The disambiguation of Australian English vowels in lateral-final syllables

Tünde Orsolya Szalay

MA, Eötvös Loránd University, 2015

Department of Linguistics

Faculty of Human Sciences

Macquarie University

A thesis submitted in fulfilment of the requirements for the degree

Master of Research (Linguistics)

10th October, 2016

Contents

Abstract

Discriminability of Australian English vowel pairs followed by coda /d/ or coda /l/ was examined using a lexical decision task. 30 native listeners categorized 10,800 Australian English words either with /hVd/ or with /hVl/structure in a binary forced choice task. Both in the coda /d/ and the coda /l/ condition, 16 target words differing only in the nuclear vowel were presented as auditory stimuli, paired with each of the remaining 15 words as competitors. Accuracy and reaction time were measured. Results of the /d/ condition show that vowels are intrinsically similar in perception when vowels of the target and competitor pair share spectral similarity but differ in length. Comparison of the coda /d/ and /l/ condition shows that coda /l/ makes vowel disambiguation harder across the board and it also makes the disambiguation of intrinsically similar target-competitor pairs even harder. The results also show that coda /l/ might reduce the contrast between /ʉː-ʊ/, /aeɔ-ae/, and /əʉ-ɔ/ pairs to such an extent that contextually conditioned mergers may occur in perception. The potential context-based mergers suggest that the rimes containing /l/ might only allow a subset of Australian English vowels to appear in the nucleus, therefore the relationship between a nucleus and coda /l/ might be stronger than the relationship between a nucleus and a coda /d/.

Declaration

Declaration

I hereby declare that this thesis has not been submitted for a higher degree to any other university or institution. I have made every effort to clearly indicate the sources of information used and acknowledge the extent to which the work of others has been used in the text. The research presented in this thesis has been approved by the Macquarie University Faculty of Human Sciences Research Ethics Sub-Committee (ref: 5201600061).

Tünde Orsolya Szalay

10. 10. 2016

Acknowledgements

There are several people to whom I would like to express my gratitude. First of all, this thesis could not have been written without my supervisors: Michael Proctor, Felicity Cox, and Titia Benders, all of whom provided immense help during all stages of the process. I would also like to thank the Phonetics Lab of the Linguistics Department for their insight on the project. I am grateful for the Writing Group, and Lauren Brooks, the writing mentor of the Linguistics Department, who provided feedback and comments on various draft versions of this thesis. I would also like to thank the participants of the experiment that forms the main part of this thesis.

I would like to thank the developers of the statistical analysis software R and the typesetting system LaTeX who keep on developing amazing software for free. I am especially grateful for the two being compatible with each other.

Finally, this thesis was funded by the ARC grant DE150100318 and the International Research Training Pathway Scholarship from Macquarie University.

1 Introduction

In syllables ending in /l/, /l/ interacts with the nucleus in several ways in many varieties of English, including Australian English (AusE). This interaction has been described in articulation, for example in General American English the articulatory gestures of coda /l/ overlap with the articulatory gestures of the vowel that precedes it (Proctor & Walker 2012), and coda /l/ can also lead to a schwa insertion between the nucleus and the coda (Gick & Wilson 2003). Similar schwa-insertion has been observed in British English (Krämer 2008, Wells 1982). Coda /l/ was also observed to lower and retract the vowel it follows (Altendorf & Watt 2008, Cruttenden 2001, Upton 2008, Wells 1982). This lowering and retracting effect was studied in the acoustics of AusE (Cox & Palethorpe 2004, Palethorpe & Cox 2003).

As it is coda /l/ that interacts with its nucleus, syllable-based explanations have been proposed to account for this phenomenon. Syllable-based explanations usually raise the question if postvocalic and pre-consonantal or word-final /l/ may be analysed as a part of the nucleus and not as a coda (Proctor & Walker 2012) or if it may be moraic (Lavoie & Cohn 1999). Other studies, although they did not propose alternative syllabic representations, also questioned the coda status of postvocalic and pre-consonantal or word-final /l/ as a coda (e.g. Campbell et al. 2010, Sproat & Fujimura

1993), showing that the position of coda /l/[1] is ambiguous and unstable in the syllable.

Despite the wealth of research on coda /l/, little is known about how the interaction of coda /l/ and the nucleus is reflected in speech perception. In AusE, only the merger of /e/ and /æ/ in a pre-lateral environment has been studied for perception (e.g. Loakes et al. 2010a; 2014b). Therefore this thesis sets out to provide an empirical answer to the following questions in AusE:

1. Does the interaction between coda /l/ and the nucleus make word-disambiguation more difficult for words ending in coda /l/ than for words ending in an obstruent coda?

2. Are there such vowel-pairs which are affected to a greater extent by coda /l/ in perception than other vowel pairs?

The thesis is organised into six chapters. Chapter 1 briefly outlines the aim of the research and the organisation of the thesis. Chapter 2 reviews the literature on speech perception, AusE, and phonological models of syllable structure to provide background for the perception of AusE lateral final rimes. Chapter 3 presents the design and the procedure of the perception experiment. Chapter 4 begins with the statistical methods and then presents descriptive and inferential statistics. The results are discussed in Chapter 5, in the light of speech perception and syllable structure. Chapter 5 also marks the limitations of this thesis and provides ideas for future research. Lastly, Chapter 6 presents the conclusion.

[1] In this thesis, we use "coda /l/" to refer to /l/ that is postvocalic and either word-final or pre-consonantal; however, we do not wish to exclude the possibility of alternative representations of /l/ in the syllable structure.

2 Literature review

This chapter provides an overview of the literature on word- and vowel perception in general, on the perceptual challenges of the AusE vowel space, and on models of syllable to give a necessary background for the perception of AusE lateral final rimes. Firstly, the importance of speech perception is addressed the importance of speech perception by briefly presenting two models of perception and experimental studies on word and vowel perception. The AusE vowel space might present special challenges to the listener, therefore the AusE vowel space is presented in terms of its phonological organisation and acoustic characteristics. AusE vowels are further affected by coda /l/, as coda /l/ might lead to context-based vowel mergers (Palethorpe & Cox 2003), which may add extra difficulties to vowel disambiguation. As the effect of coda /l/ has been motivated by its position in the syllable structure (Palethorpe & Cox 2003), a brief overview of the concepts of syllable theory is presented, focusing on the position of /l/ in the syllable to propose possible explanations for why coda /l/ influences AusE vowels.

2.1 Factors influencing word and vowel perception

The perception of the speech signal, including the perception of individual words and vowels, is accounted for in two main theoretical paradigms (Goldinger 1996b). The first is the abstractionist paradigm in which the idiosyncratic features of the speech signal, such as dialectal and social variation, and variation caused by phonetic context are removed by the listeners. As a result, the signal is normalised to an abstract lexical representation that contains nothing but the distinctive features of that word or vowel (e.g. Lahiri & Marslen-Wilson 1991). In contrast with the abstractionist view, the episodic view holds that the idiosyncratic features are preserved in the mental lexicon of the speaker-listener. As a result, in the exemplar representation of a word, e.g. *bat* /bæt/ consists of several items in the lexicon, e.g. /ˈbæt/ and /ˈbæʔ/ (Docherty & Foulkes 2014, Goldinger 1996b). Hybrid representations are also postulated in which idiosyncratic features are retained in the lexicon, but speaker-listeners still arrive to abstract representations (Goldinger 2007, Pierrehumbert 2016). As both high level, abstract phonological units, such as the syllable, and low-level changes in the phonetic signal may be part of the mental representation of words, both might need to be considered in word- and vowel recognition and disambiguation.

2.1.1 Word recognition

A large body of research has examined word recognition, and the many factors that are involved. This section reviews the role of three major factors influencing the accuracy and speed of word recognition: lexical frequency, which refers to how frequently a word occurs in a given language (Meunier & Segui 1999); the temporal nature of the acoustic signal, which means that listeners have a different amount of information at different points in time; and subphonemic variation, which refers to the non-contrastive variation between different tokens of a single phoneme.

Several studies have shown that lexical frequency facilitates the accuracy and speed of word recognition by showing that words with high word frequency are recognised more quickly and accurately than less frequent words (Forster & Chambers 1973, Meunier & Segui 1999, Rubenstein et al. 1970, Segui et al. 1982). Besides lexical frequency, individual listeners' familiarity with a word also facilitates word recognition, as listeners will be more accurate and quick in recognising words with which they are familiar (Connine et al. 1990).

When listeners hear a word, be it frequent or infrequent, familiar or unfamiliar, they may identify the word differently according to the amount of information that is available at a given time. This is shown by the gating paradigm (Grosjean 1980) and the visual world paradigm (Allopenna et al. 1998, Grosjean 1980, McQueen & Viebahn 2007). In Grosjean (1980)'s gating experiment, the listeners were first presented with the first 30 ms of the word (the first gate), then the first 60 ms of the word (second gate), and so on, until the whole word was presented. Listeners were asked to identify the word after each gate. These data provide insight into the point in time at which listeners could make an accurate decision, and which part of the word carried the decisive information.

Data from visual world paradigm experiments also provide evidence that the amount of available information affects word disambiguation. In this paradigm, the listeners eye-movement is recorded while they are presented with the audio stimulus of the target word and with a visual stimulus of a cohort consisting of the target and plausible competitors to gain information on language processing (Huettig et al. 2011). Eye-movement is used in the visual world paradigm, because the visual world paradigm adopts the linking-hypothesis that suggests eye-movement and language processing are systematically linked (Tanenhaus et al. 2000) because the audio stimulus draws listeners attention to the visual stimulus, and consequently their gaze as well (Huettig et al. 2011). As the listeners hear the target word, the target word and phonetically similar words are activated in the lexicon, therefore listeners gaze at the target and at competitors similar to the target, and as more acoustic information becomes available, the listeners exclude the competitors as a possible match for the audio stimulus and fixate upon the target word (Allopenna et al. 1998, Tanenhaus et al. 1995; 2000). Therefore both the gating paradigm and the visual world paradigm provide information on how listeners identify a target word presented aurally, but the visual world paradigm provides continuous temporal information. Both paradigms use the cohort model (first developed by Marslen-Wilson & Welsh (1978), for an overview see Cutler (2012)), which proposes that when listeners hear a target word, they do not only activate the mental representation of the target but of similar words too.

When presented with a word, or part of a word, listeners perceive and rely on subphonemic details to facilitate the recognition of words and non-words, as lexical access is sensitive to within-category gradient variation of phonetic factors. For example, voice-onset time, a major cue to the voiced-voiceless contrast, is perceived by listeners both across phonetic categories (voiced-voiceless) and within (voiced-voiced, voiceless-voiceless) (Pisoni & Tash 1974). Also, the bigger the voicing onset time difference is, the quicker listeners can identify consonants as belonging to different categories (Pisoni & Tash 1974). Listeners can also perceive details such as coarticulation. For example, listeners can distinguish a C$\tilde{\text{V}}$NC sequence from a CVC sequence on the basis of the nasalisation of the vowel before hearing the nasal consonant, and the less prominent the presence of the nasal, the more listeners rely on the nasalised vowel (Beddor et al. 2013).

Listeners' reliance on coarticulation is shown through the negative impact of mismatching coarticulatory information on perception. Listeners are slow to identify a CV sequence when the coarticulatory information on the consonant does not match the following vowel (Whalen 1991). Listeners are also slow to recognise a (C)V_1CV_2 sequence if the coarticulatory information on V_1 does not match V_2 (Martin & Bunnell 1981, Fowler 2005).

2.1.2 Vowel perception with and without context

Listeners are not only able to perceive gradient phonetic variation when they access complete words or non-words, but also when they are required to identify vowels whose formants have been synthetically modified (Kewley-Port & Watson 1994, Mermelstein 1978). That is, formant-discrimination studies have shown that listeners can discriminate between vowels which belong to the same phonemic category but have different formant structures. This is supported by studies in which vowel identification was tested with isolated vowels and with context, and under ideal and ordinary listening conditions. Under ideal conditions, when the uncertainty about what the stimulus is had been minimised, the threshold for a noticeable change in formant of an isolated vowel can be as low as a 12–36 Hz (Kewley-Port & Watson 1994) or 33–92 Hz (Mermelstein 1978). When the vowel was presented in a phonemic context, in the CVC environment, the threshold of noticeable change might become higher (Mermelstein 1978) or stay the same unless the consonant is /m/ or /l/ (Kewley-Port 1995). The consonants /m/ and /l/ degrade the threshold of formant discrimination because they shorten the steady-state part of the vowel (Kewley-Port 1995).

In addition to displaying sensitivity to the steady-state phase of the vowel, listeners can also notice inherent spectral changes in vowels (Nearey

& Assmann 1986). Vowel inherent spectral change refers to the change in the vowel formant structure that is not the result of consonantal context and it characterizes monophthongs, as well as diphthongs (Nearey & Assmann 1986). By manipulating the vowel inherent spectral changes of monophthongs, temporal information and vowel length were shown to be important in the perception of Canadian English vowels, even though it is not contrastive (Nearey & Assmann 1986). That is, listeners rely on the unfolding of the acoustic signal in time when disambiguating vowels.

Collectively, these data suggest that when listeners disambiguate words and their sounds, they rely on several levels of information: lexical, temporal, phonemic, and subphonemic to access their mental representation of words. As the lexicon, and the temporal, phonemic, and subphonemic systems of languages differ, this information might be used differently by the listeners of different languages.

2.2 The Australian English vowel space

While much is known about mechanisms of lexical disambiguation in American English, fewer studies have examined the details of these processes in AusE. Most of this work has focused on perception at the level of the word and the segment (Taft 1986). AusE presents special challenges to the listener because it uses a vowel inventory containing 18 stressed vowels (plus schwa), as presented in Figure 2.1 and 2.2. The AusE vowel inventory can be considered large, because an average-sized vowel inventory contains 5-6 vowels (Maddieson 2013). AusE also incorporates phonemic vowel length contrast for certain spectrally similar pairs. For example, pairs of vowels such as /ɐ-ɐː, e-eː/, and /ɪ-iː/ have almost identical spectral quality but contrast in length, although the onglide of /iː/ leads to spectral differences as well

(Cox et al. 2014). AusE also has short monophthong-diphthong pairs such as /æ-æɪ, ɔ-ɑe/ and /æ-æɔ/ in which the monophthong is related to the first element of the diphthong, and /ʉː-əʉ, ɔ-æɔ/, in which the monophthong is related to the second element of the diphthong (Cox 1999). This differs from General American English whose vowel inventory consists of 15 vowels and does not use length as a contrastive feature (Labov et al. 2008). As listeners rely on the unfolding of the acoustic signal in time when disambiguating vowels, a vowel space that uses temporal cues to differentiate between spectrally similar vowels makes intrinsic vowel similarities especially important in vowel disambiguation.

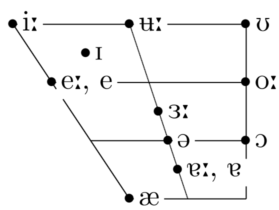

Figure 2.1: **AusE short and long monophthongs in the vowel map** Vowels are placed on the map according to their place of articulation. (Cox 2012)

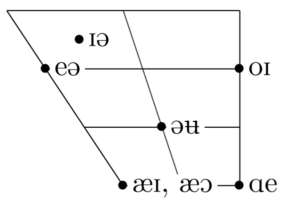

Figure 2.2: **Schematic representation of AusE diphthongs in the vowel map** Diphthongs are placed according to the place of articulation of their first elements. Figure based on Cox (2012)

In addition to the inherently complex vowel space of AusE, vowels followed by a coda /l/ are modified in their acoustic qualities. An acoustic study on AusE spoken in NSW showed that the formant structures of AusE vowels are substantially affected by a final /l/ because monophthongs are centralised, especially /iː, e, ʉː/ and /ɜː/, and the second element of front rising diphthongs is reduced (Cox & Palethorpe 2004, Palethorpe & Cox 2003). Back vowels are affected to a much smaller extent than front vowels (Cox & Palethorpe 2004).

The changes in the formant structure caused by final /l/ may lead to reduced vowel contrast and context-based vowel mergers or near-mergers in the pre-lateral environment. To date, the near merger of the vowel pairs /ʊ-ʉː, æ-æɔ/ and /ɔ-əʊ/ has been observed in acoustic studies. The vowels /ʉː/ and /ʊ/ are nearly merged before a coda /l/, as the first formant of the vowels coincide, but the second formant of /ʉː/ remains higher than the second formant of /ʊ/ (Cox & Palethorpe 2004, Palethorpe & Cox 2003). The vowels /æ-æɔ/ also undergo a near-merger, as /l/ is absorbed in the second element of the diphthong, but the first element shows acoustic differences in its F1 from the monophthong (Palethorpe & Cox 2003). The vowel /əʊ/ is completely merged acoustically with /ɔ/ due to the lowering of its F2 (Palethorpe & Cox 2003).

Contrary to the near mergers observed in the acoustics of pre-lateral vowels, an articulatory study by Lin et al. (2012) did not find a difference in the tongue movement of pre-lateral and pre-obstruent vowels corresponding to the patterns observed in acoustics. However the lack of result might be caused by that the articulatory study was restricted to a single speaker.

The change in the acoustics of pre-lateral vowels might reflect a sound change, namely towards the contextual merger of /ʊ-ʉː, æ-æɔ/ and /ɔ-əʊ/. To fully identify if it is a sound change in progress, and if it is, what motivates this change in AusE, perceptual studies are crucial (Blevins 2006b, Ohala 1981). To date, only the /e-æ/ merger in the Melbourne dialect of AusE has been studied perceptually (Loakes et al. 2014a;b; 2012; 2011;

2010a;b;c). Loakes and her colleagues used word identification in which the target words were minimal pairs differing in /el/ and /æl/, and their results show that /e/ and /æ/ are undergoing a perceptual merger before /l/ in the Melbourne dialect of AusE. However, the /el -æl/ merger of Melbourne has been studied in perception (e.g. Loakes et al. 2010a; 2014b) and acoustics (Cox & Palethorpe 2004), but not in articulation.

In summary, more research is needed on the effect of coda /l/ on vowels in AusE, as some near-mergers have been studied in acoustics, some in articulation and some in perception. Listeners' perception is important because an acoustic near-merger may mean a complete merger in perception but it may also mean that listeners perceive the vowels in words such as *pool* and *pull* as different (Docherty & Foulkes 2014). Moreover, little is known about the perception of AusE vowels in general. The majority of the AusE vowel space is unexplored with respect to the mechanisms of vowel disambiguation in general and in the pre-lateral environment in particular. Comparing the perception of vowels followed by coda /d/ and coda /l/ will be important for a better understanding of the intrinsic difficulties of vowel disambiguation and the difficulties caused by /l/ in AusE.

2.3 Coda /l/ and the syllable

As it was described in Section 2.2, coda /l/ might lead to contextual vowel mergers, therefore it might only allow a subset of AusE vowels in the nucleus. This may suggest that coda /l/ is more closely linked to the nucleus than an obstruent coda because the lateral might impose a phonotactic restriction on the nucleus which the obstruent does not. Furthermore, the fact that AusE laterals exhibit an onset-coda allophony between a clear onset [l] and a dark, velarised coda [ɫ] shows that researching lateral-final rimes may also have tentative implications for syllable structure (Borowsky 2001).

To interpret the relationship between coda /l/ and the nucleus, generative and moraic models of syllable are presented out of the several models which have been proposed in different phonological theories (for an overview on the history of syllables in phonology, see Goldsmith 2014). The syllabic affiliation of individual segments (i.e. the position of a segment with respect to the syllable) is relevant for phonological models of the syllable and can be studied using phonetic experiments. Therefore the first part of this section is an overview of the generative and the moraic model of the syllable and then the syllabification of individual segments is discussed. In the second part, former research is presented which indicates that coda /l/ might not fit these models.

2.3.1 Syllables

The syllable is represented both in the generative and the moraic models as a unit comprising a number of segments, presented in Figure 2.3 and 2.4 respectively. In these models the top node is the syllable node to which lower level constituents are linked.

In the generative model, the subsyllabic constituents are onset and rime, and the rime is further divided into nucleus and coda. Each constituent, including the syllable are maximally binary branching. Individual segments are linked to onset, nucleus, and coda directly, but never directly to the higher level nodes. This model accounts for the strong relationship between those segments which are directly dominated by the same low level constituents and the weak relationship between the segments dominated indirectly by high level nodes. For example, there are many phonotactic restrictions on what clusters are possible within the onset node and within the coda node in English. Additionally, if a nucleus dominates more than one elements (i.e. it is a branching nucleus), there are phonotactic restrictions on what segments are possible in a branching nucleus. In contrast, the fact that there are less restrictions between the nucleus and the coda is captured by the structure such that the segments in the nucleus and the coda are not dominated immediately by the same constituent, but only indirectly by the rime. Lastly, there are few phonotactic restrictions between the onset and the rime in English which is captured through that only the top node, which is the syllable node, dominates the segments in both the onset and the rime.

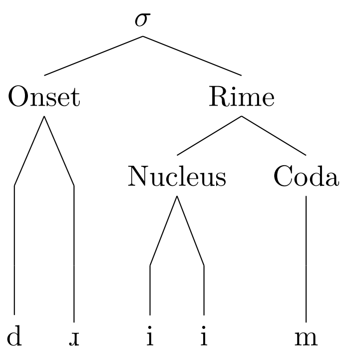

Figure 2.3: **Generative model of syllable.** Top node: syllable. Subsyllabic constituents: onset, rime. Rime consists of nucleus and coda (based on Blevins 1995)

In the moraic model, the syllable node is linked to segments or moras. Moras are units of syllable weight and timing (Ewen & Hulst 2001). Those segments that contribute to syllable weight are linked to a mora, and those that do not are linked directly to the syllable node Zec (2007). The syllable initial consonant (the onset) does not carry a mora in any language, therefore they are always linked to the syllable node, whereas vowels carry a mora in all languages, therefore they are always linked to at least one mora. Language specific rules determine whether syllable-final consonants (codas) carry a mora and the maximal number of moras in a syllable (Broselow et al. 1997, Hayes 1989). In English, vowels are inherently moraic, and coda consonants may carry a mora (Trommelen & Zonneveld 1999) when a moraic coda is necessary to meet the minimum word requirement of a bimoraic word (Lavoie & Cohn 1999).

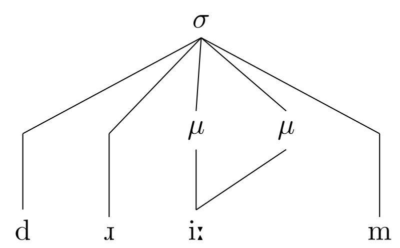

Figure 2.4: **The moraic model of the syllable.** Top node: syllable. Subsyllabic constituents: moras. (based on Zec 2007)

The segments are dominated by syllabic constituents in the generative model, and are linked to syllabic constituents in the moraic model. Therefore both models need to determine the syllabic affiliation of segments by linking each segment unambiguously to one and only one constituent, be that constituent an onset, a nucleus or a coda in the generative model or a mora or a syllable in the moraic model. To this purpose, they use the Sonority Sequencing Principle (SSP) to determine which segments can be linked to the nucleus and the syllabic affiliation of the segments in consonant clusters (Parker 2011) and use onset maximisation (Blevins 2006a) that requires syllables to have onsets.

The SSP states that segments are preferred to have a rising sonority in the onset, a sonority peak in the nucleus, and falling sonority hierarchy in the coda. SSP has two consequences. The first is that it determines which segments can be linked to the nucleus, as languages impose a minimum sonority requirement on what segments are acceptable as a nucleus, for example, English allows /l/ in the nucleus, but not /d/ (Blevins 2006a). The second consequence of the SSP concerns the syllabification of consonant clusters, as onset must have a rising, and codas a falling sonority.

However, syllabifying *Al.bert* as *alb.ert* would satisfy SSP. Additionally, when it comes to the syllabification of a rising sonority cluster, such as /bl/ in *doublet* SSP is satisfied both by *dou.blet* and *doub.let*.

To capture that *Albert* is syllabified as *al.bert* and *doublet* as *dou.blet*, the onset maximisation principle is used. The onset maximisation principle states that all consonants must be in the onset as long as the SSP allows it, because onsets are cross-linguistically preferred over codas (Blevins 2006a). A closer look at language-typology reveals that there is not only a cross-linguistic preference for onsets over codas, but there is also a preference for simple onsets (onsets consisting of a single segment) as opposed to complex onsets (onsets consisting of a cluster) (Zec 2007).

However, the fact remains that the syllabic affiliation of a given segment in a given language is expected to be unambiguously determined by SSP and onset maximisation, and phenomena such as ambisyllabicity and ambiguous syllabic affiliation pose problems for these models.

2.3.2 The syllabic affiliation of word-final /l/

The models presented in 2.3.1 appear to have difficulties when the results of phonetic experiments need to be incorporated, both when it comes to deciding whether an intervocalic /l/ is an onset or a coda and whether a postvocalic /l/ is a coda or a nucleus. Currently postvocalic /l/ is modelled as being in the coda, as it is presented in Figure 2.5. Data from phonetic experiments on word-final /l/ are concurrent with the idea that this segment might be ambisyllabic or its syllabic affiliation might be gradient, and it is ambiguous between forming part of the nucleus, the coda, or the onset of the following vowel-initial word. Incorporating the idea that postvocalic /l/ is part of the nucleus, is problematic for the generative model, and incorporating the idea that /l/ bears a mora is problematic for the moraic model when the nucleus is a long vowel or a diphthong. These problems are illustrated by Figure 2.6. In the generative model, a nucleus with a long monophthong or a diphthong is already a binary branching nucleus, therefore it cannot also accommodate postvocalic /l/ as a constituent of the nucleus, because that would make the nucleus ternary branching. In the moraic model, a long monophthong or a diphthong already carries two moras, therefore it cannot assign a mora to a postvocalic /l/ if English disprefers trimoraic syllables (Borowsky & Horvath 1997).

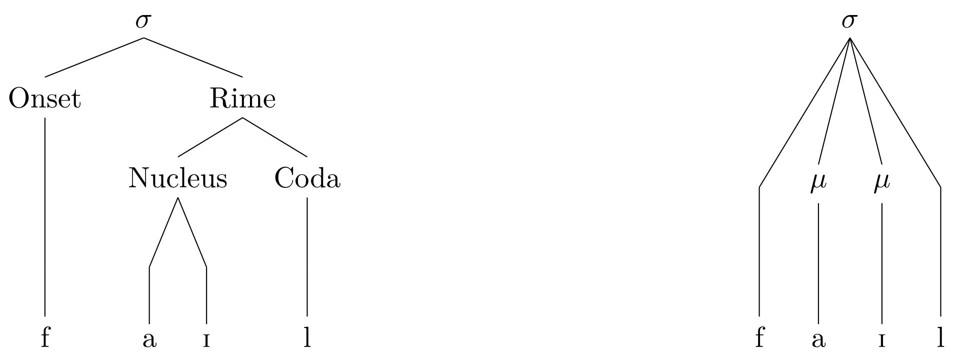

Figure 2.5: **Representations of the word *file* in a generative and a moriac model of the syllable.** Left panel: generative model in which /l/ is linked to the coda. Right panel: moraic model in which /l/ is non-moraic.

Although it is problematic to analyse postvocalic /l/ as part of the nucleus, the presented phonetic studies might be consistent with this idea. Lavoie & Cohn (1999) collected syllable count judgements for obstruent-final and /l/-final CVt and CVl words, and CVCV words. They found that obstruent-final words with a single vowel are judged as monosyllabic, and CVCV words are judged as bisyllabic. In contrast, words with a single vowel

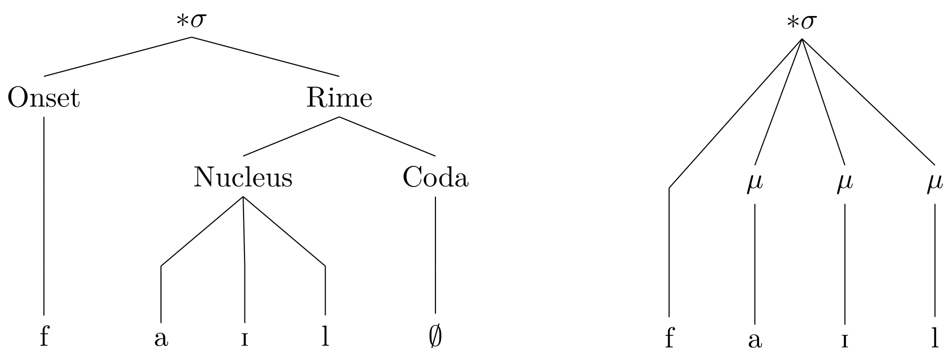

Figure 2.6: **Representations of the word *file* in a generative and a moraic model of the syllable.** Left panel: generative model in which /l/ is linked to the nucleus, violating binary branching. Right panel: moraic model with a mora-bearing /l/, violating the bimoraic constraint.

and a final /l/ are ambiguous because they are judged as having one syllable when the vowel is a short monophthong or two syllables when the vowel is short monophthong or a diphthong. Additionally, Tilsen et al. (2014) showed that /l/-final words judged as having more than one syllable have longer rimes in speech production. The results of Lavoie & Cohn (1999) and Tilsen et al. (2014) may point to that final /l/ forms a nucleus and is syllabic in a postvocalic position as well (not only in a post-consonantal position). In this case, it could be a second adjacent nucleus (i.e. part of a separate syllable) rather than part of the syllable nucleus, but this would lead to an onsetless second syllable, which is also a dispreferred structure (Blevins 2006a).

Further evidence corroborating the analysis of a rime-final /l/ as a nucleus is the close relationship between the /l/ and the preceding vowel. Articulatory, acoustic, and perceptual, studies have shown that rime-final /l/ is related closely to the vowel it follows. For example, Proctor & Walker (2012) found that vowels are coarticulated strongly with a coda liquid in General American.

This strong relationship of vowel and coda /l/ appears in AusE too. The near mergers and mergers triggered by coda /l/ in AusE (described in 2.2) may be related to the idea that a long vowel and a sonorous coda create a trimoraic syllable although English prefers bimoraic syllables (Borowsky & Horvath 1997, Lavoie & Cohn 1999). In order to reduce a trimoraic syllable to a bimoraic one, speakers can either reduce an inherently bimoraic long vowel or a diphthong to make it monomoraic or they can reduce the coda /l/ to remove a mora (Palethorpe & Cox 2003, Cox personal communication).

In contrast, gestural conflict has also been proposed to explain the effect of /l/ on the preceding vowel, by arguing that the articulatory gestures of /l/ are in an inherent conflict with the gestures of the vowel (Gick & Wilson 2003). However, ultrasound study on a single speaker did not find corresponding results in AusE Lin et al. (2012).

In addition, data from articulatory studies also point to a strong relationship between vowels and postvocalic /l/, which can be captured in Articulatory Phonology (Browman & Goldstein 1988) by proposing that syllables are organisations of articulatory units (Krakow 1989). In this account, prevocalic and postvocalic /l/ can be considered separate units (Giles & Moll 1975) that are related to the nucleus by different organisational patterns of gestures (Browman & Goldstein 1988; 1995). Although it is beyond the scope of the dissertation to consider the rich literature on sub-segmental accounts of lateral organization, data from articulatory studies are overviewed, as their results might be concurrent with the idea that coda /l/ cannot always be resyllabified and it might be difficult to determine whether final /l/ is part of the preceding vowel nucleus or the syllable coda. There is articulatory evidence for word final prevocalic /l/ sharing the gestural characteristics of a preconsonantal /l/ (Campbell et al. 2010, Gick & Campbell 2003). That is, a word-final prevocalic and word-final preconsonantal /l/ are articulated with the same gestures, therefore /l/ does not undergo resyllabification. Final /l/ was found to undergo partial resyllabification only, depending on the strength of the morpheme boundary which follows it (Sproat & Fujimura 1993). That is, the syllabic affiliation of /l/ is partially determined by its morphological environment. These results may question that /l/ is in the coda, because coda consonants can undergo resyllabification (Campbell et al. 2010, Gick & Campbell 2003) and may support the analysis that final /l/ is a nucleus. However, the results of phonetic studies do not shed light on whether it forms a branching nucleus with the vowel it follows or a separate one.

Based on the data presented in Section 2.3.1, words ending in /l/ seem to have an ambiguous and unstable syllabic structure. The ambiguity results from two possible structural affiliations. In the first, /l/ is linked to the coda, and in the second, /l/ is linked to a branching nucleus. It is unstable, because a rime containing a long vowel and a coda /l/ might be ternary branching, which is disallowed, or trimoraic, which is dispreferred in English. Thus researching the perception of words with rime-final /l/ may shed more light on the relationship between the vowel and the /l/, and on how /l/ is represented in models of the syllable.

Synthesising what is known about the AusE vowel space, the effect of coda /l/ on AusE vowels, and the ambiguous and unstable nature of coda /l/ raises several questions. If coda /l/ forms a stronger relationship with its nucleus than a coda /d/, /l/ will impose a constraint on the nucleus, allowing only a subset of AusE vowels to appear in pre-lateral environments. This might lead to reduced vowel contrasts affecting the perception and disambiguation of AusE vowels. Therefore, the aim of this thesis is to answer the following questions:

1. Are vowel contrasts reduced in perception in the pre-lateral environment?

2. If vowel contrasts are reduced in perception, which vowels or vowel pairs are affected?

3. How does the possible reduction of AusE vowel contrasts in the pre-lateral environment contribute to our understanding of the affiliation of /l/ in models of syllable?

To answer these questions a perception experiment was designed to compare the disambiguation of pre-lateral and pre-obstruent vowels.

3 Methods

This study is a preliminary investigation into vowel disambiguation and confusability in AusE in pre-obstruent and pre-lateral environments. The motivation for the experiment is that reduced vowel contrasts triggered by /l/ and the relatively large AusE vowel inventory may lead to difficulties in vowel disambiguation and to confusability of vowel pairs (see Section 2.2). The goal of this study is to investigate the perception of AusE vowels in new detail. Vowel discriminability before obstruents and laterals was compared to examine the influence of lateral codas on vowel quality, and to shed more light on the structure of lateral-final rimes.

In this chapter first a brief overview of the experiment design is provided. Secondly, the demographics of the participant pool is described. Lastly, a detailed description of the stimuli and the procedure to motivate the choices made during the experiment design are described.

3.1 Experiment design

The experiment consisted of a vowel disambiguation task using a binary forced-choice word recognition paradigm (Cutler 2012, Goldinger 1996a, Marslen-Wilson 1980, Rubenstein et al. 1970). Participants were presented with a target word as an auditory stimulus, and asked to identify the word by selecting one of two response options presented orthographically on a computer screen. Participants were instructed to select the word they heard as quickly as they could. The experiment had two conditions: the /d/ condition, in which participants were exposed to words ending in coda /d/, and the /l/ condition, in which participants were exposed to words ending in coda /l/. Each participant only participated in a single condition.

The stimulus set consisted of recordings of single word utterances of the form /hVd/ and /hVl/. Both conditions contained 16 target words. The stimulus words were paired exhaustively, leading to $16x15 = 240$ trials. Response options for the audio stimulus were presented orthographically to elicit participants' responses to audition. A trial in the /l/ condition is illustrated in Figure 3.1.

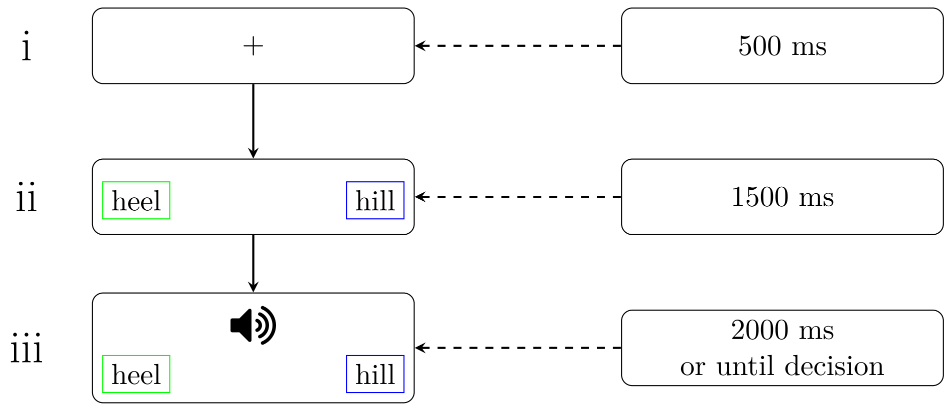

Figure 3.1: **An example of a trial** i. Fixation; ii. Orthographic presentation of response options; iii. Audition task

3.2 Participants

The experiment had 30 participants in total: 15 participants (14 females) carried out lexical decision in the coda /d/ condition and 15 participants (15 females) carried out lexical decision in the coda /l/ condition. All participants were native speakers of Australian English who were born in Australia or immigrated to Australia before the age of 2. Participants were undergraduate students of linguistics at Macquarie University. Participants had training in linguistics, but they were naive to the purpose of the experiment. They received course credit for participation.

Participants' age was between 19 and 47 years (mean: 22.5) in the /d/ condition, and between 19–56 (mean: 25.5) in the /l/ condition. There was one left-handed participant in the /d/ condition and two in the /l/ condition. None of the participants reported any hearing, speaking or reading disorders.

Language background data and information on residential history were collected with a language background questionnaire (see Appendix A). In the coda /d/ condition 7 participants were monolinguals and 8 were bilinguals. In the coda /l/ condition, 4 participants were monolinguals and 11 were bilinguals. Other languages spoken were Arabic, Croatian, Danish, German, French, Italian, Japanese, Korean, Spanish, Swedish, Tagalog, Teochew and Vietnamese. Of the 19 bilinguals 14 were simultaneous bilinguals who learnt their second languages at home from their parents, and 5 were sequential bilinguals whose first language was English and learnt their second language at school. Five of the participants in the /d/ condition and 4 in the /l/ condition were born in Australia to Australian-born parents.

3.3 Stimulus

The stimulus used to test vowel confusability in AusE consisted of two sets of words and non-words. One set ended in /d/, and the other set ended in /l/. The words were presented orthographically and aurally. The stimulus had three components: the word component, the acoustic component and the orthographic component, which are presented in the following sections.

3.3.1 Target words

AusE has 18 stressed vowels (Cox 2012), out of which 16 appear in monosyllabic words before /l/. This is because the vowels /ɪə/ and /eː/ (also pronounced as [eə]) do not appear before /l/. These sounds historically derive from R-influenced vowels, and word-final /ɹl/ clusters were rare in AusE. In the *Global Web-based Corpus of English* only 16 occurrences of orthographic word-final <rl> was found (Davies 2013), 14 of which is pronounced with /ɜː/ and two with /ɐː/.

16 vowels were selected for the target words. The target word always had the hVd or hVl structure in order to keep the stimulus uniform. When a vowel did not yield an existing English word in the h_l or in the h_d environment, the corresponding non-word was used.

Table 3.1 shows the carrier words ending in /l/ and /d/. It also shows their lexical frequency per one million words in the AusE part of the *Global Web-based Corpus of English* (Davies 2013). When the stimulus was a non-word, lexical frequency is given as zero.

3.3.2 Target vowels

The sound files used in the experiment were collected earlier for the Australian Voices project (Cox & Palethorpe 2010). The speaker of the recording is a monolingual female university student born in Australia to Australian-born parents. She was 21 year old at the time of recording.

The stimulus was collected in 2006 in the recording studio in the Department of Linguistics at Macquarie University using an AKG C535 EB microphone, Cooledit audio capture software via M-Audio delta66 soundcard, onto a Pentium 4 PC at 44.1kHz sampling rate. All stimuli were amplitude-

Vowel	/l//	Frequency for hVl words	/d/	Frequency for hVd words
iː	*heel*	7.05	*heed*	3.61
ɪ	*hill*	50.85	*hid*	3.87
e	*hell*	62.61	*head*	258.84
æ	*hal*	0.00	*had*	2210.3
ʉ	*hule*	0.00	*hude*	0.00
ɜː	*hurl*	0.74	*herd*	5.86
ɐː	*harl*	0.00	*hard*	339.68
ɐ	*hull*	4.49	*hud*	0.00
ʊ	*hool*	0.00	*hood*	5.67
oː	*hall*	42.64	*horde*	1.12
ɔ	*holl*	0.00	*hod*	0.00.
æɪ	*hail*	4.93	*hade*	0.00
ɑe	*hile*	0.00	*hide*	29.71
ɔɪ	*hoil*	0.00	*hoyd*	0.00
æɔ	*howl*	1.13	*howd*	0.00
əʉ	*hole*	36.35	*hode*	0.00

Table 3.1: **Carrier words for the 16 target vowels.** Left columns: carrier words and their lexical frequencies in the /l/ condition. Right columns: carrier words and lexical frequencies/ in the d/ condition.

normalized, truncated to common temporal landmarks, and digitized as 16 bit WAV files.

The stimulus words were coded manually in the acoustic analysis software Praat (Boersma & Weenink 2013), using a 15 ms window length to find maximum 3 formants in the range of 20 Hz to 4000 Hz. Each stimulus word was coded for three temporal landmarks, as illustrated in Figure 3.2. The first landmark was the onset of the /h/ (T0), aligned with the beginning of friction. The second was the onset of the vowel, aligned with the end of the friction of /h/ (T1). The third was the offset of the vowel, aligned with the beginning of the closure in the /d/ condition, and with the end of the formant transitions from the vowel to /l/ in the /l/ condition (T3).

When separating the first and the second element of the diphthongs, the steady-state formants were used to align the first and the second ele-

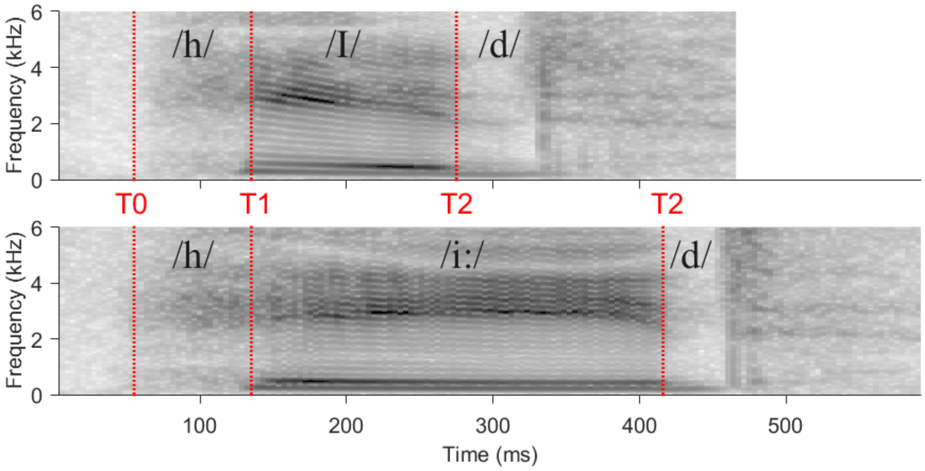

Figure 3.2: **Acoustic landmarks for vowel onset and offset**. Upper panel: stimulus word *hid*. Lower panel: stimulus word *heed*. T0: onset of /h/. T1: onset of the vowel. T2: offset of the vowel.

ment with the spectrograms; the transitional phase was marked separately. The first and the second formants of the vowels in the stimulus words were measured manually. For monophthongs, the formants were measured at the midpoint of the steady-state part of the vowels. For diphthongs, F1 and F2 were measured at the midpoint of the steady-state of both the first and the second element. Acoustic information on the stimulus can be seen in Table 3.2 for monophthongs and Table 3.3 and 3.4 for diphthongs.

3.3.3 Orthography of the response options

Response options were presented orthographically. Existing monomorphemic English words were presented with their correct spelling. The word /hɜːd/ is a homophone for *heard* and *herd*. As *heard* is a morphologically complex word, *herd* was chosen as a response option. Morphologically complex words were avoided, because morphological complexity has a complicated effect on lexical access depending on (among other things) phonetic and semantic

	Coda /d/			Coda /l/		
Vowel	F1 (Hz)	F2 (Hz)	vowel length (ms)	F1 (Hz)	F2 (Hz)	rime length (ms)
iː	357	2940	263.8	461	2750	422.7
ɪ	482	2722	136.3	502	2551	345.6
e	712	2298	219	884	2050	368.4
æ	1040	1838	215.9	1061	1848	419.6
ʉː	365	2334	308.4	450	1025	499.7
ɜː	658	1895	313.6	699	1733	490.4
ɐː	1018	1590	423.3	964	1463	450.8
ɐ	1080	1491	144.4	873	1421	582.4
ʊ	472	1027	158	477	961	432.6
oː	541	986	331.6	627	1047	444.5
ɔ	815	1271	190.8	766	1250	383.2

Table 3.2: **Formant frequencies and durations of monophthongs in acoustic stimuli.** Left three columns: F1, F2 and duration of monophthongs before coda obstruents. Right three columns: F1, F2 and rime duration of monophthongs with coda laterals.

transparency between the base form and the inflected form (Vannest et al. 2011). Therefore *who'll, who'd*, and *how'd* were respelled as *hule, hude*, and *howd* respectively.

Respelling these words also lead to representing the phoneme /h/ with its more usual spelling, the grapheme ⟨h⟩ in each stimulus word. This is considered as a further advantage because dominant (i.e. more typical and frequent) spellings lead to quicker and more accurate word recognition than subdominant (i.e. less typical and frequent) spelling of the same phoneme (Ziegler et al. 2004). The non-words, such as *harl* and *hud* were spelled according to the rules and traditions of English spelling and judged by native speakers of Australian English for transparency.

	First element		Second element		Diphthong
	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)	total length (sm)
æɪ	870	2165	402	2795	277.6
ɑe	1030	1518	693	2313	313.3
ɔɪ	636	1052	394	2789	271.6
æɔ	1059	1899	968	1609	307.7
əʉ	684	1568	415	2170	279.8

Table 3.3: **Formant frequencies and durations of diphthongs before coda obstruents in acoustic stimuli.** Left two columns: F1 and F2 of the first element of the diphthong. Middle two columns: F1 and F2 of the second element of the diphthongs. Last column: duration of the whole diphthong.

	First element		Second element		Diphthong
	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)	rime length (ms)
æɪ	858	2189	635	2421	427.5
ɑe	1146	1561	926	1873	408.7
ɔɪ	737	1142	626	2247	381.1
æɔ	1036	1885	997	1665	459
əʉ	770	1124	629	1070	471.3

Table 3.4: **Formant frequencies and durations of diphthongs before coda laterals in acoustic stimuli.** Left two columns: F1 and F2 of the first element of the diphthong. Middle two columns: F1 and F2 of the second element of the diphthongs. Last column: rime duration.

3.4 Procedure

3.4.1 Familiarisation phase

Prior to the experiment, there was a familiarisation and a training part. During familiarisation, participants listened to all the target words and were instructed to pay attention to the words and told that the training would be followed by a test. A target word was presented orthographically for 4000 ms. If the target was a non-word, a rhyming helper word was also presented orthographically on the same slide. The target word was presented auditorily at the midpoint of the visual presentation interval (at 2000 ms). After 4000 ms, the procedure automatically moved on to the next trial. Participants were not asked to make lexical decisions during familiarisation.

3.4.2 Training phase

During training, participants were tested on only the non-words. Participants were presented with two non-words visually on a computer screen, one on the left side of the screen, and one on the right side of the screen. The non-words were paired by the researcher to provide a task that was neither too easy nor too difficult. There were 14 pairs in the /d/ condition and 16 in the /l/ condition, (see in Appendix B). Each pair was presented once and all the participants received the same pairs. The target word was presented on the right side of the screen in 50% of the trials, and on the left side of the screen in the other 50% of the trials, and the trials were presented in a random order.

One out of the two words was presented auditorily on headphones. The participants were instructed to select the word they heard. During the training phase, participants saw two words on the screen for 1000 ms, then one word started playing, and participants had 2000 ms from the beginning of the audio to make their decision as to which word they heard. If a correct response was given, participants received the feedback “correct”, and testing moved to the second non-word pair. If participants responded inaccurately or did not respond in 2000 ms, they received the feedback “incorrect” or “too slow” and were presented with the same pair again until they gave the correct answer.

3.4.3 Experimental phase

The familiarisation and the training was followed immediately by the experiment. In the experiment, participants were presented with two words visually on a computer screen, and one of them auditorily on their headphones. Participants were instructed to select the word they heard. In a trial, illustrated in Figure 3.1, participants first were presented with a fixation cross for 500 ms in silence. After the fixation cross, participants were presented with the word-pair orthographically for 1500 ms. As Figure 3.1 shows, the word on the left was placed in a green box, and the word on the right was placed in a blue box. After 1500 ms, one word started playing and participants had 2000 ms from the beginning of the audio to make their decision. If participants replied, the experiment moved on to the next trial irrespective of the answer being correct. If the participants replied before the audio ended, the audio was aborted. If the participants did not answer within 2000 ms, a warning message was given to let them know that they were too slow and they were instructed to press a button to continue. The warning message remained on the screen and the experiment did not proceed to the next trial until the participants responded. The 2000 ms cap was chosen on the basis of pilot studies. During the pilot studies, there was no cap, and participants' mean response time plus two standard deviation was approximately 1900 ms.

Participants were tested on the exhaustive pairing of the 16 target words and 15 competitor words, leading to an experiment design of $16x15 = 240$ trials. The 240 pairs were repeated three times with a 10000 ms break between the repetitions. The breaks could not be skipped by the participants, and at the end of the break, participants had to start the next repetition by pressing a button. The target word was presented on the right side of the screen in 50% of the trials, and on the left side of the screen in the other 50% of the trials in a random order. The order of the individual trials was also randomised within the repetitions.

After the experiment, participants were asked to fill in an exit interview to collect feedback and a self-evaluation of the participants' performance. Participants were also asked whether they found any of the words particularly difficult. The exit interview can be seen in Appendix A.

The experiment took place in the Speech Perception Lab of the Australian Hearing Hub, Macquarie University. Participants were tested individually without the experimenter being present. The stimuli were presented with the software Psychology Software Tools (2012) on an Asus laptop with a display refresh rate of 60 Hz. The audio stimulus was presented via Sennheiser 380 Pro headphones, and participants were allowed to adjust the volume prior to the experiment. Responses were recorded with a button box. Participants were instructed to press the leftmost button on the button box if the word they heard was on the left side and to press the rightmost button on the button box if the word was on the right side. The rightmost and leftmost buttons were coloured green and blue in accordance with the visual stimulus to make the link between the words and the buttons easier to memorise.

4 Results

4.1 Data processing

Observations were excluded due to errors in stimulus presentation. Nine observations were excluded in the /d/ condition (three for each of the first three participants), and 54 observations were excluded in the /l/ condition (six for the first 3 participants and 3 for the last 12 participants).

Accuracy and RT data were examined to determine whether any participants should be excluded based on their accuracy or RT. Participants were at ceiling with respect to accuracy, with a mean accuracy of 97.7% (range: 93%–100%) in the /d/ condition, and 96.5% (range: 93%-98%) in the /l/ condition. Therefore no participants were excluded based on their accuracy rate.

Participants' individual mean RT was in the range of the overall participant mean±1 SD of all participants in both conditions (644.8 ms in the /d/ condition and 701.14 ms in the /l/ condition). Therefore no participants were excluded based on RT.

Further data were excluded from only the RT analyses. First, incorrect responses were excluded to minimize the effect of the speed-accuracy tradeoff on RT analysis. The speed-accuracy tradeoff is the inverse relationship between the accuracy and the speed of a response (Wickelgren 1977, Ruthruff

1996, Osman et al. 2000). As a result of the speed-accuracy tradeoff, the RT of a guess might be faster than the RT of a non-guess (Osman et al. 2000), which supports the exclusion of incorrect answers. 240 inaccurate observations were excluded in the /d/ condition and 406 in the /l/ condition.

Secondly, responses with an RT faster than the onset of word-initial /h/+210 ms were also excluded from the RT analysis. The minimum valid response time was calculated as the time of the /h/ onset plus a response latency of 210 ms (Woods & Reed 2015). Word initial /h/ may carry coarticulatory information that listeners may respond to (see Section 2.1), therefore the onset of the stimulus was defined as the onset of /h/ (see T0 on Figure 3.2). As according to Woods & Reed (2015) it takes 210 ms to respond to the stimulus, 210 ms was added to the onset of the /h/ to set the lower limit for RT. (However, for an extensive overview of RT latencies, see Silverman (2010)). Based on this criterion, 5 responses were excluded in the /d/ condition and 3 in the /l/ condition.

No upper limit was set on RT observations to be included, although excessive RT may have been caused by participants' lack of attention or their failure to reach a decision (Ratcliff 1993). Several methods for trimming outliers have been tested, and suggestions have been made to either eliminate a fixed percentage of responses or use an absolute cutoff point, such as eliminating all responses above 1000 or 1500 ms (Ratcliff 1993). The present experiment had an inbuilt cutoff point at 2000 ms, because participants received a time-out message on screen after 2000 ms (see Section 3). Trimming outliers in the range below 2000 ms would have been possible in principle; however, outliers may carry effects, in which case eliminating outliers decreases power (Ratcliff 1993). In this experiment, it was assumed that specific vowels or vowel-pairs would have a long RT as they are hard to disambiguate. The aim of the experiment was to find these vowels and vowel-pairs, therefore positive outliers have not been trimmed.

Raw data of all participants were transformed. Firstly, accuracy data of all responses were transformed to percentage of inaccurate responses by target word to determine which was the least accurately identified. Percentage of inaccurate responses by target word and competitor was determined to see which word-pair yielded the most inaccurate responses. Secondly, RT data was refined by adjusting it to two landmarks in the sound stimulus. The first landmark was the onset of the vowel, as marked by the end of the frication interval of /h/. The second landmark was the offset of the vowel as marked by the closure of the /d/ in the /d/ condition or by the point at which the formants reached the target of /l/. The landmarks are shown in Figure 3.2. Two RTs were calculated for each trial relative to the landmark vowel onset and the landmark vowel offset: RT from vowel onset is the time from the beginning of the vowel to the response, and RT from vowel offset is the time from the end of the vowel to the response. This was required to compensate for the intrinsic length differences of the stimulus, because for short stimulus words all information on the vowel is available sooner than for long stimulus words, which may lead to short RT for short stimulus words.

4.2 Statistical analysis

The main aim of data analysis was to answer the following questions :

1. Are vowel contrasts reduced in perception in the pre-lateral environment?

2. If vowel contrasts are reduced in perception, which vowels or vowel pairs are affected?

If vowel contrasts are reduced in perception in the pre-lateral environment, the disambiguation of vowels ending in /l/ has an overall lower rate of a accuracy and longer RT. If vowel contrasts are reduced between certain vowel-pairs, the members of these pairs have a low pairwise accuracy rate and long pairwise RT.

To explore if /l/ lowers overall accuracy and increases overall RT, we provided visual representation of the data using the percentage of incorrect answers and the mean RT across all repetitions of all participants in both coda condition. To explore if certain vowels are more affected than others, and if the contrast between certain vowel pairs is reduced, we provided visual representation by target vowel and confusion matrices for each target-competitor pair.

To confirm visual analysis, General Linear Model (GLM) in the lme4 package (Bates et al. 2015) of the statistical software R (R Core Team 2015, Team 2015) was used (GLM scripts are included in Appendix C). For the analysis of the binary accuracy data (1=correct, 0=incorrect), logistic regression was used, because accuracy is categorical data. For the analysis of RT data, GLM was used in order to handle the right-skewed RT distribution. As shown on Figure 4.1, RT measured from vowel onset and offset is not normally distributed, but has a long tail on the right. The distribution of the presented RT data is in line with that RT data is often right-skewed, but it violates the assumption of normal distribution of data and its residuals of linear models (Baayen & Milin 2010). RT data and its residuals were compared to three continuous right-skewed distribution: Weibull, Gamma, and lognormal distribution. Figure 4.1 shows that out of the three, RT follows Gamma distribution the most closely, therefore GLM with the family Gamma was used. Gamma transformation results in a the inversion of estimates, that is when an independent variable increases RT, the estimate is shown as positive, and when it decreases RT, the estimate is shown as positive.

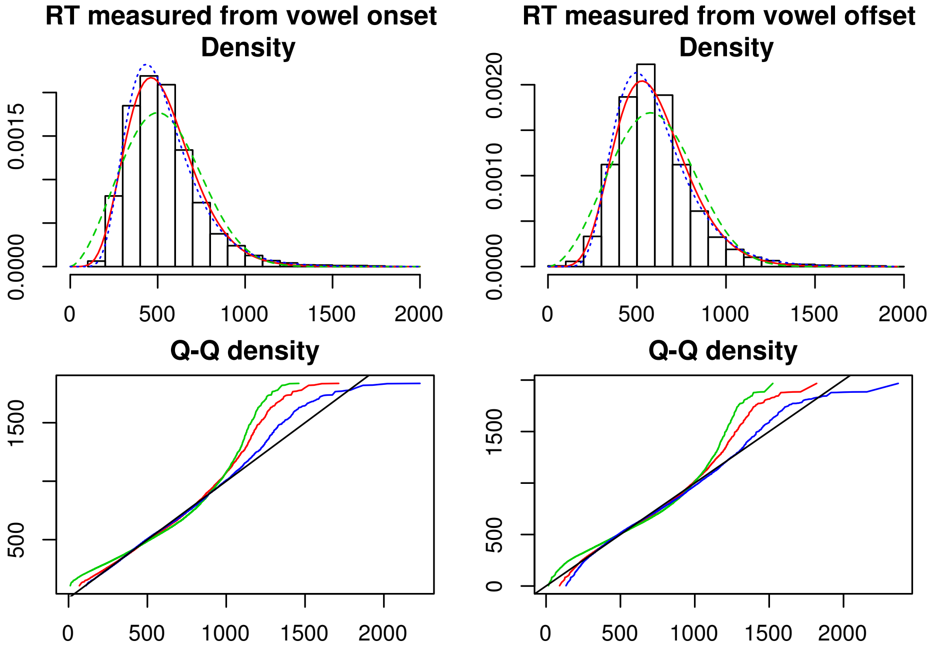

Figure 4.1: **Comparison of the distribution of RT data (black line) to three distributions.** Left panels: RT measured from vowel onset. Right panels: RT measured from vowel offset. Upper panels: density of RT. Lower panels: Q-Q plots. Red line: Gamma distribution. Green line: Weibull distribution. Blue line: lognormal distribution.

In the GLM models, the dependent variables were *accuracy* and *RT*. The independent variables were used differently in different models depending on the tests, therefore the independent variables are presented in detail in Section 4.4. The random effect structure was random intercept by participant and it was kept consistent in all models. Although Barr et al. (2013) suggest using maximal random effect structure with random intercepts and slopes, a maximal model was not viable, because models did not converge when all the independent variables were added to the random effect structure.

In order to find the best-fitting models which only contain those independent variables that have a significant effect on the dependent variables, backward elimination of independent variables was used. First, a maximal model was constructed with all the relevant independent variables. Secondly, independent variables were removed from the maximal model one by one, and the reduced model was compared with the maximal model with the likelihood ratio test (Johnson 2008). The likelihood ratio test was also used to obtain p values and χ^2 values for dropping variables. If the likelihood ratio test returned significant results ($p < 0.05$), the independent variable was kept in the model; if it did not, the independent variable was dropped.

In order to further explore the effect of the independent variables, hierarchical cluster analysis (HCA) was used on the accuracy and RT data both in the coda /d/ and coda /l/ condition, whose results are presented in four tree diagrams in Section 4.4. The input for HCA were four confusion matrices containing the percentage of incorrect answers for all the target-competitor pairs in both coda conditions, and the mean RT of each target-competitor pair in both coda conditions. The confusion matrices contain cells in which the target and the competitor vowel are the same, although a vowel was never its own competitor in the experiment design. These cells were assigned 100% confusion in the accuracy matrices, and 2000 ms RT in the RT matrices, as 2000 ms was the inbuilt cutoff point in the experiment design. Squared Euclidean distances for each pair in every confusion matrix were calculated. The distances were used to carry out agglomerative HCA. During agglomerative HCA, each individual vowel was a separate cluster and vowels (clusters) were joined when the members of two clusters were similar to each other (Everitt 2006, R Core Team 2015). Similarity was determined by Ward's method, which at each steps selects the two clusters to be merged in such a way that members of a clusters are maximally similar to each other (Ward 1963). The results of HCA are shown on tree diagrams.

Figure 4.2 is an example for a tree diagram that contains five single member clusters grouped into one cluster containing all five elements. The tree diagram shows the five single member cluster, *a,b,c,d,* and *e* as its leaves. The leaves are merged at clades (i.e. the "nodes"), and the leaves merged at one common clade form clusters with multiple members. The vertical axis shows the distance between the leaves: the lower a clade is located, the more similar are its leaves to each other. That is, in Figure 4.2 *a* and *b* form a cluster at 0.5, therefore they are more similar to each other than to *c* with which they are merged at 1.5 and they are the least similar to the cluster *e-d* with which they are only merged at the final step. An arbitrary line at 1 was added to create three clusters.

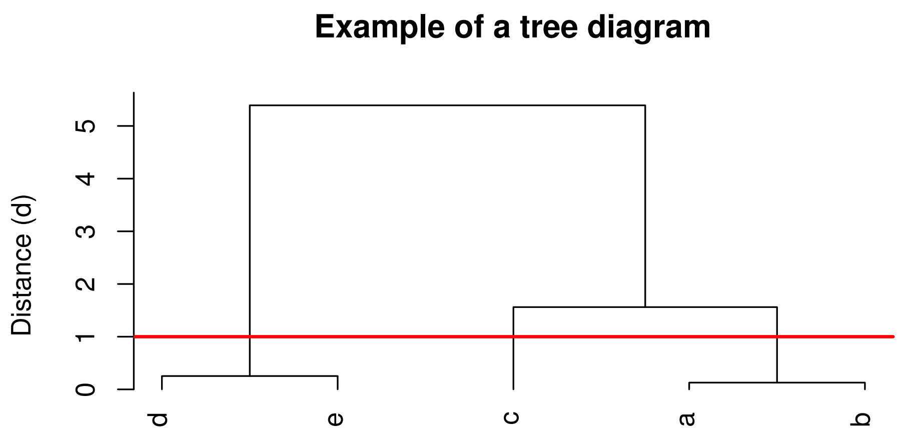

Figure 4.2: **Example of a tree diagram.** *A, b, c, d, e:* five single member clusters. Top clade: one cluster containing all five individuals. Red line: an arbitrary cutoff-point to create three clusters: *d-e, c,* and *a-b.*

To sum up, GLM was used to test if the main effect of the independent variables are statistically significant. Due to the fact that 16 target vowels were tested against 15 competitor vowels, HCA and visual representation can be more informative and detailed than GLM concerning the effect of target- and competitor vowel on accuracy and RT. Therefore the three methods can support each other to give a comprehensive picture on the disambiguation of lateral-final rimes.

4.3 Descriptive results

Figure 4.3 shows the percentage of inaccurate responses for each target vowel in both coda conditions; the target vowels are in ascending order according to the percentage of inaccurate responses in the /d/ condition. Figure 4.3 shows that not all vowels are equally easy to identify. In the /d/ condition, /ɜː/ was responded inaccurately in 1% of the trials, whereas /ɑe/ in 4% of the trials. Figure 4.3 shows that the inaccuracy rates differ between the /d/ and /l/ condition. Vowels, except for /ɪ, æɪ, oɪ,/ and /ɑe/, have higher inaccuracy rates in the /l/ condition than in the /d/ condition. Figure 4.3 shows that the difference between the inaccuracy rates in the /d/ and /l/ condition are not equal for all the vowels. The difference is the largest for /ʊ, æ, ʉː/ and /ɔ/, small for /iː/ and /e/, and reversed (i.e. more inaccurate in the /d/ condition) for four vowels. The least accurate vowel is not the same in the two conditions. Inspecting Figure 4.3 shows that in the /l/ condition, the inaccuracy rates of targets do not follow the order of targets in the /d/ condition. For example, the target vowel with the highest rate of inaccurate responses is /ɑe/ in the /d/ condition and /ɔ/ in the /l/ condition. (For the different order of the target vowels between the /d/ and /l/ condition, see Figure D.1 and D.2 in Appendix D.) Figure 4.3 indicates that different target vowels might be affected differently by coda /l/.

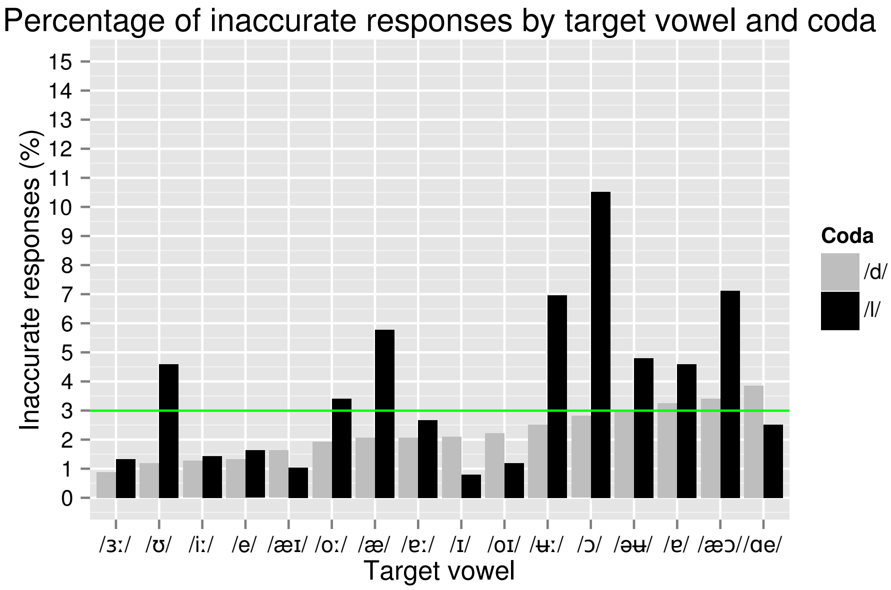

Figure 4.3: **Inaccurate responses (%) by target vowel and coda.** X axis: target vowels ordered according to the percentage of inaccurate responses in the /d/ condition. Y axis: percentage of inaccurate responses in the /d/ condition (gray columns) and in the /l/ condition (black columns). The green line marks the mean percentage of inaccurate responses for all target vowels across conditions.

Figures 4.4 and 4.5 show the mean RT of the disambiguation of each target vowel against all its competitors in the /d/ and /l/ condition. Figure 4.4 shows RT measured from the onset of the vowel (marked as T1 in figure 3.2), whereas Figure 4.5 shows RT measured from the offset of the vowel (T2 in Figure 3.2). Figures 4.4 and 4.5 show that RT was consistently longer in the /l/ condition than in the /d/ condition; the sole exception was when the target vowel was /əʉ/ with RT measured from the offset of the vowel. These figures also show that target vowels were responded to with markedly

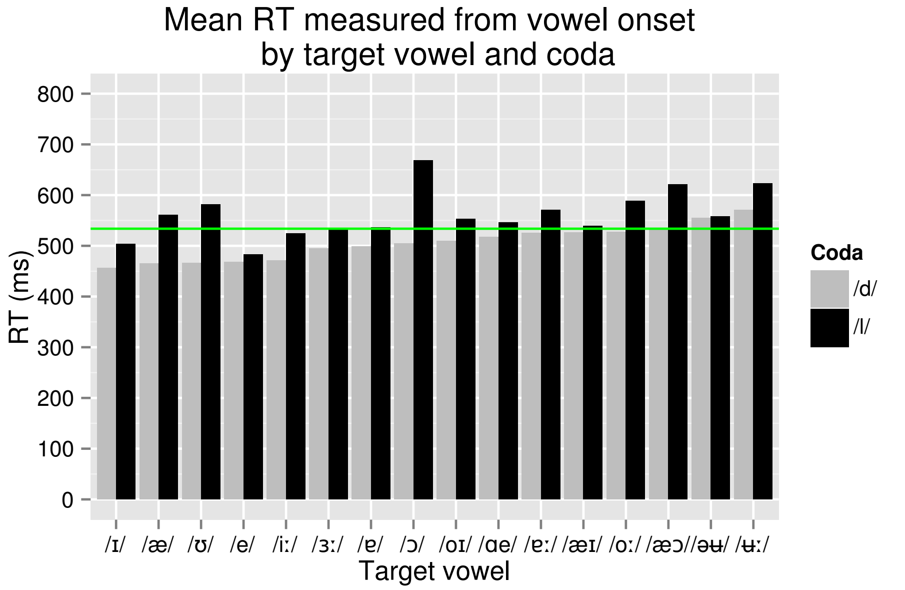

Figure 4.4: **RT of responses measured from vowel onset by target vowel and coda.** X: target vowels in ascending order according to RT in the /d/ condition. Y axis: RT (ms) in the /d/ condition (gray columns) and in the /l/ condition (black columns). The green line marks the mean RT of responses for all target vowels across conditions.

different RTs; comparing the two figures shows that short vowels are located on the left side of Figure 4.4, and on the right side of Figure 4.5. That is, short vowels have short RT when RT is measured from vowel onset, but they have long RT when RT is measured from vowel offset. Figures 4.4 and 4.5 also show that the RT to the target vowels is affected differently by coda /l/. For example, the RT to the target /æɪ/ is almost the same in both conditions, whereas greater difference between the RT of the two conditions can be observed in the RT given for /ɔ/.

To sum up, inspection of the accuracy data in Figure 4.3 and RT data in Figures 4.4 4.5 supports the hypothesis that coda /l/ hampers disambiguation, as it decreases accuracy and increases RT of word disambiguation

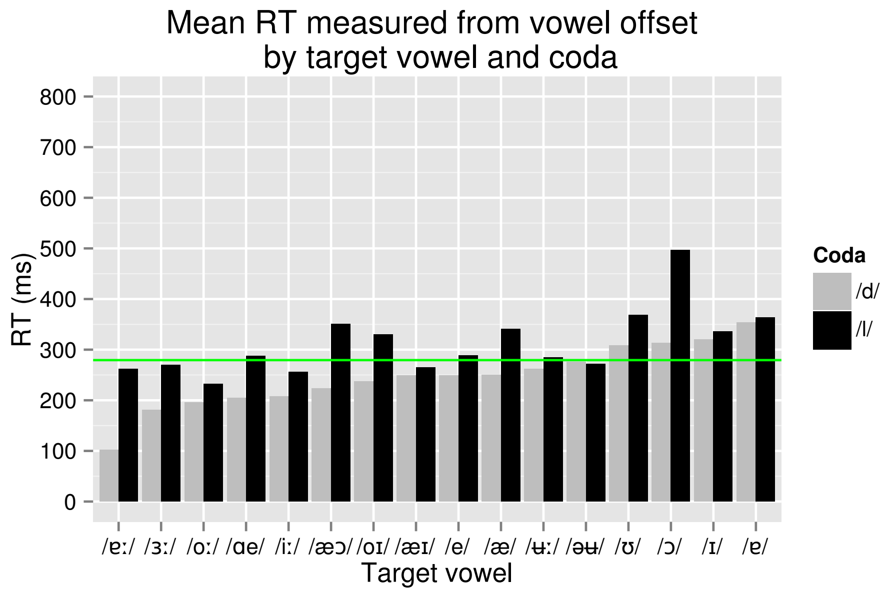

Figure 4.5: **RT of responses measured from vowel onset by target vowel and coda.** X: target vowels in ascending order according to RT in /d/ condition. Y axis: RT (ms) in the /d/ condition (gray columns) and in the /l/ condition (black columns). The green line marks the mean RT of responses for all target vowels across conditions.

compared to coda /d/. Vowels are not equal in difficulty of disambiguation, and more importantly are not equally affected by /l/.

4.4 Confirmatory results

4.4.1 Main effect of coda

To test whether the higher inaccuracy rates and longer RT of the /l/ condition are statistically significant, a maximal model was constructed with the following independent variables: *coda, target vowel, competitor vowel* and *real word.* Coda was an independent variable with two levels (/d/: -1; /l/: $+1$) which tested whether /l/ lowers accuracy rates and increases RT. Target vowel and competitor vowel were independent variables with 16 levels each (coded with IPA symbols, see Table 3.1) which tested whether different target and competitor vowels have different accuracy rates and responded to with different RTs. Real word status of the target word was an independent variable (non-word target:-1, real-word target: $+1$) testing whether non-word targets are disambiguated less accurately or more slowly. A further factor that is known to affect RT is lexical frequency, (see Section 2.1); however, when a third independent variable with either 32 levels or as a continuous variable was added, the GLM model did not converged due to the number of levels in the factors.

The independent variables might have interacted, for example, coda might affect accuracy rates and RT differently across the vowels (see Figures 4.3–4.5). However, a model with two 16-level independent variables and interacting factors did not converge due to the high number of levels. Therefore the maximal models in this section include all the relevant independent variables but not their interactions. Interactions are tested on the subset of the data in Section 4.4.2 and 4.5.

Table 4.1 shows that removal of coda, target word and competitor word each returned a significant effect($p < 0.05$) for accuracy rates, but removal of real word did not. The significant negative estimate of coda ($\beta = -0.35, se = 0.12$) shows that accuracy (0=inaccurate, 1=accurate) is significantly lower in the/l/ condition than in the /d/ condition (cf Figure 4.6). The significant main effects of target vowel and competitor vowel show that different target and competitor vowels have different accuracy rates. Because target and competitor vowel are factors with 16 levels, their effect is described in detail in Section 4.4.2 and 4.5. The null effect of the factor real word indicates that non-word targets were responded as accurately as real word targets. That is, the different accuracy rates between the two conditions and between the target vowels is not caused by the carrier word.

<table>
<tr><td colspan="6">Maximal model:
Accuracy~Coda+TargetVowel+CompetitorVowel+RealWord+(1|Subject)</td></tr>
<tr><td>Dropped variable</td><td>AIC</td><td>χ^2</td><td>p</td><td>Estimate</td><td>Standard error</td></tr>
<tr><td>Coda</td><td>5380.4</td><td>8.10</td><td>0.004</td><td>−0.357412</td><td>0.120130</td></tr>
<tr><td>Target vowel</td><td>5502.2</td><td>157.91</td><td>< 0.001</td><td colspan="2">values for the 16 individual levels are not presented</td></tr>
<tr><td>Competitor vowel</td><td>5488.2</td><td>143.91</td><td>< 0.001</td><td colspan="2">values for the 16 individual levels are not presented</td></tr>
<tr><td>Real word</td><td>5372.4</td><td>0.16</td><td>0.689</td><td>0.02243</td><td>0.05572</td></tr>
</table>

Table 4.1: **Effect of the independent variables on the accuracy of word disambiguation.** First row: maximal model. Leftmost column: independent variables. The p values and χ^2 values, estimate, and standard error belong to the independent variable in the given row. p values and χ^2 values were obtained by comparing the maximal model to each reduced model.

Tables 4.2 and 4.3 show that removing independent variables returns the same results when RT is measured from vowel onset and vowel offset. Removing coda approached the threshold of statistical significance ($p = 0.06$) when RT was measured from vowel offset. Removing target word, competitor word, and real word returned a significant effect ($p < 0.001$) at both landmarks. The negative estimate of coda shows that participants had

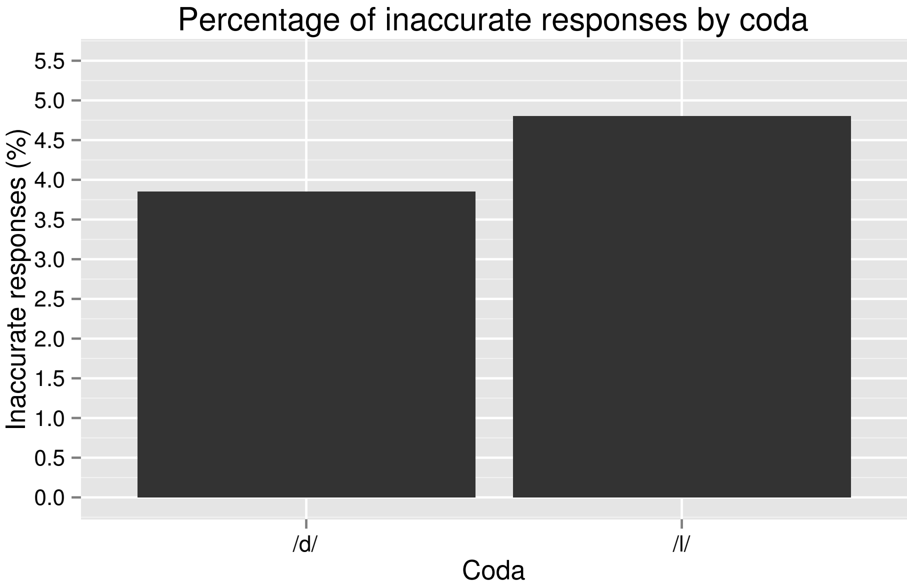

Figure 4.6: **Inaccurate responses (%) by coda** Left bar: /d/ condition. Right bar: /l/ condition. Y axis: Percentage of inaccurate responses.

a non-significant tendency to be slower to respond in the /l/ condition.[1] Different target and competitor vowels are responded with different RTs but because target and competitor vowel are factors with 16 levels, their effect is described in detail in Section 4.4.2 and 4.5. The positive estimate of real word ($\beta = 3.291e - -05, se = 5.83e - -06$ at vowel onset and $\beta = 4.636e - -05, se = 4.609e - -06$ at vowel offset) shows that participants were significantly quicker to respond when the target was a real word.[2] The results that real word decreases and coda /l/ may increase RT are shown in Figure 4.7. For the RT of the disambiguation of individual target vowels, see Figures D.3-D.6 in Appendix D.

[1]Although /d/ was encoded as -1, and /l/ was encoded as +1, the negative estimate means that RT was higher in the /l/ condition, because Gamma transformation reverses the direction of estimates. Therefore participants were slower in the /l/ condition.

[2]Although non-word was encoded as -1 and real word was encoded as +1, the positive estimate means that RT was lower in the real-word condition, because Gamma transformation reverses estimates. Therefore participants were quicker to respond to real words.

Maximal model: RTtoVowelOnset~Coda+TargetVowel+CompetitorVowel+RealWord+ (1|Subject)					
Dropped variable	AIC	χ^2	p	Estimate	Standard error
Coda	270010	2.6903	0.101	$-1.235e-04$	$7.330e-05$
Target vowel	270549	569.04	< 0.05	values for the 16 individual levels are not presented	
Competitor vowel	270071	90.502	< 0.05	values for the 16 individual levels are not presented	
Real word	270042	33.783	< 0.05	$3.291e-05$	$5.583e-06$

Table 4.2: **Effect of the independent variables on the RT of word disambiguation.** RT was measured from the onset of the vowel. First row: maximal model. Leftmost column: independent variables. The p values and χ^2 values, estimate, and standard error belong to the independent variable in the given row. p values and χ^2 values were obtained by comparing the maximal model to each reduced model.

Maximal model: RTtoVowelOffset~Coda+TargetVowel+CompetitorVowel+RealWord+ (1|Subject)					
Dropped variable	AIC	χ^2	p	Estimate	Standard error
Coda	271859	3.5046	0.06	$-1.128e-04$	$5.814e-05$
Target vowel	273474	1644.9	< 0.05	values for the 16 individual levels are not presented	
Competitor vowel	271913	83.973	< 0.05	values for the 16 individual levels are not presented	
Real word	271954	96.968	< 0.05	$4.635e-05$	$4.609e-06$

Table 4.3: **Effect of the independent variables on the RT of word disambiguation.** RT was measured from the offset of the vowel. First row: maximal model. Leftmost column: independent variables. The p values and χ^2 values, estimate, and standard error belong to the independent variable in the given row. p values and χ^2 values were obtained by comparing the maximal model to each reduced model.

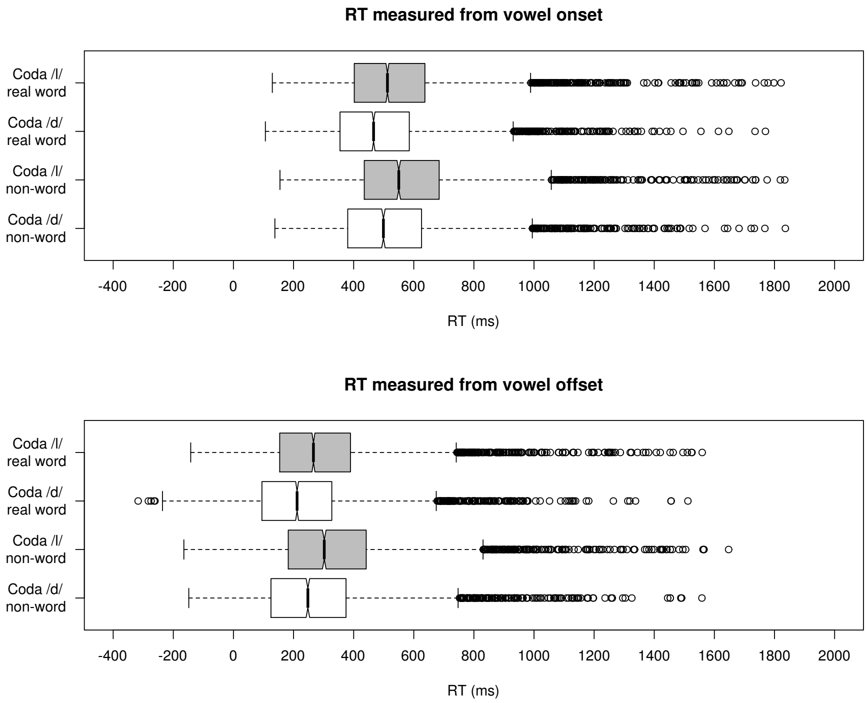

Figure 4.7: **RT to target words according to real-word and non-word targets and coda.** Upper panel: RT measured from vowel onset. Lower panel: RT measured from vowel offset. White boxplots: /d/ condition. Gray boxplots: /l/ condition.

4.4.2 The interaction of target and competitor vowels

To examine whether disambiguation is more difficult for target and competitor vowels that are similar, it was tested whether vowels that share phonological features have lower accuracy rates and longer RT. In order to further examine the effect of target and competitor vowel and their possible interaction, target and competitor vowels were assigned features based on their places of articulation. The features were binary features following the system of *Sound Pattern of English* (Chomsky & Halle 1968) and using the values appropriate for AusE (Cox 2012). The features used were ±high, ±low, ±front, ±back and ±long. Although +high and +low or +front and +back are mutually exclusive, all four features were needed to capture mid vowels, which are −high and −low, and to capture central vowels which are −front and −back. Diphthongs were classified according to the place of articulation of their first element. This was motivated by the fact that the first element of each AusE diphthong coincides with an AusE monophthong (Cox 1999), and studies done in the gating paradigm (Grosjean 1980) and in the visual world paradigm (Allopenna et al. 1998, Tanenhaus et al. 1995) showed that listeners disambiguate spoken stimulus as the acoustic signal unfolds in time. Moreover, listeners in a previous study confused monophthongs with inherent spectral change based on their first part (Nearey & Assmann 1986). These observations combined lead to the expectation that the first element of the diphthong is likely to be responsible for the confusion. The classification of vowels is shown in Table 4.4.

To see if target and competitor vowels that share features have lower accuracy rates and longer RT, three GLM models were created. The dependent variables were accuracy and RT measured from vowel onset and offset. There were five independent variables for the target vowel features

Vowel	Front	Back	High	Low	Long
iː	+	-	+	-	+
ɪ	+	-	+	-	-
e	+	-	-	-	-
æ	+	-	-	+	-
ʉː	-	-	+	-	+
ɜː	-	-	-	-	+
ɐː	-	-	-	+	+
ɐ	-	-	-	+	-
ʊ	-	+	+	-	-
oː	-	+	-	-	+
ɔ	-	+	-	-	-
æɪ	+	-	-	+	+
ɑe	-	+	-	+	+
æɔ	+	-	-	+	+
əʉ	-	-	-	-	+
oɪ	-	+	-	-	+

Table 4.4: **Classification of AusE vowels according to their binary features based on place of articulation.** The system of binary feature follows Chomsky & Halle (1968) with values appropriate for AusE (Cox 2012). Diphthongs were classified according to place of articulation of their first elements.

and five independent variables for the competitor vowel features. The independent variables of features of target and competitor vowel interacted only for the same feature. That is, target frontness interacted with competitor frontness, target height with competitor height, and target length with competitor length, but target frontness did not interact with target backness. Additionally, coda and real word were kept as independent variables, because backward elimination showed that they affect accuracy and RT respectively. Table 4.5 presents the model and reports the results on accuracy and RT measured from the offset of the vowel only, as the models return the same results for RT measured from both vowel onset and offset.

The significant negative estimate for coda ($\beta = -0.3654, se = 0.1221$) shows that accuracy decreases in the /l/ condition. The significant positive estimate for real word ($\beta = 0.1885, se = 0,0446$ for accuracy and $\beta = 6.677e - -05, se = 3.678e - -06$ for RT) shows that real word targets are identified more accurately and quickly. The significant negative estimates for the interaction terms Front:CFront ($\beta = -0.6432, se = 0.0513$ for accuracy, $\beta = -1.103e - -04, se = 5.813e - -05$ for RT), High:CHigh ($\beta = -0.5864, se = 0.0492$ for accuracy, $\beta = -2.164e, se = 4.643e - -06$ for RT), Low:CLow ($\beta = -0.3997, se = 0.0451$ for accuracy, $\beta = -2.724e - -05, se = 3.676e - -06$ for RT) show that accuracy decreased and RT increased when target and competitor vowels shared the respective features of front, high, or low. Additionally, the significant positive estimates for Back:CBack ($\beta = 0.1817, se = 0.0492$) and Long:CLong ($\beta = 0.1619, se = 0,0467$) show that accuracy increases when both target and competitor vowel are back or long. To sum up, accuracy decreases and RT increases when target and competitor vowels agree in terms of their articulatory features.

Front*CFront+Back*CBack+High*CHigh+Low*CLow +Coda+RealWord+(1|Subject)						
	Accuracy			RT from offset		
Variable	Estimate	Standard error	Pr(>|z|)	Estimate	Standard error	Pr(>|z|)
Coda	−0.36543	0.12213	0.002	$-1.103e-04$	$5.813e{-}05$	0.057
RealWord	0.18855	0.04464	< 0.001	$6.677e-05$	$3.678e{-}06$	< 0.001
Front:CFront	−0.64329	0.05037	< 0.001	$-2.238e-05$	$3.762e{-}06$	< 0.001
Back:CBack	0.18173	0.04922	0.002	$-6.564e-06$	$3.996e{-}06$	0.100
High:CHigh	−0.58645	0.05836	< 0.001	$-2.164e-05$	$4.643e{-}06$	< 0.001
Low:CLow	−0.39977	0.04513	< 0.001	$-2.724e-05$	$3.676e{-}06$	< 0.001
Long:CLong	0.16199	0.04679	< 0.001	$3.614e-06$	$3.560e{-}06$	0.310

Table 4.5: **Effects of the interaction of the binary articulatory features of target and competitor vowel on the accuracy and RT of target disambiguation.** First row: the independent variables in the maximal model. Leftmost column: independent variables. Columns 2–5: the effect of the independent variables on accuracy. Columns 5–8: effect of the independent variables on RT measured from vowel offset.

4.4.3 Pairwise analysis of target- and competitor vowels

The next set of analyses aimed at assessing the pairwise accuracy and RT of target vowels and competitor vowels and at shedding light on the effect of feature sharing. In order to asses the pairwise discriminability of target vowel and competitor vowel, four confusion matrices were created for accuracy data and RT data both in the /d/ and in the /l/ conditions to serve as an input to the HCA analysis. The confusion matrices are presented in Tables 4.6–4.9. The confusion matrices show the target vowels in the rows and the competitor vowels in the columns. In the accuracy confusion matrices (Tables 4.6 and 4.8), each cell contains the percentage of misidentified targets for a given vowel pair in the /d/ and /l/ condition. For example, in Table 4.6, the number 11 in the cell in the row /iː/ and column /ɪ/ indicates that /iː/ was mistakenly identified as /ɪ/ in 11% of the comparisons when /iː/ was the target and /ɪ/ was the competitor. In contrast, the target /iː/ was never misidentified when the competitor was /ʉː/.

In the RT based confusion matrices (Tables 4.7 and 4.9), each cell contains the mean RT to a given target and competitor pair, as measured from the offset of the vowel in the /d/ and /l/ condition respectively. Only RT measured from the vowel offset was used, as GLM models consistently returned the same results for RT measured from vowel onset and RT measured from vowel offset. For example, the number 323 in row /iː/ and column /ɪ/ of Table 4.7, indicates that the mean RT measured from the offset of /iː/ was 323 ms when /iː/ was the target, and /ɪ/ was the competitor. In contrast, when /iː/ was the target, and /ʉː/ was the competitor, mean RT measured from the offset of the target was only 159 ms. In Figure 4.6–4.9, the quantile of the strongest competitors for each target vowel was highlighted in gray. The pairwise comparisons show target-competitor asymmetries, be-

Target vowel	Competitor vowel: iː	ɪ	e	æ	ʉː	ɜː	ɐː	ɐ	ʊ	oː	ɔ	æɪ	ɑe	æɔ	əʉ	oɪ
iː	–	11	2	0	0	0	0	2	0	0	0	2	0	2	0	0
ɪ	11	–	4	2	0	2	0	0	2	0	2	4	2	0	2	0
e	2	0	–	2	2	0	2	0	2	0	2	4	2	0	0	0
æ	0	0	0	–	0	2	0	2	2	2	0	2	2	16	2	0
ʉː	2	0	0	0	–	2	0	7	9	2	4	0	0	2	9	0
ɜː	0	0	2	0	2	–	2	2	0	2	0	2	0	0	0	0
ɐː	0	2	0	2	0	0	–	11	0	0	0	13	0	2	0	0
ɐ	0	2	0	2	4	7	20	–	0	0	0	9	0	2	0	2
ʊ	0	0	0	0	2	0	0	0	–	4	0	0	0	4	2	4
oː	0	2	0	0	2	0	2	0	2	–	4	0	0	2	2	11
ɔ	0	0	0	2	0	2	2	9	2	9	–	2	0	0	2	11
æɪ	4	2	2	2	4	2	22	11	2	0	0	–	0	0	4	0
ɑe	2	0	7	2	2	0	0	2	4	0	0	2	–	2	0	0
æɔ	0	0	0	20	2	2	0	2	0	0	7	2	4	–	0	11
əʉ	0	0	0	0	0	4	2	0	2	4	9	2	0	9	–	11
oɪ	0	2	0	0	2	2	0	0	2	9	4	0	0	9	2	–

Table 4.6: **Confusion matrix of inaccurate responses (%) in the /d/ condition for all target-competitor pairs.** Rows: targets. Columns: competitors. The quantile of the strongest competitors for each target is highlighted in gray.

cause it does not follow that if Vowel_1 is a strong competitor for Vowel_2, then Vowel_2 is a strong competitor for Vowel_1. For example, in the /d/ condition, /ʉː/ was misidentified as /ɔ/ in 4% of the comparisons, but /ɔ/ was never misidentified as /ʉː/. (For the overall rate of inaccuracy and RT of the target vowels, see the figures in Appendix D.)

The data in the confusion matrices were used to carry out HCA and plot target and competitor vowels in tree diagrams based on accuracy and RT data for /d/ and /l/ separately. The tree diagrams are presented on Figures 4.8–4.9. The vowels tested in the experiment are represented as the leaves of the tree diagrams. The lower the merging point between two leaves, the more confused the vowels were within the pair, and the higher their pairwise RT. That is, low clades indicate that disambiguation between the members of the cluster was difficult, irrespective of whether the members are targets or competitors. Therefore, a tree diagram masks the asymmetries between

		Competitor vowel															
		iː	ɪ	e	æ	ʉː	ɜː	ɐː	ɐ	ʊ	oː	ɔ	æɪ	ɑe	æɔ	əʉ	oɪ
Target vowel	iː	–	323	221	204	159	223	198	238	197	175	192	199	228	181	229	191
	ɪ	335	–	320	310	343	316	358	322	348	300	334	284	331	326	304	277
	e	349	234	–	252	204	256	234	254	250	261	243	283	198	232	269	226
	æ	258	241	251	–	237	231	272	234	231	273	212	242	223	418	224	223
	ɜː	194	159	206	188	193	–	224	173	169	190	155	214	169	179	141	160
	ʉː	257	223	258	212	–	247	256	286	367	312	231	318	182	306	231	253
	ɐː	109	82	132	82	96	58	–	206	68	90	63	105	185	115	102	60
	ɐ	332	334	346	385	358	318	402	–	332	323	368	365	398	385	328	347
	ʊ	280	274	312	254	343	304	344	321	–	313	328	292	256	285	360	357
	oː	181	162	171	178	175	168	147	157	250	–	222	167	179	235	266	298
	ɔ	316	319	280	252	321	244	331	308	348	343	–	334	273	278	404	359
	æɪ	219	250	232	287	244	284	260	254	258	254	217	–	232	287	233	220
	ɑe	179	195	192	225	168	161	330	229	195	170	196	192	–	180	255	227
	æɔ	261	139	218	272	228	235	256	204	217	192	229	237	224	–	264	181
	əʉ	222	284	281	241	287	273	256	258	292	320	275	285	219	309	–	344
	oɪ	213	252	200	234	222	251	188	215	237	342	207	245	187	350	238	–

Table 4.7: **Confusion matrix of mean RT of responses (ms) in the /d/ condition.** RT was measured from the offset of the target vowel for all target-competitor pairs. Rows: targets. Columns: competitors. The quantile of the strongest competitors for each target is highlighted in gray.

		Competitor vowel															
		iː	ɪ	e	æ	ʉː	ɜː	ɐː	ɐ	ʊ	oː	ɔ	æɪ	ɑe	æɔ	əʉ	oɪ
Target vowel	iː	–	11	2	0	2	0	0	0	0	0	0	0	2	2	0	0
	ɪ	NA	–	4	0	0	0	2	0	0	0	0	0	0	0	0	4
	e	0	4	–	0	0	0	0	0	2	4	2	4	2	0	4	0
	æ	2	0	2	–	0	4	9	0	2	4	0	2	2	56	0	2
	ʉː	0	0	0	2	–	0	0	11	76	0	7	0	0	4	4	0
	ɜː	0	0	4	0	2	–	0	0	0	4	2	4	2	0	0	0
	ɐː	0	0	0	4	0	0	–	20	0	4	0	0	4	2	0	4
	ɐ	0	0	4	2	7	2	20	–	4	7	2	4	7	2	7	0
	ʊ	0	2	0	0	29	0	0	16	–	13	7	0	0	0	2	0
	oː	2	2	0	2	7	0	0	13	2	–	7	2	2	0	2	9
	ɔ	0	2	0	11	9	7	2	27	2	9	–	4	4	11	64	4
	æɪ	0	0	0	2	0	0	2	0	2	2	0	–	2	0	2	2
	ɑe	0	2	2	0	7	0	9	9	0	2	0	4	–	0	0	2
	æɔ	0	0	2	49	0	2	7	2	0	20	7	4	2	–	7	4
	əʉ	2	0	2	0	0	2	0	9	2	7	24	0	2	9	–	11
	oɪ	0	0	0	2	0	2	0	0	0	0	0	2	0	2	9	–

Table 4.8: **Confusion matrix of inaccurate responses (%) in the /l/ condition for all target-competitor pairs.** Rows: targets. Columns: competitors. The quantile of the strongest competitors for each target is highlighted in gray.

		Competitor vowel															
		iː	ɪ	e	æ	ʉː	ɜː	ɐː	ɐ	ʊ	oː	ɔ	æɪ	ɑe	æɔ	əʉ	oɪ
Target vowel	iː	–	338	247	226	306	218	229	248	225	303	188	242	328	242	234	268
	ɪ	NA	–	382	323	375	342	319	355	346	302	319	311	397	309	314	316
	e	321	277	–	327	295	275	285	271	250	303	274	335	290	302	280	248
	æ	298	337	349	–	333	322	413	391	312	357	308	361	353	411	304	309
	ʉː	238	279	219	228	–	326	294	376	452	271	278	289	216	304	334	299
	ɜː	262	249	235	276	307	–	303	271	260	275	290	260	286	271	241	262
	ɐː	205	191	208	302	271	261	–	403	210	299	243	206	336	331	249	252
	ɐ	295	321	356	400	325	453	400	–	403	389	372	356	354	379	315	344
	ʊ	325	327	350	275	561	404	340	412	–	445	369	307	331	321	418	414
	oː	170	186	187	264	183	219	248	292	228	–	319	235	179	217	282	292
	ɔ	410	396	497	517	554	488	469	622	534	501	–	412	432	579	586	558
	æɪ	237	262	276	296	258	263	276	270	224	340	227	–	268	314	228	229
	ɑe	266	268	270	255	300	248	373	317	303	258	286	310	–	280	290	308
	æɔ	308	259	397	316	384	313	432	335	334	417	405	346	303	–	357	361
	əʉ	209	245	207	280	261	301	195	301	308	303	393	213	244	317	–	348
	oɪ	296	339	315	313	313	366	335	318	335	354	341	332	316	336	338	–

Table 4.9: **Confusion matrix of RT (ms) in the /l/ condition.** RT was measured from the offset of the vowel for all target-competitor pairs. Rows: targets. Columns: competitors. The quantile of the strongest competitors for each target is highlighted in gray.

target and competitor. For example, in the /d/ condition, short /ɐ/ target was mistakenly identified as /ɐː/ in 20% of the /ɐ-ɐː/ comparisons, but long /ɐː/ target was only identified as /ɐ/ in 10% of the /ɐː-ɐ/ comparisons. Instead of showing target-competitor asymmetries like a confusion matrix, a tree diagrams shows a generalisation that /ɐ/ and /ɐː/ are hard to disambiguate from each other.

It can be seen on Figures 4.8–4.9 that the vowels that cluster together share their articulatory features but differ in length. This holds for the vowel clusters based on accuracy and RT, both in the /d/ and /l/ conditions. Therefore the results of HCA correspond to the results of the GLM models showing that accuracy decreases and RT increases when the target vowel and the competitor vowel agree in their articulatory features (see Table 4.5 for the GLM results).

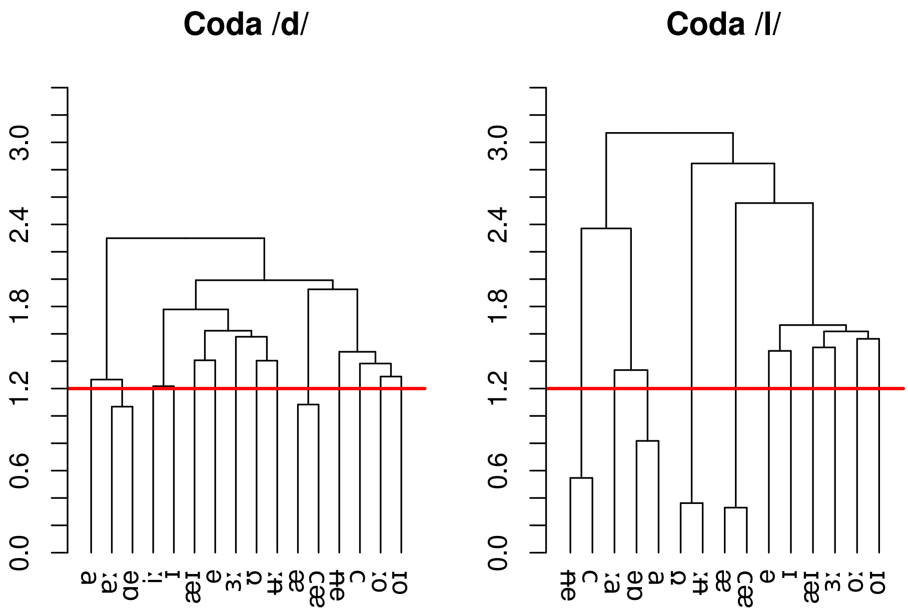

Figure 4.8: **Vowel confusablity in the /d/ and /l/ conditions based on accuracy data** The red line is an arbitrary cutoff line at 1.2.

Additionally, comparison of the accuracy tree diagrams for /d/ and /l/ on Figure 4.8 shows that the clades are located consistently lower in the /l/ than in the /d/ condition. Likewise, the clades are located consistently lower for /l/ than for /d/ in the RT tree diagrams on Figure 4.9. Therefore Figures 4.8 and 4.9 suggest that coda /l/ decreases the accuracy and increases the RT of vowel disambiguation. The HCA results based on accuracy data (represented in Figure 4.8) and RT data (Figure 4.9) correspond to the results of GLM showing that accuracy is significantly lower in the /l/ condition and RT has a non-significant tendency to be slower in the /l/ condition. (For the GLM results see Tables 4.1, 4.2 and 4.3).

Further comparison of the tree diagrams between codas show that not all of the vowel-pairs are affected equally by the change in the coda condition. For example, the clades dominating the pair /oː - oɪ/ and the pair /ʉː-

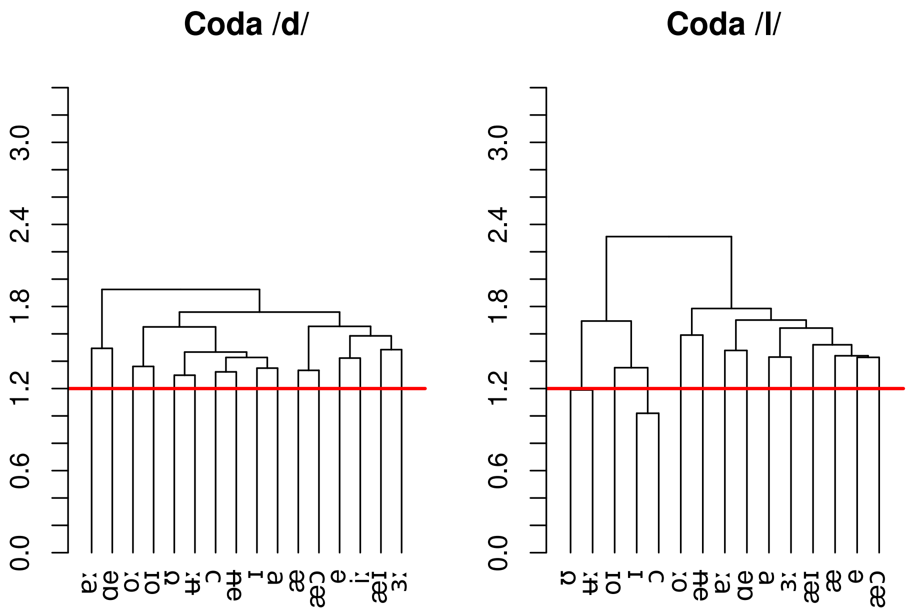

Figure 4.9: **Vowel confusablity in the /d/ and /l/ condition based on RT data.**The red line is an arbitrary cutoff line at 1.2.

ʊ/ are located at the same, slightly above average, height (around 1.4) in the coda /d/ condition both with respect to accuracy (Figure 4.8) and RT (Figure 4.9). In the /l/ condition, however, the clade dominating /ʉː-ʊ/ is located at 0.5 both with respect to accuracy and RT, whereas the clade dominating /oː - oɪ/ does not move as compared to the /d/ condition. This suggests an interaction between coda, target vowel, and competitor vowel; however, a cluster analysis does not reveal whether there is a significant interaction. Figures 4.8 and 4.9 also suggest that coda /l/ particularly affects the accuracy and RT of the pairs /ʉː-ʊ/, /æɔ-æ/ and /əʉ-ɔ/. In Figure 4.8, showing the clusters based on the accuracy data in the /l/ condition, the clades dominating these pairs are located below 0.6 on an arbitrary scale of 0 to 3, whereas the other two-member clusters are merged in the range of 1.2–1.6. The results of HCA suggest that the vowels /ʉː-ʊ/, /æɔ-æ/ and

/əʉ-ɔ/ are harder to disambiguate in the /l/ condition than the members of any other-vowel clusters. The disambiguation of the pairs /ʉː-ʊ/, /æɔ-æ/ and /əʉ-ɔ/ also seem to be made more difficult by coda /l/ than any other vowel pairs. This might indicate a reduced vowel contrast in the pre-lateral environment and a potential context-based vowel merger between the members of these pairs.

These results presented in Sections 4.5 and 4.4 partially supported the hypothesis that coda /l/ reduces accuracy and increases RT, as coda /l/ significantly reduces the accuracy of word disambiguation, but RT only shows a non-significant tendency to increase. The analysis have shown that the target-competitor pairs that agree in their articulatory-based SPE features have lower accuracy rates and longer RTs, therefore they are harder to disambiguate than those which do not. Additionally, HCA also showed that target and competitor vowel interact in such a way that long-short vowel pairs, which only differ in length and not in the other features are hard to disambiguate. Lastly, coda /l/ seems to most strongly affect the disambiguation of the pairs /ʉː-ʊ/, /æɔ-æ/ and /əʉ-ɔ/.

4.5 Exploratory results: the pairs /ʉː-ʊ/, /æɔ-æ/ and /əʉ-ɔ/

On the basis of the accuracy confusion matrices and the HCA of accuracy data, the vowel contrast between /ʉː-ʊ, æɔ-æ/ and /əʉ-ɔ/ seems to be reduced in the pre-lateral environment. RT data also appears to support this for /ʉː-ʊ, æɔ-æ/; only /əʉ/ clusters with /oː/ with respect to RT in the /l/ condition. This result is in line with Palethorpe & Cox (2003) who showed that the acoustic contrast between /ʉː-ʊ, æ- æɔ/, and /əʊ-ɔ/ is reduced, and might indicate a potential vowel merger triggered by /l/. Therefore, in this section the spectrogram of the stimulus of these six words are compared to diagnose whether acoustic similarity of the two stimuli could have caused slow RT, although in each trial participants only heard an individual word, not a pair. The RT of these target-competitor pairs are also compared with the RT of these 6 targets against all other competitors.

Figures 4.10, 4.12, and 4.14 show the first, second, and third formants of the vowels, plotted against time, and the spectrogram of the target words, all extracted from Praat (Boersma & Weenink 2013) for /ʉː-ʊ/, /æɔ-æ/ and /əʉ-ɔ/ respectively. Due to the fact that listeners responded as the acoustic information unfolded, the RT of the pairs /ʉː-ʊ/, /æɔ-æ/, and /əʉ-ɔ/ was plotted. These plots are presented in Figure 4.11, 4.13, and 4.15.

Figure 4.10 shows that in the /d/ condition, the F2 of /ʉː/ shows a central-front vowel with a long steady F2 (vowel length: 308.4 ms, F2 at midpoint: 2317 Hz), whereas the F2 of /ʊ/ shows a short back vowel (steady state vowel: 96.3 ms, F2 at midpoint of the steady state vowel: 990 Hz). In contrast, in the /l/ condition, the the formant structures of/ʉː/ and /ʊ/ only differ in the height of F3, as F3 is higher for /ʊ/. The F2 of /ʉː/ and /ʊ/ is highly similar. With respect to length, /ʉː/ is longer than /ʊ/ in both conditions. That is, /l/ reduced the spectral differences but not the length difference between /ʉː/ and /ʊ/ in the stimuli.

Figure 4.11 shows the RT to /ʉː/ and /ʊ/ in both coda conditions, when they were paired with each other as opposed to when they were paired with any other vowel. This figure shows that pairing /ʉː/ and /ʊ/ did not make disambiguation slower in the /d/ condition. However, pairing them increased RT in the /l/ condition. Coda /l/ made disambiguation slower for the /ʉː/ and /ʊ/ targets when /ʉː/ and /ʊ/ were paired with each other.

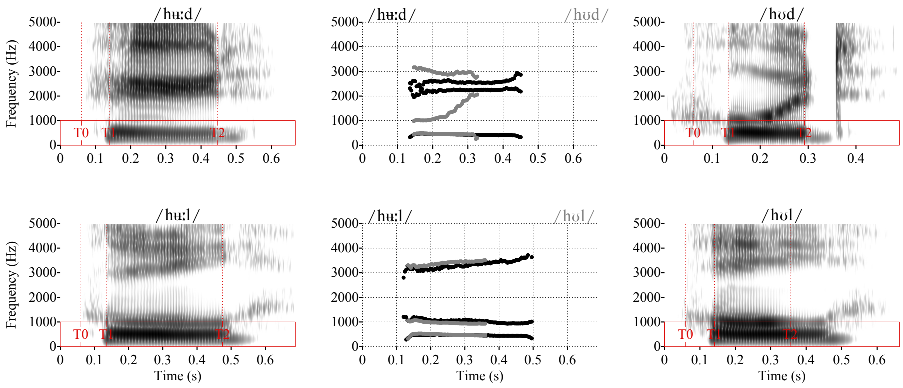

Figure 4.10: **Formant analysis of the vowels /ʉː/ and /ʊ/ in the /d/ and /l/ conditions.** Left panel: spectrograms of the carrier words of /ʉː/. Right panel: spectrograms of the carrier words of /ʊ/.Central panel: F1, F2, and F3 of the vowels /ʉː/ (black) and /ʊ/ (gray) plotted against real time. Upper panel: /d/ condition. Lower panel: /l/ condition.

Figure 4.12 shows that the F1 of /æ/ starts transitioning to the /d/ sooner than the F1 of /æɔ/. The F2 of the monophthong also starts rising due to the coarticulation with /d/ (Delattre et al. 1955) at the same point when the diphthong's F2 starts falling towards its second element, which has a low F2. In the /d/ condition, the length of the short monophthong (215.9 ms) and the diphthong (307.7) also differ in length. In contrast, in the /l/ condition, the formants of /æɔ/ and /æ/ are almost indistinguishable from each other: /æɔ/ does not show the acoustic characteristics of its second element, but the monophthong is shorter. Coda /l/ reduced the spectral but not the length difference between /æɔ/ and /æ/ in the stimuli.

Figure 4.13 shows the RT to /æ/ and /æɔ/ in both condition, when they were paired with each other as opposed to when they were paired with any other vowel. This shows that in the /d/ condition, disambiguation was slower when these two vowels were paired against each other than when they were paired against any other vowel. Pairing /æɔ/ and /æ/ made RT slower in the /l/ condition as well. Interestingly, the coda seems to matter more

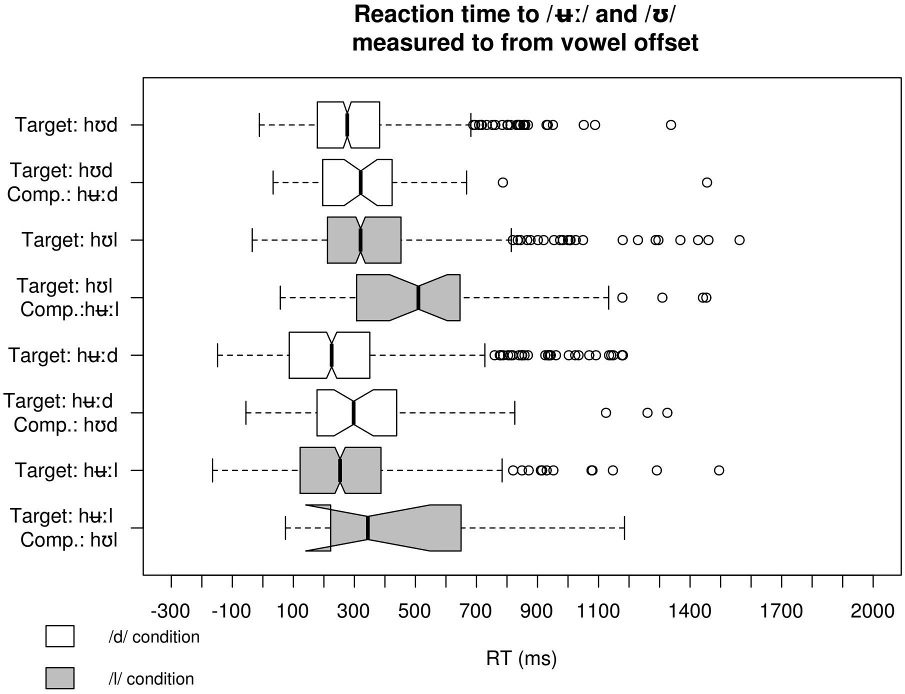

Figure 4.11: RT to the vowels /ʉː/ and ʊ/ when they were presented against each other as opposed to being presented against all other vowels in the coda /l/ (gray boxplots) and /d/ condition (white boxplots).

for those pairs in which the target vowels /æɔ/ and /æ/ competed against any other vowel than when they competed with each other.

Figure 4.14 shows that /əʉ/ and /ɔ/ differ spectrally in the /d/ condition, as the diphthong /əʉ/ has a higher F2 and the monophthong /ɔ/ has a higher F3. The formants of /ɔ/ show a back mid vowel (F1 and F2 at the midpoint of the steady-state vowel: 827 and 1242 Hz) and a transition to /d/. The F1 and the F2 of the diphthong show the transition and the target of the second element /ʉː/ with a low F1 and a high F2. In contrast, the formants coincide in the /l/ condition, in which the second segment of the diphthong appears to be lost and the formants only show the transition to /l/. The length difference between the diphthong and the monophthong does not

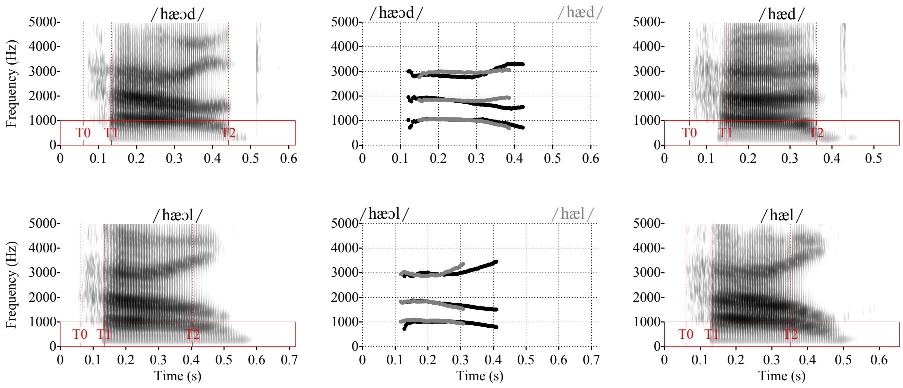

Figure 4.12: **Formant analysis of the vowels /æɔ/ and /æ/ in the /d/ and /l/ conditions.** Right panel: spectrograms of the carrier words of /æɔ/. Left panel: spectrograms of the carrier words of /æ/. Central panel: F1, F2, and F3 of the vowels /æɔ/ (black) and /æ/ (gray) plotted against real time. Upper panel: /d/ condition. Lower panel: /l/ condition.

change between the conditions. That is, /l/ reduced the spectral, but not the length difference for /ʉː/ and /ʊ in the stimulus.

Figure 4.15 shows the RT to /əʉ/ and /ɔ/ in both condition, when they were paired with each other as opposed to when they were paired with any other vowel. These RT results show that pairing /əʉ/ and /ɔ/ in the /d/ condition does not make disambiguation slower compared to the RT of the disambiguation from any other vowel. Coda /l/ appears to have made disambiguation slower even in the easy pairs, but this increase in RT seems to be bigger when /əʉ/ and /ɔ/ were paired with each other.

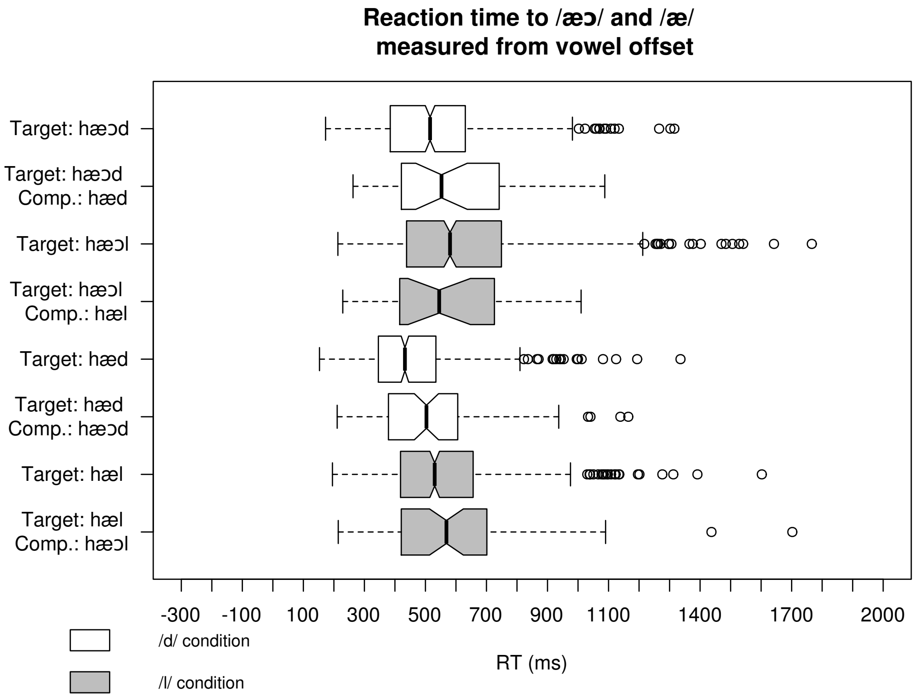

Figure 4.13: RT to the vowels /æɔ/ and /æ/ when they were presented against each other as opposed to being presented against all other vowels in the coda /l/ (gray boxplots) and /d/ condition (white boxplots).

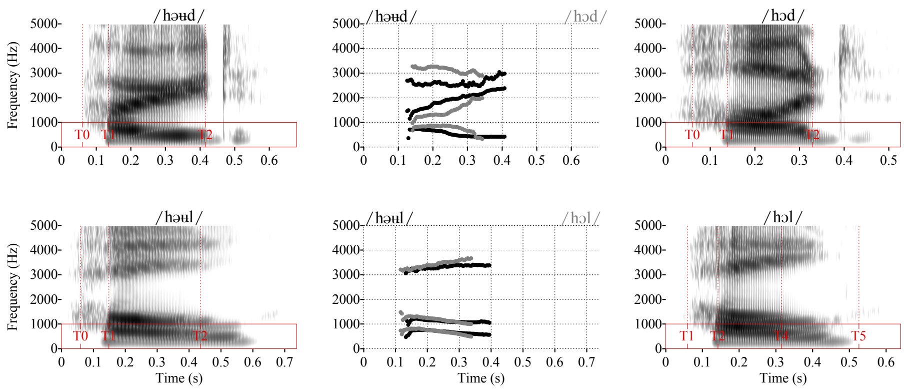

Figure 4.14: **Formant analysis of the vowels /əʉ/ and /ɔ/ in the /d/ and /l/ conditions.** Right panel: spectrograms of the carrier words of /əʊ/. Left panel: spectrograms of the carrier words of /ɔ/. Central panel: F1, F2, and F3 of the vowels /ʉ/ (black) and /ɔ/ (gray) plotted against real time. Upper panel: /d/ condition. Lower panel: /l/ condition.

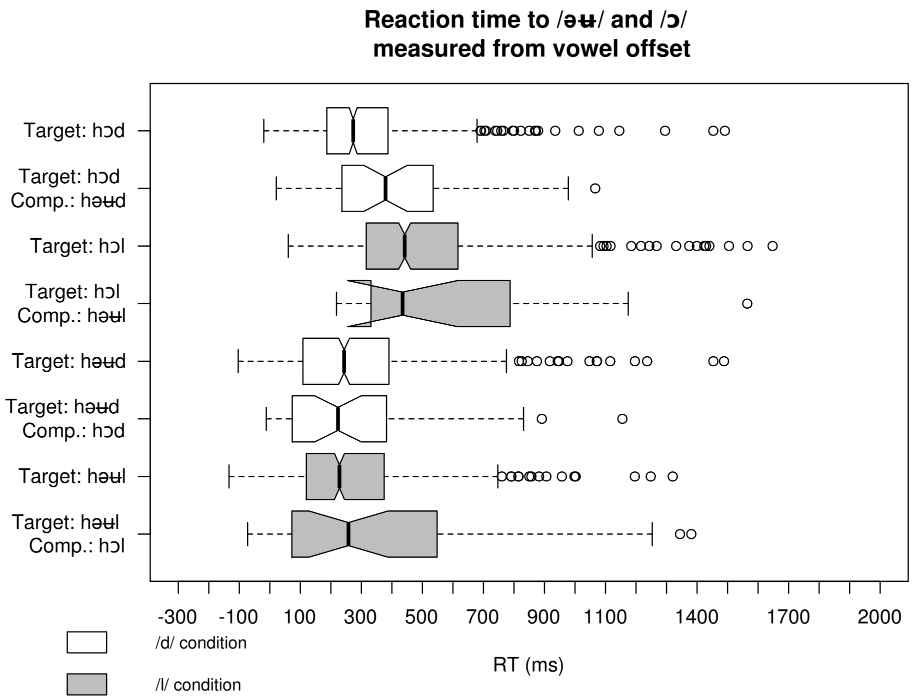

Figure 4.15: RT to the vowels /əʉ/ and /əʉ/ when they were presented against each other as opposed to being presented against all other vowels in the coda /l/ (gray boxplots) and /d/ condition (white boxplots).

4.6 Summary of findings

The results described in Sections 4.3, 4.4 and 4.5 can be summarised as the following:

1. Accuracy is lower and RT is slower for spectrally and articulatorily similar target-competitor pairs.

2. Accuracy of word disambiguation is significantly lower and RT is non-significantly slower for lateral-final than for obstruent-final words.

3. The vowel-pairs /ʉː-ʊ/, /æɔ-æ/ and /əʉ-ɔ/ show reduced vowel contrast.

5 Discussion

The aim of this study was to examine in new detail the influence of coda laterals on disambiguation of vowel contrasts in AusE. If the perception of vowel contrast is shown to be compromised in the /l/ condition, which vowel pairs show potential perceptual mergers and why? Hindered vowel disambiguation and reduced vowel contrast was hypothesised to be evident through decreased accuracy and increased RT. As predicted, these data revealed that vowel disambiguation is slower, and significantly less accurate in pre-lateral environments. These data are consistent with the hypothesis that coda /l/ reduces vowel contrast. On the basis of low-pairwise accuracy and long pairwise RT in the /l/ condition, three spectrally similar target and competitor pairs were discovered that show a potential context-based vowel merger in AusE: /ʉː-ʊ/, /æɔ- æ/ and /əʉ-ɔ/ in the pre-lateral environment. In addition, the data also revealed that target- competitor pairs that share spectral similarities and differ in length, such as /ɐː-ɐ/ and /iː-ɪ/, are disambiguated less accurately and more slowly than spectrally different target-competitor pairs. In this section, we consider the results of this experiment in light of previous findings, and discuss the implications for our understanding of phonological and lexical representation, with reference to exemplar phonology. The implications of the results for the representation of postvocalic /l/ in the models of syllables presented in Section 2 are also described. Lastly, the limitations of this study are discussed.

5.1 Intrinsic perceptual similarities in AusE vowel disambiguation: long-short pairs

The data gained in the /d/ condition show that whereas spectrally different vowels are easily disambiguated from each other, the members of the long-short vowel pairs are intrinsically easier to confuse and take the longest to disambiguate. That is, vowels which were perceived as similar to each other in this experiment can be grouped together in the vowel space, as in Figure 5.1. Figure 5.1 shows a vowel map of AusE in which vowels are placed according to their place of articulation (Cox 2012). Vowels are colour-coded according to their pairwise confusability on the basis of HCA of accuracy data in the /d/ condition in such a way that the tree diagram in Figure 4.8 was cut at the lowest point where there were no single-member clusters. Figure 5.1 shows groups according to vowel frontness, because front vowels are grouped with front vowels, but never with non-front vowels. Within the front vowels, there are three perceptually distinct groups made of the high front vowels (/iː-ɪ/) and two groups of non-high front vowels (/e-æɪ/ and /æ- æɪ/). Central vowels do not form a separate group, but they pattern with back vowels according to their height. That is, there are three group consisting of central and back vowels: the high-mid group (/ʉː-ɜː-ʊ/), the mid-group (/əʉ-ɔ-oː-oɪ/) and the low group (ɐː-ɐ-ɑe)). Long and short vowels, however, do not form two opposing groups, but are dispersed between the front and non-front groups according to their place of articulation. The diphthongs pattern on the basis of the place of articulation of their first element, as it is shown by the tree diagrams in Section 4.4. The diphthongs /əʉ/ and /oɪ/ pattern with the mid monophthongs and /ɑe/ patterns with the low monophthongs. The three diphthongs beginning with /æ/ pattern with the front monophthongs, however, front-rising /æɪ/ patterns with front /e/, and back-rising /æɔ/ patterns with /æ/. The patterning of perceptually similar vowels in AusE corresponds to the acoustic descriptions of the AusE long-short vowel pairs, including the short monophthong-diphthong pairs.

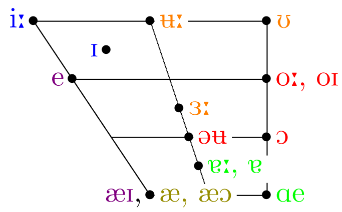

Figure 5.1: **16 AusE vowels in a vowel map.** Monophthongs are placed according to their place of articulation. Diphthongs are placed according to the place of articulation of their first element. All vowels are colour-coded according to their clustering in Figure 4.8.

The fact that spectrally similar vowel pairs that differ in length are less easily disambiguated may be explained with respect to the acoustic similarity of the vowels, the processing of the vowels, and the mental representation of the vowels.

Firstly, within each condition, listeners were exposed to the acoustic signal of all the words in the experiment in the full paradigm for that condition, because they were trained on the stimulus prior to the experiment, and because all the words were used as targets and competitors. Vowel discrimination studies can help us explain the results of the lexical decision task used here. In order to make correct lexical choices, listeners had to be able to discriminate between the vowels they were exposed to. Listeners are sensitive to formant changes above a threshold of 12–36 Hz (Kewley-Port &

Watson 1994, Kewley-Port & Zheng 1999) or 33–92 Hz (Mermelstein 1978), and participants did have spectral cues in the stimulus. If listeners rely on spectral information that is available early, the long RT can be explained by the longer time required to process the information and make a decision when the differences are smaller. Low accuracy rates raise the question of whether the acoustic difference between two stimulus vowels reach the threshold of formant discrimination in this experiment.

Spectral differences might have been small between target and competitor such as between /ɐː and /ɐ/, but the target and competitor pairs differed in length, a property of vowels to which listeners are sensitive (Nearey & Assmann 1986). If listeners rely on length cues, the acoustic information necessary to make a decision is not available until the acoustic signal of a short target reaches the coda, or the long target exceeds the length of a short vowel. This leads to slower RT because the listeners need to wait until the point of disambiguation, whereas listeners can make decisions early on for spectrally different vowels.

Secondly, studies conducted using the cohort model found that an acoustic stimulus activates acoustically similar competitors but does not activate dissimilar competitors (Allopenna et al. 1998, Grosjean 1980, Tanenhaus et al. 1995; 2000). Applying the results of the cohort model to the present experiment means that the orthographic forms activate the pronunciation of the target and competitor word, and the acoustic signal is compared with the representation of the pronunciation of the target- and competitor word. As a result, the size of the difference between the pronunciation of the target words and competitor words affects the accuracy and speed of disambiguation. The smaller the difference between target and competitor, the more strongly the target activates the competitor, leading to reduced accuracy and increased response latency. The fact that listeners were less accurate and slower in the /l/ condition accords with the idea that vowels followed by /l/ are more similar to each other. That is, coda /l/ might reduce listeners' ability to disambiguate certain vowels thereby potentially reducing contrast.

Thirdly, the acoustic stimulus might not only activate the mental representation of a similar competitor, but the stimulus can also be ambiguous between the mental representation of the two candidates. The token of /æɔ/ in *howl* might be equally similar to the mental representation of /æɔ/ and /æ/, causing the stimulus to overlap with the mental representation of the competitor. The spectral qualities and the length of a stimulus might overlap with the mental representation of another vowel as it is stored in the listeners' lexicon. In exemplar phonology, listeners' mental representation of a phoneme emerges from perceptual experience of multiple tokens (Docherty & Foulkes 2014), therefore the exemplar clouds of two vowels might overlap. Both the overlap between the signal and the competitor's mental representation and the overlap between the mental representation of the target and the competitor are important considerations in accounting for these results in this model.

In summary, disambiguation of pairs of spectrally similar vowels that only differ in length as target and competitor requires time for length cues to unfold and requires time for listeners to process disambiguating spectral information. Spectral similarity between target and competitor also leads to strong lexical activation of the competitor during processing. In addition, spectral similarity may also lead to possible overlaps between the stimulus and listeners' mental representations. All of these factors can contribute to the lower accuracy rates and to slower RT of spectrally similar target and competitor pairs compared to the spectrally different pairs.

5.2 Disambiguation of AusE vowels before laterals

5.2.1 Accounting for changes in vowel disambiguation in pre-lateral environments

The possible explanations proposed in Section 5.1 for the increased difficulty listeners have disambiguating spectrally similar target-competitor pairs are relevant for the /l/ condition as well. In addition to the inherent perceptual similarities, coda /l/ causes words to be overall significantly less accurately disambiguated and show a tendency to be more slowly disambiguated than words ending in coda /d/. The difference between the two conditions might be explained by coda /l/ modifying the acoustic qualities of the nucleus vowel. Acoustic changes in the pre-lateral vowels might have two consequences: reduced contrast between the formant structures of target and competitor, and an overlap between the stimulus and the mental representation of target and competitor.

Reduced acoustic contrast between the vowels was found in the stimulus words of the experiment (see Figures 4.10, 4.12, 4.14), corresponding to the acoustic research on pre-lateral vowels in AusE by Cox & Palethorpe (2004) and Palethorpe & Cox (2003). Reduced vowel contrast caused by coarticulation can contribute to the lower performance of listeners' in the /l/ condition. Listeners, however, are also known to be able to compensate for variation in the acoustic signal caused by phonetic context (Kewley-Port 1995, Ohala 1981). However, our results show that listeners had high rates of confusion in the /l/ condition for the vowel pairs /ɔː-ʊ/, /æɔ-æ/ and /əʉ-ɔ/, indicating that compensation for coarticulation in the /l/ context was a challenge.

The 76% confusion rate of /ʉː/ with /ʊ/ for example shows that the listeners in this experiment did not recognise the vowel in *hule* as /ʉː/, although listeners did recognise the vowel in *hude* as /ʉː/, as shown by confusion rate in the /d/ condition (9%). The vowel in *hule* was identified as /ʊ/. This accuracy result might have been affected by the fact that both *hule* and *hool* were non-words, but a non-word effect has difficulties in explaining the asymmetry of the confusion rates for the /l/ condition: /ʉː/ was identified as /ʊ/ in 76% of the trials, but /ʊ/ was identified as /ʉː/ in only 29% of the trials. This result is compatible with the idea that the formant differences between /ʉː/ and /ʊ/ did not reach the threshold for formant-frequency discrimination under the conditions of the present experiment. However, the present thesis cannot confirm that the formants of /ʉː/ and /ʊ/ do not reach the threshold of discriminability in the /l/ context.

The confusion rate of /ʉː/ with /ʊ/ is also compatible with the idea that the acoustic signal of vowels is discriminable from their competitor in the /l/ context, but the acoustic signal of the vowel in /hʉːl/ overlaps with the mental representation of /ʊ/. When listeners were presented with the acoustic signal of *hule*, they compared it to their mental representation of /ʉː/ and /ʊ/, and they found that it is more similar to /ʊ/. As there was no evidence for the contrary (e.g. semantic or grammatic context), listeners identified the word as *hool*. In addition, in exemplar phonology, the mental representation of /ʉː/ and /ʊ/ may contain pre-obstruent tokens and pre-lateral tokens as well, and the pre-lateral tokens of the two vowels might overlap.

The present results are compatible with the explanation that the vowels' formants are below the threshold of formant frequency discrimination or were below the threshold under the circumstances of the experiment. It is also compatible with the explanation that the stimulus of target words presented individually overlap with the mental representation of the target and the competitor, and that the mental representations of the target and competitor overlap. It is beyond the scope of this thesis to answer the questions whether the results are caused by listeners' difficulties in formant discrimination, in processing, or by listeners' overlapping mental representations.

5.2.2 Contrast reduction in pre-lateral vowels /l/

The accuracy and the RT results of the vowel pairs /ʉː-ʊ/, /æɔ-æ/ and /əʉ-ɔ/ allows us to draw conclusions on potential AusE vowel mergers in the pre-lateral context. The confusion rates of these vowel pairs were presented in Table 4.8, and are repeated here as Table 5.1 for the convenience of the reader. The high percentage of inaccurate responses provides evidence for a reduced vowel contrast in the pre-lateral environment. Table 5.1 also shows that the relationship between the members within a pair might be different: /ʉ-ʊ/ and /əʉ-ɔ/ show asymmetric confusion rates, whereas /æɔ-æ/ show symmetric confusion rates in the /l/ condition.

		Competitor	
		ʉː	ʊ
Target	ʉː	-	76
	ʊ	29	-

		Competitor	
		æɔ	æ
Target	æɔ	-	49
	æ	56	-

		Competitor	
		əʉ	ɔ
Target	əʉ	-	24
	ɔ	64	-

Table 5.1: **Inaccurate responses (%) for the target-competitor pairs /ʉː-ʊ/, /æɔ-æ/ and /əʉ-ɔ/ in the /l/ condition**. Rows: targets. Columns: competitors. Data repeated from Table 4.8

For /ʉ-ʊ/, the long vowel is identified as the short vowel in 76% of the cases, that is, listeners almost consistently select /ʊ/ instead of /ʉː/. In contrast, short /ʊ/ is only identified as /ʉː/ in 26% of the cases. This suggests that when listeners were exposed to either /ʉː/ or /ʊ/, they selected /ʊ/ for both targets. RT data from the /l/ condition shows, however, that when /ʉː/ and /ʊ/ were paired, listeners took a longer time to identify both /ʉː/ and /ʊ/ than when these vowels were paired with any other competitors. That is, just because listeners seem to default to /ʊ/ based on the accuracy data, the RT data shows that they do not default to /ʊ/ quickly. The RT data also shows that pairing the two vowels with each other in the /d/ condition does not make disambiguation slower, that is vowel contrast is not reduced across the board. RT data also shows that /l/ makes disambiguation slower when the vowels are paired with each other, but not across the board. Therefore accuracy and RT data suggest that there might be a conditional /ʉː-ʊ/ merger in the direction of /ʊ/ in the prelateral environment that slows down the disambiguation of both members of the /ʉː-ʊ/ pair.

When it comes to the /æɔ- æ/ pair, there is no asymmetry in the accuracy data, as both the long and the short vowel were identified at chance level when they were paired with each other. RT data in Figure 4.13 shows that pairing the two vowels made disambiguation slower in both the /d/ and the /l/ condition, that is the vowels /æɔ- æ/ appear to be intrinsically similar in perception. Accuracy and RT data suggest that the disambiguation of the intrinsically similar vowels /æɔ/ and /æ/ is made harder by coda /l/ but do not suggest a potential merger into the direction of any of the members of the pair.

In the case of the /əʉ-ɔ/ pair, accuracy data again suggest an asymmetry because the identification of the target /ɔ/ is worse than chance when it is paired with /əʉ/ (64%), but /əʉ/ is only confused with its short pair in 24% of the comparisons. RT data is in line with this results, as /ɔ/ seems to be disambiguated more slowly than /əʉ/ in Figure 4.15 in the /l/ condition.

That is, coda /l/ makes the identification of the short vowel harder across the board. This result can be influenced by the carrier words, as the carrier word for /əʉ/ was a real word (*hole)*, and for /ɔ/ was a non-word (*holl*). The result can also signal a potential /əʉ-ɔ/ merger towards /əʉ/.

Thus coda /l/ can make the disambiguation of a vowel harder across the board, and it can make two dissimilar as well as two similar vowels hard to disambiguate. All of which might result in potential mergers in the pre-lateral context, and in the pre-lateral context only, as listeners' accuracy was at ceiling in the /d/ condition.

5.3 Implications for the syllable structure of /l/-final rimes

As it was discussed in Section 4, in general, coda /l/ increases vowel confusability compared to coda /d/. That is, words with lateral codas are harder to disambiguate for listeners across the board. Additionally, coda /l/ leads to potential vowel mergers for /ʉː-ʊ/, /æɔ-æ/ and /əʉ-ɔ/ in perception. These results suggest that a smaller subset of AusE vowels may be robustly contrastive in pre-lateral environments, compared to the full set of 18 stressed vowels that contrast before coda /d/. The fact that /l/ and /d/ behave differently might indicate that coda /l/ has a stronger relationship with its nucleus than coda /d/.

This difference between lateral codas and obstruent codas is not captured by generative and moraic syllable theories, because in neither model do different types of codas have a different structural relationship with the nucleus per se. Figures 5.2 and 5.3 show that neither the generative nor the moraic modelling of coda /l/ captures the difference between /l/ and /d/.

In the generative model, in Figure 5.2, /l/ and /d/ occupy the same coda position, and they are equally distant from the vowel in the nucleus. In the moraic model, in Figure 5.3, neither /d/ nor /l/ carries a mora because consonants are not inherently moraic and neither /d/ nor /l/ is assigned a mora after a long vowel (Hayes 1989).

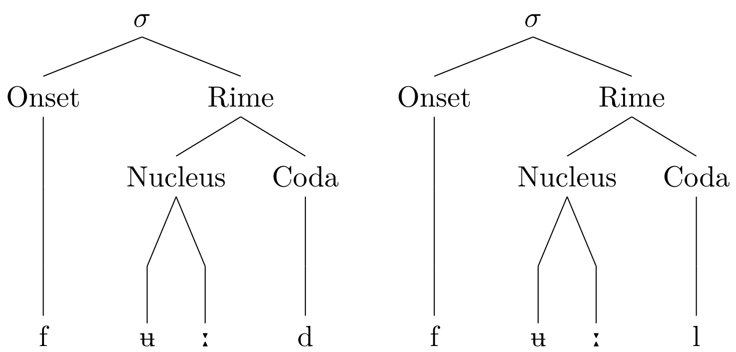

Figure 5.2: **Coda /d/ and coda /l/ in generative model of the syllable.** Left panel: *food*. Right panel: *fool*

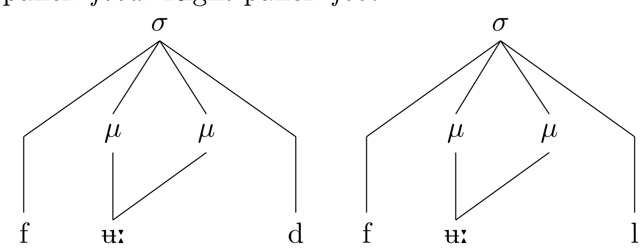

Figure 5.3: **Coda /d/ and coda /l/ in moraic model of the syllable.** Left panel: *food*. Right panel: *fool*

Despite having the same structural representation, obstruent-final rimes and lateral-final rimes differ with respect to articulation (Campbell et al. 2010, Gick & Campbell 2003), acoustics (Palethorpe & Cox 2003), syllable count judgements (Lavoie & Cohn 1999, Tilsen et al. 2014) and according to the result of this thesis, in perception too. Cumulative evidence from these studies can lead to the question of whether lateral codas should be represented differently from obstruent codas. A possible representation in a generative model would be to propose that /l/ forms a nucleus instead of the coda, as in Figure 5.4. In the moraic model, Lavoie & Cohn (1999) proposed that coda /l/ carries a mora, whereas coda /d/ has not , as in Figure 5.5.

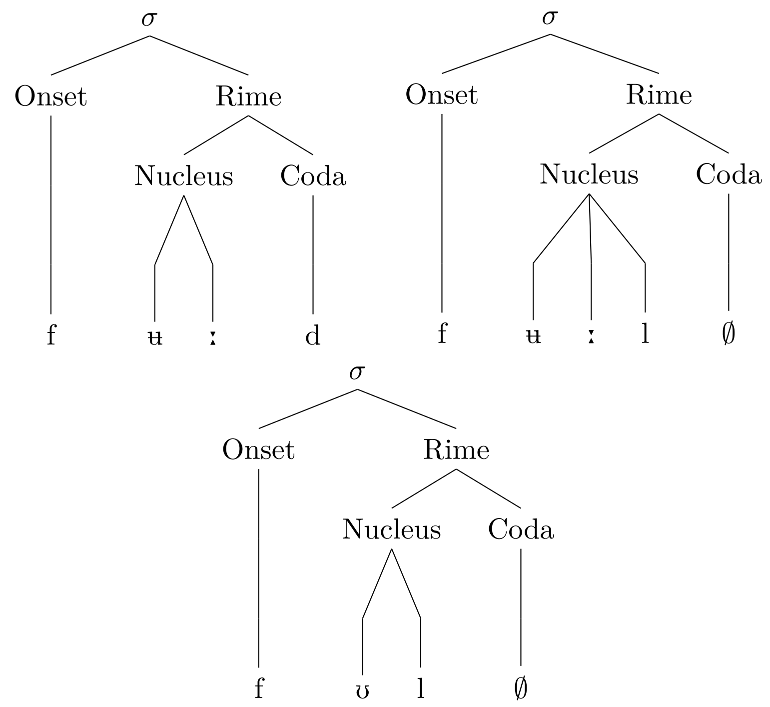

Figure 5.4: **Coda /d/ and nucleus /l/ in the generative model of the syllable.** Upper left panel: *food.* Upper right panel: *fool.* Lower panel: *full*

Both the representations in Figure 5.4 and 5.5 can capture aspects of the perceptual difference between postvocalic /l/and /d/ by representing the different codas with different structures. The model presented in Figures 5.4 predicts a reduced vowel contrast in the pre-lateral environment, because it shows that the nucleus of the words *fool* and *full* have the same second element. The model presented in Figure 5.4 predicts differences between lateral-final and obstruent-final syllables by assigning a mora after a long vowel to /l/ but not to /d/, although a mora-bearing coda is not

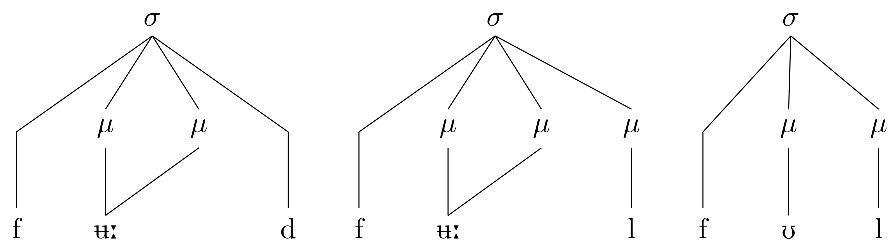

Figure 5.5: **Non-moraic /d/ and moraic /l/ in the moraic model of the syllable.** Left panel: *food*. Central panel: *fool*. Right panel: *full*

necessary after a long vowel to meet the bimoraic minimum word requirement (see Lavoie & Cohn 1999, Hayes 1989)). These representations also have the additional benefit of offering insights into the way in which /l/ is often vocalised, forms a peak of the nucleus (Borowsky 2001, Horvath & Horvath 2002), as well as an account of the tendency of laterals to resist resyllabification (Gick & Wilson 2003). The model in Figure 5.4 motivates /l/ vocalisation because vowels are preferred as nuclei over sonorant consonants (Blevins 2006a) and does not predict the resyllabification of /l/ because a segment in a nucleus is not expected to resyllabify. The moraic model in Figure 5.5 predicts a vocalised after long vowels /l/ which is consistent with the fact that long vowel facilitates /l/ vocalisation in AusE (Borowsky 2001, Horvath & Horvath 1997; 2002).

However, in the generative model, proposing that /l/ is a nucleus leads to a ternary branching nucleus when the vowel is long, which is disallowed. In the moraic model, adding an extra mora to a syllable when the vowel is long leads to a trimoraic syllable, which is again dispreffered in English (Lavoie & Cohn 1999). The fact that pre-lateral long /ʉː/ is merged with short /ʊ/ is concurrent with the hypothesis that a ternary branching nucleus or a trimoraic syllable is dispreferred. This is due to the fact that if postvocalic /l/ is assigned a nuclear status or a mora, long /ʉː/ and /l/ leads to a ternary branching nucleus, but short /ʊ/ does not, as presented in Figure 5.4. In the moraic model presented in Figure 5.5, analysing /l/ as mora-bearing unit only leads to a trimoraic syllable, if the vowel is long /ʉː/. That is, merging the long vowel with the short vowel makes the syllabic structure of the word grammatical. In contrast, the potential merger between /æɔ/ and /æ/ does not show that the direction of the merger is towards the short vowel. The potential merger between /əʉ/ and /ɔ/ seems to be going to the direction of the long vowel, as /ɔ/ was more often identified as /əʉ/ than the other way around. Therefore, the overall data is not easily reconciled with either of these standard models of syllable structure.

The effect of coda /l/ on its nucleus can also be accounted for in terms of coarticulation, without proposing alternative representations for lateral-final rimes in the structure of syllable (e.g. Gick & Wilson 2003). Coarticulation can also contribute to vowel disambiguation becoming harder in the pre-lateral environment when it reduces vowel contrast. As listeners can predict and compensate for coarticulation up to a point (Beddor et al. 2013, Fowler 2005, Kewley-Port 1995), the question is to what extent does the coarticulatory influence of coda laterals affect vowel production in AusE, and when does it exceed the limits of listeners' ability to compensate for coarticulation. In this experiment listeners could expect and predict the coda, as listeners were only exposed to one condition. That is, the experiment had low stimulus uncertainty, which aids vowel disambiguation (Kewley-Port & Watson 1994). Despite listeners general sensitivity to coarticulation and the low stimulus uncertainty, coda /l/ still led to a significantly lower accuracy rate and to a non-significantly longer RT.

Thus, the results of this experiment bring more data to bear on the complex issue of the extent to which lateral final rimes differ from obstruent final rimes. However, it is still an open question whether models of the syllable should capture this, and if so, how.

5.4 Limitations of the study

Testing 16 vowels pairwise led to an exhaustive testing of the disambiguation of AusE vowels in the pre-lateral environment, but also set limitations on the methodology of this experiment.

Testing 16 vowels comprehensively led to gaining a better overall understanding of the AusE vowel space, but it also generated a complex dataset which made it difficult to examine all details of perception and interactions between factors of interest. As a result of testing 16x15=240 comparisons, the independent variables target and competitor vowels each had 16 levels, which did not allow for the analysis of interactions between target and coda with 30 participants. The effects of /l/ on individual vowels and vowel-pairs might have been masked by the high number of target and competitor pairs.

When selecting the target words non-words were mixed with high- and low-frequency real words to address lexical gaps in AusE. This might have affected the results, because lexical frequency and real-word status affect RT word recognition (Forster & Chambers 1973, Meunier & Segui 1999, Rubenstein et al. 1970, Segui et al. 1982). Although an independent binary variable coding lexical status (word/non-word) was introduced to the statistical models, lexical frequency of the target and the competitor words was not included in the analysis. Additionally, it is an open question if listeners built a mental representation for the non-words during this experiment (see Escudero et al. 2008).

During the aural presentation of the target words, natural recordings were used. This led to a loss of control over the exact parameters that might cause difficulties. The recordings came from a single speaker, which limits the generalizability of the results. Lexical candidates in the decision task were presented orthographically, which might have affected listeners' performance, especially because non-words do not have an established orthographic form. Morphologically complex words were also respelled as non-words, increasing the number of non-words and creating a mismatch between the real-word audio and the non-word orthography.

Using a binary choice task in which listeners only hear the target word but not the competitor word does not allow to tell whether listeners can hear the difference between nearly identical vowels in the /l/ condition. It only allows to tell that listeners misidentify certain targets when they compare their mental representation of the two words presented orthographically to the audio stimulus. That is, it is impossible to say whether the high confusion rates are caused by listeners having the same representation for *hule* as for *hool* or by that the signal was ambiguous between two overlapping representations.

A final, central limitation of this study is that both accuracy and RT data have their limitations as metrics providing insights into the complex range of phenomena being explored here. A limitation of accuracy data is that participants were at a ceiling in the overall task due to the high number of target-competitor pairs, although they were at or below chance in the hard target-competitor pairs. That is, the accuracy data of participants who are at ceiling might provide poor data. The limitation of RT is that it is only a crude estimate of participants reaction to linguistic stimulus. RT includes stimulus detection time and movement initiation (Woods & Reed 2015), and it is not possible to tell how much time participants spent on detecting the stimulus, making a meta-linguistic judgement and then configuring the movement to press the button. Moreover, accuracy and RT might not be independent from each other due to the speed-accuracy tradeoff; however, some models propose that guesses are faster than answers based on sufficient evidence, whereas other models propose that guesses are slower (Osman et al. 2000).

As a result, further studies are required that target specific vowel pairs in AusE to avoid masking by the high number of comparisons. This also allows better testing of more specific hypothesis on the reduced contrast between /ʉː-ʊ, æɔ-æ, ʉ-ɔ/ in pre-lateral environments. This reduced contrast might have implications for conditional sound change, if the same phenomena were observed across more speakers and a broader range of linguistic data.

This can be researched by using different methods and paradigms. For example, eye-tracking can overcome the limitations of RT, as it does not require metalinguistic judgement and makes time-locking reaction to the acoustics of the stimulus easier. Using the ABX paradigm instead of a two-way alternative forced-choice might provide information on whether listeners' inaccuracy was caused by the word recognition task used in the present experiment. In addition, articulatory studies can provide insight on potential articulatory causes for the contrast reduction found in perception and on /l/ vocalisation as well.

Thus, several further studies can be built on the results of this thesis that are relevant for the better understanding of AusE and of the representation of coda /l/.

6 Conclusion

The aim of this thesis was to examine how lateral codas affect perception and identification of Australian English vowels. These data provide the first systematic study of vowel disambiguation across the full Australian English vowel space, in lateral and non-lateral rimes. The results of this thesis therefore contribute both to the literature on AusE and on coda /l/.

Firstly, the results provide evidence that vowel identification becomes harder across the board for each AusE vowel in the pre-lateral environment. In addition, three vowel-pairs, /ʉː-ʊ, æɔ-æ, əʉ-ɔ/, have been identified which may be undergoing a context-based merger in perception. Reduced contrast for these vowel-pairs is absent from the pre-obstruent environment, indicating that they are contextual vowel mergers triggered by coda /l/. Identifying these vowel-pairs empirically supports the results of acoustic studies which found little spectral differences between the members of these vowel-pairs (Palethorpe & Cox 2003) and adds new information to our understanding of AusE, because a merger in acoustics does not necessarily imply a merger in perception.

Secondly, finding evidence for reduction of contrast between /ʉː-ʊ, æɔ-æ, ʉ-ɔ/ before lateral codas suggests that coda /l/ reduces the phonetic difference compared to coda /d/. This provides an additional example in which coda /l/ interacts with its nucleus differently from an obstruent coda.

The representation of coda /l/ is the subject of ongoing discussion in the literature. Research has shown that postvocalic /l/ displays ambiguous and unstable syllabic affiliation between two syllables or between nucleus and coda (e.g. Campbell et al. 2010, Cox & Palethorpe 2004, Gick et al. 2002, Lavoie & Cohn 1999, Loakes et al. 2010a; 2014b, Sproat & Fujimura 1993, Tilsen et al. 2014, Palethorpe & Cox 2003). These perceptual data contribute to the growing body of evidence that coda /l/ behaves differently from coda /d/ within the syllable rime.

A Questionnaires used in the experiment

A.1 Language background questionnaire

Perception of laterals in Australian English - language background questionnaire

Name: ____________________

E-mail: ____________________

Mobile: ____________________

Gender: M/F

Date of Birth: ____________________

Place of Birth: ____________________

Residential history: *(Please list every city and country where you have lived for at least a year, and your approximate ages in each place):*

____

____

____

____

Primary school(s): ____________________

____

Secondary school(s): ____________________

____

Main language: *What language(s) do you mainly speak at home?*

____

____

Other languages spoken: *What other language(s) can you speak or understand?*

____

____

Mother's place of birth: ____________________

Mother's language(s): ____________________

Mother's occupation: ____________________

Father's place of birth: ____________________

Father's language(s): ____________________

Father's occupation: ____________________

Hearing: *Do you have, or have you ever had, any hearing problems?*

____

Speech: *Do you have, or have you ever had, any speech problems?*

____

Reading: *Do you have, or have you ever had, any reading problems?*

____

A.2 Exit interview

Perception of laterals in Australian English - exit interview

I paid attention to all the words I heard.

○ ○ ○ ○ ○ ○ ○

Strongly disagree Strongly agree

I read all the printed words.

○ ○ ○ ○ ○ ○ ○

Strongly disagree Strongly agree

Most of the times, I was just guessing the words.

○ ○ ○ ○ ○ ○ ○

Strongly disagree Strongly agree

I made mistakes I noticed right after pressing the button.

○ ○ ○ ○ ○ ○ ○

Strongly disagree Strongly agree

I found the task difficult

○ ○ ○ ○ ○ ○ ○

Strongly disagree Strongly agree

I found the task slow paced.

○ ○ ○ ○ ○ ○ ○

Strongly disagree Strongly agree

I found the task fast paced.

○ ○ ○ ○ ○ ○ ○

Strongly disagree Strongly agree

I found the task boring.

○ ○ ○ ○ ○ ○ ○

Strongly disagree Strongly agree

B Trials in the test phase

Vowel 1	Response option 1	Vowel 2	Response option 2 3
æɪ	hade	ɔɪ	hoyd
æɪ	hade	æɔ	howd
ɔ	hod	ʉː	hude
ɔ	hod	ɔɪ	hoyd
əʉ	hode	æɔ	howd
əʉ	hode	ɐ	hud
æɔ	howd	æɪ	hade
æɔ	howd	əʉ	hode
ɔɪ	hoyd	æɪ	hade
ɔɪ	hoyd	ɔ	hod
ɐ	hud	ʉː	hude
ɐ	hud	əʉ	hode
ʉː	hude	ɐ	hud
ʉː	hude	ɔ	hod

Table B.1: **Response pairs in the training prior to the experiment in the /d/ condition.** The participants were presented with the audio stimulus of one of the words and were instructed to select the word from the response options.

Vowel 1	Response option 1	Vowel 2	Response option 2 3
æ	hal	ɔ	holl
æ	hal	ʊ	hool
ɐː	harl	ʉː	hule
ɐː	harl	ɑe	hile
ɑe	hile	ɐː	harl
ɑe	hile	ɜː	hurl
ɔɪ	hoil	ɜː	hurl
ɔɪ	hoil	ʊ	hool
ɔ	holl	ʉː	hule
ɔ	holl	æ	hal
ʊ	hool	ɔ	holl
ʊ	hool	æ	hal
ʉː	hule	ɐː	harl
ʉː	hule	ɔ	holl
ɜː	hurl	ɑe	hile
ɜː	hurl	ɔ	hoil

Table B.2: **Response pairs in the training prior to the experiment in the /l/ condition.** The participants were presented with the audio stimulus of one of the words and were instructed to select the word from the response options.

C R scripts

C.1 Script for getting the main effects

```
#Tunde Szalay, 2016
#This is the script used to get main effects
#I always use General Linear Model (function: glmer)
#with Subject as a random factor.
#I always use anova to compare two models:
#one without a given factor and with a full model with all the factor

#Step 0: Loading the libraries
library(car)
library(lme4)
#Step 1 Analysing accuracy results
#Step1.1 creating a maximal model without an interaction
accuracy_max=
glmer(TargetSoundOut.ACC~Coda+TargetVowe+CompetitorVowel
+RealWord+(1|Subject),
family = binomial(), data = alldata_clean3)
summary(accuracy_max)
```

```
#Step 1.2 dropping factors one by one and comparing the models
#removing Coda
accuracy3=
glmer(TargetSoundOut.ACC~TargetVowel+CompetitorVowel+RealWord+(1|Subject),
family = binomial(), data = alldata_clean3)
summary(accuracy3)
anova(accuracy_max, accuracy3) #significant -> Coda

#removing TargetVowel
accuracy3=glmer(TargetSoundOut.ACC~Coda+CompetitorVowel
+RealWord+(1|Subject),
family = binomial(), data = alldata_clean3)
anova(accuracy_max, accuracy3) # significant -> keep TargetVowel

#removing CompetitorVowel
accuracy4=glmer(TargetSoundOut.ACC~Coda+TargetVowel+
RealWord+(1|Subject),
family = binomial(), data = alldata_clean3)
anova(accuracy_max, accuracy4) # significant -> keep competitor
vowel

#removing RealWord
accuracy5=glmer(TargetSoundOut.ACC~Coda+TargetVowel+CompetitorVowel+(1|Subjec
family = binomial(), data = alldata_clean3)
anova(accuracy_max, accuracy5) #not significant, drop RealWord
```

```
#Step 2: analysing RT measured from vowel offset
#As RT measured from vowel offset had negative RT,
# a constant of 320 was added
#Step 2.1 creating a maximal model without interaction

rt_max=
glmer(RTtoVowelOffset_pos~Coda+TargetVowel+CompetitorVowel
+RealWord+(1|Subject),
family = Gamma(), data = alldata_nooutlier_clean3)
summary(rt_max)

#Step 2.2 removing factors one by one and comparing models with anova
#removing Coda
rt1=glmer(RTtoVowelOffset_pos~TargetVowel+CompetitorVowel
+RealWord+(1|Subject),
family = Gamma(), data = alldata_nooutlier_clean3)
anova(rt_max, rt1) #not significant -> drop Coda

#removing TargetVowel
rt2=glmer(RTtoVowelOffset_pos~CompetitorVowel+RealWord+(1|Subject),
family = Gamma(), data = alldata_nooutlier_clean3)
anova(rt1, rt2) # significant -> keep TargetVowel

#removing CompetitorVowel
rt3=glmer(RTtoVowelOffset_pos~TargetVowel+RealWord+(1|Subject),
family = Gamma(), data = alldata_nooutlier_clean3)
anova(rt1, rt3) #significant -> keep CompetitorVowel
```

```
#removing RealWord
rt4=glmer(RTtoVowelOffset_pos~TargetVowel+CompetitorVowel+(1|Subject),
family = Gamma(), data = alldata_nooutlier_clean3)
anova(rt1, rt4) #significant -> keep RealWord


#Step 3 analysing RT measured from vowel onset
#Step 3.1 creating a maximal model without interactions
rt_max2=
glmer(RTtoVowelOnset~Coda+TargetVowel+CompetitorVowel+RealWord+(1|Subject),
family = Gamma(), data = alldata_nooutlier_clean3)
#Step 3.2 removing factors one by one and comparing models with anova
#removing Coda
rt5=glmer(RTtoVowelOnset~TargetVowel+CompetitorVowel+RealWord+(1|Subject),
family = Gamma(), data = alldata_nooutlier_clean3)
anova(rt_max2, rt5) #not significant -> drop Coda

#removing TargetVowel
rt6=glmer(RTtoVowelOnset~CompetitorVowel+RealWord+(1|Subject),
family = Gamma(), data = alldata_nooutlier_clean3)
anova(rt5, rt6) # significant -> keep TargetVowel

#removing CompetitorVowel
rt7=glmer(RTtoVowelOnset~TargetVowel+RealWord+(1|Subject),
family = Gamma(), data = alldata_nooutlier_clean3)
anova(rt5, rt7) #significant -> keep CompetitorVowel
```

```
#removing RealWord
rt8=glmer(RTtoVowelOnset~TargetVowel+CompetitorVowel+(1|Subject),
family = Gamma(), data = alldata_nooutlier_clean3)
anova(rt5, rt8) #significant -> keep RealWord

#Conclusion: best model is rt5
```

C.2 Script for getting the effects of the features of target and competitor vowel

```
#Tunde Szalay, 2016
#This is the script used to get the effects
#of interaction of the features of target- and competitor vowel
#I always use General Linear Model (function: glmer)
#with Subject as random factor

#Step 0: I load the library
library(car)
library(lme4)
library(beepr)

#Step 1 analysing accuracy data
#Step 1.1 creating a maximal model in which
#target- and competitor features interact
accuracyfeat_max=
  glmer(TargetSoundOut.ACC~Coda+Front*CFront+Back*CBack
```

```
+High*CHigh+Low*CLow+Long*CLong
+RealWord+(1|Subject),
family = binomial(), data = alldata_clean3)

#significant negative interaction for
#+Front & +CFront, +High & +CHigh, +Low & +CLow,
#significant positive interaction for +Long & +CLong



#Step 2 analysing RT measured from the onset of the vowel
#Step 2.1: creating a maximal model in which
#the features of target and competitor vowels are interacting factors
rtfeat_max1=
glmer(RTtoVowelOnset~Coda+Front*CFront+Back*CBack
+High*CHigh+Low*CLow+Long*CLong
+RealWord+(1|Subject),
family =Gamma(), data = alldata_nooutlier_clean3)
summary(rtfeat_max1)

#Target and competitor features have a significant
#negative interaction, when both features are +1
#except for +Long, which has a non-significant positive interaction
#Coda1 has an almost significant negative effect
#RealWord has a significant positive effect

#Step 3 analysing RT measured from the offset of the vowel
#Step 3.1: creating a maximal model in which
```

```
#the features of target and competitor vowels are interacting factors
rtfeat_max2=
  glmer(RTtoVowelOffset_pos~Coda+Front*CFront+Back*CBack+High*CHigh+Lo
       +RealWord+(1|Subject),
       family = Gamma(), data = alldata_nooutlier_clean3)
summary(rtfeat_max2)

#Target and competitor features have a significant negative interactio
#when both features are +1
#except for +Long, which has a non-significant negative interaction
#Coda1 has an almost significant negative effect
#RealWord has a significant positive effect
```

D Additional figures

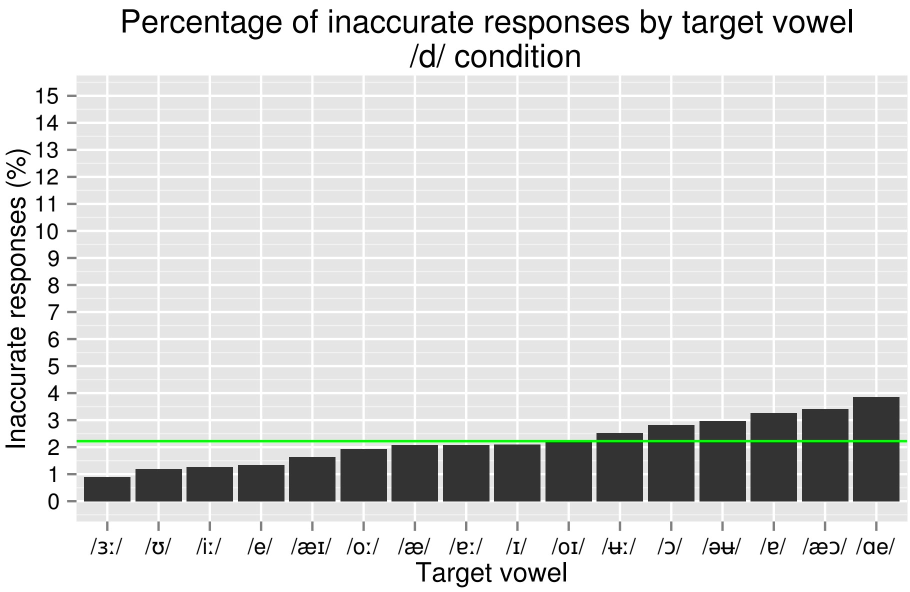

Figure D.1: **Inaccurate responses (%) by target vowel in the /d/ condition.** X axis: target vowels ordered according to the percentage of inaccurate responses. Y axis: percentage of inaccurate responses. Green line: grand mean.

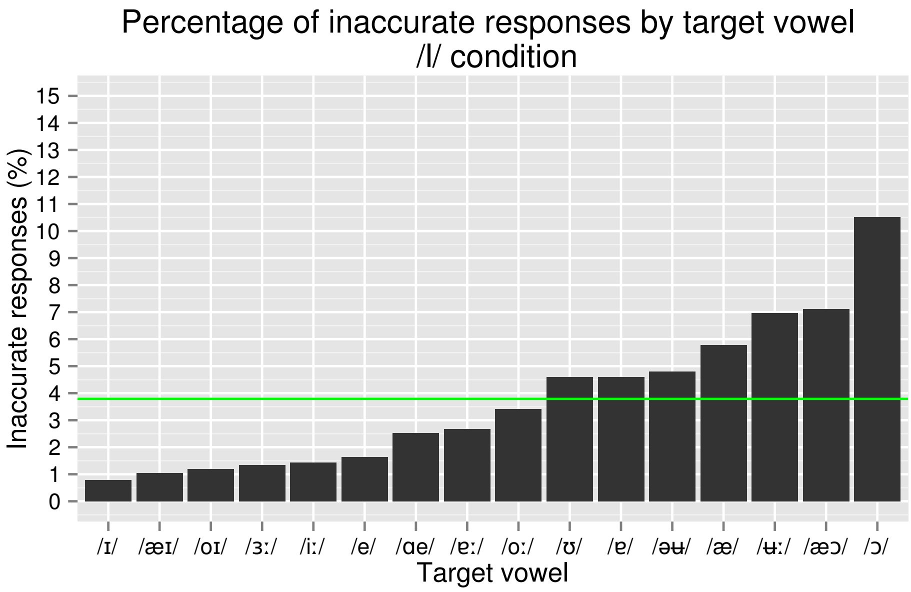

Figure D.2: **Inaccurate responses (%) by target vowel in the /l/ condition.** X axis: target vowels ordered according to the percentage of inaccurate responses. Y axis: percentage of inaccurate responses. Green line: grand mean.

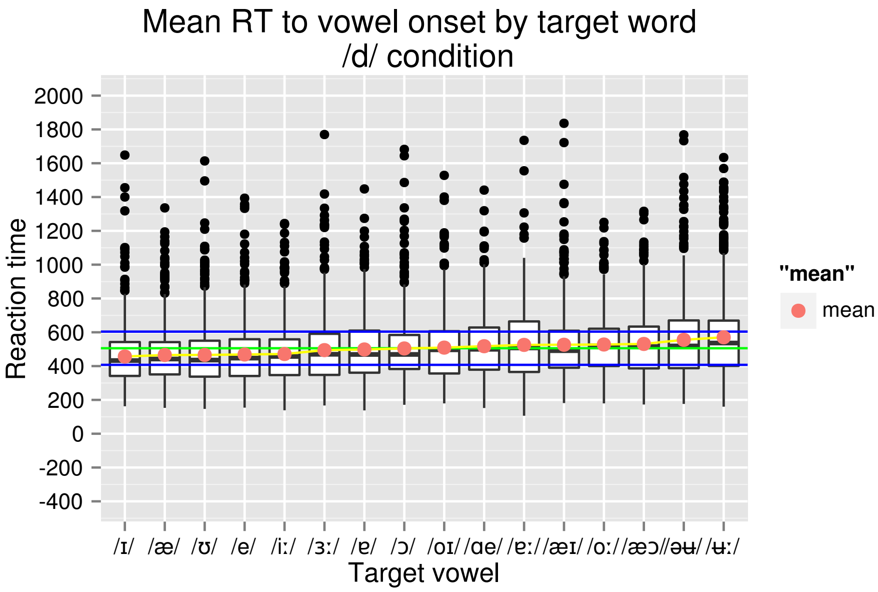

Figure D.3: **RT measured from vowel onset by target vowel in the /d/ condition.** X axis: target vowels. Y axis: Mean RT (ms) measured form vowel onset. Green line marks the grand mean and blue lines mark +0.5SD and -0.5SD

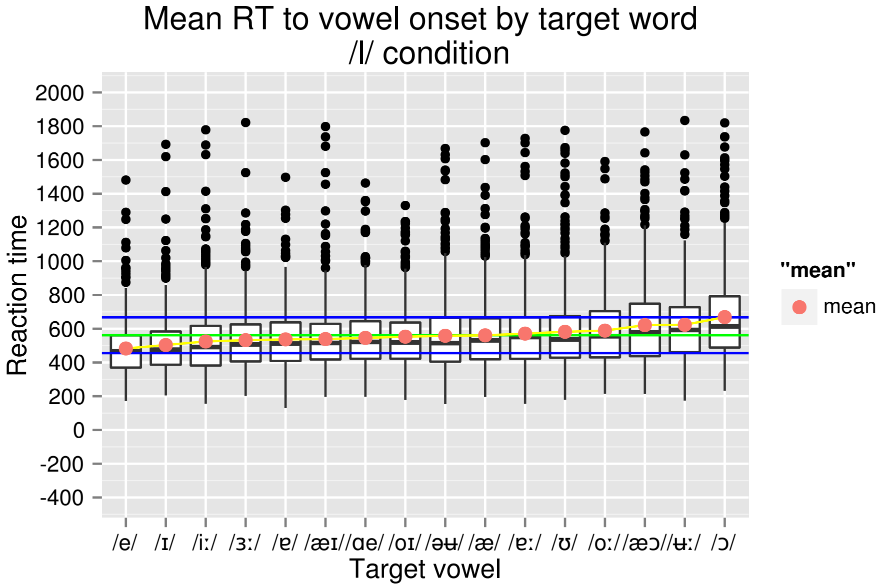

Figure D.4: **RT measured from vowel onset by target vowel in the /l/ condition.** X axis: target vowels. Y axis: Mean RT (ms) measured form vowel onset. Green line marks the grand mean and blue lines mark +0.5SD and -0.5SD

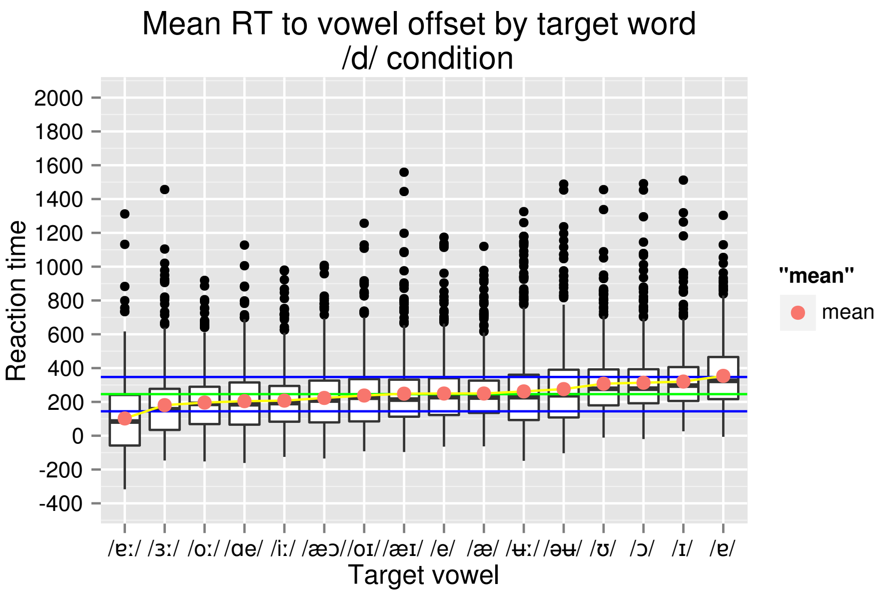

Figure D.5: **RT measured from vowel offset by target vowel in the /d/ condition.** X axis: target vowels. Y axis: Mean RT (ms) measured form vowel offset.Green line marks the grand mean and blue lines mark +0.5SD and -0.5SD

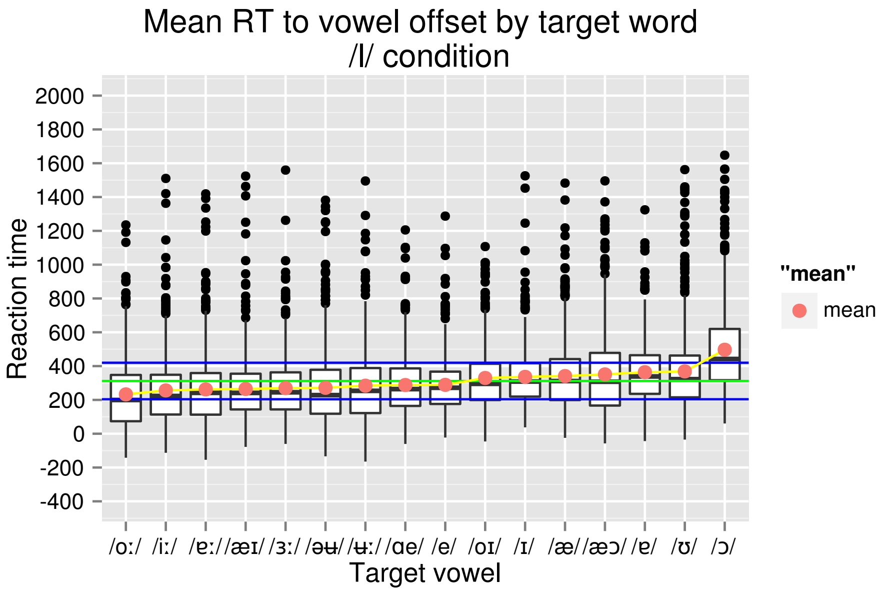

Figure D.6: **RT measured from vowel offset by target vowel in the /l/ condition.** X axis: target vowels. Y axis: Mean RT (ms) measured form vowel offset. Green line marks the grand mean and blue lines mark +0.5SD and -0.5SD

E Ethics approval

Dear Dr Proctor,

Re: "Perception of laterals in Australian English" (5201600061)

Thank you very much for your response. Your response has addressed the issues raised by the Faculty of Human Sciences Human Research Ethics Sub-Committee and approval has been granted, effective 18th March 2016.

This email constitutes ethical approval only.

This research meets the requirements of the National Statement on Ethical Conduct in Human Research (2007). The National Statement is available at the following web site:

http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72.pdf

The following personnel are authorised to conduct this research:

Dr Michael Proctor

Ms Tünde Orsolya Szalay

Yours sincerely,

Dr Anthony Miller Chair Faculty of Human Sciences Human Research Ethics Sub-Committee

---

Faculty of Human Sciences - Ethics Research Office Level 3, Research HUB, Building C5C Macquarie University NSW 2109

Ph: +61 2 9850 4197

Email: fhs.ethics@mq.edu.au

http://www.research.mq.edu.au/

Bibliography

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, (38):419–439.

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

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278.

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

Blevins, J. (1995). Syllable in phonological theory. In Goldsmith, J., (Ed.), *The handbook of phonological theory*. Blackwell, first edition.

Blevins, J. (2006a). Syllable typology. In Brown, K., (Ed.), *Encyclopedia of language and linguistics*, volume 12, pp. 333–337. Elsevier, second edition.

Blevins, J. (2006b). A theoretical synopsis of evolutionary phonology. *Theoretical Linguistics*, (32):117–165.

Boersma, P. & Weenink, D. (2013). Praat 5.3.42.

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

Borowsky, T. & Horvath, B. (1997). L-vocalisation in australian english. In Hinskens, F., van Hout, R., & Wetzels, W. L., (Eds.), *Variation, Change and Phonological Theory.*, pp. 101–123. Amsterdam, Benjamins.

Broselow, E., Chen, S.-I., & Huffman, M. (1997). Syllable weight: convergence of phonology and phonetics. *Phonology*, 14(1):47–82.

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

Browman, P. C. & Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica*, 45:140–55.

Campbell, F., Gick, B., Wilson, I., & Vatikiotis-Bateson, E. (2010). Spatial and temporal properties of gestures in North American English /r/. *Language and Speech*, 53:49–69.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*, New York, Harper & Row.

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. (2012). *Australian English Pronunciation and transcription*, Cambridge University Press.

Cox, F. & Palethorpe, S. (2004). The border effect: Vowel differences across the NSW - Victorian border. In Moskovsky, C., (Ed.), *Proceedings of the 2003 Conference of the Australian Linguistic Society*, pp. 1–14.

Cox, F. & Palethorpe, S. (2010). Broadness variation in australian english speaking females. In Tabain, M., Fletcher, J., Grayden, D., Hajek, J., & Butcher, A., (Eds.), *Proceedings of the 13th Australasian International Conference on Speech Science and Technology.*

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.

Cruttenden, A. (2001). *Gimson's pronunciation of English*, London, Arnold.

Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*, Cambridge, MA, MIT Press.

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

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):769–773.

Docherty, G. J. & Foulkes, P. (2014). An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua*, 142:42–56.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36(2):345 – 360.

Everitt, B. S. (2006). *The Cambridge dictionary of statistics*, Cambridge University Press, third edition.

Ewen, C. J. & Hulst, H. v. d. (2001). *The phonological structure of words*, Cambridge, Cambridge University Press.

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

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

Gick, B. & Campbell, F. (2003). Intergestural timing in english /r/. In *Proceedings of the XVth International Congress of Phonetic Sciences*, pp. 911–1914. Barcelona, Universitat Autonoma de Barcelona.

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):357–371.

Gick, B. & Wilson, I. (2003). Excrescent schwa and vowel laxing: cross-linguistic responses to conflicting articulatory targets. In *Papers in Lab-*

oratory Phonology VIII: Varieties of Phonological Competence, pp. 635–660. New York, Mouton de Gruyter.

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

Goldinger, D. S. (1996a). Auditory lexical decision. *Language and Cognitive Processes*, 6(11):559–567.

Goldinger, S. D. (1996b). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 5(22):1166–1183.

Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In *Proceedings of the 17th International Congress of Phonetic Sciences*, pp. 49–54.

Goldsmith, J. (2014). The syllable. In Goldsmith, J., Riggle, J., & Yu, A. C. L., (Eds.), *The handbook of phonological theory*. Blackwell, second edition.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28(4):267–283.

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20(2):253–306.

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. (2002). The geolinguistics of /l/ vocalisation in Australia and New Zealand. *Journal of Sociolinguistics*, 6/3:319–346.

Huettig, F., Joost, R., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica*, 137(2):151–171.

Johnson, K. (2008). *Quantitative methods in linguistics*, Blackwell Publishing.

Kewley-Port, D. (1995). Thresholds for formant-frequency discrimination of vowels in consonantal context. *Journal of the Acoustical Society of America*, 5(97):3139–3146.

Kewley-Port, D. & Watson, C. S. (1994). Formant frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America*, 1(95):485–496.

Kewley-Port, D. & Zheng, Y. (1999). Vowel formant discrimination: Towards more ordinary listening conditions. *Journal of the Acoustical Society of America*, 106(5):2945–2958.

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

Krämer, M. (2008). English schwa insertion before liquids and phonological opacity. *Proceedings of CLS*, 41.1:267–282.

Labov, W., Sharon, A., & Boberg, C. (2008). *The Atlas of North American English, Phonetics, Phonology and Sound Change*, Berlin, Boston, Gruyter Mouton.

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

Lavoie, L. M. & Cohn, Abigail, C. (1999). Sesquisyllables of English: The structure of vowel-liquid syllables. In *Proceedings of the XIVth International Congress of Phonetic Science*, pp. 109–112.

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*. Sydney.

Loakes, D., Clothier, J., Hajek, J., & Fletcher, J. (2014a). Identifying /el/-/æl/: A comparison between two regional Australian towns. In Hay, J. & Parnell, E., (Eds.), *15th Australasian International Conference on Speech Science and Technology*.

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 *Australasian International Conference on Speech Science and Technology*, volume 14. Macquarie University.

Loakes, D., Hajek, J., & Fletcher, J. (2010a). The /el/-/æl/ sound change in Australian English: A preliminary perception experiment. In *Australian Linguistic Society Conference*. Australian Linguistics Society.

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 *13th Australasian International Conference on Speech Science and Technology*. Australasian Speech Science and Technology Australia (ASSTA).

Loakes, D., Hajek, J., & Fletcher, J. (2010c). (Mis)perceiving /el/-/æl/ in Melbourne English: a micro-analysis of sound perception and change. In *13th Australasian International Conference on Speech Science and Technology*. Australasian Speech Science and Technology Australia (ASSTA).

Loakes, D., Hajek, J., & Fletcher, J. (2011). /æl/-/el/ transposition in Australian English: Hypercorrection or a competing sound change? In *The 17th International Congress of Phonetic Sciences*. City University of Hong Kong.

Maddieson, I. (2013). Vowel quality inventories. In Dryer, M. S. & Haspelmath, M., (Eds.), *The World Atlas of Language Structures Online*. Leipzig, Max Planck Institute for Evolutionary Anthropology.

Marslen-Wilson, W. & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology*, 10:29–63.

Marslen-Wilson, W. D. (1980). *Speech Understanding as a Psychological Process*, pp. 39–67. Dordrecht, Springer Netherlands.

Martin, J. G. & Bunnell, T. H. (1981). Perception of anticipatory coarticulation effects. *The Journal of the Acoustical Society of America*, 69:559–567.

McQueen, J. M. & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, 60(5):661–671.

Mermelstein, P. (1978). Difference limens for formant frequencies of steady-state and consonant-bound vowels. *Journal of the Acoustical Society of America*, 2(63):572–580.

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

Nearey, Terrance, M. & 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.

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

Osman, A., Lou, L., Muller-Gethmann, H., Rinkenauer, G., Mattes, S., & Ulrich, R. (2000). Mechanisms of speed–accuracy tradeoff: evidence from covert motor processes. *Biological Psychology*, 51(2–3):173 – 199.

Palethorpe, S. & Cox, F. (2003). Vowel modification in pre-lateral environments. In *International Seminar on Speech Production*. Sydney.

Parker, S. (2011). Sonority. In van Oostendorp, M., Ewen, C. J., Hume, E., & Rice, K., (Eds.), *The Blackwell companion to phonology*, volume 2, pp. 1160–1184. Blackwell Publishing.

Pierrehumbert, J. B. (2016). Phonological representation: beyond abstract versus episodic. *The Annual Review of Linguistics*, (2):33–52.

Pisoni, D. B. & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15(2):285–290.

Proctor, M. & Walker, R. (2012). Articulatory bases of sonority in English liquids. In Parker, S., (Ed.), *The sonority controversy*, pp. 285–312. Berlin, New York, Mouton de Gruyter.

Psychology Software Tools, I. (2012). E-prime 2.0.

R Core Team (2015). R: A language and environment for statistical computing.

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

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.

Ruthruff, E. (1996). A test of the deadline model for speed-accuracy tradeoffs. *Perception & Psychophysics*, 58(1):56–64.

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

Silverman, I. W. (2010). Simple reaction time: It is not what it used to be. *The American Journal of Psychology*, 123(1):39–50.

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

Taft, M. (1986). Lexical access codes in visual and auditory word recognition. *Language and Cognitive Processes*, 4(1):297–308.

Tanenhaus, M. K., Magnuson, S. J., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, 29(6):557–580.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, pp. 1632–1634.

Team, R. (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.

Tilsen, S., Cohn, A., & Ricciardi, E. (2014). Syllable count judgments and durations of liquid rimes in English. *Cornell Working Papers in Phonetics and Phonology*.

Trommelen, M. & Zonneveld, W. (1999). English. In van der Hulst, H., (Ed.), *Word Prosodic Systems in the languages of Europe*, pp. 478–49. Mouton de Gruyter.

Upton, C. (2008). Received pronunciation. In Kortmann, B. & Upton, C., (Eds.), *Varieties of English*, volume 1, pp. 237–252. Berlin, New York, Mouton de Gruyter.

Vannest, J., Newport, E. L., Newman, A. J., & Bavelier, D. (2011). Interplay between morphology and frequency in lexical access: The case of the base frequency effect. *Brain Research*, 1373:144–159.

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

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

Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception and Psychophysics*, 50(4):351–360.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41(1):67 – 85.

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

Zec, D. (2007). The syllable. In de Lacy, P., (Ed.), *The Cambridge handbook of phonology*, pp. 161–194. Cambrdige, New York, Cambridge University Press.

Ziegler, J. C., Ferrand, L., & Montant, M. (2004). Visual phonology: The effects of orthographic consistency on different auditory word recognition tasks. *Memory & Cognition*, 32(5):732–741.