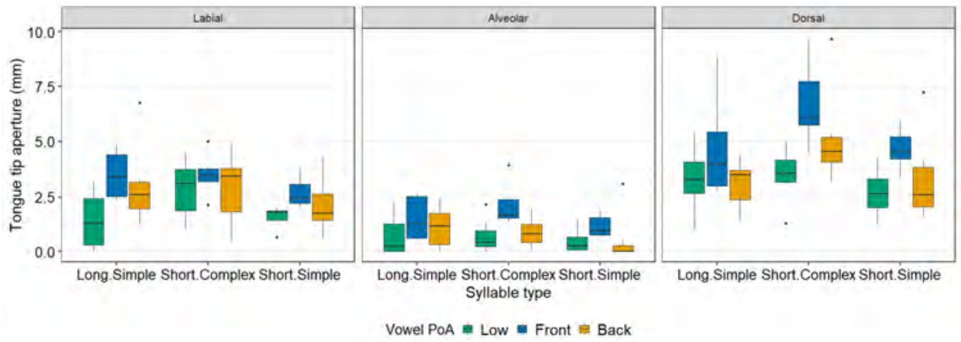
(b) W3


Labial
Alveolar
Dorsal
Tongue tip aperture (mm)
10
5
0
Long.Simple Short.Complex Short.Simple
Syllable type
Vowel PoA Low Front Back

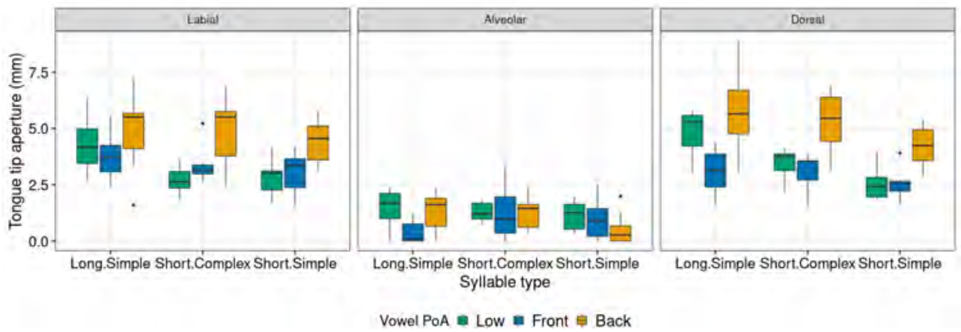
(c) W2



(d) W7



(e) W5



(f) W4

Figure 6.14: Minimum TT aperture by participant. Left panel: labial consonant context. Middle panel: alveolar consonant context. Right panel: dorsal consonant context. Green: low vowel context. Blue: front vowel context. Yellow: low vowel context. X-axis: Syllable type.

TT aperture was significantly decreased in the alveolar consonant context for all participants except W10. TT aperture was decreased in the long nucleus context for W3, W7, and W10. TT aperture was increased after a long nucleus for W4; however, the negative estimate in the alveolar-long nucleus interaction suggests that an alveolar consonant opposed the effect of a long nucleus (**Table 6.4**).

Results so far have revealed opposing main effects on the TT aperture, which decreased in the alveolar context and increased in the complex coda and in the back vowel contexts. It appears that the effect of the alveolar consonant is overall stronger: TT aperture also decreased when the target contained an alveolar consonant in a complex coda for W2, W5, and W7 and when the target contained an alveolar consonant and a back vowel for W2, W4, and W10. Main effects showed that TT aperture significantly decreased in the long nucleus context; this mitigated the effect of back vowel for W10 and the effect of dorsal consonant for W5 (**Table 6.4**).

The effect of front vowel context on TT aperture differed between participants: W4 and W10 produced a smaller TT aperture, W3 and W5 produced a significantly larger, and W2 and W7 produced a non-significantly larger TT aperture in the front vowel context. Interactions show that for W3 and W5, who both produced a larger TT aperture in the front vowel context, a following alveolar consonant mitigated the effect of front vowel (significantly for W3 and non-significantly for W5). These interaction between front vowels and alveolar consonant is consistent with the overall pattern of the alveolar consonant having a stronger effect. For W3 and W5, a following dorsal strengthened the effect of a front vowel. However, the effects of front vowel and complex coda context, both of which increase TT aperture, did not strengthen each other for W3. Interactions show that for W4 and W10, who produced a smaller TT aperture in the front vowel context, the front vowel effect did not strengthen that of the alveolar consonant; however, it mitigated the effect of a dorsal consonant for W4 and W10 and that of a complex coda for W10. Interactions between front vowel context and long nucleus context show that the facilitating effects did not strengthen each other for W10; however, front vowel context mitigated the effect of long nucleus for W4, who produced larger TT apertures in the long nucleus context (**Table 6.4**).

Table 6.4: TT aperture estimates for all participants. Asterisks indicate statistical significance based on p-values calculated with Satterthwaite's degrees of freedom method. *p<0.1; **p<0.05; ***p<0.01

		W2	W3	W4	W5	W7
	Intercept	4.8***	9.5***	2.9***	2.6***	4.4***
Facilitating	Dorsal	2.2***	3.2***	1.0***	1.5***	2.0***
	Short.Complex	1.4***	2.2***	0.03	0.6***	1.2***
	Dorsal:Short.Complex	0.3	1.2**	0.01	0.5**	0.8**
	Back	0.2	0.1	0.7***	−0.1	0.3
	Back:Dorsal	0.7**	−0.3	0.5***	−0.1	0.3
	Back:Short.Complex	−0.04	−0.2	0.1	−0.03	0.4
Inhibiting	Alveolar	−3.4***	−6.9***	−1.8***	−1.6***	−3.6***
	Long.Simple	0.3	−1.2***	0.4***	−0.1	−0.5**
	Alveolar:Long.Simple	−0.4	0.7	−0.4***	0.2	0.5
Opposing	Alveolar:Short.Complex	−0.7**	−0.8	0.2	−0.3*	−0.6*
	Alveolar:Back	−0.5	−0.1	−0.8***	−0.1	−0.3
	Long.Simple:Back	−0.03	−0.3	−0.02	0.1	−0.1
	Long.Simple:Dorsal	0.3	−0.6	0.3**	−0.3*	−0.4
Varied	Front	0.3	1.6***	−0.5***	0.8***	0.01
	Front:Alveolar	0.4	−1.9***	0.3**	−0.2	0.1
	Front:Long.Simple	0.6**	0.6	−0.4**	−0.1	0.4
	Front:Dorsal	−0.7**	1.0*	−0.5***	0.4**	−0.5
	Front:Short.Complex	−0.4	−1.2**	0.1	0.2	−0.6*
Note:					*p<0.1; **p<0.05; **	

6.4 Discussion

The aim of the current study was to examine the effect of coronal contact in AusE coda /l/ using articulatory methods. Hypothesis 1 predicted that coronal lenition would be larger from onset /l/ to coda /l/ than from onset /d/ to coda /d/. Hypothesis 1 did not hold across the six speakers tested in this experiment; however, closer inspection of interspeaker variability showed that two out of six lenited coda /l/ more than coda /d/ in simpleton codas followed by word-onset /h/ (**Figure 6.7**). Lack of increased coda /l/ lenition can be attributed to the fact that both coda /d/ and coda /l/

were followed by /h/, and speakers have been shown to lenite /l/ less before /h/ than before an obstruent (Scobbie & Pouplier 2010), which is why this context was selected for these initial analyses. When we examined coda /l/ lenition in the pre-obstruent context, we found that most participants produced coda /l/ with a lenited coronal gesture in at least some contexts. For instance, W10 and W4 produced /l/ with a larger TT aperture only when /l/ was preceded by a back vowel and followed by a dorsal consonant tautosyllabic with /l/ (*polk* in **Figures 6.8, 6.13**), while W3 and W2 produced /l/ with larger TT aperture in most contexts (**Figures 6.9, 6.10**).

Context effects on TT aperture trajectories and targets show similar patterns between speakers. Hypothesis 2 predicted that speakers would lenite /l/ less before a following alveolar consonant. This part of Hypothesis 2 held, as the only context in which all participants consistently achieved alveolar closure in the analysis window was before a following alveolar consonant. As speakers achieved a coronal closure for /t/, the tongue tip also achieved a canonical /l/ closure, even when the preceding vowel might facilitate /l/-vocalisation. Thus a following /t/ inhibits /l/-lenition even when the vowel context might facilitate it. **Figure 6.15** shows that coda /l/ is more lenited in *pol peep* compared to *puhl peep*, consistent with the effect of back vowel, but shows no lenition in *polt heap*. Although alveolar /t/ prevents /l/-lenition when the vowel context would facilitate it, an alveolar /t/ did not strengthen the effect of preceding vowel inhibiting /l/-lenition. This is presumably due to the fact that we measured /l/-lenition by TT aperture, which cannot go below 0: if TT aperture reached near-0 both in the alveolar context and in the front vowel or long nucleus context, then the inhibitive effects cannot be additive. In the current analyses, it is impossible to separate whether the tongue tip closure is primarily attributed to /l/ or /t/; to answer this question, the length of the coronal gesture in an /lt/ cluster and in coda /l/ and the timing relationships and gestural coordination between the dorsal gesture and the coronal gesture in an /lt/ cluster need to be examined.

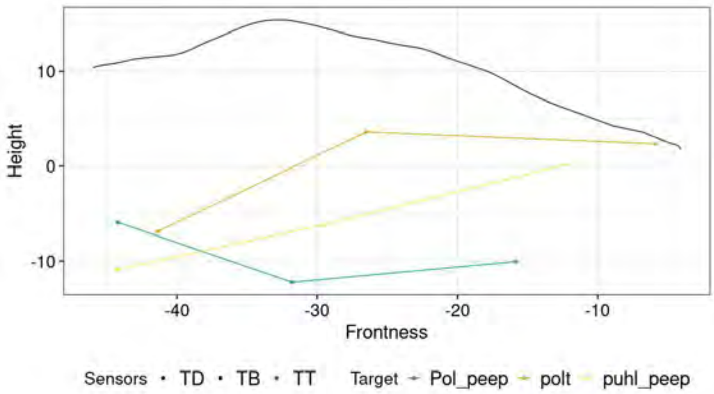


Figure 6.15: Effects of a following alveolar consonant and a preceding back vowel on the coronal target of coda /l/. Coronal target in *puhl peep*, *pol peep* and *polt heap* produced by W3. Yellow: *puhl peep*, repetition 6. Green: *Pol peap*, repetiton 5. Orange: *polt*, repetition 2. Star: TT sensor. Square: TB sensor. Circle: TD sensor.

Hypothesis 2 predicted that speakers would lenite /l/ more after a back vowel due to the articulatory similarity between back vowels and the tongue dorsum gesture of /l/. Hypothesis 2 did not hold, as only two speakers lenited /l/ more after a back vowel. However, all speakers lenite /l/ more before a following dorsal consonant. A following dorsal consonant is not gesturally similar to the dorsal gesture of /l/ as it has a smaller constriction degree and a higher constriction location. Yet, it has a back place of articulation, which apparently can drag the tongue tip away from its canonical alveolar target. In addition, we found that when the vowel context facilitates /l/-lenition, such as the back vowel for W4 and W10, or the front vowel for W3 and W5, the vowel strengthened the effect of the dorsal context, leading to increased /l/-lenition. **Figure 6.16** shows overall larger coronal lenition before /k/ than before /p/, and a larger coronal lenition in *polk* (/pɔlk/) than in *pulk* (/pɐlk/). The findings that a following dorsal consonant facilitates /l/-lenition while a preceding back vowel has a limited influence indicate that /l/-lenition is facilitated by coarticulatory backing of the tongue rather than by articulatory similarity between the adjacent segments and the dorsal gesture of /l/.

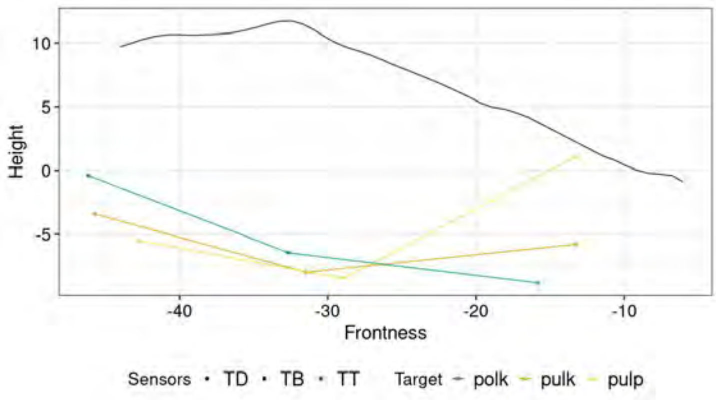
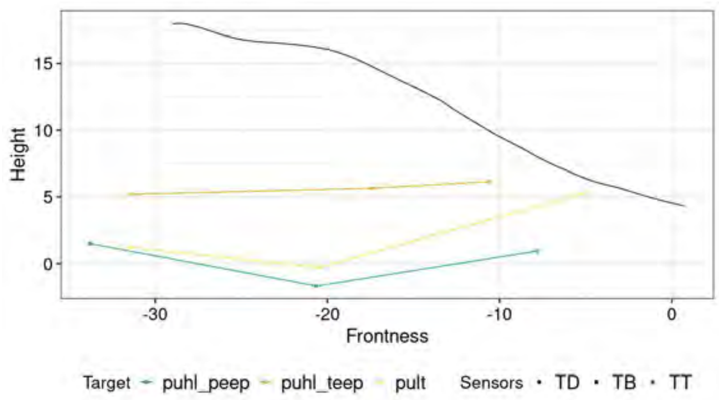


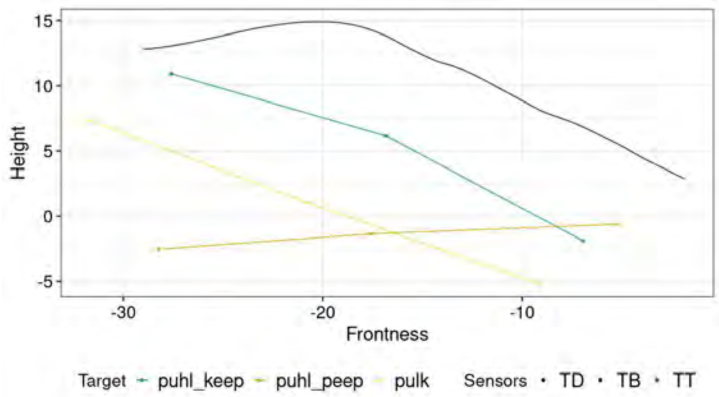
Figure 6.16: Effects of a following dorsal consonant and a preceding back vowel on the coronal target of coda /l/. Coronal target in *pulp*, *pulk*, and *polk* produced by W10. Yellow: *pulp*, repetition 4. Orange: *pulk*, repetition 1. Green: *polk*, repetition 4. Star: TT sensor. Square: TB sensor. Circle: TD sensor.

Speakers showed the largest interspeaker variability with respect to /l/-lenition in the context of a preceding front vowel: out of six speakers, a preceding front vowel facilitated /l/-lenition for two, inhibited for two speakers, and did not affect two speakers. These differences can be attributed to different speaker strategies. Speakers who lenited less in the front vowel context potentially did so due it to coarticulatory reasons: as the front vowel requires tongue tip fronting, the tongue tip is moving closer to its canonical /l/ target, decreasing the probability or degree of lenition. The patterns produced by these two speakers are in line with Borowsky's (2001) result, who also found that speakers vocalise less after a front vowel. Speakers who lenited more in the front vowel context potentially did so because of perceptual reasons: in the front vowel context, dorsum lowering is sufficient to create the percept of /l/, while tongue tip closure is required in the low vowel context (Hardcastle & Barry 1989).

Hypothesis 3 predicted that speakers would lenite more in a complex coda or after a long nucleus. Hypothesis 3 partially held, as speakers lenited more before a complex coda but lenited less after a long nucleus. These results partially contrast with Borowsky's (2001) who found that speakers vocalise more when /l/ is in a complex coda compared to a simple coda and when /l/ is preceded by long nucleus compared to a short nucleus. Although our results on the facilitating effect of coda cluster align with Borowsky's (2001), they cannot be interpreted without considering the place of articulation of the consonant following /l/ in the cluster. When the following consonant in the cluster was alveolar, /l/-lenition decreased, as an alveolar inhibited /l/-lenition in general, and even more so when it was tautosyllabic with /l/. In contrast, when the following consonant was dorsal, /l/-lenition increased as a dorsal facilitated /l/-lenition in general, and even more so when it was tautosyllabic with /l/. **Figure 6.17a** shows that coda /l/ is less lenited in *puhl teep* than in *puhl peep*, and the least lenited in *pult heap*. **Figure 6.17b** shows that coda /l/ is more lenited in *puhl keep* than in *puhl peep*, and the most lenited in *pulk heap*. That is, the effect of the following consonant was always strengthend when the consonant was in a tautosyllabic cluster. This can be attributed to the different gestural coordination between a tautosyllabic /lC/ cluster and a heterosyllabic /l#C/ sequence. In a tautosyllabic cluster, consonantal gestures are coordinated with the nuclear vowel and are shortened, therefore the effect of the obstruent is strengthened, while in a heterosyllabic /l#C/ sequence, the obstruent is coordinated with the nucleus of the following word and affects the realisation of the coronal gesture of /l/ less (Browman & Goldstein 1988). The differing results on the effect of long nucleus can be attributed to interspeaker variability as out of six speakers, one speaker did lenite /l/ more after a long nucleus, and another produced a non-significantly larger TT aperture.



(a) Effects of a following alveolar consonant on the coronal target of coda /l/ in /Vl#t/ and /Vlt/ syllables. Coronal target in *puhl peep*, *puhl teep*, and *pult* produced by W2. Green: *puhl peep*, repetition 2. Orange: *puhl teep*, repetition 6. Green: *pult*, repetition 6. Star: TT sensor. Square: TB sensor. Circle: TD sensor.



(b) Effects of a dorsal consonant on the coronal target of coda /l/ in /Vl#k/ and /Vlk/ syllables. Coronal target in *puhl peep*, *puhl keep* and *pulk* produced by W7. Orange: *puhl peep*, repetition 4. Green: *puhl keep*, repetition 3. Yellow: *pulk*, repetition 1. Star: TT sensor. Square: TB sensor. Circle: TD sensor.

Figure 6.17: Effects of a following consonant on the coronal target of coda /l/ in /Vl#C/ and /VlC/ syllables.

The lack of TT closure observed in the pre-obstruent context is consistent with coronal lenition in coda /l/. Coda /l/-lenition can stem from coronal constriction reduction and from the delay of the coronal gesture, when minimum TT aperture is achieved after the end of the voiced interval belonging to /l/ (Strycharczuk & Scobbie 2019). In Standard British English, gestural delay without gestural reduction was not found: when the minimum TT aperture was achieved after the end of the voiced interval, coronal closure was not achieved (Strycharczuk & Scobbie 2019). Our results showing a lack of TT closure are consistent with both coronal constriction reduction and gestural delay. The lack of TT closure in the analysis window might indicate spatial reduction, but the TT aperture trajectories decrease until the end of the analysis window, which might indicate that minimum TT aperture was achieved after the end of the voiced interval associated with coda /l/. Further analyses of a longer stretch of articulatory trajectories could give insight into the extent to which coronal constriction is reduced in delayed gestures; however, such analyses are beyond the scope of this dissertation.

The lack of observed TT closure is consistent with both coronal constriction reduction and deletion of the coronal gesture. Closer inspection of tongue positions in single tokens measured at minimum TT aperture indicates that both reduction and deletion might be present in the data (**Figures 6.15–6.17**). Some tokens, such as *puhl peep* produced by W3 and W7 (**Figures 6.15** and **6.17b** respectively), show a tongue tip raising gesture that does not achieve coronal closure. In these tokens, coronal closure might only be delayed and might be achieved after the end of the voiced interval. Other tokens, such as *polk heap* produced by W10 and *pulk heap* produced by W7 (**Figures 6.16** and **6.17b** respectively), show no tongue tip raising gesture. In these tokens, coronal contact is less likely to be achieved later on and the coronal gesture might be deleted. In a potentially deleted gesture, both gestural reduction and gestural delay are problematic to discuss: without a gesture, nothing is reduced or delayed. The observed variation between gestural reduction, possible delay, and gestural deletion in AusE is similar to patterns of /l/-lenition in Southern British English: Southern British English speakers who lost tongue tip contact show variation between gradient reduction of tongue tip contact and between gradient reduction of the tongue tip raising gesture (Strycharczuk & Scobbie 2019). Further analyses of tongue positions could answer the questions whether the coronal gesture is deleted or lenited and whether the choice between lenition and deletion is affected by phonetic context. If the coronal gesture is lenited, further analyses of trajectories could answer the question how spatial reduction and temporal delay contribute to lenition across phonetic contexts.

6.5 Conclusion

We found that the magnitude of /l/-lenition varied between speakers; however, speakers showed similar patterns in the same phonetic contexts. We found that the key factor in coda /l/-lenition is coarticulation between the coronal gesture of /l/ and the phonetic context: coda /l/-lenition is more likely when adjacent segments are articulated with tongue backing. Our findings on the facilitating effect of dorsal consonant are consistent both with the auditory-impressionistic analysis of Borowsky (2001) and the articulatory analysis of Hardcastle & Barry (1989) who found the same effect of following post-alveolar consonant. Our findings on the facilitating effect of a preceding back vowel are consistent with Borowsky's (2001) results, who reported more vocalised tokens in the context of back vowels. In contrast, Hardcastle & Barry (1989) reported less lenited tokens in the back vowel context, which they attributed to perceptual reasons, arguing that coronal contact is required to create the percept of /l/ in the back vowel context. Our results indicate that the coarticulatory influence of tongue backing facilitates coronal lenition and do not provide support for the perceptual reasoning of Hardcastle & Barry (1989).

Our findings on the inhibiting effect of alveolar consonant are consistent both with the auditory-impressionistic analysis of Borowsky (2001) and the articulatory analysis of Hardcastle & Barry (1989) who found the same effect of following alveolar. However, our findings on the effect of preceding front vowel are contradictory: some speakers lenite less, as (Borowsky 2001) reported, which we can attribute to the coarticulatory influence of the front vowel. Some speakers lenite more, as Hardcastle & Barry (1989) reported, which we can attribute to perceptual reasons, as the dorsal gesture contrast with the front vowel gesture and can create the percept of /l/ without tongue tip contact.

Lastly, we found that coarticulatory influence of the following obstruent is stronger when the obstruent is tautosyllabic with /l/, which we attribute to the lateral and obstruent gestures being coordinated with the same nucleus as /l/ in clusters. This finding provides new insight to Borowsky's (2001) report who found that /l/-vocalisation is more likely in coda clusters compared to singleton codas without considering the place of articulation of the following consonant.

6.6 Acknowledgements

This research was supported in part by iMQRTP 2015144 and ARC DE150100318 grants. We thank Louise Ratko for her help in data collection, Peter Humburg for his help in the statistical analysis, and the members of the Phonetics Lab, the Centre for Language Sciences and the ARC Centre of Excellence in Cognition and its Disorders at Macquarie University for their comments, feedback and support.

6.7 Appendix

6.7.1 Result of GAMs examining the effect of phonetic context on TT aperture trajectories for each participants

Table 6.5: Parametric and smooth terms for TT aperture produced by W2

A. parametric coefficients	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	8.2	0.3	28.6	< 0.0001
LongNucleus	1.5	0.3	5.3	< 0.0001
ComplexCoda	1.2	0.3	4.0	0.0001
Alveolar	-1.5	0.3	-5.2	< 0.0001
Dorsal	-0.6	0.3	-2.1	0.0356
Front	-3.1	0.3	-11.0	< 0.0001
Back	0.3	0.3	1.1	0.278
LongNucleus:Alveolar	0.5	0.4	1.3	0.1837
ComplexCoda:Alveolar	-0.8	0.4	-2.2	0.0318
LongNucleus:Dorsal	1.5	0.4	4.1	< 0.0001
ComplexCoda:Dorsal	2.1	0.4	5.4	< 0.0001
LongNucleus:Front	1.9	0.4	5.0	< 0.0001
ComplexCoda:Front	0.9	0.4	2.4	0.0156
LongNucleus:Back	-0.8	0.4	-2.0	0.0431
ComplexCoda:Back	-0.1	0.4	-0.3	0.7656
Alveolar:Front	0.9	0.4	2.5	0.0136
Dorsal:Front	0.8	0.4	2.3	0.0223
Alveolar:Back	-1.3	0.4	-3.6	0.0003
Dorsal:Back	-0.6	0.4	-1.7	0.0822
LongNucleus:Alveolar:Front	-1.6	0.5	-3.0	0.0025
ComplexCoda:Alveolar:Front	0.7	0.5	1.4	0.1682
LongNucleus:Dorsal:Front	-3.0	0.5	-5.4	< 0.0001
ComplexCoda:Dorsal:Front	-0.7	0.5	-1.3	0.1928
LongNucleus:Alveolar:Back	0.1	0.5	0.2	0.8583
ComplexCoda:Alveolar:Back	2.7	0.5	5.0	< 0.0001
LongNucleus:Dorsal:Back	2.4	0.5	4.6	< 0.0001
ComplexCoda:Dorsal:Back	-1.2	0.5	-2.4	0.0171
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Time):ShortNucleus.SimpleCoda	7.5	10.2	8.9	< 0.0001
s(Time):LongNucleus	1.0	1.0	12.7	0.0004
s(Time):ComplexCoda	1.2	1.4	3.0	0.0941
s(Time):Labial	3.5	4.9	1.7	0.1265
s(Time):Alveolar	8.0	10.7	13.6	< 0.0001
s(Time):Dorsal	5.2	7.5	6.3	< 0.0001
s(Time):Low	4.3	6.3	2.4	0.0247
s(Time):Front	1.7	2.2	9.8	< 0.0001
s(Time):Back	1.0	1.0	4.0	0.0447

Table 6.6: Parametric and smooth terms for TT aperture produced by W3

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	11.8	0.6	18.5	< 0.0001
LongNucleus	1.5	0.6	2.4	0.0167
ComplexCoda	4.9	0.7	7.5	< 0.0001
Alveolar	-2.6	0.7	-3.9	0.0001
Dorsal	3.0	0.7	4.5	< 0.0001
Front	1.6	0.6	2.8	0.0054
Back	6.3	0.7	9.0	< 0.0001
LongNucleus:Alveolar	4.1	0.8	4.9	< 0.0001
ComplexCoda:Alveolar	-3.1	0.8	-3.8	0.0001
LongNucleus:Dorsal	-2.6	0.8	-3.4	0.0007
ComplexCoda:Dorsal	-2.4	0.8	-3.0	0.0031
LongNucleus:Front	-1.9	0.7	-2.8	0.0057
ComplexCoda:Front	-5.4	0.8	-7.0	< 0.0001
LongNucleus:Back	-5.5	0.8	-6.9	< 0.0001
ComplexCoda:Back	-7.7	0.8	-9.4	< 0.0001
Alveolar:Front	-4.7	0.7	-6.6	< 0.0001
Dorsal:Front	-4.7	0.8	-6.3	< 0.0001
Alveolar:Back	-5.8	0.8	-7.2	< 0.0001
Dorsal:Back	-6.4	0.8	-7.9	< 0.0001
LongNucleus:Alveolar:Front	-3.9	1.0	-3.9	0.0001
ComplexCoda:Alveolar:Front	5.0	1.0	4.8	< 0.0001
LongNucleus:Dorsal:Front	3.5	1.0	3.6	0.0004
ComplexCoda:Dorsal:Front	5.0	1.1	4.7	< 0.0001
LongNucleus:Alveolar:Back	0.4	1.01	0.4	0.7221
ComplexCoda:Alveolar:Back	6.2	1.1	5.9	< 0.0001
LongNucleus:Dorsal:Back	7.6	1.1	7.2	< 0.0001
ComplexCoda:Dorsal:Back	5.9	1.0	5.7	< 0.0001
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Time):ShortNucleus.SimpleCoda	1.0	1.0	8.4	0.0038
s(Time):LongNucleus	0.0	0.0	0.1	0.9980
s(Time):ComplexCoda	1.4	1.7	167.9	< 0.0001
s(Time):Labial	1.0	1.0	10.7	0.0011
s(Time):Alveolar	4.2	5.8	22.9	< 0.0001
s(Time):Dorsal	5.1	7.0	5.0	< 0.0001
s(Time):Low	3.8	5.5	4.2	0.0006
s(Time):Front	9.1	12.1	12.8	< 0.0001
s(Time):Back	4.6	6.3	3.0	0.0059

Table 6.7: Parametric and smooth terms for TT aperture produced by W4

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	8.1	0.3	28.9	< 0.0001
LongNucleus	1.9	0.3	5.7	< 0.0001
ComplexCoda	-0.3	0.3	-0.8	0.4020
Alveolar	-1.7	0.3	-5.1	< 0.0001
Dorsal	-0.8	0.3	-2.2	0.0268
Front	-2.6	0.3	-7.7	< 0.0001
Back	0.0	0.3	0.0	0.9960
LongNucleus:Alveolar	-0.3	0.4	-0.6	0.5704
ComplexCoda:Alveolar	0.6	0.4	1.3	0.2127
LongNucleus:Dorsal	1.4	0.5	3.1	0.0023
ComplexCoda:Dorsal	0.1	0.5	0.3	0.7720
LongNucleus:Front	-0.2	0.5	-0.3	0.7326
ComplexCoda:Front	1.8	0.5	4.0	0.0001
LongNucleus:Back	-1.5	0.4	-3.6	0.0004
ComplexCoda:Back	2.0	0.5	4.5	< 0.0001
Alveolar:Front	0.5	0.5	1.1	0.2644
Dorsal:Front	0.6	0.5	1.2	0.2208
Alveolar:Back	-0.5	0.5	-1.2	0.2424
Dorsal:Back	1.9	0.5	4.3	< 0.0001
LongNucleus:Alveolar:Front	-1.1	0.6	-1.7	0.0882
ComplexCoda:Alveolar:Front	-1.	0.6	-1.5	0.1252
LongNucleus:Dorsal:Front	-1.6	0.7	-2.5	0.0122
ComplexCoda:Dorsal:Front	-0.6	0.7	-0.9	0.3956
LongNucleus:Alveolar:Back	0.1	0.6	0.1	0.9362
ComplexCoda:Alveolar:Back	-1.9	0.6	-3.1	0.0019
LongNucleus:Dorsal:Back	-1.6	0.6	-2.6	0.0101
ComplexCoda:Dorsal:Back	-5.0	0.7	-7.6	< 0.0001
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Time):ShortNucleus.SimpleCoda	6.1	8.3	7.4	< 0.0001
s(Time):LongNucleus	1.0	1.0	15.5	0.0001
s(Time):ComplexCoda	0.5	0.9	0.29	0.6645
s(Time):Labial	1.09	1.0	9.7	0.0019
s(Time):Alveolar	1.0	1.0	1.2	0.2681
s(Time):Dorsal	1.0	1.0	19.9	< 0.0001
s(Time):Low	9.9	13.0	17.0	< 0.0001
s(Time):Front	8.1	11.0	13.1	< 0.0001
s(Time):Back	7.8	10.5	12.7	< 0.0001

Table 6.8: Parametric and smooth terms for TT aperture produced by W5

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
(Intercept)	5.7	0.3	20.4	< 0.0001
LongNucleus	2.3	0.4	6.2	< 0.0001
ComplexCoda	0.6	0.3	1.8	0.0770
Alveolar	-0.4	0.3	-1.3	0.2121
Dorsal	1.1	0.3	3.3	0.0011
Front	-0.2	0.3	-0.7	0.4897
Back	0.6	0.3	2.0	0.0508
LongNucleus:Alveolar	-0.1	0.5	-0.2	0.8226
ComplexCoda:Alveolar	-1.5	0.4	-3.5	0.0004
LongNucleus:Dorsal	-1.0	0.5	-2.1	0.0350
ComplexCoda:Dorsal	-0.4	0.4	-1.1	0.2959
LongNucleus:Front	-0.5	0.4	-1.1	0.2668
ComplexCoda:Front	0.4	0.4	0.9	0.3873
LongNucleus:Back	-1.1	0.4	-2.5	0.0137
ComplexCoda:Back	0.3	0.4	0.7	0.4675
Alveolar:Front	-0.6	0.4	-1.4	0.1582
Dorsal:Front	1.0	0.4	2.4	0.0158
Alveolar:Back	0.2	0.4	0.6	0.5834
Dorsal:Back	-0.4	0.4	-1.0	0.3433
LongNucleus:Alveolar:Front	-0.0	0.6	-0.0	0.9662
ComplexCoda:Alveolar:Front	2.8	0.6	4.7	< 0.0001
LongNucleus:Dorsal:Front	-0.3	0.6	-0.5	0.6155
ComplexCoda:Dorsal:Front	-0.3	0.6	-0.6	0.5809
LongNucleus:Alveolar:Back	-1.4	0.6	-2.3	0.0197
ComplexCoda:Alveolar:Back	-1.0	0.6	-1.7	0.0904
LongNucleus:Dorsal:Back	1.1	0.6	1.9	0.0540
ComplexCoda:Dorsal:Back	1.0	0.6	1.7	0.0820
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Time):ShortNucleus.SimpleCoda	5.8	7.9	3.5	0.0005
s(Time):LongNucleus	7.7	10.4	4.3	< 0.0001
s(Time):ComplexCoda	1.0	1.0	1.0	0.3167
s(Time):Labial	7.6	10.6	9.7	< 0.0001
s(Time):Alveolar	2.3	3.1	6.2	0.0003
s(Time):Dorsal	11.0	14.6	17.9	< 0.0001
s(Time):Low	2.9	3.9	9.9	< 0.0001
s(Time):Front	2.8	4.2	4.0	0.0027
s(Time):Back	1.0	1.0	28.8	< 0.0001

Table 6.9: Parametric and smooth terms for TT aperture produced by W7

A. parametric coefficients	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	7.4	0.3	25.4	< 0.0001
LongNucleus	0.4	0.4	1.1	0.2845
ComplexCoda	2.6	0.3	7.7	< 0.0001
Alveolar	-1.8	0.3	-4.1	< 0.0001
Dorsal	-0.2	0.3	-0.6	0.5749
Front	-0.1	0.4	-0.2	0.8086
Back	0.8	0.4	2.2	0.0313
LongNucleus:Alveolar	0.9	0.5	1.9	0.0598
ComplexCoda:Alveolar	-3.1	0.5	-6.7	< 0.0001
LongNucleus:Dorsal	1.1	0.4	2.5	0.0139
ComplexCoda:Dorsal	0.6	0.4	1.4	0.1613
LongNucleus:Front	-0.9	0.5	-2.0	0.0591
ComplexCoda:Front	-1.8	0.5	-4.0	0.0001
LongNucleus:Back	-0.4	0.5	-1.0	0.3722
ComplexCoda:Back	-1.3	0.5	-2.9	0.0043
Alveolar:Front	-2.6	0.5	-5.6	< 0.0001
Dorsal:Front	-1.1	0.4	-2.7	0.0079
Alveolar:Back	-1.1	0.4	-2.5	0.0118
Dorsal:Back	1.2	0.4	2.8	0.0053
LongNucleus:Alveolar:Front	1.3	0.7	2.0	0.0427
ComplexCoda:Alveolar:Front	3.2	0.7	4.5	< 0.0001
LongNucleus:Dorsal:Front	-1.0	0.6	-1.5	0.1253
ComplexCoda:Dorsal:Front	0.8	0.7	1.2	0.2438
LongNucleus:Alveolar:Back	-1.3	0.7	-2.0	0.0487
ComplexCoda:Alveolar:Back	3.8	0.7	5.7	< 0.0001
LongNucleus:Dorsal:Back	-3.1	0.6547	-4.7	< 0.0001
ComplexCoda:Dorsal:Back	-0.8	0.6087	-1.3	0.1819
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Time):ShortNucleus.SimpleCoda	4.3	6.0	1.5	0.1772
s(Time):LongNucleus	1.0	1.0	0.1	0.7327
s(Time):ComplexCoda	4.3	6.13	5.6	< 0.0001
s(Time):Labial	3.7	5.4	2.0	0.0729
s(Time):Alveolar	8.0	10.7	9.8	< 0.0001
s(Time):Dorsal	1.0	1.0	3.1	0.0775
s(Time):Low	4.8	6.5	2.2	0.0369
s(Time):Front	6.3	8.4	3.5	0.0003
s(Time):Back	2.6	3.5	1.7	0.1585

Table 6.10: Parametric and smooth terms for TT aperture produced by W10

A. parametric coefficients	**Estimate**	**Std. Error**	**t-value**	**p-value**
(Intercept)	7.3	0.4	20.7	< 0.0001
LongNucleus	1.6	0.3	4.9	< 0.0001
ComplexCoda	1.7	0.3	5.5	< 0.0001
Alveolar	1.2	0.3	3.7	0.0002
Dorsal	-1.5	0.3	-4.7	< 0.0001
Front	-7.0	0.3	-22.4	< 0.0001
Back	1.7	0.3	5.5	< 0.0001
LongNucleus:Alveolar	-1.3	0.4	-3.1	0.0019
ComplexCoda:Alveolar	-1.0	0.4	-2.6	0.0083
LongNucleus:Dorsal	1.9	0.4	4.9	< 0.0001
ComplexCoda:Dorsal	1.9	0.4	5.2	< 0.0001
LongNucleus:Front	0.1	0.4	0.2	0.8119
ComplexCoda:Front	-1.2	0.4	-3.2	0.0012
LongNucleus:Back	-2.4	0.4	-6.4	< 0.0001
ComplexCoda:Back	-2.6	0.4	-7.1	< 0.0001
Alveolar:Front	0.9	0.4	2.5	0.0115
Dorsal:Front	3.3	0.4	8.3	< 0.0001
Alveolar:Back	-1.2	0.4	-3.	0.0011
Dorsal:Back	0.8	0.4	2.0	0.0410
LongNucleus:Alveolar:Front	0.1	0.5	0.2	0.8762
ComplexCoda:Alveolar:Front	1.6	0.5	3.2	0.0013
LongNucleus:Dorsal:Front	-3.1	0.6	-5.7	< 0.0001
ComplexCoda:Dorsal:Front	-2.2	0.5	-3.8	0.0001
LongNucleus:Alveolar:Back	2.0	0.6	3.7	0.0002
ComplexCoda:Alveolar:Back	3.4	0.5	6.2	< 0.0001
LongNucleus:Dorsal:Back	-0.8	0.5	-1.6	0.1220
ComplexCoda:Dorsal:Back	2.4	0.9	2.8	0.0047
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Time):ShortNucleus.SimpleCoda	7.7	10.5	8.5	< 0.0001
s(Time):LongNucleus	7.4	10.0	5.2	< 0.0001
s(Time):ComplexCoda	1.0	1.0	5.7	0.0173
s(Time):Labial	7.6	10.2	5.4	< 0.0001
s(Time):Alveolar	4.8	6.5	4.2	0.0002
s(Time):Dorsal	6.5	8.7	7.0	< 0.0001
s(Time):Low	6.8	9.1	519.4	< 0.0001
s(Time):Front	0.0	0.0	0.0	0.9996
s(Time):Back	7.7	10.2	592.1	< 0.0001

6.7.2 Results of linear models examining the effect of phonetic context on TT aperture targets for each participants

Table 6.11: Effect of phonetic context on TT aperture targets produced by W2

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	4.8	0.1	33.8	0.0
Dorsal	2.2	0.2	11.2	0.0
Alveolar	-3.4	0.2	-16.7	0.0
Back	0.2	0.2	0.8	0.4
Front	0.3	0.2	1.3	0.2
Short.Complex	1.4	0.2	7.2	0.0
Long.Simple	0.3	0.2	1.5	0.1
Dorsal:Back	0.7	0.3	2.5	0.0
Alveolar:Back	-0.5	0.3	-1.6	0.1
Dorsal:Front	-0.7	0.3	-2.5	0.0
Alveolar:Front	0.4	0.3	1.5	0.1
Dorsal:Short.Complex	0.3	0.3	1.1	0.3
Alveolar:Short.Complex	-0.7	0.3	-2.4	0.0
Dorsal:Long.Simple	0.3	0.3	1.0	0.3
Alveolar:Long.Simple	-0.4	0.3	-1.5	0.1
Back:Short.Complex	-0.0	0.3	-0.1	0.9
Front:Short.Complex	-0.4	0.3	-1.6	0.1
Back:Long.Simple	-0.0	0.3	-0.1	0.9
Front:Long.Simple	0.6	0.3	2.0	0.0
Dorsal:Back:Short.Complex	0.0	0.4	0.0	1.0
Alveolar:Back:Short.Complex	0.4	0.4	1.1	0.3
Dorsal:Front:Short.Complex	-0.3	0.4	-0.7	0.5
Alveolar:Front:Short.Complex	0.2	0.4	0.6	0.5
Dorsal:Back:Long.Simple	0.4	0.4	1.1	0.3
Alveolar:Back:Long.Simple	-0.2	0.4	-0.5	0.6
Dorsal:Front:Long.Simple	0.1	0.4	0.1	0.9
Alveolar:Front:Long.Simple	-0.2	0.4	-0.4	0.7

Table 6.12: Effect of phonetic context on TT aperture targets produced by W3

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	9.5	0.3	33.6	0.0
Dorsal	3.2	0.4	8.4	0.0
Alveolar	-6.9	0.4	-17.6	0.0
Back	0.1	0.4	0.3	0.8
Front	1.6	0.4	4.2	0.0
Short.Complex	2.2	0.4	5.6	0.0
Long.Simple	-1.2	0.4	-3.1	0.0
Dorsal:Back	-0.3	0.5	-0.5	0.6
Alveolar:Back	-0.1	0.5	-0.2	0.8
Dorsal:Front	1.0	0.5	1.8	0.1
Alveolar:Front	-1.9	0.5	-3.4	0.0
Dorsal:Short.Complex	1.2	0.5	2.2	0.0
Alveolar:Short.Complex	-0.8	0.6	-1.4	0.2
Dorsal:Long.Simple	-0.6	0.5	-1.0	0.3
Alveolar:Long.Simple	0.7	0.6	1.2	0.2
Back:Short.Complex	-0.2	0.5	-0.3	0.8
Front:Short.Complex	-1.2	0.6	-2.1	0.0
Back:Long.Simple	-0.3	0.5	-0.5	0.6
Front:Long.Simple	0.6	0.5	1.1	0.3
Dorsal:Back:Short.Complex	1.5	0.8	1.9	0.1
Alveolar:Back:Short.Complex	0.4	0.8	0.6	0.6
Dorsal:Front:Short.Complex	-0.7	0.7	-1.0	0.3
Alveolar:Front:Short.Complex	1.4	0.8	1.8	0.1
Dorsal:Back:Long.Simple	1.1	0.8	1.5	0.1
Alveolar:Back:Long.Simple	-0.7	0.7	-0.9	0.4
Dorsal:Front:Long.Simple	-0.1	0.7	-0.1	0.9
Alveolar:Front:Long.Simple	-0.8	0.8	-1.1	0.3

Table 6.13: Effect of phonetic context on TT aperture targets produced by W4

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.9	0.1	38.8	0.0
Dorsal	1.0	0.1	8.9	0.0
Alveolar	-1.8	0.1	-17.5	0.0
Back	0.7	0.1	6.8	0.0
Front	-0.5	0.1	-4.8	0.0
Short.Complex	0.0	0.1	0.3	0.8
Long.Simple	0.4	0.1	3.8	0.0
Dorsal:Back	0.5	0.1	3.6	0.0
Alveolar:Back	-0.8	0.1	-5.1	0.0
Dorsal:Front	-0.5	0.2	-3.0	0.0
Alveolar:Front	0.3	0.1	2.3	0.0
Dorsal:Short.Complex	0.0	0.2	0.1	0.9
Alveolar:Short.Complex	0.2	0.1	1.1	0.3
Dorsal:Long.Simple	0.3	0.2	2.1	0.0
Alveolar:Long.Simple	-0.4	0.2	-2.8	0.0
Back:Short.Complex	0.1	0.1	1.0	0.3
Front:Short.Complex	0.1	0.2	0.9	0.4
Back:Long.Simple	-0.0	0.1	-0.1	0.9
Front:Long.Simple	-0.4	0.2	-2.5	0.0
Dorsal:Back:Short.Complex	-0.0	0.2	-0.1	0.9
Alveolar:Back:Short.Complex	-0.1	0.2	-0.5	0.6
Dorsal:Front:Short.Complex	-0.0	0.2	-0.2	0.9
Alveolar:Front:Short.Complex	0.0	0.2	0.2	0.8
Dorsal:Back:Long.Simple	0.0	0.2	0.2	0.8
Alveolar:Back:Long.Simple	0.3	0.2	1.4	0.2
Dorsal:Front:Long.Simple	-0.1	0.2	-0.6	0.5
Alveolar:Front:Long.Simple	-0.1	0.2	-0.5	0.6

Table 6.14: Effect of phonetic context on TT aperture targets produced by W5

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	2.6	0.1	26.0	0.0
Dorsal	1.5	0.1	10.9	0.0
Alveolar	-1.6	0.1	-11.3	0.0
Back	-0.1	0.1	-0.9	0.4
Front	0.8	0.1	5.8	0.0
Short.Complex	0.6	0.1	4.0	0.0
Long.Simple	-0.1	0.1	-0.7	0.5
Dorsal:Back	-0.1	0.2	-0.7	0.5
Alveolar:Back	-0.1	0.2	-0.3	0.8
Dorsal:Front	0.4	0.2	2.2	0.0
Alveolar:Front	-0.2	0.2	-1.2	0.2
Dorsal:Short.Complex	0.5	0.2	2.3	0.0
Alveolar:Short.Complex	-0.3	0.2	-1.8	0.1
Dorsal:Long.Simple	-0.3	0.2	-1.7	0.1
Alveolar:Long.Simple	0.2	0.2	1.0	0.3
Back:Short.Complex	-0.0	0.2	-0.2	0.9
Front:Short.Complex	0.2	0.2	0.8	0.4
Back:Long.Simple	0.1	0.2	0.3	0.8
Front:Long.Simple	-0.1	0.2	-0.3	0.8
Dorsal:Back:Short.Complex	0.4	0.3	1.3	0.2
Alveolar:Back:Short.Complex	-0.1	0.3	-0.5	0.6
Dorsal:Front:Short.Complex	0.2	0.3	0.7	0.5
Alveolar:Front:Short.Complex	0.1	0.3	0.5	0.6
Dorsal:Back:Long.Simple	-0.4	0.3	-1.4	0.2
Alveolar:Back:Long.Simple	0.1	0.3	0.4	0.7
Dorsal:Front:Long.Simple	-0.2	0.3	-0.6	0.6
Alveolar:Front:Long.Simple	-0.1	0.3	-0.4	0.7

Table 6.15: Effect of phonetic context on TT aperture targets produced by W7

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	4.4	0.2	27.8	0.0
Dorsal	2.0	0.2	8.9	0.0
Alveolar	-3.6	0.2	-16.1	0.0
Back	0.3	0.2	1.4	0.2
Front	0.0	0.2	0.0	1.0
Short.Complex	1.2	0.2	5.0	0.0
Long.Simple	-0.5	0.2	-2.2	0.0
Dorsal:Back	0.3	0.3	0.9	0.4
Alveolar:Back	-0.3	0.3	-1.1	0.3
Dorsal:Front	-0.5	0.3	-1.6	0.1
Alveolar:Front	0.1	0.3	0.2	0.8
Dorsal:Short.Complex	0.8	0.3	2.4	0.0
Alveolar:Short.Complex	-0.6	0.3	-1.8	0.1
Dorsal:Long.Simple	-0.4	0.3	-1.2	0.2
Alveolar:Long.Simple	0.5	0.3	1.4	0.2
Back:Short.Complex	0.4	0.3	1.3	0.2
Front:Short.Complex	-0.6	0.3	-1.8	0.1
Back:Long.Simple	-0.1	0.3	-0.2	0.8
Front:Long.Simple	0.4	0.3	1.2	0.2
Dorsal:Back:Short.Complex	0.2	0.4	0.4	0.7
Alveolar:Back:Short.Complex	-0.3	0.5	-0.6	0.6
Dorsal:Front:Short.Complex	0.1	0.5	0.3	0.8
Alveolar:Front:Short.Complex	0.2	0.5	0.5	0.6
Dorsal:Back:Long.Simple	-0.3	0.4	-0.8	0.4
Alveolar:Back:Long.Simple	-0.3	0.4	-0.7	0.5
Dorsal:Front:Long.Simple	-0.0	0.4	-0.0	1.0
Alveolar:Front:Long.Simple	0.2	0.4	0.4	0.7

Table 6.16: Effect of phonetic context on TT aperture targets produced by W10

	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**
(Intercept)	1.3	0.1	13.1	0.0
Dorsal	0.6	0.1	4.5	0.0
Alveolar	-0.0	0.1	-0.2	0.9
Back	0.5	0.1	3.5	0.0
Front	-0.8	0.1	-5.6	0.0
Short.Complex	1.1	0.1	7.9	0.0
Long.Simple	-0.4	0.1	-2.4	0.0
Dorsal:Back	0.5	0.2	2.1	0.0
Alveolar:Back	-0.4	0.2	-2.2	0.0
Dorsal:Front	-0.8	0.2	-4.0	0.0
Alveolar:Front	0.7	0.2	3.7	0.0
Dorsal:Short.Complex	0.6	0.2	2.9	0.0
Alveolar:Short.Complex	0.1	0.2	0.4	0.7
Dorsal:Long.Simple	0.0	0.2	0.1	0.9
Alveolar:Long.Simple	0.0	0.2	0.2	0.8
Back:Short.Complex	0.6	0.2	3.0	0.0
Front:Short.Complex	-1.1	0.2	-5.4	0.0
Back:Long.Simple	-0.7	0.2	-3.5	0.0
Front:Long.Simple	0.7	0.2	3.5	0.0
Dorsal:Back:Short.Complex	1.5	0.3	4.7	0.0
Alveolar:Back:Short.Complex	-0.6	0.3	-2.2	0.0
Dorsal:Front:Short.Complex	-0.9	0.3	-3.2	0.0
Alveolar:Front:Short.Complex	0.3	0.3	1.1	0.3
Dorsal:Back:Long.Simple	-0.8	0.3	-2.8	0.0
Alveolar:Back:Long.Simple	0.4	0.3	1.2	0.2
Dorsal:Front:Long.Simple	0.2	0.3	0.6	0.6
Alveolar:Front:Long.Simple	0.1	0.3	0.5	0.6

6.8 References

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Borowsky, T. (2001). The vocalisation of dark *l* in Australian English. In D. Blair & P. Collins (Eds.), *English in Australia* (p. 69–87). Philadelphia, Amsterdam: John Benjamins Publishing Company.

Borowsky, T. & Horvath, B. (1997). L-vocalisation in Australian English. In F. Hinskens, R. van Hout & W. L. Wetzels (Eds.), *Variation, change and phonological theory.* (p. 101–123). Amsterdam: Benjamins.

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

Browman, C. P. & Goldstein, L. M. (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica*, *45*, 140–55.

Browman, C. P. & Goldstein, L. M. (1995). Gestural syllable position effects in American English. In F. Bell-Berti & J. L. Raphael (Eds.), *Producing speech: contemporary issues.* New York: AIP Press.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of phonetics*, *24*(2), 209–244.

Byrd, D., Lee, S., Riggs, D. & Adams, J. (2005). Interacting effects of syllable and phrase position on consonant articulation. *The Journal of the Acoustical Society of America*, *118*(6), 3860–3873.

Cox, F. & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association*, *37*(3), 341–350.

Draxler, C. & Jansch, K. (2017). *Speechrecorder 3.10.0.*

Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. *Phonetic interpretation: papers in laboratory phonology VI*, 222–236.

Giles, S. & Moll, K. (1975). Cinefluorographic study of selected allophones of English /l/. *Phonetica*, *31*(3-4), 206–227.

Hardcastle, W. & Barry, W. (1989, 7). Articulatory and perceptual factors in /l/ vocalisations in English. *Journal of the International Phonetic Association*, *15*, 3–17.

Horvath, B. M. & Horvath, R. J. (1997). The geolinguistics of a sound change in progress: /l/ vocalization in Australia. *U. Penn Working Papers in Linguistics*, *4*(1), 109–124.

Horvath, B. M. & Horvath, R. J. (2001). A multilocality study of a sound change in progress: The case of /l/ vocalization in New Zealand and Australian English. *Language Variation and Change*, *13*, 37–57.

Horvath, B. M. & Horvath, R. J. (2002). The geolinguistics of /l/ vocalisation in Australia and New Zealand. *Journal of Sociolinguistics*, *6/3*, 319–346.

Keating, P., Cho, T., Fougeron, C. & Hsu, C.-S. (2004). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden & R. A. M. Temple (Eds.), *Papers in laboratory phonology VI: Phonetic interpretation* (p. 145–163). Cambridge: Cambridge University Press.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford, UK; Cambridge, Mass.: Blackwell.

Lee-Kim, S.-I., Davidson, L. & Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology*, *4*(2), 475–511.

Lin, S., Beddor, P. & Coetzee, A. (2014). Gestural reduction, lexical frequency, and sound change: a study of post-vocalic /l/. *Laboratory Phonology*, *5*(1), 9–36.

Narayanan, S. S., Alwan, A. A. & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. part I. The laterals. *The Journal of the Acoustical Society of America*, *101*(2), 1064–1077.

Proctor, M., Walker, R., Smith, C., Szalay, T., Narayanan, S. & Goldstein, L. (2019). Articulatory characterization of English liquid-final rimes. *Journal of Phonetics*, *77*, 100921.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Schiel, F. (1999, August). Automatic Phonetic Transcription of Non-Prompted Speech. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey (Eds.), *Proceedings 14th international congress of phonetic sciences* (p. 607–610). San Francisco, CA, USA.

Scobbie, J. M. & Pouplier, M. (2010). Syllable structure and external sandhi: an EPG study of vocalisation and retraction of word-final English /l/. *Journal of Phonetics*, *38*, 240–259.

Scobbie, J. M., Pouplier, M. & Wrench, A. A. (2007). Conditioning factors in external sandhi: an EPG study of English /l/vocalisation. In *Proceedings of the 16th international congress of phonetic sciences* (p. 441–444). Saarbrücken, Germany.

Scobbie, J. M. & Wrench, A. A. (2003). An articulatory investigation of word final /l/ and //l/-sandhi in three dialects of English. In *Proceedings of the 15th international congress of phonetic sciences.* Barcelona, Spain.

Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*(3), 291–311.

Strycharczuk, P., Derrick, D. & Shaw, J. (2018). How vocalic is vocalised /l/? Evidence from lateralisation. In *Conference of the association for laboratory phonology.* Lisbon, Portugal.

Strycharczuk, P. & Scobbie, J. M. (2019). Gestural delay and gestural reduction. Articulatory variation in /l/-vocalisation in Southern British English. In A. Prezwozny, C. Viollain & S. Navarro (Eds.), *Advances in the study of contemporary spoken English: Multifocal analyses of variation.* Edinburgh: Edinburgh University Press.

Tiede, M. (2005). Mview: software for visualization and analysis of concurrently recorded movement data. *New Haven, CT: Haskins Laboratories.*

Turton, D. (2017). Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalization in varieties of English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1).

van Rij, J., Wieling, M., Baayen, R. H. & van Rijn, H. (2017). *itsadug: Interpreting time series and autocorrelated data using GAMMs.* (R package version 2.3)

Wells, J. C. (1982). *Accents of English.* Cambridge: Cambridge University Press.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116.

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J. & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience*, *9*(131), 1–12.

Ying, J., Carignan, C., Shaw, J. A., Proctor, M., Derrick, D. & Best, C. (2017). Temporal dynamics of lateral channel formation in /l/: 3D EMA data from Australian English. In *Proceedings of the 18th annual conference of the international speech communication association.* Stockholm, Sweden.

7 The percept of vocalised laterals: articulatory correlates and conditioning context

This chapter is based on the following paper, which is being prepared for submission:

Szalay, T., Benders, T., Cox, F., & Proctor, M. (in preparation). The percept of vocalised laterals: articulatory correlates and conditioning context.

I certify that I was responsible for the development of the concept of this paper, in discussion with my supervisors/co-authors. I took leadership in conducting the research, and was responsible for the construction of the stimuli, all data collection, the majority of the articulatory and all of the statistical analyses, and the writing of all parts of the paper. Data was collected as a part of a larger project: methodological innovations are reported in **Chapter 5**, detailed articulatory analyses are reported in **Chapter 6** and perceptual-articulatory analyses are reported in the current chapter. My co-authors provided advice to improve the experimental design and methods, the analyses, the interpretation of the data, as well as the presentation of the written component.

Abstract

The lateral approximant /l/ has been observed to undergo vocalisation in coda position in Australian English (Horvath & Horvath 1997, Borowsky 2001). /l/-vocalisation can be identified both articulatorily as the lenition of the coronal gesture and the loss of the coronal contact (Giles & Moll 1975, Strycharczuk & Scobbie 2019) and auditorily, as the realisation of /l/ resembling a back voicoid (Ash 1982, Hall-Lew & Fix 2012). However, it is still unclear what the relationship is between the percept of vocalised laterals and articulatory characteristics of /l/. The aim of this study was to discover how coronal lenition of coda /l/ affect listeners' percept of /l/ as vocalised

or non-vocalised across phonetic contexts. First we tested whether phonetically trained listeners can reliably identify /l/-vocalisation. We found that listeners were inconsistent in their classification of coda /l/ as vocalised or non-vocalised. Secondly, we tested whether coronal lenition correlates with an increased percept of /l/-vocalisation across phonetic contexts. We found that coronal constriction reduction in coda /l/ does not strongly correlate with an increased percept of /l/-vocalisation, but vocalised percepts were observed more often in contexts that facilitate coronal constriction reduction, such as a following dorsal consonant or in coda clusters. These results indicate that the percept of /l/-vocalisation is a multi-faceted phenomenon involving the complex interaction of more factors than have previously been considered.

7.1 Introduction

The lateral approximant /l/ is prone to vocalisation in coda position in several dialects of English, including Australian English (AusE) (Wells 1982, Horvath & Horvath 1997, Borowsky 2001). The lateral approximant is a multi-gestural segment articulated with a coronal closure, tongue dorsum lowering and retraction and tongue lateralisation (Giles & Moll 1975, Ladefoged & Maddieson 1996). In vocalised /l/, the reduction of the tongue tip gesture leads to a percept of back vocoid (e.g. *milk* [mɪʊk] or [mɪɤk], *Saul* [ʃoː]) due to the articulatory similarity between the tongue dorsum gesture of the lateral and back vowels (Giles & Moll 1975, Hardcastle & Barry 1989, Gick 1999, Gick et al. 2002). A vocalised /l/ can be identified by articulatory methods, as in observing the lenition of the coronal gesture and the loss of the coronal contact (Hardcastle & Barry 1989, Strycharczuk & Scobbie 2019). Some studies have examined /l/-vocalisation by characterising the tongue tip gesture using articulatory methods such as electropalatography (e.g. Giles & Moll 1975, Hardcastle & Barry 1989, ?, Scobbie & Pouplier 2010). A vocalised /l/ can also be identified by auditory-impressionistic methods, as in observing whether /l/ is perceived as a back voicoid (Ash 1982, Borowsky 2001, Hall-Lew & Fix 2012). Listeners have been consistent in their perception of canonically vocalised /l/ (Hall-Lew & Fix 2012); however, it is not known what factors influence listeners' percept of /l/ as vocalised or non-vocalised.

In AusE, coda /l/ vocalisation and its conditioning factors have been identified using auditory-impressionistic ratings of /l/ as vocalised or consonantal (Borowsky & Horvath 1997, Horvath & Horvath 1997, Borowsky 2001, Horvath & Horvath 2001; 2002). Our aim was to discover factors conditioning listeners' percept of /l/ as vocalised or non-vocalised, in particular to link listeners' auditory-impressionistic ratings of /l/-vocalisation to coronal lenition in some key phonetic and prosodic contexts. In addition, we provide a detailed characterisation of the differences in the articulation of /l/ associated with the differences in its perception from consonantal to vocalised.

7.1.1 Variation between clear, dark and vocalised /l/

English /l/ is characterised by great variation between clear, dark, and vocalised /l/ (Giles & Moll 1975, Turton 2017). Dark [ɫ] differs from clear [l] in three respects: its less raised and less fronted tongue tip gesture; its more lowered and retracted tongue dorsum gesture; and the delayed tongue tip gesture, as the tongue tip gesture follows the tongue dorsum gesture in dark [ɫ], but precedes it in clear [l] (Giles & Moll 1975, Sproat & Fujimura 1993, Browman & Goldstein 1995, Gick 2003, Proctor et al. 2019, Ying et al.

2017). As a result of the different coordination between the tongue dorsum and the tongue tip gestures in dark [ɫ] compared to clear [l], the tongue dorsum gesture is temporally closer to the vowel, while the tongue tip gesture is farther both in clear and dark [l] (Sproat & Fujimura 1993). Lateralisation differences may be regionally variable, or may be variable on a speaker by speaker basis: in New Zealand English, clearer /l/ sounds were found to be more lateralised than darker /l/ sounds (Strycharczuk et al. 2018). In AusE, the timing and degree of lateralisation were not significantly different between clear [l] and dark [ɫ] (Ying et al. 2017).

Pre-pausal or pre-consonantal coda /l/, and to a lesser extent word-final prevocalic coda /l/ can undergo /l/-vocalisation (Giles & Moll 1975, Scobbie & Pouplier 2010). /l/-vocalisation arises from the lenition of the tongue-tip gesture (i.e. less constricted closure), delay of the tongue tip gesture, and lenition of tongue lateralisation (i.e. a flatter tongue shape) (Giles & Moll 1975, Hardcastle & Barry 1989, Browman & Goldstein 1995, Wrench & Scobbie 2003, Scobbie et al. 2007, Scobbie & Pouplier 2010, Strycharczuk et al. 2018, Strycharczuk & Scobbie 2019). The lenition of the tongue tip gesture and lateralisation are correlated but not directly causally linked (Strycharczuk et al. 2018). The lenition of the tongue tip and tongue-lateralisation gestures can be understood in terms of gestural lenition in coda position. Importantly, during /l/-vocalisation, the tongue tip and the tongue lateralisation gestures are lenited and the tongue dorsum gesture creates the percept of a back vocoid due to the articulatory similarities between the tongue dorsum gesture of /l/ and back vowels (Strycharczuk et al. 2018, Gick et al. 2002).

/l/-vocalisation is subject to prosodic, segmental, lexical, and stylistic factors, leading to a great amount of intraspeaker variation (Hardcastle &

Barry 1989, Scobbie & Pouplier 2010, Lin et al. 2014 and **Chapter 6**). In connected speech, /l/-vocalisation is the most likely in preconsonantal context, and more likely in pre-pausal than in pre-vocalic context (Scobbie & Wrench 2003, Scobbie et al. 2007 and **Chapter 6**).

/l/-vocalisation is inhibited by articulatory similarity between the gestures of /l/ and adjacent segments (Hardcastle & Barry 1989). Vocalisation of preconsonantal /l/ in coda clusters is inhibited by a following alveolar consonant that shares the coronal gesture of /l/ and by a preceding back vowel that shares the dorsal gesture of /l/ (Hardcastle & Barry 1989). The effect of alveolar consonant is attributed to alveolars being homoorganic with the tongue tip gesture of /l (Hardcastle & Barry 1989). In contrast, the effect of a back vowel is attributed to perceptual, rather than articulatory reasons: the tongue dorsum gesture of /l/ is similar to that of the preceding back vowel which makes achieving tongue tip contact necessary for creating different percepts for the back vowel and the /l/ (Hardcastle & Barry 1989). In contrast, lowering the tongue dorsum after a front vowel is enough to create different vowel and /l/ percepts and the tongue tip contact contributes less to the percept of /l/ (Hardcastle & Barry 1989).

/l/-vocalisation is also subject to intraspeaker variability (Hardcastle & Barry 1989, Strycharczuk & Scobbie 2019) The exact realisation of a vocalised /l/, such as the extent of coronal constriction reduction, temporal delay of the coronal gesture, or the deletion of the coronal gesture varies between speakers (Strycharczuk & Scobbie 2019). The overall frequency of /l/-vocalisation and the frequency of /l/-vocalisation in different phonetic context also vary between speakers (Hardcastle & Barry 1989, Horvath & Horvath 1997).

7.1.2 /l/-vocalisation in Australian English

AusE is known for having /l/-vocalisation, but most of the data on its conditioning factors come from impressionistic analyses of audio recordings collected in several locations across Australia (Wells 1982, Borowsky & Horvath 1997, Borowsky 2001, Horvath & Horvath 1997; 2001; 2002). Similarly to other varieties of English, the most important phonological factor in /l/-vocalisation is the syllabic affiliation of /l/: dark, syllabic or coda [ɫ] can be vocalised, but clear onset [l] cannot (cf. *metal, steel* and *light*) (Borowsky 2001). Word-final /l/ is thus the least likely to be vocalised in the l#V environment, where it can be resyllabified. Word-final /l/ is less likely to be vocalised before a pause than before a consonant (Borowsky 2001, Horvath & Horvath 2001). It is unclear whether /l/ in a coda cluster is more likely to be vocalised than word-final /l/, as Borowsky (2001) found more vocalisation in clusters while Horvath & Horvath (2001) found more vocalisation in word-final position. However, these studies analysed both connected speech and individual words, and it is unclear whether their "word-final" categories distinguish between prevocalic, preconsonantal, and pre-pausal environments (e.g. *feel angry* vs. *feel good*).

Another factor conditioning /l/-vocalisation in AusE is articulatory similarity between /l/ and segments adjacent to /l/. In AusE, as in other dialects of English, articulatory similarity between the following consonant and the tongue tip gesture of /l/ inhibits /l/-vocalisation in coda clusters: /l/-vocalisation is least likely before an alveolar, and more likely before a velar than a bilabial consonant (Borowsky 2001, Hardcastle & Barry 1989). Borowsky's (2001) impressionistic analyses of the effect of consonantal context on /l/-vocalisation in AusE provide the same results as articulatory analyses on British English (Hardcastle & Barry 1989). However, in AusE, articulatory similarity between the tongue dorsum gesture of /l/ and the preceding vowel facilitates /l/ vocalisation (Borowsky 2001), while Hardcastle & Barry (1989) found that a preceding back vowel inhibits /l/-vocalisation. These contradictory results can be explained by methodological differences or by differences in speakers' dialect. Hardcastle & Barry (1989) measured tongue tip contact using EPG, while (Borowsky 2001) relied on acoustic-impressionistic encoding of tokens of /l/. If front vowels facilitate /l/-vocalisation because the tongue dorsum gesture without the tongue-tip gesture is sufficient to create the percept of /l/ after a front vowel (Hardcastle & Barry 1989), then the percentage of vocalised /l/ following front vowels may be underestimated in impressionistic analyses. The difference can also be attributed to speakers' dialects: perhaps, speakers of AusE (as tested by Borowsky (2001)) are more likely to vocalise in the context of back vowels, whereas speakers of British English (which constituted four out of five speakers in Hardcastle & Barry's (1989) study) are more likely to vocalise in the context of front vowels.

The third factor to be considered in /l/-vocalisation in AusE is the length of the preceding vowel: a preceding long monophthong or a diphthong facilitates /l/-vocalisation compared to a preceding short monophthong. This might be explained by the differing syllable structure of /l/ final rimes differs with long and short vowels: /l/ tends to be syllabic after long vowels in AusE (Borowsky 2001).

7.1.3 Aims

The studies of Borowsky & Horvath (1997), Borowsky (2001) and Horvath & Horvath (1997; 2001; 2002) are invaluable in identifying numerous key factors in /l/ vocalisation. However, these studies are limited in scope, as the data were coded for /l/ vocalisation, by a single person, auditorily and impressionistically. Although listeners appear to be consistent in their acoustic-impressionistic coding of canonically vocalised /l/ (Hall-Lew & Fix 2012), acoustic-impressionistic coding does not provide any information on the coronal gesture. It is not known how listeners' perception relates to coronal lenition of coda /l/ and what factors condition listeners' percepts.

The aim of this study was to discover how coronal lenition of coda /l/ affect listeners' percept of /l/ as vocalised or non-vocalised across phonetic contexts. To do so, we used a corpus consisting of synchronised acoustic and articulatory data acquired simultaneously. The corpus was designed to elicit coda /l/ produced with different degrees of coronal reduction by manipulating the phonetic context in which coda /l/ occurred (**Chapter 6**). In the corpus, target phrases manipulated the place of articulation of the adjacent vowel (preceding front, back, or low vowel), adjacent consonant (following glottal, labial, alveolar, or dorsal consonant) and syllable type (syllables containing short nucleus and simple coda, long nucleus and simple coda, or short nucleus and complex coda). In the current study, we established whether expert listeners could provide reliable ratings of tokens of /l/ as vocalised or non-vocalised. Then, we explored relations between listeners' ratings and the articulatory characteristics of coda /l/ on the one hand and different phonetic environments on the other hand. Lastly, we compared the articulatory characteristics of a subset of tokens that were identified as vocalised and non-vocalised.

7.2 Methods

7.2.1 Collection of the articulatory corpus

Detailed articulatory analyses of the data, including the effects of phonetic and prosodic contexts and individual variation have been reported in **Chapter 6**.

Participants

Data from six female native speakers of AusE (mean age = 23.4, range = 20–27) were analysed. Participants were born and raised in New South Wales (NSW). All but one participant had two NSW-born parents; W2 had one NSW-born and one Victoria-born parent. Participants received course credit and/or $40/hour for participation. None of the participants reported any current or past reading, hearing, or speaking disorders.

Experimental materials

The stimuli consisted of 33 three-word phrases containing coda /l/ in the second word, across three vowel contexts, four consonantal contexts, and three syllable types (**Table 7.1**). Coda /l/ was preceded by a high, low, or back vowel and followed by a glottal, labial, alveolar, or dorsal voiceless consonant at the onset of the next word or in a coda cluster. Pre-pausal /l/ was not elicited. To create 3 different syllable types, we manipulated vowel length and coda complexity: target words in 12 phrases contained a short vowel and a simple coda /l/; target words in 12 phrases contained a long vowel and a simple coda /l/; and target words in 9 phrases contained a short vowel and a complex coda. Phonotactically illicit words containing a long vowel and a complex coda or a coda cluster of /l/ followed by a glottal consonant were not tested. For words containing a coda /l/ in a simple coda, the following consonant was identified as the first consonant of the following word; for clustered /l/, it was the following segment of the cluster. To provide a consistent phonetic frame of reference, non-words were used when necessary (**Table 7.1**).

Target words were placed in a carrier phrase with antagonistic vowel contexts; that is, words with a front vowel were placed in the carrier phrase "far ____ harp" and words with low or back vowels were placed in the carrier phrase "fee ____ heap". To minimise lingual coarticlation, all non-target consonants in the stimuli were labial /f, p/ and glottal /h/. The task also included /dVp, pVd/, and /lVp/ words which are analysed in **Chapter 6**.

Table 7.1: Target words containing coda /l/ followed by the last word of the carrier phrase. Targets containing single coda /l/ are represented by words with long and short vowels. Targets containing coda clusters are represented by words with short vowels only, as /VːlC/ clusters, with the exception of /iːld/, are phonotactically illicit. The /h/ context is only represented by targets with simple coda /l/, as an /lh/ coda cluster is phonotactically illicit.

Vowel Context	Word-final				Cluster		
	Glottal	Bilabial	Alveolar	Dorsal	Bilabial	Alveolar	Dorsal
Front	*peel hVp*, *pill hVp*	*peel pVp*, *pill pVp*	*peel tVp*, *pill tVp*	*peel kVp*, *pill kVp*	*pilp hVp*	*pilt hVp*	*pilk hVp*
Back	*Paul hVp*, *pol hVp*	*Paul pVp*, *pol pVp*	*Paul tVp*, *Pol tVp*	*Paul kVp*, *Pol kVp*	*polp hVp*	*polt hVp*	*polk hVp*
Low	*parl hVp*, *puhl hVp*	*parl pVp*, *puhl pVp*	*parl tVp*, *puhl tVp*	*parl kVp* *puhl kVp*	*pulp hVp*	*pult hVp*	*pulk hVp*

Procedure

Participants were instructed to read the phrases aloud while seated approximately 150 cm from a computer screen. They were introduced to the task and the experimental materials with a short practice block. Each trial began with a blank screen for 500 ms, followed by a stimulus presented orthographically for 2000 ms. After 2000 ms, the experiment automatically moved on to the next trial. Phrases were divided into two blocks. One block contained the 6 /pVl#h/ targets (together with the /lVp, dVp, pVd/ targets). The other contained the remaining 27 phrases (pV(ː)l#p, pV(ː)l#t, pV(ː)l#k, pVlp, pVlt, pVlk). Targets were randomised within blocks and the order of the blocks was counterbalanced between participants. Blocks were repeated 8 times, eliciting a total of 264 phrases per participant.

Articulatory data were acquired using electromagnetic articulography (EMA). EMA records the movement of articulators over time in an electromagnetic field by tracking sensors attached to the participant. Eleven sensors were used. Five sensors were attached to the tongue to track lingual articulation: the tongue tip (TT), tongue body (TB), tongue dorsum (TD), and the left lateral and right lateral sensors; data from the parasagittal sensors were not analysed because they were not reliable. Two sensors were attached to the lips (upper- and lower lips) to track lip aperture and lip rounding. One sensor was attached to the gumline below the lower incisor to measure jaw movement. There were three reference sensors to correct for head movement (nasion, left mastoid, right mastoid). The occlusal plane was located with a bite trial and the palate was traced with a palate probe. The intersection of the occlusal plane and the incisors was defined as the origin for all sensor measurements: vertical displacement is expressed relative to the occlusal plane and horizontal displacement relative to the upper incisors.

Audio was acquired using two microphones located 150 cm from the lips and offset by 15°. A Røde NTG-1 was connected through a Focusrite OctoPre MkII preamplifier to the NDI Wave system, recording synchronised acoustic data simultaneously with the spatial data from the sensor coils. A

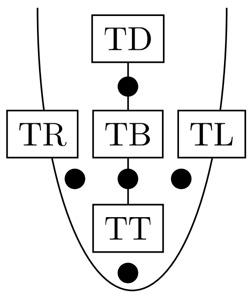


Figure 7.1: Tongue sensor placements viewed from top.

second microphone (Røde NT1-A) was connected through a separate Focusrite OctoPre MkII preamplifier to the computer presenting the experimental stimuli, capturing the utterance as a series of WAV files sampled at 44100 Hz using SpeechRecorder (Draxler & Jansch 2017).

7.2.2 Articulatory characterisation of the experimental corpus

Articulatory analysis

Tokens were excluded if they were misread (68 tokens), if the audio file was corrupted (50 tokens), or if the sensors were tracked incorrectly (110 tokens). Tokens were also excluded when a sensor fell off and had to be reattached, leading two incomparable sensor configuration (177 tokens). In total, 1179 tokens were included in the analyses.

Customarily, articulatory targets are identified as the point in time when the relevant articulator achieves maximum displacement and minimum velocity. However, in this dataset, the coronal target of /l/ could not be identified reliably based on velocity profiles either due to TT constriction reduction or gestural overlap between adjacent segments. Instead, an analysis window was identified between the acoustic onset of the vowel (T0 in **Figure 7.2**) and offset of the /l/ (T1 in **Figure 7.2**) using MAUS (Schiel

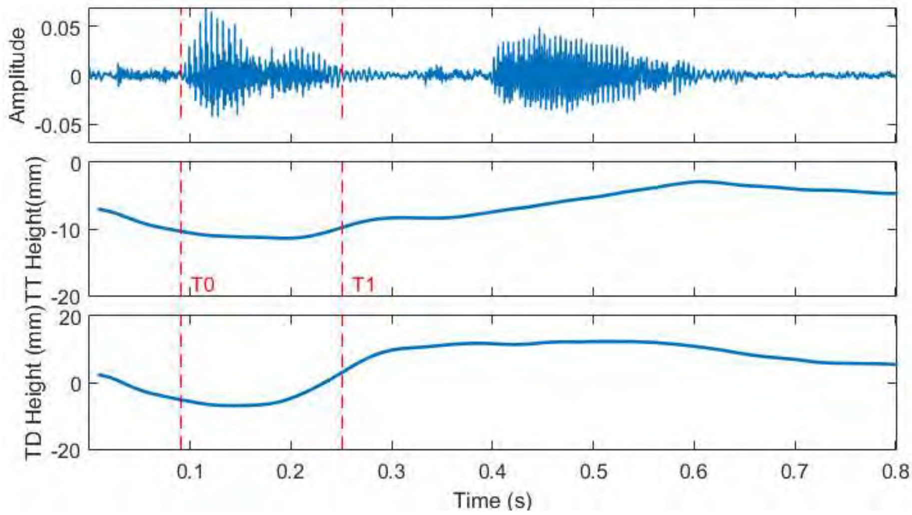


Figure 7.2: Identifying analysis window in coda laterals without clear articulatory targets: *pol keep* (W2, 2nd repetition). Top panel: acoustic speech waveform. Middle panel: vertical tongue tip trajectory. Bottom panel: vertical tongue dorsum trajectory. T0: start of the analysis window identified at the acoustic vowel onset. T1: end of the analysis window identified at the acoustic offset of /l/.

1999). The TT aperture trajectory was calculated as the tangential distance between the TT sensor and the nearest point on the palate at every 10 ms in the analysis window. The coda /l/ target was defined at the minimum TT aperture within the analysis window, at which point TT aperture and lower lip fronting were extracted.

Patterns of TT constriction reduction in the corpus

Articulatory analyses of the corpus (**Chapter 6**) has revealed that the degree of coronal constriction varied between phonetic contexts and participants. In the pre-glottal context, participants reduced the coronal constriction in coda /l/ compared to onset /l/, but not more than they reduced the coronal constriction in coda /d/. Participants achieved a consistent TT closure in the pre-alveolar context. In the pre-dorsal and pre-labial contexts, only W10 achieved a consistent closure, other participants did not. The effect of consonant context was larger when the consonant was tautosyllabic with /l/, i.e. TT closure was more likely to be achieved and was achieved earlier in /pɪlt# hɐːp/ than in /pɪl# tɐːp/ and TT aperture was less likely to be achieved in /pɪlp/ and /pɪlk/ than in /pɪl# pɐːp/ and /pɪl# kɐːp/. The effect of the preceding vowel context varied between participants.

7.2.3 Auditory-impressionistic ratings of the corpus

Participants

Four phonetically trained expert listeners rated the audio-recorded tokens of /l/ produced in the EMA experiment for /l/ vocalisation on a scale ranging from 0 (not vocalised) to 3 (1-3: increasing vocalisation with 3 representing "maximally vocalised"). Listeners were native speakers of AusE, members of the Department of Linguistics at Macquarie University with varying levels of experience in phonetic research. Listener 1 was a postgraduate student of phonetics. Listeners 2 and 4 hold a Ph.D. in phonetics and have taught AusE phonetics and phonology at a university level. Listener 3 has also taught AusE phonetics and phonology at a university level and holds an IPA Certificate of Proficiency in the Phonetics of English. None of the listeners reported any hearing, reading, or speaking disorders. Listeners received $200 for their time. Only Listener 4 was familiar with the experiment design.

Material

Listeners were presented with the last two words of the three-word phrases (e.g. *parl peap*) produced by participants W2, W3, W4, W5, W7, and W10, using the audio captured via SpeechRecorder. The audio was amplitude normalised and truncated after the first word of the three-word phrase and

0.3 s of silence was added to the beginning. Each listener rated 1184 unique tokens; in addition, 10% of the tokens were repeated to measure intrarater reliability. Tokens covered the 33 targets produced 2-8 times by the 6 participants of the production experiment in 3 blocks. The number of tokens varied for each targets, because the number of excluded tokens varied between targets. For instance, the token *polk heap* was more prone to being misread than the token *pill harp*, therefore more *polk heap* tokens were excluded than *pill harp*.

Procedure

Prior to the task, listeners were informed that the audio they were about to hear was recorded during an EMA experiment, and contained some amount of background noise produced by the EMA machine and potentially "unusual" articulations caused by the speakers having sensors on their tongues. Listeners were introduced to the task with a short practice session, listening to audio recordings of ten words and rating them. The audio for the practice session was taken from the audio recordings of W8, an excluded EMA participant, to match rest of the stimuli in audio quality. Data produced by W8 was excluded due to technical difficulties in the articulatory data collection which did not affect the quality of audio recordings.

Listeners were seated in front of a computer monitor located at eye height at a distance of 50 cm and wore Sennheiser 380 Pro headphones with the volume adjusted to listeners' comfortable listening level. Participants were instructed to respond as accurately as possible. To begin each trial, a fixation cross was displayed in the centre of the screen. After 500 ms, the response options appeared and simultaneously the target phrase started playing. Listeners entered their rating using a button box. Participants only heard each phrase once and were not allowed to change their answers. Audio was presented and ratings and reaction time of ratings were collected with Expyriment (Krause & Lindemann 2014).

Tokens were divided into three blocks; each block contained audio from one vowel context. The order of the blocks was randomised between listeners and items were randomised within blocks. As each block took approximately an hour to complete, the blocks were conducted on 3 separate days within a maximum of four days.

Prior to the task, listeners were asked to provide their definition of vocalised and non-vocalised /l/. After each block, listeners were asked to fill out a questionnaire to describe the cues they listened for and their rating criteria. Listeners were also asked about the difficulties they faced in the task. After the last block listeners were asked to fill out a questionnaire on whether their definition of vocalised and non-vocalised /l/ has changed during the experiment and whether their listening strategies varied across phonetic contexts or speakers.

7.3 Results

7.3.1 Listeners' strategies: qualitative results

Listeners were asked how they define /l/-vocalisation. After each block, they were asked what cues they listened for when rating a token as non-vocalised (0) vs. vocalised (1-3) and how they distinguished between levels of vocalisation (1-3). They were also asked what difficulties they had. After the experiment, they were asked whether they thought their understanding of /l/-vocalisation and their criteria changed throughout the experiment, or between different speakers and phonetic contexts.

Listener 1 defined vocalised /l/ in articulatory terms as an /l/ that lacks TT contact and is potentially articulated with lip rounding. Listeners 2 and 3 provided auditory-perceptual definitions, stating that vocalised /l/ creates the "percept of a vowel" (Listeners 2 and 3), "as in *foot*" (Listener 3). Listener 4 provided both an articulatory and an auditory-perceptual definition, naming both the lack of TT contact and the vowel-like percept in their definition (**Table 7.2**).

Listeners 1 and 4, who provided articulatory definitions, reported that they listened for the presence of TT contact and lip rounding to differentiate vocalised and non-vocalised /l/, although Listener 4 mentioned that they were aware that TT contact can be perceived "only to an extent". Listener 2, who provided a perceptual definition, listened for longer duration, and a syllabic [u]-like percept to distinguish vocalised /l/ from non-vocalised /l/. Although Listener 2 did not report explicitly that they listened for lip rounding, the fact that they rated a token as vocalised when it sounded as round [u] might indicate that they were aware of lip rounding. In contrast to the other three listeners, Listener 3, who holds an IPA Certificate of Proficiency in the Phonetics of English, reported that they selected "not vocalised" for tokens which they "would transcribe as dark /l/" and "vocalised" for tokens that could be "transcribed as a vowel". As Listener 3 indicated that they had listened for different cues compared to the other three, we analysed listeners' rating both with and without Listener 3 (**Table 7.2**).

When rating a vocalised token from 1 to 3, Listeners 1 and 2 indicated that their choice was motivated by their increasing confidence in the token being vocalised. For instance, both Listeners 1 and 2 said they gave a rating of 1 when they thought the token might have been vocalised and they gave a rating of 3 when they were almost certain that the token was vocalised or that there was no TT contact. Similarly, Listener 3 rated a token as 1 when they would have accepted both an [l] and a vowel in a transcription, and rated a token as 2 when they would have only accepted a vowel. In contrast, Listener 4 had different phonetic criteria for distinguishing between 1, 2, and 3: they rated a token as 1 when they perceived it as having a "lenited tongue tip contact", as 2, when they perceived the token as having no TT contact, and as 3, when they perceived no TT contact coupled with lip rounding (**Table 7.2**).

Table 7.2: Listeners' definition criteria for rating tokens

(a) Listeners' definition of vocalisation

Rating	Definition	Listener 1	Listener 2	Listener 3	Listener 4
Non-vocalised: 0	TT contact	✓			✓
	percept of /l/		✓	✓	✓
Vocalised: 1-3	no TT contact	✓			✓
	lip rounding	✓		?([u]-like)	✓
	percept of vowel		✓	✓	✓

(b) Listeners' criteria and perceptual cues for decision making

Rating	Listener 1	Listener 2	Listener 3[1]	Listener 4
Non-vocalised: 0	TT contact	shorter duration	transcribed as /l/	TT contact
Vocalised: 1-3	no TT contact, rounding	longer duration, u-like	transcribe as a vowel	no TT contact, rounding
Vocalised: 1	least certain		transcribed as vowel or /l/	lenited TT
Vocalised: 2	more certain		transcribed as vowel	no TT contact
Vocalised: 3	most certain		NA	no TT contact, rounding

When asked about the consistency of their rating criteria, listeners reported using the same cues across speakers and phonetic contexts. However, all listeners, including Listener 3, who rarely rated a token as vocalised, were aware that some speakers in the corpus vocalised more than others. Listeners were also aware that the number of tokens they perceived as vocalised varied across phonetic context. Both Listeners 1 and 4 reported hearing a

[1]Listener 3 holds the IPA Certificate of Proficiency in the Phonetics of English.

lower number of vocalised tokens in the context of a preceding low vowel. Listener 2 reported perceiving a lower number of tokens as vocalised when the preceding vowel was short, and more when the preceding vowel was long and high /iː/. Listener 4 reported hearing a higher number of vocalised tokens following a back vowel. Listener 1 did not comment on the number of tokens they perceived as vocalised, but they reported finding the pre-alveolar and the post-back vowel contexts more difficult then the others. In the alveolar context, they were unsure whether they perceived the TT contact in /l/ or in /t/, and in the post-back vowel context they were unsure whether they perceived lip rounding due to /oː, ɔ/ being rounded or due to a rounded vocalised /l/.

7.3.2 Reliability of auditory-impressionistic ratings

Listeners rated tokens of /l/ as not vocalised (0) or vocalised (1-3). Gradient ratings of vocalised /l/ (1-3) were merged into one group to achieve categorical ratings of non-vocalised versus vocalised /l/ (**Figure 7.3**). Intra- and interrater reliability was measured both on gradient and categorical ratings; gradient ratings were treated as ordinal and categorical ratings as nominal data. Intrarater and interrater agreement between gradient and categorical ratings was tested using Krippendorff's alpha (Krippendorff 2011). Krippendorff's alpha calculates agreement between two or more datasets both on ordinal data, such as the gradient ratings, and nominal data, such as the categorical ratings. In the case of ordinal data, disagreements are weighted differently (e.g. rating the same token as 2 and 3 or as 0 and 1 provide a higher agreement score than rating the same token maximally differently, as 0 and 3) (Krippendorff 2011). All measurements of inter- and intrarater reliability were calculated using the library *irr* in R (R Core Team 2019).

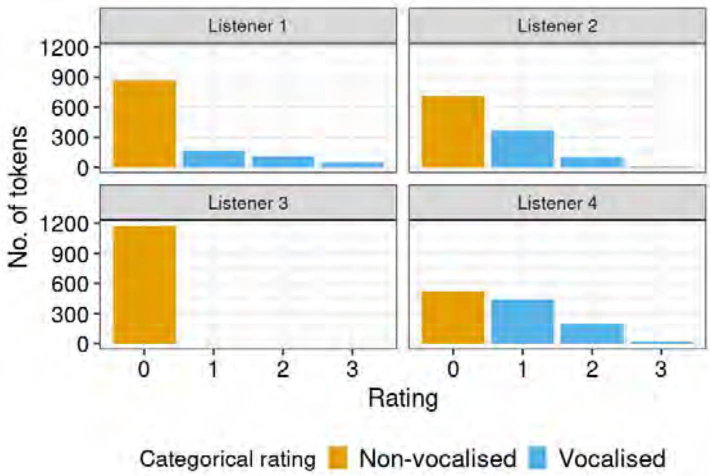


Figure 7.3: Distribution of listeners' responses.

Distribution of ratings showed that most of the tokens were rated as 0 or 1, and only a few tokens were rated as 3 (maximally vocalised) (**Figure 7.3**). Qualitative inspection of Listener 3's responses showed they only rated four tokens as weakly vocalised (1) in the dataset. None of these tokens were rated as vocalised by all four listeners: these tokens have a mean gradient rating of 0.25 0.625, 0.875, and 1 out of the maximal 3, and a mean categorical rating of 0.25, 0.5, 0.5, and 0.75 out of the maximal 1. Listener 3, in agreement with her differing criteria, showed a different rating behaviour compared to the other three.

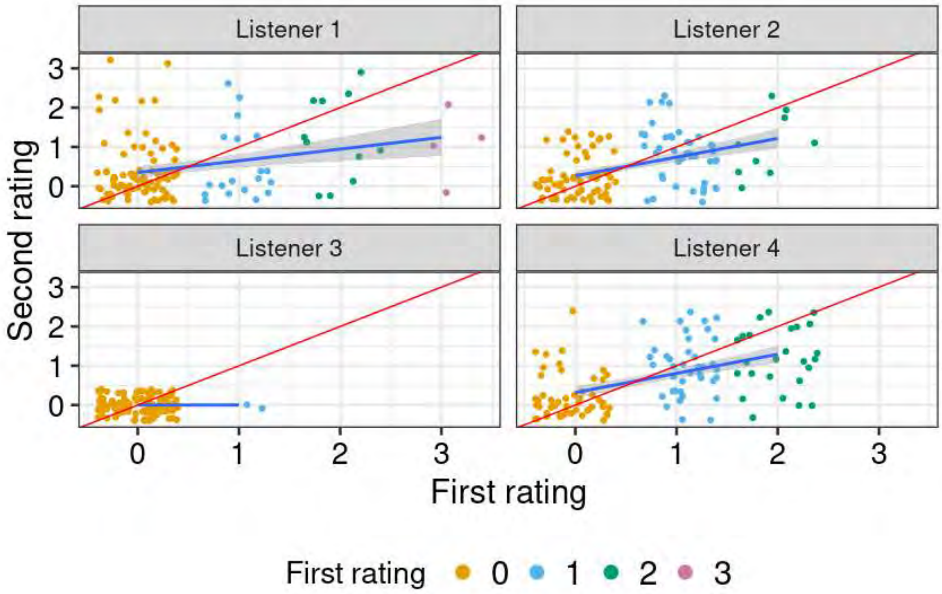
Krippendorff's alpha indicates weak to moderate intrarater reliability both on gradient and categorical ratings for all listeners (**Table 7.3**). Listener 1 showed slight agreement on gradient ratings and moderate agreement on categorical ratings. Listener 2 and 4 showed moderate agreement on both gradient and categorical ratings. For Listeners 1, 2, and 4, reliability is somewhat higher for gradient than categorical ratings, suggesting that listeners rarely gave maximally different ratings to the same token, but did not consistently perceived a token as vocalised (1-3) or non-vocalised

(0). Listener 3 showed no agreement either on gradient or on categorical ratings due to the lack of variance in her responses (**Figure 7.3**). Listener 3 was excluded from further analysis because out of the 1179 tokens they only perceived four as vocalised.

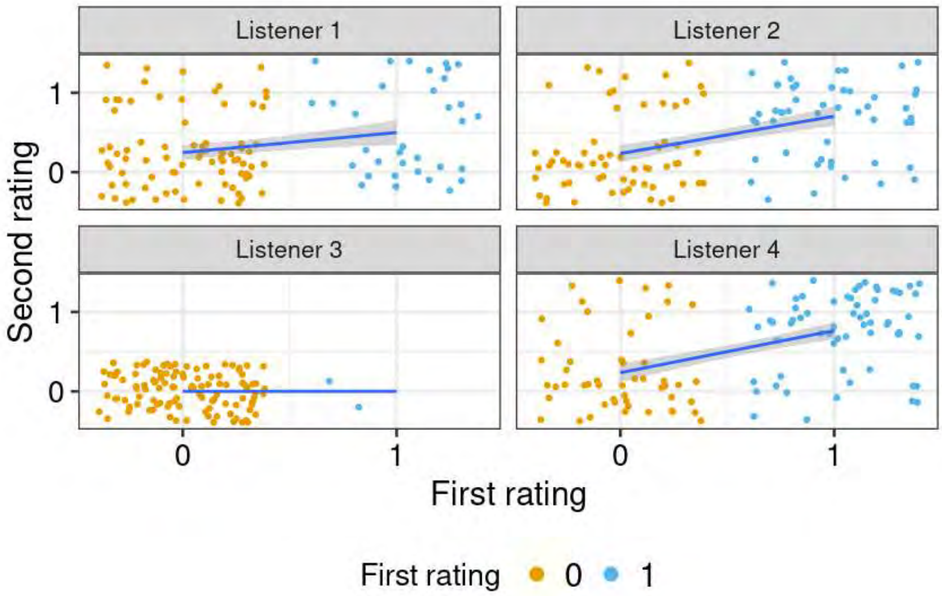
Krippendorff's alpha indicated no agreement between Listeners 1, 2, and 4, on either gradient or on categorical ratings (**Table 7.3**). Due to the lack of interrater reliability, vocalisation ratings are modelled separately for Listeners 1, 2, and 4, although the weak intrarater reliability still limits the interpretation of the following results. Implications of the weak to moderate intrarater reliability and the lack of interrater reliability for previous studies of /l/-vocalisation using auditory-impressonistic ratings are discussed in **Section 7.5**.

Table 7.3: Inter- and intrarater reliability of gradient and categorical ratings using Krippendorff's alpha

	Listeners	Gradient rating	Categorical rating
Intrarater reliability	L1	0.3	0.25
	L2	0.49	0.47
	L3	0	0
	L4	0.54	0.53
Interrater reliability	L1, L2, L4	0.06	0.05



(a) Gradient rating of /l/-vocalisation



(b) Categorical rating of /l/-vocalisation

Figure 7.4: Intrarater reliability for each listener. Top panel: Gradient ratings (0, 1, 2, or 3). Bottom panel: categorical ratings (0 or 1). X axis: first rating. Y axis: second rating. Blue regression line: observed correlation between first and second rating. Red line: 1-to-1 correlation between first and second rating, plotted as a reference. Jitter has been added to avoid overlapping datapoints.

7.3.3 Listeners' rating and its correlates

Vocalisation ratings for Listeners 1, 2, and 4 were modelled separately for each listener using penalised logistic regression (James et al. 2013) as the function of two articulatory measures (TT aperture, lower lip fronting), one acoustic measure (duration), and three contextual factors (preceding vowel's place of articulation, following consonant's place of articulation, and syllable type). Articulatory and acoustic factors were continuous. Contextual factors were binary-coded for modelling gradient responses. Contextual factors were contrast coded for modelling categorical responses: back and front vowels were compared to the baseline low vowels; alveolar, dorsal, and glottal consonants were compared to the baseline labial consonants; syllables with long nuclei and simple codas and short nuclei and complex codas were compared to the baseline syllables with short nuclei and simple coda. Independent variables were selected based on listeners' responses (**Section 7.3.1**): TT aperture was included in the model because listeners listened for TT contact and lenition. Lower lip fronting was included in the model because listeners listened for lip rounding. Duration was included in the model because Listeners 2 and 4 reported being affected by duration. Phonetic contexts were included in the model because listeners reported perceiving different number of tokens as vocalised depending on phonetic context. Listeners were modelled separately due to the lack of interrater reliability (**Section 7.3.2**).

To build a parsimonious model and minimise the effect of the correlation between phonetic characteristics and phonetic context (**Chapter 6**), we modelled listeners' responses using penalised logistic regression. Penalised logistic regression penalises a model for having too many factors and shrinks the coefficient of the less contributive variables to or toward zero (James et al. 2013). As a result, the estimates provided by penalised regression are biased, therefore no standard error or *p*-value are provided (Simon et al. 2011, Archer & Williams 2012). A zero coefficient indicates that a variable increases model complexity without substantially improving accuracy, whereas a non-zero coefficient indicates that a variable improves accuracy and shows the size and the direction of a variable's effect. To evaluate model fit, we compared observed values to values predicted from the model. We analysed gradient ratings using the package *glmnetcr* (Archer & Williams 2012) and categorical ratings using the package *glmnet* (Simon et al. 2011) in R (R Core Team 2019).

Factors affecting listeners' rating

The patterns for the categorical ratings were similar to those for the gradient ratings; however some factors had an effect on either gradient or categorical ratings (**Table 7.4**). For modelling categorical ratings, the labial consonant, low vowel, short nucleus and simple coda contexts served as the baseline , therefore they did not provide estimates. Some factors had different effects on listeners (**Table 7.4**).

The ratings of Listeners 1, 2, and 4 increased as TT aperture increased, indicating more vocalised percept when the TT aperture is larger. Comparison of estimates indicate that TT aperture has a comparatively smaller effect. Listeners 1, 2, and 4 rated /l/ as more vocalised when the nucleus was long; however, for Listeners 1 and 4 this was only evident in the categorical data. The ratings of Listeners 1 and 4 also increased in the dorsal context, but only in categorical ratings, indicating a more vocalised percept in the context of a following dorsal consonant (**Table 7.4**).

The ratings of Listeners 1, 2, and 4 decreased in the glottal context, indicating a less vocalised percept; however, for Listener 4 this is only evident in the categorical data. Gradient ratings of Listeners 1 and 4 decreased in the context of short nucleus simple coda syllable type, indicating a less vocalised percept. Listener 1's gradient ratings somewhat decreased in the context of a following labial consonant (**Table 7.4**). Listener 2's gradient ratings somewhat decreased when the lower lip was more fronted and when the vowel was low, indicating a less vocalised percept. Only Listener 2 was affected by lower lip fronting, although all listeners mentioned rounding as a characteristic o vocalised /l/ either in their definitions or when they were asked about the cues they listened for.

The effect of other factors varied between listeners with Listeners 1 and 4 showing similar patterns with respect to acoustic duration and the complex coda syllable type; and Listeners 2 and 4 showing similar patterns with respect to the front and back vowel contexts. The ratings of Listeners 1 and 4 indicate a less vocalised percept as duration increased; in contrast, Listener 2's ratings indicate a more vocalised percept as duration increases. Listeners 1 and 4 provided higher, more vocalised ratings when the coda was complex; in contrast, Listener 2 provided lower ratings when /l/ was in a complex coda. The ratings of Listeners 4 and 2 increased in the front vowel context, whereas Listener 1's decreased. Categorical ratings of Listeners 2 and 4 increased in the back vowel context, whereas Listener 1's decreased. Listener 1's ratings decreased in the alveolar context, while Listener 4's increased, and Listener 2's ratings were not affected. None of the coefficients were shrunk to zero for all listeners.

Table 7.4: Listeners' coefficients for gradient and categorical ratings. Coefficients shrunk to 0 for all listeners are not reported.

	Coefficient	L1		L2		L4	
		Gr	C	Gr	C	Gr	C
Vocalised percept	TT Aperture	0.10	0.13	0.04	0.03	0.11	0.13
	Dorsal	0	0.49	0	0	0	0.05
	Long Simple	0	1.29	1.50	1.64	0	0.29
Non-vocalised percept	Glottal	−0.7	−0.53	−3.32	−0.42	0	−0.23
	Short Simple	−1.1	–	0	–	−0.27	–
	Lower lip fronting	0	0	−0.01	0	0	0
	Labial	−0.1	–	0	–	0	–
	Low	0	–	−0.56	–	0	–
Varied percept	Duration	−6.11	−8.55	4.45	5.71	−6.28	−8.71
	Short Complex	0.28	1.6	−0.17	−0.22	0.93	1.66
	Alveolar	0.54	1.1	0	0	−0.1	−0.08
	Front	−0.3	−0.34	1.32	2	0	0.17
	Back	0	−0.01	0	0.56	0	0.27

Model fit

For gradient ratings, comparison of predicted and observed gradient ratings indicate that model fit is poor, as the models predict lower ratings compared to observed ratings, overestimating the number of 0 responses and underestimating the number of 3 responses (**Table 7.5**). Overestimating the number of 0 responses is in line with the high number of observed 0 responses, but indicates that listeners' responses are not explained well by predictors in the models. Poor model fit indicates that there may be many, yet unknown factors influencing listeners' percept of vocalisation other than the factors that listeners reported basing their ratings on. These factors are addressed in **Section 7.5**. Despite the poor fit, the models reveal shared patterns between Listeners 1 and 4 and Listeners 2 and 4, indicating that the models are useful in understanding listeners' strategies.

Listeners' categorical responses can be predicted with higher accuracy than their gradient responses. Comparing predicted values to observed values shows that Listener 1's responses can be predicted with 75% accuracy, Listener 2's and 4's with 69% accuracy.

Table 7.5: Comparison of Listeners' responses (columns) and predicted responses (rows) for gradient rating

(a) Listener 1

	0	1	2	3
0	860	162	96	36
1	3	2	3	3
2	2	4	3	5
3	0	0	0	0

(b) Listener 2

	0	1	2	3
0	641	210	38	2
1	68	155	57	6
2	0	0	2	0
3	0	0	0	0

(c) Listener 4

	0	1	2	3
0	436	203	57	6
1	80	210	116	6
2	4	24	32	5
3	0	0	0	0

7.3.4 Articulatory comparison of /l/ perceived as vocalised and non-vocalised

A large-scale quantitative comparison between the articulatory characteristics of tokens perceived as vocalised and non-vocalised was not possible in this dataset. Articulatory characteristics differed between target phrases representing different syllable types, vowel- and consonant contexts (**Chapter 6**), therefore it was not possible to compare articulatory characteristics of tokens perceived as vocalised and non-vocalised when they represented different target phrases. None of the target phrases had enough occurrences with both vocalised and non-vocalised percepts (**Figure 7.5**).

Instead of a quantitative analysis, we identified tokens representing those articulatory and acoustic characteristics and phonetic contexts that affected listeners percept as vocalised and non-vocalised, first for individual listen-

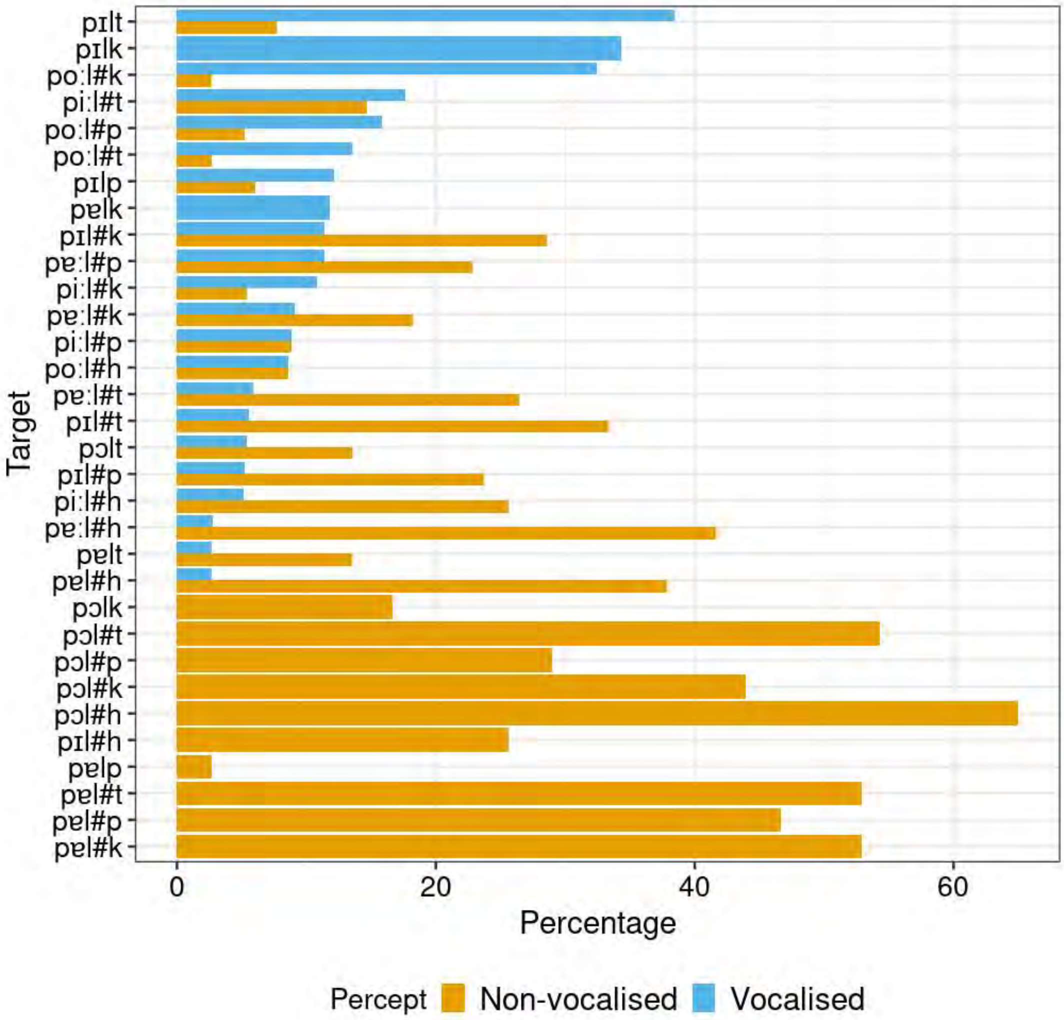


Figure 7.5: Percentage of vocalised and non-vocalised percepts by target. Blue bar: percentage of tokens perceived as vocalised (1-3) by Listeners 1, 2, and 4. Orange bar: percentage of tokens perceived as non-vocalised (0) by Listeners 1, 2, and 4. Percentage of tokens rated unequivocally (0-3) by Listeners 1, 2, and 4 is not shown.

ers and then across Listeners 1, 2, and 4. For each comparison, syllable type, vowel- and consonant context were used to select the target phrase; TT aperture, acoustic duration and listeners' ratings were used to select a specific token. To illustrate /l/ perceived as vocalised we selected contexts in which /l/ was more likely to be perceived as vocalised, such as after a long nucleus, and selected a target representing those contexts, and selected a specific tokens that showed articulatory-acoustic characteristics associated with vocalised percept, such as larger TT aperture (**Table 7.4**). To illustrate /l/ perceived as non-vocalised we selected contexts in which /l/ was less likely to be perceived as vocalised, such as before a glottal consonant, and selected a target representing those contexts, and selected a specific token that showed articulatory-acoustic characteristics associated with non-vocalised percept, such as smaller TT aperture (**Table 7.4**). The same method of token selection was used for individual listeners and across listeners.

Tokens selected for comparing articulatory characteristics of /l/ perceived as vocalised and non-vocalised by listener are presented in **Table 7.6**. Listeners disagreed on the rating of vocalised tokens, but mostly agreed on the rating of non-vocalised tokens (**Table 7.6**).

Table 7.6: Tokens selected for comparing articulatory characteristics of /l/ perceived as vocalised and non-vocalised. Rows show token ratings by all listeners. **Bold** ratings indicate that a token was selected to illustrate a given listener.

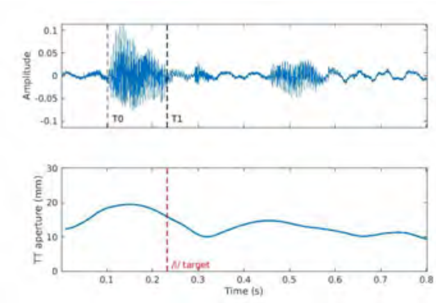
Speaker	Target	Item Number	Listener 1	Listener 2	Listener 4
W3	*pulk heap*	1	**3**	0	1
W10	*pill harp*	1	**0**	0	1
W3	*peel karp*	7	0	**3**	1
W10	*puhl heap*	5	0	**0**	0
W3	*pilk harp*	4	3	0	**2**
W10	*puhl heap*	3	0	1	**0**
W2	*pilt harp*	1	**2**	**2**	**2**
W2	*pilt harp*	5	**0**	**0**	**0**
W2	*parl heap*	3	**1**	**1**	**2**
W2	*parl heap*	7	**0**	**0**	**0**

Tokens illustrating individual listeners' percepts

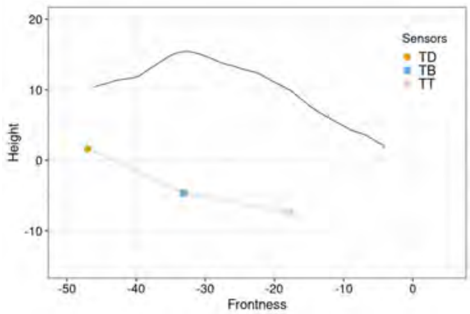
For Listener 1, we selected a token of *pulk heap* with large TT Aperture and short duration, representing the low vowel, alveolar consonant, short nucleus-complex coda context to exemplify a token perceived as vocalised (**Figure 7.6**). To exemplify a token perceived as non-vocalised, we selected a token of *pill harp*, with small TT aperture and long duration, representing the front vowel, glottal consonant, short nucleus - simple coda context (**Figure 7.7**).

For Listener 2, we selected a token of *peel karp* with large TT Aperture and long duration, representing the front vowel, long nucleus-simple coda context to exemplify a token perceived as vocalised (**Figure 7.8**). To exemplify a token perceived as non-vocalised, we selected a token of *puhl heap* with small TT aperture and short duration, representing the low vowel, glottal consonant context (**Figure 7.9**).

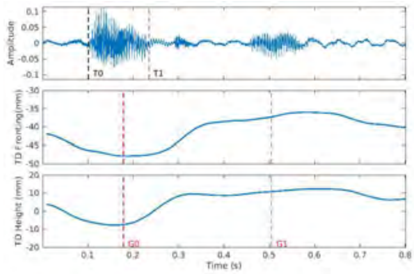
For Listener 4, we selected a token of *pilk harp* with large TT Aperture and short duration, representing the front vowel, dorsal consonant, short



(a) Vocalised percept: *pulk heap*



(b) Vocalised percept: *pulk heap*

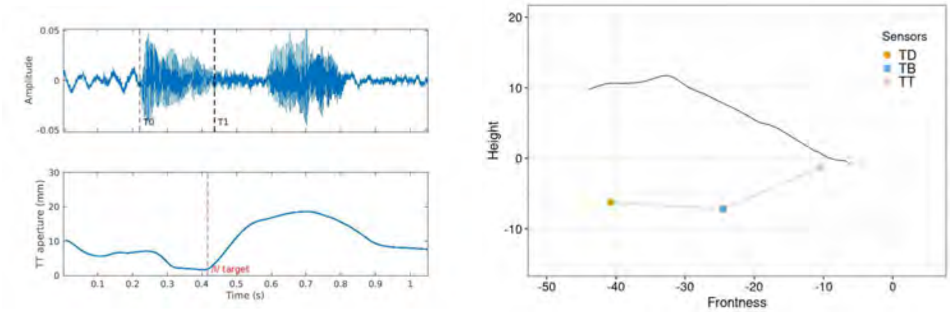


(c) TD gesture for /k/ in *pulk heap*. G0: Gestural onset for /k/. G1: gestural offset for /k/.

Figure 7.6: *pulk heap* perceived as vocalised by Listener 1. Top left: Acoustic waveform aligned with tongue tip aperture trajectories. T0: Onset of analysis window. T1: Offset of analysis window. Red line: minimum TT aperture in the last 40% of the analysis window used to identify /l/ target. Top right: Tongue shape measured at minimum TT aperture in the analysis window. Bottom: Acoustic waveform aligned with tongue dorsum raising and fronting trajectories in /k/. G0: Gestural onset of /k/. G1: Gestural offset of /k/.

nucleus-complex coda context to exemplify a token perceived as vocalised (**Figure 7.10**). To exemplify a token perceived as non-vocalised, we selected a token of *puhl heap* with small TT aperture and long duration, representing the low vowel, glottal consonant, short nucleus - simple coda context (**Figure 7.11**).

The TT aperture trajectories of tokens perceived as non-vocalised achieved and maintained a local minima near 0 mm in the second half of the

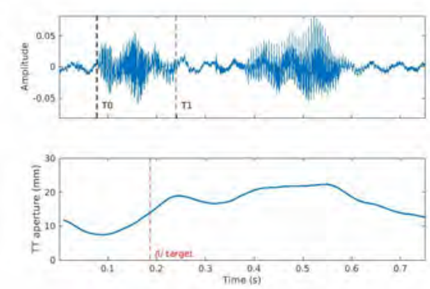


(a) Non-vocalised percept: *pill harp* **(b)** Non-vocalised percept: *pill harp*

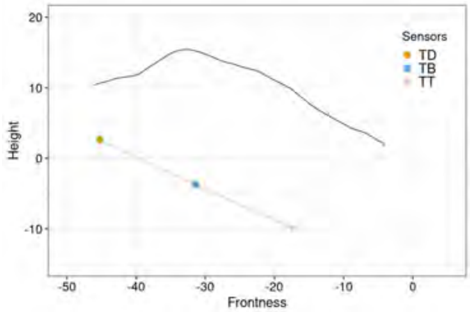
Figure 7.7: *pill harp* perceived as non-vocalised by Listener 1. Left: Acoustic waveform aligned with tongue tip aperture trajectories. T0: Onset of analysis window. T1: Offset of analysis window. Red line: minimum TT aperture in the last 40% of the analysis window used to identify /l/ target. Right: Tongue shape measured at minimum TT aperture in the analysis window.

analysis window, associated with /l/ (**Figures 7.7, 7.9** and **7.11**). These TT aperture trajectories show that coronal closure was achieved within the voiced interval and maintained until the offset of voicing. Tongue shapes measured at minimum TT aperture show a coronal closure and a backed and lowered tongue dorsum position, consistent with a non-vocalised /l/ (**Figures 7.7b, 7.9b, 7.11b**).

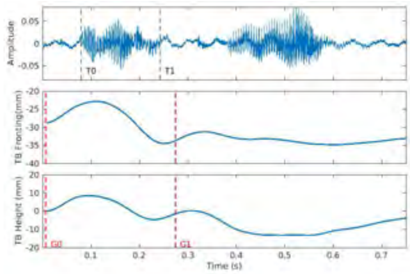
In contrast, the TT aperture trajectories of tokens perceived as vocalised do not achieve a near zero TT aperture within the analysis window (**Figures 7.6, 7.8**, **7.10**). The TT aperture trajectory in *pulk heap* shows a TT aperture minimum of 10 mm which coincides with the release burst of /k/ (**Figure 7.6a**). **Figure 7.6c** shows that the minimum TT aperture during the /k/ release burst is most likely to be caused by the raising of the tongue due to the raised tongue dorsum gesture of /k/. The TT aperture trajectories in the tokens *peel karp* and *pilk harp* show TT aperture minima in the first half of the analysis window, which is associated with the front vowel (**Figures 7.8a, 7.10a**). **Figures 7.8c** and **7.10c** show that the minimum



(a) Vocalised percept: *peel karp*



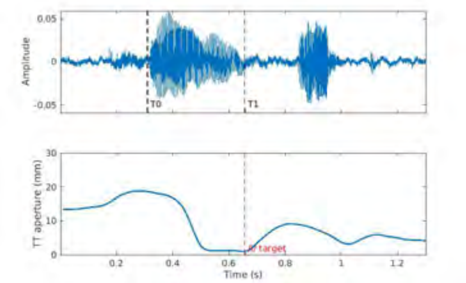
(b) Vocalised percept: *peel karp*



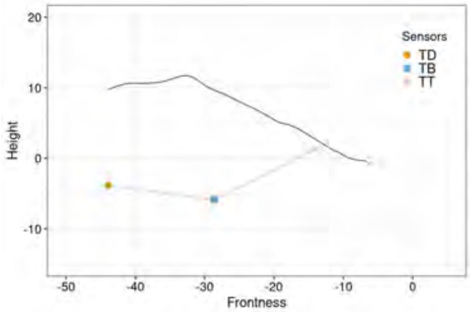
(c) TB gesture for /iː/ in *peel karp*. G0: Gestural onset for /iː/. G1: gestural offset for /iː/.

Figure 7.8: *peel karp* perceived as vocalised by Listener 2. Top left: Acoustic waveform aligned with tongue tip aperture trajectories. T0: Onset of analysis window. T1: Offset of analysis window. Red line: minimum TT aperture in the last 40% of the analysis window used to identify /l/ target. Top right: Tongue shape measured at minimum TT aperture in the analysis window. Bottom: Acoustic waveform aligned with tongue body raising and fronting trajectories in /iː/. G0: Gestural onset of /iː/. G1: Gestural offset of /iː/.

TT aperture in the first half of the analysis window is most likely to be caused by the raising and fronting of the tongue body due to the tongue body gestures of /iː/ and /ɪ/ respectively. TT aperture trajectories in the tokens *peel karp* and *pilk harp* do not show TT aperture minima beyond the analysis window (**Figures 7.8a, 7.10a**). These TT aperture trajectories in *pulk heap, pill karp* and *pilk harp* show that TT closure was not achieved either during the voiced interval or beyond the voiced interval, indicating



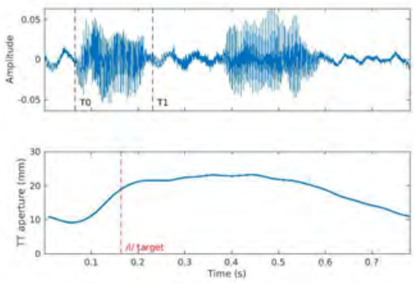
(a) Non-vocalised percept: *puhl heap*



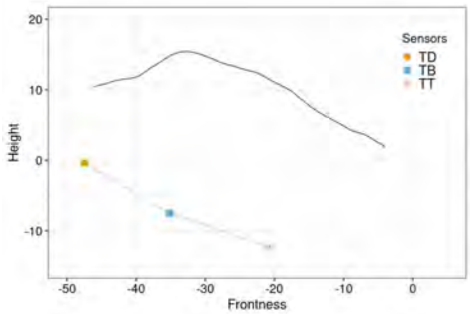
(b) Non-vocalised percept: *puhl heap*

Figure 7.9: *puhl heap* perceived as non-vocalised tokens by Listener 2. Left: Acoustic waveform aligned with tongue tip aperture trajectories. T0: Onset of analysis window. T1: Offset of analysis window. Red line: minimum TT aperture in the last 40% of the analysis window used to identify /l/ target. Right: Tongue shape measured at minimum TT aperture in the analysis window.

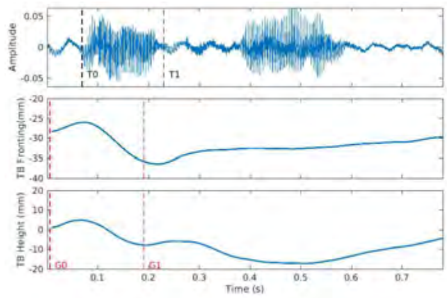
spatial reduction or deletion of the TT gesture. Tongue shapes measured at the lowest TT aperture in the second half of the analysis window show a downward pointing tongue tip and a raised tongue dorsum (**Figures 7.6b, 7.8b, 7.10b**).



(a) Vocalised percept: *pilk harp*

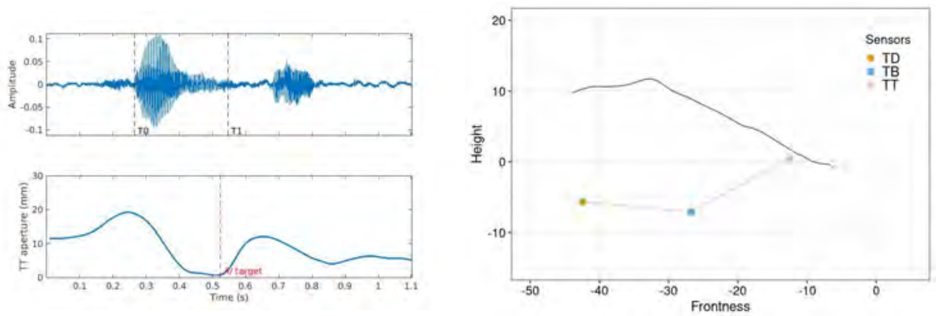


(b) Vocalised percept: *pilk harp*



(c) TB gesture for /ɪ/ in *pilk harp*. G0: Gestural onset for /ɪ/. G1: gestural offset for /ɪ/.

Figure 7.10: *pilk harp* perceived as vocalised by Listener 2. Top left: Acoustic waveform aligned with tongue tip aperture trajectories. T0: Onset of analysis window. T1: Offset of analysis window. Red line: minimum TT aperture in the last 40% of the analysis window used to identify /l/ target. Top right: Tongue shape measured at minimum TT aperture in the analysis window. Bottom: Acoustic waveform aligned with tongue body raising and fronting trajectories in /ɪ/. G0: Gestural onset of /ɪ/. G1: Gestural offset of /ɪ/.



(a) Non-vocalised percept: *puhl heap*

(b) Non-vocalised percept: *puhl heap*

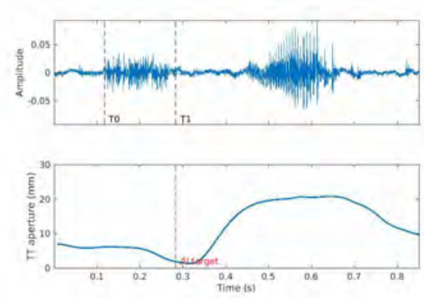
Figure 7.11: *puhl heap* perceived as non-vocalised tokens by Listener 4. Left: Acoustic waveform aligned with tongue tip aperture trajectories. T0: Onset of analysis window. T1: Offset of analysis window. Red line: minimum TT aperture in the last 40% of the analysis window used to identify /l/ target. Right: Tongue shape measured at minimum TT aperture in the analysis window.

Unequivocally rated tokens

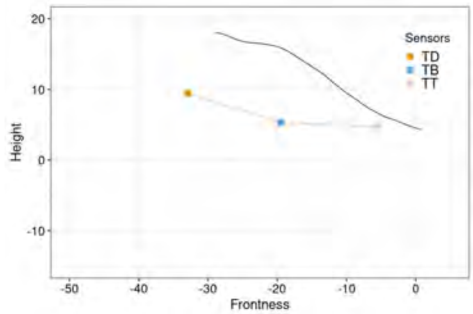
We identified the targets that were mostly identified as vocalised and those mostly identified as non-vocalised by Listeners 1, 2, and 4 (**Figure 7.5**). 38% of *pilt harp* tokens were perceived as vocalised (rated as 1-3), while 8% were perceived as non-vocalised (rated as 0); in contrast, 42% of *parl heap* tokens were perceived as non-vocalised (rated as 0), and 3% as vocalised (rated as 1-3). We selected two *pilt harp* tokens, one perceived as vocalised and one perceived as non-vocalised, as well as two *parl heap* tokens, one perceived as vocalised and one perceived as non-vocalised. All unequivocally rated tokens were produced by W2.

TT aperture trajectories of both *pilt harp* tokens show that TT aperture minima near 0 was achieved and maintained, therefore coronal closure was achieved in both tokens (**Figure 7.12**). TT aperture trajectories show that the coronal closure was released after the end of the voiced interval (**Figure 7.12**). These TT closures can be attributed both to /l/ and the following /t/. However, in the token perceived as non-vocalised TT closure was achieved before the end of the voiced interval, whereas in the token perceived as vocalised, TT closure was achieved at the end of the voiced interval. A TT closure achieved beyond the voiced interval is consistent with a delayed coronal closure (Strycharczuk & Scobbie 2019). In addition, tongue shapes measured at the TT minimum show that the *pilt harp* token perceived as non-vocalised has a lowered TD, whereas the token perceived as vocalised has a raised tongue dorsum.

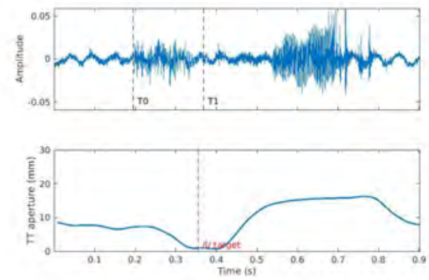
In both of the *parl heap* tokens, TT aperture trajectories achieved local minima within the voiced interval. However, the minimum TT aperture is near 0 mm in the token perceived as non-vocalised, and near 6 mm in the token perceived as vocalised (**Figures 7.13a, 7.13c**). TT closure is achieved during the voiced interval in the token perceived as non-vocalised and not achieved either during or beyond the voiced interval in the token perceived as vocalised, indicating spatial reduction of the coronal constriction.



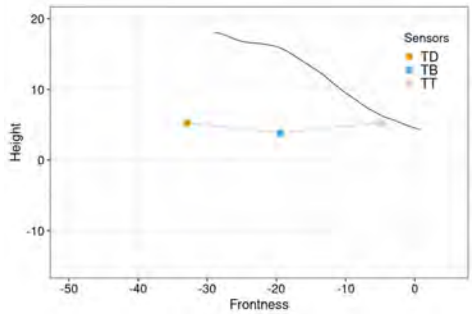
(a) Vocalised percept. Target: *pilt harp*. Token number: 1.



(b) Vocalised percept. Target: *pilt harp*. Token number: 1.

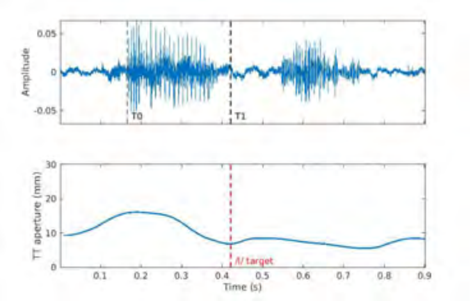


(c) Non-vocalised percept. Target: *pilt harp*. Token number: 7.

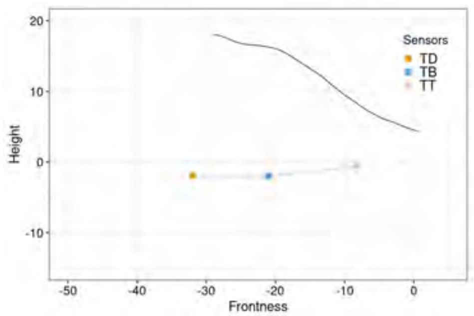


(d) Non-vocalised percept. Target: *pilt harp*. Token number: 7.

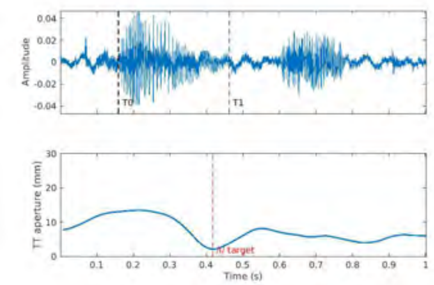
Figure 7.12: Tokens perceived as vocalised and non-vocalised by Listeners 1, 2, and 4. Top: *pilt harp*, repetition 1, produced by W2, perceived as vocalised. Bottom: *pilt harp*, repetition 5, produced by W2, perceived as non-vocalised. Left: Acoustic waveform aligned with tongue tip aperture trajectories. T0: Onset of analysis window. T1: Offset of analysis window. Red line: minimum TT aperture in the last 40% of the analysis window used to identify /l/ target. Right: Tongue shape measured at minimum TT aperture in the analysis window.



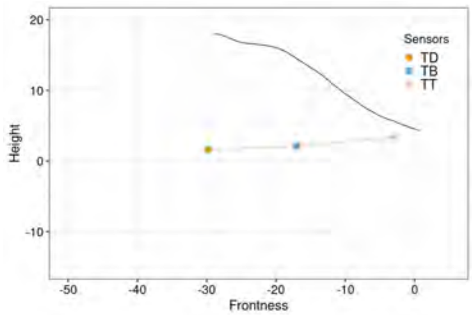
(a) Vocalised percept. Target: *parl heap*. Token number: 3.



(b) Vocalised percept: *parl heap*



(c) Non-vocalised percept: *parl heap*



(d) Non-vocalised percept: *parl heap*

Figure 7.13: Tokens perceived as vocalised and non-vocalised by Listeners 1, 2, and 4. Top: *parl heap*, repetition 3, produced by W2, perceived as vocalised. Bottom: *parl heap*, repetition 7, produced by W2, perceived as non-vocalised. Left: Acoustic waveform aligned with tongue tip aperture trajectories. T0: Onset of analysis window. T1: Offset of analysis window. Red line: minimum TT aperture in the last 40% of the analysis window used to identify /l/ target. Right: Tongue shape measured at minimum TT aperture in the analysis window.

7.4 Summary of Results

1. Listeners were inconsistent in classifying coda /l/ as vocalised and non-vocalised.

2. Listeners tended to perceive a coda /l/ with larger TT aperture as more vocalised compared to a coda /l/ with smaller TT aperture.

3. Listeners' percepts of coda /l/ as vocalised or non-vocalised are affected by the phonetic context of coda /l/; however, effects of phonetic contexts vary between listeners.

4. Contrasting articulatory characteristics of tokens perceived as vocalised and non-vocalised indicates that in tokens perceived as non-vocalised, TT closure is achieved and maintained during the voiced interval, whereas in tokens perceived as vocalised, TT closure is not achieved within the voiced interval.

7.5 Discussion

The aim of the present study was to establish whether listeners can reliably differentiate vocalised and non-vocalised /l/ and to explore articulatory and acoustic correlates of listeners' percept.

7.5.1 Reliability

Listeners showed low to moderate intrarater reliability, and no interrater reliability. When analysing the effects of articulatory characteristics and phonetic context on listeners' percept, listeners showed a large discrepancy between observed and predicted responses in gradient rating. The discrepancy is attributed to listeners rating the majority of tokens as non-vocalised, which resulted in the model overestimating the number of non-vocalised responses. The large discrepancy is not surprising in light of listeners' reported uncertainties. Predicted and observed responses aligned more when listeners responses were treated as categorical. The effects of articulatory characteristics and phonetic context of /l/ on listeners' percept show similar patterns between categorical and gradient ratings, but some effects were only observable when the ratings were treated as categorical.

Low intrarater reliability and the lack of interrater reliability contrasts with the results of (Hall-Lew & Fix 2012) who found that phoneticians can reliably identify vocalised /l/ using auditory-impressionistic methods. The different results can be attributed to the different stimuli: Hall-Lew & Fix (2012) selected tokens which the authors rated unequivocally for /l/ vocalisation on a scale ranging from 1 ("definitely consonantal") to 4 ("definitely vocalised") based on acoustic and auditory observations. Careful stimulus selection could have potentially maximised the perceptual difference between vocalised and non-vocalised tokens. In contrast, in the current study, all tokens produced by the speakers were presented to the listeners. As a result, our study might not have contained enough tokens canonically perceived as vocalised, leading to the large number of non-vocalised responses and therefore to poor model fit on the gradient rating.

Low interrater reliability can be attributed to the blocks and trials being randomised between listeners. As some listeners might have been first exposed to the least-vocalised tokens while other listeners to the most-vocalised tokens, their decision criteria could have been affected by their initial perception of the range of vocalisation present in the dataset. Randomisation between tokens and speakers might have had smaller effect on Hall-Lew & Fix's (2012) results due to the presence of canonically vocalised tokens and to the larger number of listeners. However, the number of listeners is too low in the current study to further investigate the effects of order of presentation on listeners' percept.

Task difficulty can also account for low intrarater and interrater reliability: listeners reported that they thought the task was long and repetitive. We used several non-words in the task; the pronunciation of the stimuli was somewhat unnatural due to the sensors attached to speakers' tongue; the recordings contained some amount of background noise from the EMA system, all of which could have made the task difficult.

7.5.2 Correlates of listeners' categorical ratings

Articulatory observations indicated prevalent coronal constriction reduction in the stimuli and listeners indicated that they consciously listened to the acoustic cues of TT contact, or the lack thereof. Larger TT aperture consistently corresponds to a more vocalised percept for Listeners 1, 2, and 4. Listeners also indicated that they listened to the acoustic cues of lip rounding; however, lower lip fronting did not have a consistent effect on listeners' ratings.

Listeners reported that although they faced different difficulties and perceived a different amount of tokens as vocalised in different phonetic contexts, they still used the same rating criteria across contexts. However, listeners' ratings were strongly influenced by phonetic context. For instance, a following glottal consonant lead to a less-vocalised, whereas a preceding long nucleus or a following dorsal consonant lead to a more vocalised percept. These context effects on perception are somewhat consistent with articulatory analyses of the corpus showing that TT aperture is smaller before an /h/ and larger before a dorsal, but smaller after a long nucleus

(**Chapter 6**). The combination of the articulatory and perceptual results indicate that these phonetic contexts affected both TT constriction reduction and listeners' percept. Although phonetic contexts had a consistent effect on coronal constriction reduction and listeners' percept as vocalised /l/, the fact that phonetic context affected listeners' percept over and above TT aperture suggests that listeners are sensitive to the context itself.

All other factors affected listeners' perception differently. Two listeners (1 and 4) perceived tokens with a phonetically long vowel and /l/ sequence as less vocalised, whereas another listener (Listener 2) perceived tokens with a phonetically long vowel and /l/ sequence as more vocalised. The effect of duration on this latter listener (Listener 2) is consistent with the syllable type effects on perception: vowel-/l/ duration increases when the vowel preceding /l/ is phonologically long (e.g. /ɐːl/ in *parl peep*), and Listener 2 perceived both /l/ in a phonetically longer vowel-lateral sequence and /l/ following a phonologically long vowel as more vocalised. Similarly, vowel-l duration decreases when the vowel preceding the /l/ is phonologically short (e.g. /ɐl/ in *pulp*), and Listener 2 perceived both /l/ in a phonetically shorter vowel-lateral sequence and /l/ following a phonologically short vowel in a complex coda as less vocalised. In contrast, two listeners who perceived phonetically longer vowel-lateral sequence as less vocalised nevertheless perceived /l/ following a phonologically long nucleus as more vocalised.

Listeners differed in the effect that a following alveolar consonant and preceding back vowel had on their percept of vocalisation. The differing effects of alveolar consonant and back vowel contexts can be explained by different listener strategies in attributing acoustic cues to their articulatory source. When the consonant following /l/ is alveolar, TT contact is made at some point (**Chapter 6**). The resulting cue to alveolar closure in the acoustic signal can be attributed both to the alveolar consonant and a non-vocalised lateral. A listener who attributes the closure cues to the /l/, will perceive /l/ before an alveolar as non-vocalised (Mann & Repp 1980, Mann 1980, Kleber et al. 2012). A listener who attributes the closure cues to the following alveolar will perceive /l/ in /l(#)t/ as vocalised (Mann & Repp 1980, Mann 1980, Kleber et al. 2012). When the vowel preceding /l/ is rounded and back, the resulting cues for lip rounding in the acoustic signal can be attributed both to the vowel and to a vocalised lateral. A listener who attributes the backing and rounding cues to the preceding back vowel, will perceive /l/ after a back vowel as non-vocalised (Mann & Repp 1980, Mann 1980, Kleber et al. 2012). A listener who attributes the backing and rounding cues to /l/ will perceive back vowel-/l/ sequences as vocalised (Mann & Repp 1980, Mann 1980, Kleber et al. 2012).

Listener 1 seems to have attributed both the closure cues and the backing and rounding cues to the phonetic context, perceiving the lateral as more vocalised when it was followed by /t/ and less vocalised when it was preceded by /oː, ɔ/. In contrast, Listener 4 seems to have attributed both the closure and the backing and rounding cues to the lateral, perceiving /l/ as less vocaised when it was followed /t/ and more vocalised when it was preceded by /oː, ɔ/. Also in contrast to Listener 1, Listener 2 seems to have attributed the rounding cues to the lateral, perceiving /l/ as more vocalised when it was preceded by /oː, ɔ/. An alternative explanation for the rating patterns of Listeners 2 and 4 is that they were affected by their knowledge of phonological literature arguing that /l/-vocalisation is less likely before alveolars and more likely after back vowels in AusE (Borowsky 2001).

The findings that listeners were strongly affected by phonetic context and less affected by the speakers' actual TT aperture and lip rounding are consistent with the effect of phonetic context on perception (Mann & Repp 1980, Mann 1980) and are less consistent with listeners' self-reported reliance on acoustic cues of TT lenition and rounding. The low impact of TT aperture coupled with the impact of phonetic context questions what previous studies in which data was coded with auditory-impressionistic methods can tell us about /l/-vocalisation. One of the key defining features of /l/-vocalisation is the loss of tongue tip contact and lip rounding in production, but our data indicates that listeners' attribution of acoustic cues might be equally important. The varying effects of phonetic context questions what previous studies in which data was coded by a single listener can tell us about /l/-vocalisation. In the current study, every listener seemed to have a slightly different percept of vocalisation.

It is also quite likely, given the poor model fit, that all listeners relied on certain cues which were not captured by either the articulatory metrics or the phonetic contexts. For instance, coronal constriction reduction is accompanied by smaller F1-F2 distance, and lip rounding by lowering of formants (Lin et al. 2012). /l/-vocalisation is more likely in faster speech than in slower speech (Wright 1988), and listeners could have listened for overall speech rate which is captured poorly by the length of the vowel-lateral sequence. Acoustic analyses of the corpus are required to shed light on this but falls outside the scope of the present study.

7.5.3 Articulatory characteristics of tokens perceived as vocalised and non-vocalised

A detailed comparison of the articulatory characteristics of tokens that were perceived as vocalised or non-vocalised by either one or all listeners suggested that a vocalised percept corresponds to coronal constriction reduction. Coronal closure was not observed in tokens perceived as vocalised either within or beyond the voiced interval. We did not observe tongue tip closure beyond the voiced interval, except for one token of *pilt harp*; however, in *pilt harp*, tongue tip closure can be attributed to both /l/ and /t/. The lack of delayed closure in the selected tokens is consistent with the findings of Strycharczuk & Scobbie (2019) who did not observe gestural delay without spatial reduction; however, delayed coronal closure can be present in the larger dataset. In contrast, coronal closure was achieved and maintained during the voiced interval associated with /l/ in tokens perceived as non-vocalised. In addition, tokens perceived as vocalised often exhibited a raised tongue dorsum, whereas tokens perceived as non-vocalised showed a lowered and backed tongue dorsum. These observations hold both across each listeners' quintessential vocalised and non-vocalised tokens, as well as across the two tokens whose vocalised/non-vocalised status listeners agreed on.

Shared articulatory patterns of the tokens perceived as vocalised and of those perceived as non-vocalised might indicate that listeners could have identified tokens matching some of the articulatory definitions of /l/-vocalisation. If this is the case, low intrarater reliability can be explained if all listeners underestimated the proportion of vocalised tokens in the dataset: as listeners were affected differently by the phonetic characteristics and the phonetic context of the tokens, they identified a different subset of all lenited tokens as vocalised. This argument is supported by comparing each listeners rating for the tokens presented as vocalised in **Figures 7.6, 7.8 7.10**. Articulatory analyses of the tokens perceived as vocalised by *one listener* indicates that these tokens were articulated with a lenited tongue tip gesture. However, these tokens were perceived as non-vocalised by at least two other listeners (**Table 7.6**).

As we only examined ten tokens in which one or more listeners' percept as vocalised or non-vocalised matches articulatory analysis, we cannot conclude that listeners could reliably identify vocalised and non-vocalised tokens in the entire dataset. A large-scale quantiative comparison between the articulatory characteristics of tokens perceived as vocalised and of those perceived as non-vocalised could give insight into the extent to which listeners can reliably identify tokens with all articulatory characteristics of vocalisation.

7.6 Conclusion

This study showed that listeners are inconsistent in their percept of vocalised and non-vocalised /l/. The effects of articulatory, acoustic, and contextual factors on listeners' percept of /l/ as vocalised varies between listeners. Despite the subtly different interpretations of the rating scales and different phonetic correlates of perception, closer inspection of articulatory patterns in tokens perceived as vocalised and non-vocalised indicates that listeners vocalised percept corresponds to coronal lenition.

7.7 Acknowledgements

This research was supported in part by iMQRTP 2015144 and ARC DE150100318 grants. We thank Louise Ratko for her help in data collection, Peter Humburg for his help in the statistical analysis, and the members of the Phonetics Lab, the Centre for Language Sciences and the

ARC Centre of Excellence in Cognition and its Disorders at Macquarie University for their comments, feedback and support.

7.8 References

Archer, K. J. & Williams, A. A. (2012). L1 penalized continuation ratio models for ordinal response prediction using high-dimensional datasets. *Statistics in Medicine*, *31*, 1464–1474.

Ash, S. (1982). *The vocalization of /l/ in Philadelphia.* Unpublished doctoral dissertation.

Borowsky, T. (2001). The vocalisation of dark *l* in Australian English. In D. Blair & P. Collins (Eds.), *English in Australia* (p. 69–87). Philadelphia, Amsterdam: John Benjamins Publishing Company.

Borowsky, T. & Horvath, B. (1997). L-vocalisation in Australian English. In F. Hinskens, R. van Hout & W. L. Wetzels (Eds.), *Variation, change and phonological theory.* (p. 101–123). Amsterdam: Benjamins.

Browman, C. P. & Goldstein, L. M. (1995). Gestural syllable position effects in American English. In F. Bell-Berti & J. L. Raphael (Eds.), *Producing speech: contemporary issues.* New York: AIP Press.

Draxler, C. & Jansch, K. (2017). *Speechrecorder 3.10.0.*

Gick, B. (1999). A gesture-based accounts of intrusive consonants in English. *Phonology*, *16*(1), 29–54.

Gick, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. *Phonetic interpretation: papers in laboratory phonology VI*, 222–236.

Gick, B., Kang, M. A. & Whalen, D. H. (2002). MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics*, *30*(3), 357–371.

Giles, S. & Moll, K. (1975). Cinefluorographic study of selected allophones of English /l/. *Phonetica*, *31*(3-4), 206–227.

Hall-Lew, L. & Fix, S. (2012). Perceptual coding reliability of (l)-vocalization in casual speech data. *Lingua*, *122*(7), 794–809.

Hardcastle, W. & Barry, W. (1989, 7). Articulatory and perceptual factors in /l/ vocalisations in English. *Journal of the International Phonetic Association*, *15*, 3–17.

Horvath, B. M. & Horvath, R. J. (1997). The geolinguistics of a sound change in progress: /l/ vocalization in Australia. *U. Penn Working Papers in Linguistics*, *4*(1), 109–124.

Horvath, B. M. & Horvath, R. J. (2001). A multilocality study of a sound change in progress: The case of /l/ vocalization in New Zealand and Australian English. *Language Variation and Change*, *13*, 37–57.

Horvath, B. M. & Horvath, R. J. (2002). The geolinguistics of /l/ vocalisation in Australia and New Zealand. *Journal of Sociolinguistics*, *6/3*, 319–346.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Kleber, F., Harrington, J. & Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, *55*(3), 383–405.

Krause, F. & Lindemann, O. (2014, 1st Jun). Expyriment: A python library for cognitive and neuroscientific experiments. *Behavior Research Methods*, *46*(2), 416–428.

Krippendorff, K. (2011). Computing krippendorff's alpha-reliability.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford, UK; Cambridge, Mass.: Blackwell.

Lin, S., Beddor, P. & Coetzee, A. (2014). Gestural reduction, lexical frequency, and sound change: a study of post-vocalic /l/. *Laboratory Phonology*, *5*(1), 9–36.

Lin, S., Palethorpe, S. & Cox, F. (2012). An ultrasound exploration of Australian English /CVl/ words. In *14th Australasian International Conference on Speech Science and Technology* (p. 105–108). Sydney.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, *28*(5), 407–412.

Mann, V. A. & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*(3), 213–228.

Proctor, M., Walker, R., Smith, C., Szalay, T., Narayanan, S. & Goldstein, L. (2019). Articulatory characterization of English liquid-final rimes. *Journal of Phonetics*, *77*, 100921.

R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Schiel, F. (1999, August). Automatic Phonetic Transcription of Non-Prompted Speech. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville &

A. C. Bailey (Eds.), *Proceedings 14th international congress of phonetic sciences* (p. 607–610). San Francisco, CA, USA.

Scobbie, J. M. & Pouplier, M. (2010). Syllable structure and external sandhi: an EPG study of vocalisation and retraction of word-final English /l/. *Journal of Phonetics*, *38*, 240–259.

Scobbie, J. M., Pouplier, M. & Wrench, A. A. (2007). Conditioning factors in external sandhi: an EPG study of English /l/vocalisation. In *Proceedings of the 16th international congress of phonetic sciences* (p. 441–444). Saarbrücken, Germany.

Scobbie, J. M. & Wrench, A. A. (2003). An articulatory investigation of word final /l/ and //l/-sandhi in three dialects of English. In *Proceedings of the 15th international congress of phonetic sciences.* Barcelona, Spain.

Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. (2011). Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, *39*(5), 1–13.

Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*(3), 291–311.

Strycharczuk, P., Derrick, D. & Shaw, J. (2018). How vocalic is vocalised /l/? Evidence from lateralisation. In *Conference of the association for laboratory phonology.* Lisbon, Portugal.

Strycharczuk, P. & Scobbie, J. M. (2019). Gestural delay and gestural reduction. Articulatory variation in /l/-vocalisation in Southern British English. In A. Prezwozny, C. Viollain & S. Navarro (Eds.), *Advances in the* *study of contemporary spoken English: Multifocal analyses of variation.* Edinburgh: Edinburgh University Press.

Turton, D. (2017). Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalization in varieties of English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1).

Wells, J. C. (1982). *Accents of English.* Cambridge: Cambridge University Press.

Wrench, A. A. & Scobbie, J. M. (2003). Categorising vocalisation of English /l/ using EPG, EMA and ultrasound. In *Proceedings of the 6th international seminar on speech production* (p. 314–319).

Wright, S. (1988). The effects of style and speaking rate on/l/-vocalisation in local Cambridge English.

Ying, J., Carignan, C., Shaw, J. A., Proctor, M., Derrick, D. & Best, C. (2017). Temporal dynamics of lateral channel formation in /l/: 3D EMA data from Australian English. In *Proceedings of the 18th annual conference of the international speech communication association.* Stockholm, Sweden.

8 Summary of findings

1. Acoustics

    (a) The acoustic vowel space is reduced for pre-lateral vowels, compared to vowels produced before coronal obstruents (**Chapter 2**).

    (b) Acoustic vowel contrast is particularly reduced between the members of the pairs /iːl-ɪl, ʉːl-ʊl, æɔl-æl, əʉl-ɔl/ (**Chapter 2**).

2. Perception

    (a) Vowel disambiguation is overall more difficult before coda /l/ than coda /d/ (**Chapter 3**).

    (b) Perceptual vowel contrast is particularly reduced between the members of the pairs /ʉːl-ʊl, æɔl-æl, əʉl-ɔl/ (**Chapter 3**).

    (c) Reduced perceptual contrast between the members of the pairs /ʉːl-ʊl, æɔl-æl, əʉl-ɔl/ is mediated by lexical frequency (**Chapter 3**).

    (d) When both the speaker and the listener produce larger duration contrast between the members of the pairs /ʉːl-ʊl, æɔl-æl, əʉl-ɔl/, listeners are better at disambiguating members of the pairs (**Chapter 4**).

3. Articulation

    (a) The tongue is more elongated in /l/ compared to /d/ in both onset and coda position (**Chapter 5**).

    (b) The coronal gesture of coda /l/ is lenited as the coronal constriction is reduced (**Chapter 6**).

    (c) The magnitude of coronal constriction reduction varies between speakers (**Chapter 6**).

(d) The magnitude of coronal constriction reduction varies between phonetic contexts depending on the place of articulation of the adjacent segments: preceding back vowel and following dorsal consonant decrease coronal constriction, while following alveolar consonant increase it (**Chapter 6**).

(e) Coronal constriction reduction in coda /l/ does not predict listeners' perception of that token as vocalised. However, listeners perceive /l/ as more vocalised in phonetic contexts that reduce constriction: after a preceding back vowel and before a dorsal consonant. (**Chapter 7**).

9 General discussion

In this dissertation, we aimed to extend our knowledge of lateral-final rimes by systematically characterising aspects of their acoustic structure, of listener perception, and speaker articulation in AusE – a language that shows variation between clear, dark, and vocalised /l/. Integrating the findings from these three domains helps to provide a more comprehensive account of lateral-final rimes than has previously been undertaken in this variety. We found that prelateral vowels undergo acoustic and perceptual contrast reduction, which we attributed to the coarticulatory influence of the dorsal gesture of /l/. This is consistent with previous findings on the coarticulatory influence of coda /l/ on prelateral vowels in American English (Gick & Wilson 2006, Proctor et al. 2019). We also found that in AusE the coronal gesture of coda /l/ is lenited, – the midsagittal occlusion is, on average, less constricted than in onset /l/ – but the magnitude of lenition varies between speakers and phonetic contexts. This is consistent with auditory-impressionistic observations that coda /l/ is variably vocalised in AusE (Wells 1982, Borowsky 2001). We found that phonetic context, rather than tongue tip constriction is a better predictor of listeners' perception of /l/ as vocalised.

We begin this discussion by offering some possible explanations for prelateral vowel contrast reduction, referring to our findings on the acoustic characteristics and the perception of prelateral vowels in **Section 9.1**. In **Section 9.2**, we reflect on the characteristics of /l/-vocalisation, based on our articulatory analysis of /l/ lenition and the perception of /l/ vocalisation. In **Section 9.3**, we discuss the implications of our results for the goals of lateral production. In **Section 9.4** we consider the implications of our results for models of sound change. Finally, we address some of the limitations of this work (**Section 9.5**) and suggestions for how the studies contained in this thesis provide the impetus for future work on lateral-final rimes. **Section 9.6** presents our concluding statements.

9.1 Contrast reduction in lateral-final rimes

The acoustic and perceptual analyses of /l/-final rimes provide comprehensive, although indirect, information on the coarticulatory influences of the coda lateral on preceding vowels; in particular, these data (**Chapters 2** and **3**) suggest that spectral contrast is reduced in the prelateral vowel space both in acoustics and in perception. Members of the vowel pairs /ʉː-ʊ, əʉ-ɔ, æɔ-æ/, and to a lesser extent /iː-ɪ/ show increased contrast reduction compared to other vowel pairs both in acoustics and in perception. The acoustic analysis in **Chapter 2**, although indicative, cannot reveal whether a contrast has been reduced or neutralised. However, the perceptual confusions found in **Chapter 3** are consistent with a potential perceptual merger, as listeners were not always able to distinguish prelateral vowels on the basis of spectral and durational cues. Our finding that listeners rely on high lexical frequency in word recognition is consistent with models of speech perception proposing that the importance of top-down information increases when signal ambiguity increases (Norris & McQueen 2008).

We attributed spectral and perceptual contrast reduction between certain vowel pairs to the influence of the dorsal gesture of /l/, as the dorsal gesture has been shown to be adjacent to the vowel gesture and has been shown to influence the vowel gesture (Sproat & Fujimura 1993, Proctor et al. 2019). Although a detailed articulatory analysis of prelateral vowels and the gestural interactions between vowels and /l/ were beyond the scope of the current dissertation, we explored how tongue backing could explain the influence of coda /l/ on prelateral vowels in acoustics. The overall phonemic backing and lowering observed in AusE pre-lateral vowels (**Chapter 2**) is consistent with a pattern of production in which the lowered and retracted tongue dorsum gesture of coda /l/ coarticulates with the vowel gesture (Fant 1960). In particular, the phonetic backing of /ʉː, əʉ/ and the lowering of /oː/ observed here are consistent with the articulatory backing and lowering of these vowels observed in previous work for AusE (Lin et al. 2012). However, our results show that all vowels are affected by coda /l/, which does not align with Lin et al.'s (2012) findings that only /ʉː, əʉ, oː/ and /æɪ/ are affected by a following lateral in AusE. In addition, Ying et al. (2012) suggested that acoustic backing may coincide with lateral channel formation during /l/ production. A follow-up study on the prelateral vowel space in AusE using articulatory methods could shed further light on vowel-lateral coarticulation.

Having established spectral and perceptual ambiguity within the members of the pairs /iː-ɪ, ʉː-ʊ, əʉ-ɔ, æɔ-æ/, the question arises whether listeners use different criteria to distinguish within them in prelateral contexts compared to the pre-obstruent contexts. Listeners are known to use context-sensitive criteria for distinguishing phonemes (Mann & Repp 1980, Mann 1980). For instance when differentiating between /ʉː-ʊ/ in the fronting /s_t/ context, listeners are more likely to accept a vowel with a higher $F2$ as /ʊ/ than in the backing /w_l/ context (Kleber et al. 2012). Further investigation is required to establish that listeners due different decision criteria for identifying prelateral vowels compared to preobstruent vowels. Finding different decision criteria would indicate that listeners compensate for the coarticulatory influence of /l/ (Mann & Repp 1980, Mann 1980, Kleber et al. 2012).

Our studies also examined the influence of temporal factors in disambiguating lateral-final rimes; however, the challenge of identifying discernible boundaries in the acoustic signal between vowels and coda laterals within the rime made it difficult to examine the realisation of durational contrast. Despite such difficulties, our findings indicate that lateral-final rimes containing long vowels are longer compared to lateral-final rimes containing short vowels (**Chapters 2, 4**). In **Chapter 4**, we found that the presence of duration cues facilitated recognition of /l/-final rimes for listeners who produced a larger duration contrast, but not for listeners who produced a smaller duration contrast. The lack of an overall beneficial effect of the duration cue is consistent with previous research showing that English listeners rely less on durational cues than on spectral cues and that Australian English listeners confuse spectrally similar vowels more than those that contrast in duration (Bennett 1968, Szalay et al. 2016). The finding that only those speaker-listeners who produce larger duration contrast can benefit from the duration cue in perception sheds new light on how listener-speakers' own production affects their perception (Perkell, Guenther et al. 2004, Perkell, Matthies et al. 2004, Wade 2017).

Chapter 4 raises the question whether there is a link between spectral contrast reduction in production and accuracy of perception, and whether there are trade-off, cue weighting, or cue transferring relationships between durational and spectral contrast. Future studies exploring these relationships in more detail would be especially important because AusE differs from other varieties of English in having phonologically contrastive duration (Cox & Palethorpe 2007).

9.2 Coronal constriction reduction and /l/-vocalisation

Articulatory analysis of coda /l/ provided rich information on coronal constriction reduction in coda /l/ (**Chapter 6**). Perceptual analyses of tokens with known articulatory characteristics allow us to investigate the relationship between coronal constriction reduction and vocalised percept (**Chapter 7**). Articulatory and perceptual analyses of coda /l/-lenition indicate a discrepancy at the token-by-token level: coronal constriction reduction does not always correspond to a vocalised percept. However, the patterns of production and perception are similar, as the phonetic contexts that facilitate coronal constriction reduction in coda /l/ also increase the likelihood of a vocalised percept.

Coda /l/ lenition was observed in the articulatory data, as five out of six speakers produced coda /l/ with some reduction in midsagittal occlusion (**Chapter 6**). Finding reduced coronal constriction in coda /l/ is consistent with previous research that reported coda /l/ vocalisation in AusE (e.g. Horvath & Horvath 1997; 2001; 2002, Borowsky 2001, Cox & Palethorpe 2007). However, in **Chapter 6**, we only analysed coronal constriction within the voiced interval associated with the /l/-final rime. Lack of coronal closure within the voiced interval is consistent with temporal delay, when coronal closure is achieved beyond the voiced interval; with spatial reduction, when the tongue tip approaches the alveolar ridge without achieving closure; and with deletion, when no tongue tip gesture can be observed. In a closer observation of a few selected tokens, we did not find any tokens in which the coronal gesture was delayed but constriction was not reduced; however, we did observe coronal constriction reduction as well as deletion of the coronal gesture (**Chapter 7**). Our finding that /l/-lenition in AusE is caused by spatial reduction, rather than temporal delay, is consistent with findings on Standard Southern British English in which /l/-vocalisation was found to be primarily spatial phenomenon, as all delayed gestures were reduced and delay without reduction was not observed (Strycharczuk & Scobbie 2019). Our finding that /l/-vocalisation in AusE is caused by spatial reduction is consistent with differentiating between three allophones of /l/: clear /l/, dark /l/ articulated with delayed tongue tip, and vocalised /l/ articulated with reduced constriction (Strycharczuk & Scobbie 2019). However, categorising a token of /l/ as vocalised or non-vocalised is arbitrary, as constriction reduction is inherently gradient: one criterion to differentiate between non-vocalised and vocalised /l/ can be the presence or the absence of coronal closure, with partial closure classified as an intermediate category, while another criterion can be the presence or the absence of the coronal gesture with coronal gesture with reduced constriction classified as an intermediate category (Strycharczuk & Scobbie 2019).

Despite the prevalence of reduced coronal constriction, listeners perceived most tokens as non-vocalised (**Chapter 7**). In contrast, listeners in previous studies were able to observe /l/-vocalisation in AusE auditorily and impressionistically. One possible explanation for our findings might have been listeners underestimating /l/-vocalisation despite /l/ often being lenited. This would indicate that the differentiation between vocalised and dark /l/ is a challenge, possibly because dark and vocalised /l/ are too similar and the difference is subtle and gradient.

Another, more likely, explanation is that the reduction of the coronal constriction is a necessary but not sufficient condition for creating the percept of vocalised /l/. For instance, additional temporal delay of the coronal gesture, lip rounding or the loss of the lateral gesture have been demonstrated to be crucial in /l/-vocalisation (Strycharczuk et al. 2018, Strycharczuk & Scobbie 2019), and might be necessary for creating the percept of a vocalised /l/. In this case, the coronal constriction reduction shown in **Chapter 6** is not sufficient to argue that /l/-vocalisation is present in the corpus without additional analysis of the timing of the coronal gesture, lip rounding and lateralisation. For instance, it is possible that only tokens with a deleted coronal gesture are perceived as vocalised, while tokens with a weakened coronal gesture are not. If coronal constriction reduction is not a sufficient marker of /l/-vocalisation, then articulatory studies analysing tongue tip articulation alone could possibly have overestimated the presence of /l/-vocalisation in their dataset (e.g. Hardcastle & Barry 1989). Further investigation of what gradient articulatory characteristics might motivate listeners' percepts can aid us in drawing a less arbitrary cut-off point when classifying tokens of /l/ as vocalised or non-vocalised.

It is possible that there are better predictors for listeners' percept of coda /l/ as vocalised than articulatory characteristics of coda /l/. Vocalised /l/ might be associated with a set of acoustic cues, such as the frequency separation between the first and second formants (Lin et al. 2014). Duration of the vowel-lateral interval was already shown as an important correlate of listeners' percept (**Chapter 6**). Future analyses would benefit from exploring acoustic correlates of listeners' percepts to provide important information on the cues that listeners use to differentiate between vocalised and non-vocalised tokens.

Despite the token-by-token differences in perception and production and the fact that listeners did not associate increased tongue tip aperture strongly with increased vocalisation, speakers and listeners did show similar patterns in the same phonetic contexts. Tokens in phonetic contexts that facilitate coronal constriction reduction, such as a following dorsal consonant or a preceding back vowel, were perceived as more vocalised by some listeners compared to tokens in contexts inhibiting coronal constriction reduction. These findings indicate that listeners might be more sensitive to context than to articulatory characteristics of the tokens (Mann & Repp 1980, Mann 1980, Kleber et al. 2012). For instance, coronal constriction reduction was frequently observed in tokens in which coda /l/ was preceded by a back vowel and followed by a dorsal consonant (e.g. *polk*) because of the the overall backing of the tongue. A token with a dorsal consonant and a back vowel contains cues to overall backness. A token with a back vowel also contains cues to rounding, because all AusE back vowels are rounded. A listener could have interpreted cues to backing and rounding as cues to vocalised /l/. In contrast, coronal constriction reduction was rarely observed when /l/ was followed by /t/, because of the coarticulatory influence of the coronal stop (e.g. *pilt*). The same *pilt* token could also contain cues to overall frontness and a coronal closure, which the listeners could attribute to a non-vocalised /l/.

Listeners could rely on their explicit phonological knowledge to identify vocalising contexts: all listeners were expert phoneticians with explicit knowledge on the articulation of /l/, coarticulatory mechanisms, and potentially on the effect of phonetic context on /l/ vocalisation as described by Borowsky (2001). Therefore listeners could have been influenced by knowing that an alveolar consonant “should inhibit” while a dorsal consonant “should facilitate” /l/-vocalisation.

The high variability in listeners' percepts indicates that different participants may have been operating on different dimensions and responding in different ways. Listeners might differ in their perceptual acuity, in the cues they rely on, in their interpretation of acoustic cues, and in their explicit knowledge of phonology. Therefore it can be difficult to interpret and compare data from studies in which coda /l/ has been identified as vocalised or non-vocalised using auditory-impressionistic methods (e.g. Ash 1982, Borowsky 2001, Stuart-Smith et al. 2006, Hall-Lew & Fix 2012) without further information about the behaviour and understanding of the listeners.

Despite the low number of speakers and listeners and the high variability in listeners' percept, we identified contexts in which coronal constriction reduction and vocalised percepts are observed more frequently. Future studies should establish whether these patterns can be observed in a larger dataset. These results raise questions about additional characteristics of coda /l/ lenition in articulation, about the acoustic correlates of a lenited coda /l/, and about the acoustic cues that correlate with listeners' percept. They also raise questions about the impact of phonetic context not only on coda /l/ lenition, but also on the acoustic signal, and about how listeners attribute acoustic cues to the phonetic context or to coda /l/. Future studies should bear in mind that coronal constriction reduction alone does not necessarily correlate with vocalised percept, and make use of articulatory, acoustic, and perceptual methods simultaneously to examine /l/-vocalisation.

9.3 Implications for the goal of lateral production

Articulatory phonology proposes that the smallest contrastive unit in phonology is the articulatory gesture (Browman & Goldstein 1986); however, the primary goal of production of the lateral is still an open question (Browman & Goldstein 1989, Ying et al. 2017, Proctor et al.

2019). The simultaneous tongue tip raising and fronting and tongue dorsum lowering and retraction has been proposed as the primary goal of production for English /l/ (Browman & Goldstein 1992, Ladefoged & Maddieson 1996). This two-point displacement leads to tongue elongation and narrowing, and therefore to lateral channel formation (Browman & Goldstein 1992, Ladefoged & Maddieson 1996, Stone & Lundberg 1996). The findings that both the coronal and the dorsal gestures of /l/ show a consistent constriction location across phonetic contexts is consistent with the hypothesis that simultaneous production of these gestures is the goal of lateral production (Giles & Moll 1975, Gick et al. 2002, Proctor & Walker 2012). However, a relatively large lag was found between the coronal and dorsal gestures of /l/ in both onset and coda position: in onset position, the tongue dorsum target followed the tongue tip target by 90 ms, and in coda position it preceded the tongue tip gesture by 150 ms (Proctor et al. 2019). A larger lag between the lingual gestures raises questions whether *simultaneous* production of the coronal closure and tongue dorsum lowering and retraction is the primary goal of production.

An alternative proposal states that the primary goal of production is the lateral channel formation (Sproat & Fujimura 1993, Ying et al. 2017). The lateral channel production is subject to variation (Narayanan et al. 1997). Sproat & Fujimura (1993) argued that lateral channels are produced with tongue narrowing which results in the backing of the tongue dorsum: as the tongue is incompressible, narrowing of the tongue blade must result in tongue elongation and in the displacement of the tongue dorsum. Browman & Goldstein (1989) noted that to capture laterals, a narrow constriction shape must be specified for the tongue body gesture, but did not mention that tongue narrowing necessarily results in elongation. Ying et al. (2017)

showed that the lateral channel can be formed by the lowering of the sides of the tongue. The finding that lateral channel formation is stable across onset and coda position and across vowel contexts is consistent with the lateral channel being the primary goal of production (Ying et al. 2017). This generalisation is, however, not consistent with reduction of lateralisation in dark [ɫ], reported for New Zealand English (Strycharczuk et al. 2018).

Our findings on the coarticulatory influence of coda /l/ on the preceding vowel (**Chapter 2**) and on spatial reduction and potential temporal delay of the coronal gesture (**Chapter 6**) are consistent with a more loose coordination between the coronal and dorsal gestures of laterals (Proctor et al. 2019). The acoustic influence of coda /l/ on the preceding vowel (**Chapter 2**) is consistent with the small articulatory lag between the achievement of the vowel gesture and the dorsal gesture of the laterals (Proctor et al. 2019). The findings that the coronal constriction is reduced and potentially delayed (**Chapter 6**) are consistent with the larger lag between the coronal and the dorsal gesture in lateral (Proctor et al. 2019). Our findings are less consistent with the simultaneous production of the coronal closure and tongue dorsum lowering and retraction being the primary goal of production.

Observing tongue elongation is consistent with both the simultaneous coronal promotion and tongue dorsum lowering and retraction being the primary goal of articulation (Browman & Goldstein 1992, Ladefoged & Maddieson 1996) and with lateral channel formation produced with tongue narrowing being the primary goal of articulation (Sproat & Fujimura 1993). Our preliminary results indicate tongue elongation in both onset and coda /l/ in the speech of two speakers (**Chapter 5**), and tongue elongation was regularly observed in several tokens during data analysis for **Chapter 6**. However, we did not succeed in quantitatively capturing tongue elongation in a larger dataset because of difficulties in locating maximum tongue elongation associated with /l/.

The lateral approximant is known for showing great interspeaker variability (Narayanan et al. 1997). Future studies analysing gestural coordination and timing between the tongue dorsum and the coronal gesture as well as the dynamics of lateral channel formation would provide new insight on the goals of lateral production.

9.4 Implications for sound change

Speakers produced reduced vowel contrast between /iː-ɪ, ʉː-ʊ, æɔ-æ, əʉ-ɔ/ before a coda lateral (**Chapter 2**), and listeners confuse the members of these pairs in perception (**Chapter 3**). The observed reduced contrast between phonologically contrastive vowels in production and confusion in perception are consistent with a potential ongoing vowel merger. Speakers' production, and to a lesser extent, listeners' perception are both consistent with the loss of the coronal gesture of /l/ and /l/ vocalisation (**Chapters 6–7**).

Vowel-lateral coarticulation and coda /l/ lenition both introduce synchronic variation to speech, which can serve as the prerequisite of sound change, as recurrent sound changes are drawn from a pool of synchronic variation (Ohala 1989). However, not all variation becomes sound change, and the role of listener is crucial in the attenuation of sound change (Ohala 1993, Blevins 2006). Listeners' confusion between acoustically similar percepts in itself can initiate sound change (Blevins 2006); so can the lack of compensation for coarticulation (Ohala 1981). Our results on the confusion of prelateral vowels (**Chapter 3**) can be accounted for both by proposing that listeners confuse similar percepts and by proposing that listeners do not compensate for the effect of /l/. Perceptual confusion of key vowel pairs, such as *pool* and *pull* (shown in **Chapter 3**) is consistent with failure to differentiate vowels due to their acoustic similarity (shown in **Chapter 2**). Confusion of *pool* and *pull* is also consistent with a lack of compensation for the coarticulatory influence of /l/: listeners take a realisation of *pool* with a low $F2$ in the vowel at face value and map it to *pull* without compensating for the backing and lowering influence of /l/.

Coronal constriction reduction in coda /l/ is prevalent in production, but is not perceived as /l/-vocalisation. Perception of /l/ produced with a decreased constriction as a consonant prevents sound change for two reasons. Firstly, it indicates that listeners normalise variation in /l/ production (Ohala 1981). Listeners' normalisation is not consistent with the hypothesis that /l/-vocalisation is a perceptual-driven change, during which the acoustic effects of the coronal closure are masked, resulting in listener-speakers weakening the coronal gesture in their own speech (Tollfree 1999). Secondly, the percept of coda /l/ allows listeners to compensate for its coarticulatory influence on the preceding vowel (Gaskell & Marslen-Wilson 1998). A listener who perceives a lenited coda /l/ as consonantal can attribute the coarticulatory influence of /l/ on the preceding vowel to /l/ (Beddor 2009). For instance, a listener who perceives the /l/ in *pool* can attribute the low $F2$ frequencies to /l/ and map the word as it was intended by compensating for coarticulation. A listener who does not perceive /l/ in *pool* must attribute the low $F2$ frequencies to the vowel and map it to *pull*.

The observed acoustic, articulatory, and perceptual patterns create the necessary prerequisites for a sound change and are consistent with the change in prelateral vowels being more advanced than the change in the realisation of /l/. Therefore the results presented in this dissertation raise the question of whether there are ongoing sound changes in lateral-final rimes in AusE. A diachronic study is one method to answer this question: younger and older listener-speakers must be compared both in production and in perception. It is important to determine whether younger listener-speakers produce smaller contrast in prelateral vowels and more lenited coda laterals compared to older listener-speakers. It is also important to examine whether younger listener-speakers compensate less for the coarticulatory influence of coda /l/ on prelateral vowels and whether they perceive a lenited /l/ as vocalised more often than older listener-speakers.

Another method to answer whether there are ongoing sound changes in lateral-final rimes in AusE is examining whether the processes described in this dissertation are gradient or categorical and whether they apply at the domain of the phrase, the word or the stem (Bermúdez-Otero & Trousdale 2012, Turton 2017). During diachronic sound change, phonetically driven gradient rules enter the grammar, and stabilise as categorical rules first at the phrase level, then at the word level, and finally at the stem-level (Bermúdez-Otero & Trousdale 2012, Turton 2017). This process is called domain narrowing: a more narrow domain indicates a more advanced sound change with stem-level being the narrowest, followed by word-level, and phrase-level being the least narrow (Bermúdez-Otero & Trousdale 2012, Turton 2017). Categorical rules do not replace gradient rules, instead they can both coexist in the same grammar. Therefore the presence of categorical effects and the domain of rule application can indicate how advanced a given sound change is. For instance, /l/-darkening in Manchester English only shows a gradient darkening conditioned by length in word-final /l/, indicating an early stage, whereas /l/-darkening in RP shows both categorical differences between onset and coda /l/ and a gradient effect of duration, indicating a later stage in sound change (Turton 2017). In RP, /l/-darkening appears to be a word-level process, as word-final /l/ is always dark, even when it is followed by a vowel, as in *heal # V* (Turton 2017); in contrast, Essex English, /l/-darkening is a stem-level rule, as /l/ is dark stem-finally, e.g. in *peeling* (Turton 2014). The narrower domain of /l/-darkening in Essex English than in RP is consistent with /l/-darkening being at a more advanced state in Essex English than in RP (Turton 2014).

The same approach can be applied to the effect of coda /l/ on the preceding vowel in future studies. The domain of the influence of coda lateral on the preceding vowel need to be examined: the effect of coda /l/ is a phrase-level rule, if for instance /æɔ/ is affected by /l/ in *howl with laughter*, but not in *howl in pain*; it is a word-level rule if /æɔ/ is affected by /l/ in *howled*, but not in *howling*; and it is a stem-level rule if /æɔ/ is affected by /l/ in *howling*. In addition, it also must be examined whether the effect of coda /l/ on prelateral vowels is categorical or forms a gradient scale between *howl with laughter* and *howling*. The presence of categorical effects and a narrow domain of rule application would be consistent with a more advanced sound change. Similarly, the extent of coronal constriction reduction and /l/-vocalisation can be compared at phrase-level (*fool Anna* vs *fool Beth*), word-level (*fooling* vs *fooled*), and stem-level (*fooling*). Both diachronic analyses and synchronic analyses of domain application will shed light on ongoing sound change in AusE.

9.5 Limitations and directions for future research

This dissertation addressed the coarticulatory relationships between /l/ and the preceding vowel, and coda lenition. However, characterising tongue lateralisation and intergestural coordination between the coronal, dorsal, and lateral gestures would provide a more comprehensive account of lateral final rimes. We described the acoustic effect of a coda lateral on the preceding vowel; however, the influence of the dorsal and the lateral gesture on the articulation of the preceding vowel would provide new insight on coarticulation between the vowel-lateral gestures. We described constriction reduction in coda /l/; however, a lenited coda /l/ is characterised by spatial reduction and temporal delay of the coronal gesture, and lenition of the lateral gesture (Strycharczuk et al. 2018, Strycharczuk & Scobbie 2019). If the coronal closure is delayed, its acoustic effects might be masked by the following gesture (Browman et al. 1990, Tollfree 1999). For instance, if minimum tongue tip aperture in the coronal gesture of /l/ in *bulb* is achieved during the labial closure formation of /b/, acoustic cues to the coronal gesture of /l/ will be masked. Future studies examining temporal delay and spatial reduction of the coronal gesture, lenition of the lateral gesture, and their acoustic effects would contribute new information to our understanding of the articulatory structures that create the percept of vocalised /l/.

This dissertation examined lateral-final rimes using acoustic, articulatory, and perceptual methods. We showed that spectral vowel contrast reduction is accompanied by perceptual vowel contrast reduction (**Chapters** 2-3), but coronal lenition is not accompanied by a vocalised percept (**Chapters** 6-7). These results showed us that acoustics, and articulation, and perception of coda /l/ cannot be studied separately, as the result of one study does not necessarily predict the result of the other. Therefore an articulatory study is needed to test whether the acoustic effect of /l/ on prelateral vowels results from spatial and timing characteristics of the dorsal gesture of /l/. An articulatory-acoustic study is needed to explore the acoustic consequences of coronal lenition, and an acoustic-perceptual study is needed to explore acoustic cues for /l/-vocalisation.

Lastly, throughout this dissertation, we worked with a small and restrictive sample of the AusE-speaking population. We recruited only native speakers of AusE, who were young, female, born in Australia to Australian-born parents, or in the case of the articulatory study (**Chapter 6**) born in New South Wales to New South Wales-born parents. The motivation for this decision was to minimise dialectal and sociolectal variation in the data and achieve comparable results across the experiments to help us gain phonological and phonetic insights on lateral-final rimes. However, AusE is known to exhibit regional variation in lateral-final rimes (e.g. Bernard 1985, Oasa 1989, Bradley 2004, Cox & Palethorpe 2004, Butcher 2006, or see **Section 1.1.3** for a summary). In addition, Australia is multilingual, with 22% of all households in Australia and 38% of households in the Greater Sydney area speaking a language other than English at home (Australian Bureau of Statistics 2016). A multilingual environment can contribute to language change in unique ways as it provides a diverse feature pool to select from (Cheshire et al. 2011).

9.6 Conclusion

This thesis provided a systematic and empirical examination of lateral-final rimes, their production, their perception, and their implications for sound change in AusE. The results show that vowel-lateral coarticulation reduces spectral and perceptual vowel contrast. The findings show that the lenition of the coronal gesture of /l/ is prevalent in AusE and is affected by phonetic context. However, coronal lenition does not necessarily correspond to a vocalised percept. These studies suggest that /l/-vocalisation is a more complex and multi-faceted phenomenon than has previously been considered, and one which cannot be reduced simply to either tongue tip lenition, or perception thereof. The observed acoustic, articulatory, and perceptual patterns create the necessary prerequisites for a sound change.

9.7 References

Ash, S. (1982). *The vocalization of /l/ in Philadelphia.* Unpublished doctoral dissertation.

Australian Bureau of Statistics. (2016). *2016 census quickstats.* Retrieved from `https://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/1GSYD?opendocument`

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 785–821.

Bennett, D. C. (1968). Spectral form and duration as cues in the recognition of English and German vowels. *Language and Speech*, *11*(2), 65–85.

Bermúdez-Otero, R. & Trousdale, G. (2012). Cycles and continua: on unidirectionality and gradualness in language change. In T. Nevalainen & E. C. Traugott (Eds.), *Handbook on the history of English: rethinking and extending approaches and methods* (p. 691–720). Oxford: Oxford University Press.

Bernard, J. (1985). Some local effects of post-vocalic (l). In J. E. Clark (Ed.), *The cultivated Australian: Festschrift in honour of Arthur Delbridge* (p. 319–332). Hamburg: Buske.

Blevins, J. (2006). A theoretical synopsis of evolutionary phonology. *Theoretical Linguistics*(32), 117–165.

Borowsky, T. (2001). The vocalisation of dark *l* in Australian English. In D. Blair & P. Collins (Eds.), *English in Australia* (p. 69–87). Philadelphia, Amsterdam: John Benjamins Publishing Company.

Bradley, D. (2004). Regional characteristics of Australian English: Phonology. In K. Burridge & B. Kortmann (Eds.), *Varieties of English: The Pacific and Australasia* (p. 645–655). Berlin: Mouton de Gruyter.

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

Browman, C. P., Goldstein, L. et al. (1990). Tiers in articulatory phonology, with some implications for casual speech. *Papers in laboratory phonology I: Between the grammar and physics of speech*, 341–376.

Browman, C. P. & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook*, *3*(1), 219–252.

Browman, C. P. & Goldstein, L. M. (1992). Articulatory phonology: An overview. *Haskins Laboratories Status Report on Speech Research*, *111/112*, 23–42.

Butcher, A. (2006). Formant frequencies of /hVd/ vowels in the speech of South Australian females. In P. Warren & C. I. Watson (Eds.), *Proceedings of the 11th Australian international conference on speech science and technology* (p. 449–453). Auckland, New Zealand.

Cheshire, J., Kerswill, P., Fox, S. & Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of multicultural London English. *Journal of Sociolinguistics*, *15*(2), 151–196.

Cox, F. & Palethorpe, S. (2004). The border effect: Vowel differences across the NSW-Victorian border. In C. Moskovsky (Ed.), *Proceedings of the conference of Australian Linguistics Society.* Newcastle, Australia.

Cox, F. & Palethorpe, S. (2007). Australian English. *Journal of the International Phonetic Association*, *37*(3), 341–350.

Fant, G. (1960). Acoustic theory of speech production. *The Hague: Mouton.*

Gaskell, M. G. & Marslen-Wilson, W. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 380–396.

Gick, B., Kang, M. A. & Whalen, D. H. (2002). MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics*, *30*(3), 357–371.

Gick, B. & Wilson, I. (2006). Excrescent schwa and vowel-laxing: Corss-linguistic responses to conflicting articulatory targets. In L. M. Goldstein, D. H. Whalen & C. T. Best (Eds.), *Papers in laboratory phonology* (Vol. 8, p. 635–659). Berlin: Mouton de Gruyter.

Giles, S. & Moll, K. (1975). Cinefluorographic study of selected allophones of English /l/. *Phonetica*, *31*(3-4), 206–227.

Hall-Lew, L. & Fix, S. (2012). Perceptual coding reliability of (l)-vocalization in casual speech data. *Lingua*, *122*(7), 794–809.

Hardcastle, W. & Barry, W. (1989, 7). Articulatory and perceptual factors in /l/ vocalisations in English. *Journal of the International Phonetic Association*, *15*, 3–17.

Horvath, B. M. & Horvath, R. J. (1997). The geolinguistics of a sound change in progress: /l/ vocalization in Australia. *U. Penn Working Papers in Linguistics*, *4*(1), 109–124.

Horvath, B. M. & Horvath, R. J. (2001). A multilocality study of a sound change in progress: The case of /l/ vocalization in New Zealand and Australian English. *Language Variation and Change*, *13*, 37–57.

Horvath, B. M. & Horvath, R. J. (2002). The geolinguistics of /l/ vocalisation in Australia and New Zealand. *Journal of Sociolinguistics*, *6/3*, 319–346.

Kleber, F., Harrington, J. & Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, *55*(3), 383–405.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford, UK; Cambridge, Mass.: Blackwell.

Lin, S., Beddor, P. & Coetzee, A. (2014). Gestural reduction, lexical frequency, and sound change: a study of post-vocalic /l/. *Laboratory Phonology*, *5*(1), 9–36.

Lin, S., Palethorpe, S. & Cox, F. (2012). An ultrasound exploration of Australian English /CVl/ words. In *14th Australasian International Conference on Speech Science and Technology* (p. 105–108). Sydney.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, *28*(5), 407–412.

Mann, V. A. & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*(3), 213–228.

Narayanan, S. S., Alwan, A. A. & Haker, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. part I. The laterals. *The Journal of the Acoustical Society of America*, *101*(2), 1064–1077.

Norris, D. & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological review*, *115*(2), 357–395.

Oasa, H. (1989). Phonology of current Adelaide English. In P. Collins & D. Blair (Eds.), *Australian English: The language of a new society* (p. 271–287). St Lucia, Queensland: University of Queensland Press.

Ohala, J. J. (1981, May). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (Eds.), *Papers from the parasession on language and behaviour.* Chicago.

Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. In L. E. Breivik & E. H. Jahr (Eds.), *Language change: Contributions to the study of its causes* (p. 173–198). Mouton de Gruyter.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (p. 237–278). London: Longman.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M. et al. (2004). Cross-subject correlations between measures of vowel production and perception. *Journal of the Acoustical Society of America*, *116*(4), 2338–2344.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N. et al. (2004). The distinctness of speakers' /s/—/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of speech, language, and hearing research*, *47*(6), 1259–1269.

Proctor, M. & Walker, R. (2012). Articulatory bases of sonority in English liquids. In S. Parker (Ed.), *The sonority controversy* (p. 289–316). Berlin, New York: Mouton de Gruyter.

Proctor, M., Walker, R., Smith, C., Szalay, T., Narayanan, S. & Goldstein, L. (2019). Articulatory characterization of English liquid-final rimes. *Journal of Phonetics*, *77*, 100921.

Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*(3), 291–311.

Stone, M. & Lundberg, A. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *The Journal of the Acoustical Society of America*, *99*(6), 3728–3737.

Strycharczuk, P., Derrick, D. & Shaw, J. (2018). How vocalic is vocalised /l/? Evidence from lateralisation. In *Conference of the association for laboratory phonology.* Lisbon, Portugal.

Strycharczuk, P. & Scobbie, J. M. (2019). Gestural delay and gestural reduction. Articulatory variation in /l/-vocalisation in Southern British English. In A. Prezwozny, C. Viollain & S. Navarro (Eds.), *Advances in the study of contemporary spoken English: Multifocal analyses of variation.* Edinburgh: Edinburgh University Press.

Stuart-Smith, J., Timmins, C. & Tweedie, F. (2006). conservation and innovation in a traditional dialect: L-vocalization in Glaswegian. *English World-Wide*, *27*(1), 71–87.

Szalay, T., Benders, T., Cox, F. & Proctor, M. (2016). Disambiguation of Australian English vowels. In C. Carignan & M. D. Tyler (Eds.), *Proceedings of the 16th australasian international conference on speech science and technology* (p. 73–76). Sydney: Australasian Speech Science and Technology Association.

Tollfree, L. (1999). South east London English: discrete versus continuous modelling of consonantal reduction. In P. Foulkes & G. J. Docherty (Eds.), *Urban voices: Accent studies in the British Isles* (p. 163–184). Arnold.

Turton, D. (2014). *Variation in English /l/: Synchronic reflections of the life cycle of phonological processes.* Unpublished doctoral dissertation, University of Manchester.

Turton, D. (2017). Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalization in varieties of English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1).

Wade, L. (2017). The role of duration in the perception of vowel merger. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1).

Wells, J. C. (1982). *Accents of English.* Cambridge: Cambridge University Press.

Ying, J., Carignan, C., Shaw, J. A., Proctor, M., Derrick, D. & Best, C. (2017). Temporal dynamics of lateral channel formation in /l/: 3D EMA data from Australian English. In *Proceedings of the 18th annual conference of the international speech communication association.* Stockholm, Sweden.

Ying, J., Shaw, J., Kroos, C. & Best, T. C. (2012). Relations between the acoustic and articulatory measurements of /l/. In F. Cox et al. (Eds.), *14th australasian international conference on speech science and technology* (Vol. 14). Macquarie University.

10 Appendix

10.1 Ethics approval

Appendix (Ethics Approval) of this thesis has been removed as it may contain sensitive/confidential content