

Article

Australian Indian English: Contact-Induced Adaptation in the Perception of Vowel Categories

Olga Maxwell ^{1,*}, Elinor Payne ^{2,*}, Debbie Loakes ¹ and Mitko Sabev ³

¹ School of Languages and Linguistics, University of Melbourne, Parkville, VIC 3010, Australia; dloakes@unimelb.edu.au

² Faculty of Linguistics, Philology and Phonetics, Schwarzman Centre for the Humanities, Radcliffe Observatory Quarter, Woodstock Road, University of Oxford, Oxford OX2 6GG, UK

³ Department of Language Science and Technology, Saarland University, 66123 Saarbrücken, Germany; msabev@lst.uni-saarland.de

* Correspondence: omaxwell@unimelb.edu.au (O.M.); elinor.payne@phon.ox.ac.uk (E.P.)

Abstract

Increased global mobility has intensified contact between regional English varieties, creating new opportunities for large-scale second dialect acquisition. Australia, with its rapidly growing population due to migration, offers a particularly dynamic context for exploring such contact. This study investigates how first-generation Indian migrants in the Australian city of Melbourne perceive Australian English vowels in the lexical items *DRESS* and *TRAP*, a contrast chosen because of sound changes that are well-documented for this location. Listeners completed a vowel categorization task involving target words in non-lateral and lateral contexts. To assess contact-induced adaptation, their responses were compared with those of Australian English speakers in Australia and those of Indian English speakers in India. The results reveal that perceptual adaptation among first-generation Indian migrants in Australia is context-dependent. In the non-lateral coda context, migrant Indian English listeners (in Australia) showed intermediate responses, between those of Australian English listeners (in Australia) and Indian English listeners (in India), indicative of a relatively ‘linear’ adaptation towards Australian English. Responses to stimuli in the lateral coda context, however, revealed a more complex picture. Australian English listeners (in Australia) and Indian English listeners (in India) responded more closely to one another than migrant Indian English listeners (in Australia), with the latter instead exhibiting a substantial degree of perceptual confusion toward the endpoint of the continuum for *hell–Hal* and, to a lesser extent, for *shell–shall* and *pell–pal*. These findings suggest that in the perceptual adaptation to a second dialect, the acquisition of a wider pool of phonetic variants is mediated by the acquisition of structural knowledge.

Keywords: vowel perception; front vowels; Indian English; Australian English; second dialect acquisition; dialectal differences



Academic Editors: Donna Erickson and Anthony Pak-Hin Kong

Received: 26 June 2025

Revised: 3 April 2026

Accepted: 7 April 2026

Published: 11 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

New patterns of mobility have led to a growing body of research on multilingual migrant communities, particularly in predominantly English-speaking countries. Even when a common language is shared, when individuals relocate, they often find themselves in new environments where the majority speak a second dialect (D2), distinct from their native dialect (D1). This consistent, prolonged exposure to another dialect frequently

results in changes to their linguistic system, a process known as *second dialect acquisition* (Siegel, 2010).

Research shows that in such scenarios, listeners' resultant phonemic categorization may be influenced by their D1, and their perceptual response to dialect variation is shaped by the nature and degree of linguistic distance across those dialects, i.e., between D1 and D2 (e.g., Clopper, 2014, 2021). Furthermore, dialect perceptual categories are critically influenced by individual experience, with listeners generally exhibiting greater accuracy when classifying speech input from talkers with whom they share the same variety. However, listeners with high mobility and substantial and/or long-term exposure to multiple dialects can develop more robust perceptual representations to unfamiliar dialects. In other words, such wide-ranging exposure can facilitate phoneme discrimination, even for contrasts present in D2 but absent in D1 (Evans & Iverson, 2004, 2007; Dufour et al., 2007; Nycz, 2013; Clopper, 2014; Diskin-Holdaway et al., 2024). Importantly, exposure to multiple dialects here is not restricted to that enabled by geographic mobility but can also arise through media and education (Clopper, 2014).

A key question in current research is the extent to which, and how, a listener's perceptual responses are susceptible to change after long-term exposure to other dialects. Several studies have shown that both the loss of D1 features and the acquisition of D2 features may result in linguistic forms that are intermediate between the two dialects (Munro et al., 1999; Ziliak, 2012; Nycz, 2018; Kunkel et al., 2023). However, questions remain regarding the extent of change and which variables impact people's linguistic behavior the most. With the aim of addressing this gap, the present study examines the perception of the English front lax vowels /e/ (DRESS) and /æ/ (TRAP), for the dialect of Australian English, by first-generation Indian migrants in Melbourne. These individuals, at least on arrival, predominantly speak Indian English (among other languages), an umbrella term for the varieties of English spoken in India and, to some degree, in the diaspora (Wiltshire, 2020). Long in contact with India's Indigenous languages, Indian English is recognized for its distinct phonetic and phonological features, including its vowel system (Payne & Maxwell, 2018; Maxwell et al., 2018; Maxwell & Payne, 2018, 2023; Payne et al., 2019, 2023). However, little is currently known about perception patterns of Indian Australians and the consequences of inter-dialectal contact between Indian English and Australian English. The front lax vowels, specifically /e/ and /æ/, are of particular interest in Australian English because they are part of ongoing dynamic sound changes (Cox & Palethorpe, 2008; Cox et al., 2024).

To investigate the perceptual behavior of first-generation Indian Australians and examine the presence and degree of contact-induced adaptation, we designed a forced-choice vowel categorization task with real and nonce words. We compare their perception of the vowels /e/ and /æ/ with that of native Australian English speakers and that of native Indian English speakers in India. We also consider the length of residence (LOR) in Australia for Indian migrants in Australia. The findings of this study will contribute to a better understanding of variation in Indian English in the global context, the dynamic linguistic landscape in multicultural Australia, and the broader body of work on the effects of acoustic-phonetic variability on speech perception and perceptual adaptation.

2. Background

2.1. Australian English—Front Vowels and Merger

The vowel space of Mainstream Australian English is well-documented, primarily for large urban areas such as Sydney (Harrington et al., 1997¹; Cox, 1999; Elvin et al., 2016; Cox et al., 2024) and Melbourne (Billington, 2011; Cox & Palethorpe, 2019). Work by Cox and Palethorpe (2019) further demonstrates that monophthongal vowels produced in urban areas across Australia (Melbourne, Sydney, Adelaide and Perth) are relatively similar,

with only minor regional differences, and none pertaining to the vowels under analysis in this study.

In addition to descriptive accounts of the relative location of vowels acoustically, research has shown that urban Australian English vowels have undergone rapid changes in a relatively short period, particularly in the short front vowel system. For example, and highly relevant to the current study, researchers describe a raising of /i/, a lowering of /e/, a lowering and retraction of /æ/, and a retraction of /ɐ/ (Cox & Palethorpe, 2008; Grama et al., 2019; Cox et al., 2024), as well as other smaller changes in relative vowel position and in F1/F2 trajectories (Cox et al., 2024).

Research on perception has aligned with the findings of these production studies. For example, by comparing results of a perception experiment within a relatively small window of time (between 1998 and 2004), Mannell (2004) reported a shift upwards in the perceptual boundaries between /i/ and /e/, a shift down in the perceptual boundary between /e/ and /æ/, and a retraction of the perceptual boundary between /æ/ and /ɐ/. For example, listeners in the later period required a more retracted /ɐ/ to classify the token as /ɐ/ compared to listeners in the earlier period. More recently, Loakes et al. (2024a) analyzed vowel categorization behavior across various Australian English-speaking communities and observed significant diachronic differences between younger and older listeners, which were best explained with reference to production findings. For example, in the context of categorizing /e/ and /æ/, younger speakers needed to hear a vowel with a much more open quality (i.e., higher F1) than older speakers to identify it as /æ/, while older speakers shifted to /æ/ much earlier along the /e/-/æ/ continuum.

While regional variation in the relative position of Mainstream Australian English vowels is relatively minor (e.g., Cox & Palethorpe, 2019), another difference which potentially impacts the current study is the presence of a regionally defined vowel merger in south-eastern Australia. The merger of /e/->/æ/ has been shown to occur in both the production and perception of some speakers in the Australian state of Victoria. Its presence has been observed in the state-capital city Melbourne and in some other regional areas, particularly in the southernmost regions of the state. For example, in an especially large corpus of over 4 million spontaneous speech tokens across urban and regional areas in Australia, Coats et al. (2025) have shown that /e/->[æ] is indeed confined to the south-east of the state and is more prevalent in the southern areas, which includes the capital Melbourne. This aligns with other research indicating that the merger tends to occur in southern Victoria in production (Cox & Palethorpe, 2004; Loakes et al., 2017, 2024b) and in perception (Loakes et al., 2024a, 2024b). It should also be noted that the merger is incomplete, and when it is present, it occurs at different rates within and between speakers (see, e.g., Loakes et al., 2019, 2024a, 2024b; Schmidt et al., 2021), with some participants not displaying any merger at all and some having a fully entrenched merger (Loakes et al., 2014, 2019).

Also relevant to the current study are investigations into how listeners of other non-Mainstream Australian English varieties interpret Mainstream Australian English vowels in general, and /e/-/æ/ specifically. This is of particular interest to the present study because, as noted in the Introduction, vowel categorization is influenced by various factors, including, in the case of multilingual speakers, the acoustic properties and inter-category boundaries of vowels in the listeners' first language (L1) or dialect (D1) and their listening experience in the second language (L2) or dialect (D2). Clopper and Walker (2017) discuss the possibility that people who speak, or who have been exposed to, multiple varieties may display enhanced cross-variety processing as a result of increased awareness of category boundaries. However, it is also theoretically plausible that they have broader phoneme categories as a result of this exposure to multiple varieties, leading to the possibility of

greater perceptual confusion. Clopper and Walker's (2017) work analyzing highly mobile speakers in the U.S. showed that in fact the latter option, i.e., with greater uncertainty about phoneme mapping, was the case for their listeners.

A consequence of the reported vowel changes, including vowel merger, in the south-east of Australia, is that all people in these communities are exposed to highly variable /e-/æ/ tokens (see also Loakes et al., 2024a, 2024b). This serves as an already variable backdrop for what may be happening in D2 or L2 perception. Indeed, research has also shown that listeners for whom Australian English is an L2, or D2, have additional influences that impact the way they perceive and process Mainstream Australian English vowels, with categories from their L1 and/or D1 potentially influencing their perception behavior. The findings of Diskin-Holdaway et al. (2024) have shown that when exposed to Australian English vowels, both Irish English speakers who have migrated to Australia (and are thus D2 listeners of Australian English) and L1 Chinese speakers (who are L2 listeners of Australian English) categorize vowels in ways that indicate the importance of both exposure (i.e., a new environment) and experience (i.e., living and working or studying in Australia). In regard to experience, L2 Chinese listeners responded at random for multiple vowel contrasts in Australian English, demonstrating a difficulty in making categorical judgements about vowels in this variety of English at all. D2 Irish English listeners could distinguish them into separate categories but had completely different crossovers when compared with a control group of Australian English listeners. The study showed that their judgements about Australian English vowels were likely conditioned by the respective vowel systems of their L1/D1. Additionally, Diskin-Holdaway et al. (2024) reported that the more exposure a listener had to Australian English, the more Australian-like the patterns of their vowel categorization were. The vowel merger context was also analyzed in their study, showing little evidence of merger in perception, even for the native Australian English listeners. However, the findings revealed that the availability of fine acoustic detail from coarticulation was important for listeners, with a "downshifting" of category boundaries when a word contained a coda /l/. Having a longer availability of relevant cues from the vowel, listeners overall had a later crossover from /e/ to /æ/ (a similar result is seen in Loakes et al. (2024a) for Mainstream Australian English).

Other work involving Aboriginal English listeners has also shown that these D2 listeners, who speak and are exposed to both Mainstream Australian English and Aboriginal English, exhibit different category boundaries when compared with the listeners for whom Mainstream Australian English is their D1. For example, sound changes which are reported to apply in the mainstream variety are not as entrenched in Aboriginal English (/e/ and /æ/ are not as phonetically low) (Loakes & Gregory, 2024). As a result, Aboriginal English listeners tend to have earlier crossovers for categorizing TRAP in perception (Loakes et al., 2024a, 2024b). Additionally, and likely because of the exposure to multiple linguistic variants, Aboriginal English listeners are more varied in their responses (Loakes et al., 2024b), echoing the greater perceptual uncertainty reported by Clopper and Walker (2017).

2.2. Indian English—Front Vowels and /l/

As described above, 'Indian English' encompasses a wide range of English varieties spoken both within India and across the extensive Indian diaspora globally. While there is a substantial body of work on the phonetics and phonology of Indian English, empirical research—particularly sociophonetic investigations of vowel production and perception—remains limited (e.g., see Domange, 2020; Wiltshire, 2020; Maxwell et al., 2018, 2023; Maxwell & Payne, 2023). Earlier studies on the Indian English vowel system were primarily based on auditory analyses, and both descriptive and empirical accounts have

tended to focus on L1- or region-based variation, often examining the influence of substrate languages on English spoken in India.

Indian English, as a postcolonial variety of English, has developed over several centuries in contact with, and influenced by, numerous indigenous languages and is notable for its self-replicating nature—being taught by Indians to Indians (Kachru, 1983). It is structurally very distinct from indigenous Indian languages, which themselves vary widely. L1 effects can be subtle or absent for some features but not for others, are mediated by multiple sociolinguistic factors, and are further enhanced by increased domestic mobility and migration, resulting in growing contact among speakers of different L1/regional Indian English varieties, especially in large urban centers (Sirsa & Redford, 2013; Maxwell & Payne, 2023).

The variability in vowel systems across Indian English varieties is reported to be more pronounced than that found in consonant systems (Sharma, 2005). Mixed findings across studies—particularly in relation to the DRESS and TRAP vowel categories—is an artifact of this complexity. Several studies have found that speakers of Indian English generally maintain the DRESS and TRAP contrast, even when their L1 lacks this phonemic distinction (Wells, 1982; Trudgill & Hannah, 2008). Maxwell and Fletcher (2009), for example, report that L1 Hindi and Punjabi speakers produced this contrast, although the vowels were realized acoustically close to one another in the F1/F2 space. In that study, only one Hindi L1 speaker failed to produce the contrast. These findings contrast with earlier claims (e.g., Sethi, 1980; Hickey, 2004) that Punjabi L1 speakers commonly neutralize the /e-/æ/ distinction. Wiltshire (2005) also reported a lack of contrast among Indian English speakers whose L1 is Angami, suggesting that this feature could be variable depending on the speakers' L1 background.

Of particular relevance to the present study is Domange's (2020) apparent-time investigation of Delhi English, an emerging urban variety that diverges from external norms (Satyanath & Sharma, 2016). All speakers in that study maintained the DRESS–TRAP distinction, but a change in progress was observed: younger speakers showed lowering and fronting of DRESS and lowering and retraction of TRAP. TRAP was also found to be lower following /l/ and more retracted following obstruent-liquid clusters. This is noteworthy given that Indian English—and many Indian languages—feature a clear /l/ in all syllable positions (Wells, 1982; Gargesh, 2006), typically articulated with the tongue tip or blade at the alveolar ridge and a raised tongue body (Ladefoged & Johnson, 2015). This differs notably from Australian English, where /l/ is dark in a coda position and may also be dark in an onset position, with variation observed across speakers (Cox & Fletcher, 2017).

3. Present Study

The present study investigates how first-generation Indian Australians perceive the Australian English vowels /e/ (DRESS) and /æ/ (TRAP)—a contrast that has recently undergone diachronic change, as well as a partial merger before /l/, in Australian English. In this context, Australian English functions as both a D2 and a newly dominant variety for these speakers of Indian English. Focusing on this vowel contrast within a system in flux offers valuable insight into how contact-induced variation may affect speech perception within a dynamically evolving linguistic landscape.

Over the past two decades, Australia has seen a rapid increase in migrants from India, with the Indian diaspora there now comprising approximately 3.1% of the national population (ABS, 2021) and making a significant contribution to Australia's linguistic diversity. Furthermore, within India itself, internal mobility for education and employment exposes speakers to different varieties of Indian English, contributing further to the complexity of the situation, by increasing the variability within D1 and potentially re-shaping perceptual categories even before individuals migrate to Australia. As discussed, previous

research suggests that exposure to multiple varieties can enhance cross-dialectal processing but may also introduce perceptual instability. In particular, highly mobile individuals may experience greater uncertainty in phoneme categorization when exposed to multiple, conflicting exemplars (Clopper & Walker, 2017).

To explore these dynamics, we designed a vowel categorization task targeting the perception of English front lax vowels by Indian English speakers residing in Melbourne. Their responses were compared to those of native Australian English speakers and a baseline group of Indian English speakers residing in India. This study addresses the following research questions:

1. To what extent do first-generation Indian migrants in Melbourne adapt perceptually to Australian English as a second dialect (D2)?
2. What patterns of contact-induced variation or change are evident in their perception of the DRESS–TRAP contrast in AusE?
3. To what extent, and how, do factors such as length of residence in Australia and gender influence the perceptual behavior of Indian English migrants?

4. Materials and Methods

4.1. Participants

One hundred and twenty-four participants took part in the study across the two locations, Australia and India. Table 1 presents the number of listeners in each group, along with the breakdown by gender, age range, and median age. The Australian English group (AusE) and the Australian Indian English group (Aus-IndE) included participants aged between 18.4 and 59.7 years and between 19 and 59.7 years, respectively, while the group in India (Ind-IndE) had a narrower age range (22.6–36.7 years). This was also because varying LOR for the Aus-IndE cohort was desired, which necessitated a wider age range, and which was then controlled for in the AusE cohort. Despite these differences, the median ages across the three groups were relatively similar, indicating broadly comparable cohorts. Males were notably overrepresented in the Aus-IndE group. All participants were screened for any listening or speech disorders during the recruitment stage.

Table 1. Demographic summary of the three listener groups, including gender distribution, age range, and age median across the two locations, Australia and India.

Listener Group and Location	Total	Gender	Age Range	Median Age
AusE, Australia	34	Male: 14 Female: 20	18.4–59.7	28.5
Aus-IndE, Australia	40	Males: 24 Female: 16	19–59.7	33.8
Ind-IndE, India	50	Males: 24 Females: 26	22.6–36.7	27.9

Participants in the AusE group were all speakers of Australian English as their L1 and had been born and raised in Melbourne or the greater Melbourne area, where they were still residing at the time of data collection. Although principally monolingual, some of the Australian English-speaking participants reported some degree of foreign language knowledge acquired as part of school or tertiary education. Participants in India were recruited across multiple locations, targeting highly proficient English speakers who had attended an English-medium school and had completed at least an undergraduate degree or were enrolled in a university for further graduate study (if not employed). Controlling for participants' educational background was crucial to ensure an adequate comparison within and across groups. While none identified English as their L1 chronologically, all cited En-

English as currently either their dominant language or one of their dominant languages, and the majority of the group reported acquiring English in early childhood.

Inevitably, given the linguistic situation of India, listeners in both IndE groups (Ind-IndE and Aus-IndE) were multilingual, with L1 languages from one or the other of India's two largest language families, Indo-Aryan and Dravidian. Some L1s were more strongly represented, with a relatively balanced number of listeners—particularly those speakers located in India—speaking L1 Hindi, Bengali (both Indo-Aryan), Tamil, Telugu, Kannada, or Malayalam (all Dravidian). For Indian participants residing in Australia, length of residence (LOR) was also recorded, calculated from the month and year of their arrival to the month and year they completed the tasks of the present study (representing five LOR groups), and this is shown in Table 2 below in relation to listeners' gender and age group. LOR ranges were created to provide an overview of the data. However, during statistical analyses, LOR was treated as a continuous variable.

Table 2. Length of residence (LOR) by gender and age group among the IndE listeners in Australia.

LOR Range (In Years)	Gender	Age Group	Number of Participants
1–2	M	41–46	1
1–2	M	50–60	1
1–2	F	19–25	2
1–2	F	26–29	1
1–2	F	30–35	1
3–6	M	41–46	1
3–6	M	50–60	1
3–6	F	19–25	2
3–6	F	26–29	2
3–6	F	30–35	1
3–6	M	26–29	1
7–10	M	30–35	3
7–10	M	36–40	2
11–14	M	26–29	1
11–14	M	30–35	3
11–14	M	36–40	2
15–20	F	19–25	1
15–20	F	30–35	3
21–25	M	30–35	1
21–25	M	36–40	1

4.2. Materials

The speech samples used to generate the stimuli for the task were words produced by a 43-year-old female Australian English speaker from Brisbane (Queensland, Australia), following Diskin-Holdaway et al. (2024) and Loakes et al. (2024a), whose perception experiment stimuli were based on recordings of the same speaker. She had lived for 7 years in Melbourne prior to moving interstate and was residing in Brisbane at the time of the recording. No DRESS–TRAP merger has been reported in Queensland, and it was not evident in her recordings. This is precisely the reason we chose this speaker, so that the stimuli that listeners were exposed to were acoustically distinct at the endpoints of the DRESS–TRAP continua. See Appendix A for formant values for each target word. The recordings were made on the speaker's personal laptop after an online debriefing session (via Zoom), in addition to the written instructions outlining steps for making a good-quality recording. Five repetitions of each target word produced in isolation were recorded.

The complete word list included 40 target words (grouped into 20 pairs), with the target vowel appearing in /CVC(V)/ and /(C)VI(V)/ contexts, differing in lexical frequency and word shape (e.g., 1 vs. 2 syllables), and including two pairs of nonce words. In the present study², we focus on the /e-æ/ vowel contrast in two contexts, /CVt/ and /CVI/, thus also limiting the investigation to monosyllabic words. This approach was adopted to allow for direct comparisons with previous research on DRESS–TRAP perception conducted among Mainstream Australian English speakers, second-language English speakers, and speakers of other D2 English varieties (e.g., Diskin-Holdaway et al., 2024; Loakes et al., 2024b). Table 3 includes the six /e-æ/ contrast conditions, presented by continuum (word pair) and syllabic structure.

Table 3. /e-æ/ contrast conditions analyzed in this study.

Context	Continuum (Word Pair)	Syllable Structure
/_Vt/	<i>set–sat</i>	CVC
/_Vt/	<i>pet–pat</i>	CVC
/_VI/	* <i>pell–pal</i>	CVC
/_VI/	<i>shell–shall</i>	CVC
/_VI/	<i>hell–Hal</i>	CVC
/_VI/	** <i>skell–** skall</i>	CCVC

Very low frequency or obsolete words are marked * and nonce words are marked **.

For each target word, the most distinct tokens within the relevant pair were selected from the original recording to form the basis for creating a 7-step continuum of auditory stimuli. In the first step, the duration values of the endpoint stimuli in each word pair were equalized to the logarithmic mean of the raw durational values using the phonetic software PRAAT Version 6.3.17 (Boersma, 2001). This procedure ensured that potential perceptual bias derivable from temporal differences was minimized. While vowel duration is not known to impact listener perception of the two short vowels under analysis (Mannell, 2004; Loakes et al., 2024a), durational differences were controlled for at the word level. We acknowledge that this approach does not guarantee perceptual equivalence for stimuli duration across the steps but is consistent with the methods in vowel perception research in similar studies (e.g., Harrington et al., 2008). All word pairs have a durational ratio difference lower than 20%, which corresponds to the just-noticeable-difference (JND) threshold in human perception of vowel duration (Mauk & Buonomano, 2004). Each pair (e.g., *set–sat*) was then used to resynthesize a seven-step spectral continuum using the STRAIGHT package (Kawahara, 2006), applying a set of temporal and equidistant frequency anchor points to demarcate the phonetic-articulatory landmarks in the speech signal (e.g., vowels and liquids formant structures and the aperiodic wave of plosive consonants). The resynthesis was carried out using the default linear interpolation method, and the peak amplitude of all stimuli was automatically equalized at 1 Pascal. For each contrast pair and stimulus step, formant measurements are listed in Appendix B. This method allows for an analysis of the crossover between phonemes and is a standard way to analyze categorical perception.

In addition to the listening task, an online questionnaire was devised using Qualtrics (<https://www.qualtrics.com/>, Qualtrics, Provo, UT, USA) with the aim of collecting detailed demographic and sociolinguistic background information about the listeners. For all IndE participants (Ind-IndE and Aus-IndE groups), additional questions were included, such as languages spoken, onset of exposure to English during childhood, medium of instruction in school, and other factors. Due to space constraints and the specific focus of the present study, these variables are not explored in detail here and will be of particular relevance for examining the production data subset, based on the speech recordings from the

same participants. However, we include a summary of key points from the analysis of the production results in the Discussion (See Section 6) to further support the interpretation of findings in the present study.

4.3. Procedure

An online listening task was designed, following Loakes et al. (2024a), to elicit forced-choice auditory identification of an audio stimulus.³ The task was set up in *PsychoPy*, Version 2020.1.3 (Peirce, 2009) and hosted online using *Pavlov* (Bridges et al., 2020). Listeners were presented visually with two words on the screen (the DRESS and TRAP items from a given lexical pair), along with an audio stimulus corresponding to a different step along the continuum, and were instructed to select one of the words that best matched the stimulus they heard. Each unique audio stimulus was heard twice, once with the DRESS option on the left, and once on the right side of the screen, to control for visual biases and to help counterbalance for trial order. The listening task comprised a brief training block designed to familiarize participants with the task procedures. This was followed by seven experimental blocks, each containing 40 audio stimuli. In each trial, the two target words appeared on the screen 400 milliseconds after the audio stimulus began. Listeners then had an additional 100 milliseconds before the keyboard was activated, at which point they could make their selection. Reaction times (RTs) were also recorded for each trial. Participants were instructed to use headphones when completing the listening task.

Due to COVID-19 restrictions, data collection was carried out online. In both settings, Australia and India, a local research assistant with established links to the target listener groups recruited participants, communicated the instructions for the listening task and the questionnaire (together with the production task), and monitored completions.

4.4. Data Analysis

The collated responses were first examined for general patterns, i.e., distribution and outliers. As a result, the data from seven speakers were excluded from subsequent analyses (AusE group: three female and one male participant; Aus-IndE group: one female and one male participant; Ind-IndE group: one female participant). It is likely that these participants misunderstood the task since their responses showed a reverse pattern on the DRESS–TRAP continuum.

All statistical analyses were performed using R (version 4.5.0; R Core Team, 2025), employing Generalized Linear Mixed-Effects Models (GLMMs) from the *lme4* package (Bates et al., 2015). The primary goal was to investigate how GROUP, the acoustic STEP along a continuum, and GENDER influenced binary perceptual categorization responses. To account for repeated measures from the same individuals, LISTENER was included as a random intercept in all models. Each continuum (pair) was analyzed independently. For each pair, we used a forward stepwise model selection approach. This began with a critical evaluation of the STEP * GROUP interaction. We compared an intercept-only model, a main-effects model (STEP + GROUP), and an interaction model (STEP * GROUP) by incrementally adding more complex terms, including a quadratic term for step ($I(\text{step}^2)$) and the fixed effect of GENDER (and its interactions), retaining these only if they significantly improved the model's fit (LRT, $p < 0.05$).

Model fitting was optimized using the “bobyqa” algorithm with increased iterations and adaptive Gauss–Hermite quadrature points to ensure better convergence. Once the best-fitting model was identified for each pair, we assessed the significance of its fixed effects using Type II Wald χ^2 tests from the *car* package (Fox & Weisberg, 2019). For any models exhibiting a significant STEP * GROUP interaction, post hoc pairwise comparisons of estimated marginal means using the *emmeans* package (Lenth, 2023) were con-

ducted. A summary of model fit parameters for each continuum (word pair) is presented in Table A3 in Appendix C, including the results of the likelihood ratio tests (LRTs) between the final model and the baseline intercept-only model, as well as the concordance index *C*. *C*-values between 0.8 and 0.9 indicate excellent discrimination (Hosmer & Lemeshow, 2000; Levshina, 2015). For the final models, we report log-odds coefficients (β) and their significance tests, including odds ratios (OR) and 95% confidence intervals (See Tables A4–A9, Appendix C).

In addition to GLMMs, we employ a relatively recent approach in sociolinguistic research, namely random forest and, more specifically, conditional inference trees (Levshina, 2015), to examine the potential influence of GENDER and LOR on listener responses (see also Diskin-Holdaway et al., 2024). Similarly to logistic models such as GLMM, random forests aim to predict an outcome based on a set of predictors. However, whereas logistic regression uses a fixed mathematical equation to estimate the influence of each predictor, random forests adopt a data-driven approach. They construct conditional inference trees through a process of recursive binary splitting, where the dataset is repeatedly divided into subsets based on statistically significant associations between predictors and the response variable (Tagliamonte & Baayen, 2012; Levshina, 2015). At each step, a test of independence is conducted for each predictor—whether it be categorical (e.g., STEP) or continuous (e.g., LOR). Predictors that are statistically independent of the response are excluded from the model. If more than one predictor is found to be significant, the one with the strongest association (i.e., lowest *p*-value) is selected, and a binary split is made on that variable. This recursive process continues until no further significant splits can be made. Importantly, conditional inference trees are designed to minimize bias in variable selection, ensuring that neither continuous variables, e.g., LOR, nor categorical predictors with multiple levels are favored. This makes them well-suited to analyses involving a mix of data types (Tagliamonte & Baayen, 2012) and “particularly useful in the situations of small *n* [number of observations], large *p*” (Levshina, 2021, p. 613).

5. Results

5.1. Vowel Categorization

5.1.1. Reading the Plots

Figures 1–3 illustrate responses to the seven-step continuum for a selection of word pairs across the three listener groups. In each plot, the *x*-axis represents the stimulus step, and the *y*-axis shows the proportion of responses (i.e., observed probabilities of selecting one of the words). We begin by explaining how to read the figures, using the *pet*–*pat* contrast (Figure 1a) as an example. This plot shows the proportion of responses for *pet* at each step of the continuum. For example, a value of 1.0 at Step 1 would indicate that 100% of listeners responded with *pet* for that stimulus, while a value of 0.75 at Step 1 would mean that *pet* was chosen in 75% of responses, and *pat* for the remaining 25%. Similarly, 0% at Step 7 would mean *pet* was never selected for the most *pat*-like stimulus at Step 7, whereas the value of 0.25 at Step 7 would indicate that 25% of listeners chose *pet* at the endpoint.

Further, on the *x*-axis, Step 4 represents the acoustic midpoint for the contrast pair, while on the *y*-axis, the value of 0.50 corresponds to the crossover point—the precise center of the perceptual boundary, or transition area—between stimulus at Steps 1 and 7. For example, values above 0.50 indicate a preference for *pet*, whereas values below 0.50 indicate that listeners have perceptually ‘crossed over’ to *pat*. Responses with values at or near 0.50 suggest that the stimulus is maximally ambiguous (at least across the group). Crossing the 0.50 line indicates a shift (of greater or lesser magnitude) in what the listeners perceive, from more DRESS-like to more TRAP-like, and so behavior around the 0.5 line—determining the shape of the identification curve—is also informative. Many re-

sponses falling in a wider zone around 0.5 would suggest a high degree of uncertainty, while a steeper ‘fall’ from above 0.50 to below 0.50 suggests a more categorical perceptual response (the typical s-curve of a categorical response).

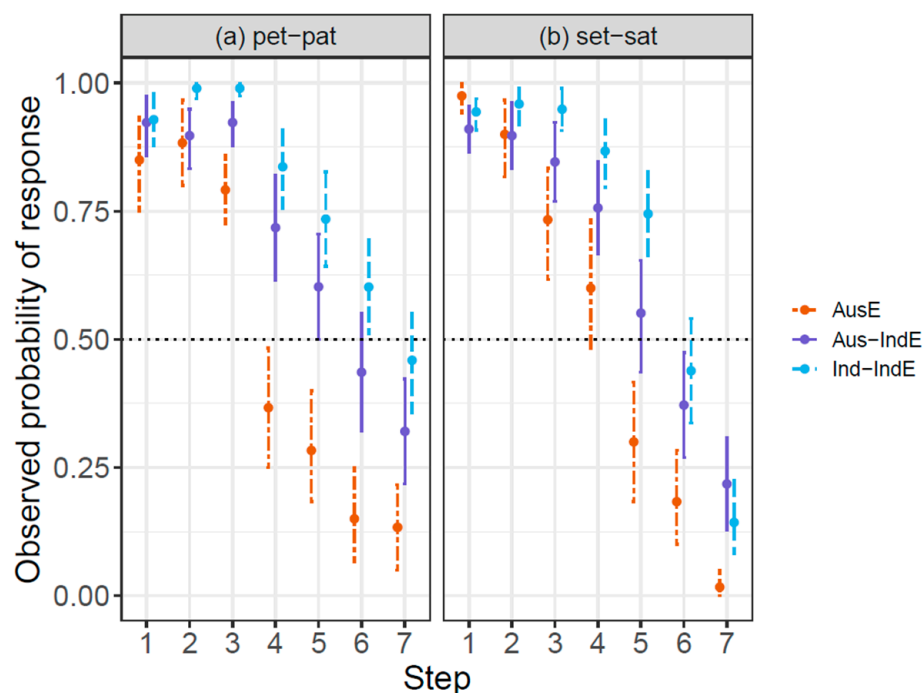


Figure 1. (a) Observed probabilities by continuum step for *pet-pat*; (b) observed probabilities by continuum step for *set-sat*; presented by group: red—AusE listeners, blue—Aus-IndE listeners, and purple—Ind-IndE listeners.

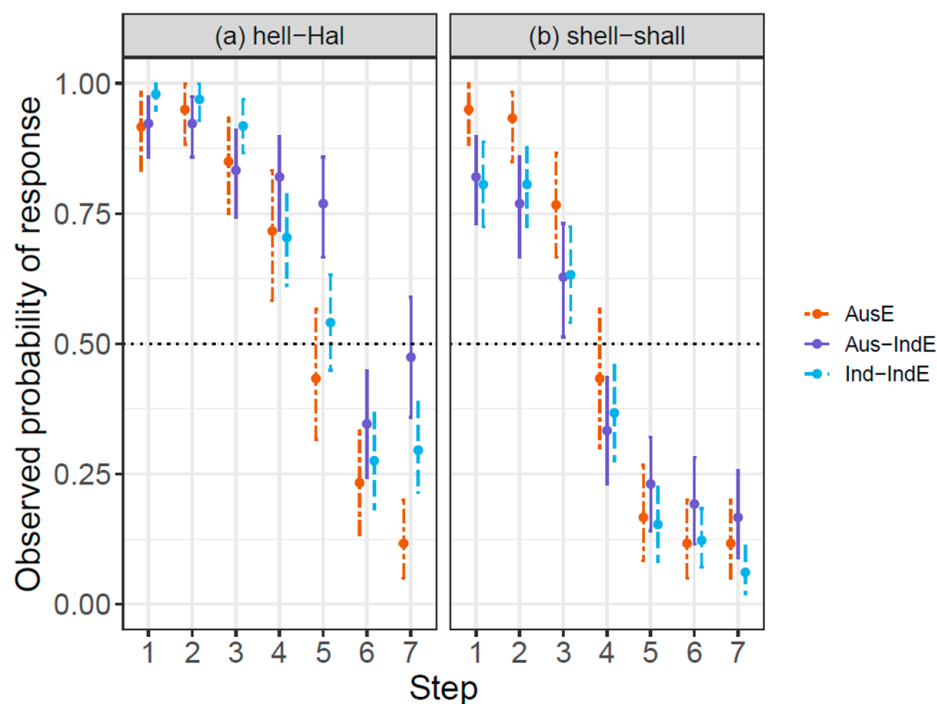


Figure 2. (a) Observed probabilities by continuum step for *hell-Hal*; (b) observed probabilities by continuum step for *shell-shall*; presented by group: red—AusE listeners, blue—Aus-IndE listeners, and purple—Ind-IndE listeners.

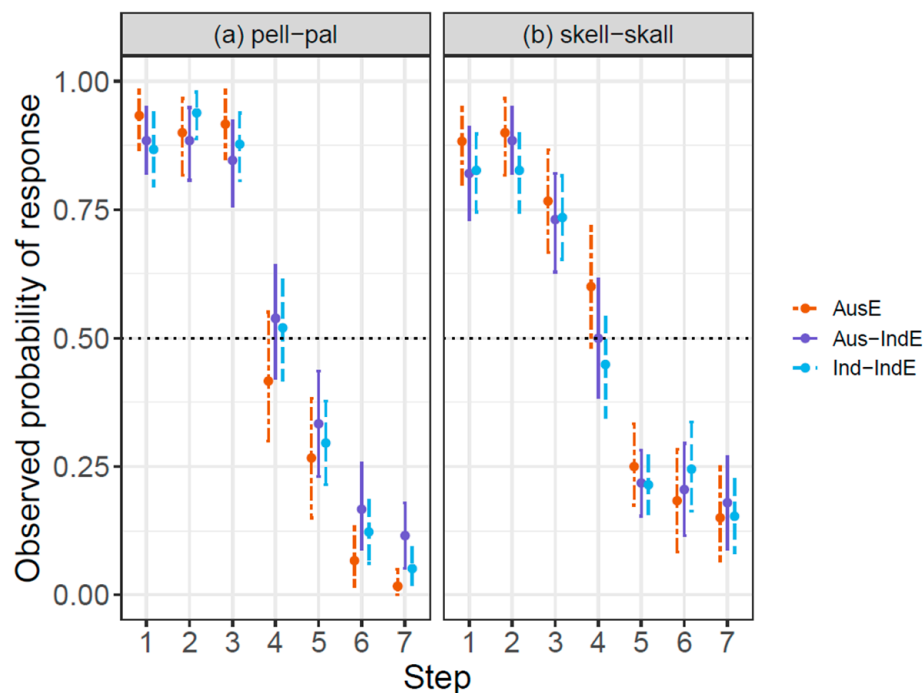


Figure 3. (a) Observed probabilities by continuum step for *pell-pal*; (b) observed probabilities by continuum step for the nonce pair *skell-skall*; presented by group: red—AusE listeners, blue—Aus-IndE listeners, and purple—Ind-IndE listeners.

5.1.2. Non-Lateral Contexts

The analysis for *pet-pat* revealed a significant effect of STEP ($\beta = -1.09$, SE = 0.10, $z = -11.00$, $p < 0.001$) and GROUP (Type II Wald $\chi^2(2) = 42.14$, $p < 0.001$), as well as a significant interaction between STEP and GROUP (Type II Wald $\chi^2(2) = 6.80$, $p = 0.033$), indicating that the effect of STEP varied by cohort. In other words, the cohorts responded differently to the continuum. A summary of the model results, including coefficients and their significance tests, can be viewed in Table A4 (Appendix C).

As illustrated in Figure 1a, at every step, both IndE groups (blue and purple trajectories) have more DRESS responses than the AusE group (red trajectory), indicative of a greater perceptual bias towards DRESS for the two IndE groups (relative to AusE speakers), although this effect is much weaker for the Aus-IndE group (purple is lower than blue). Further, the AusE group demonstrates the earliest crossover, which occurs close to the acoustic midpoint at Step 4. This earlier shift for the AusE listeners is also relatively abrupt, as evidenced by the responses to stimuli in Steps 1–3, which largely remain in the upper quartile, being followed by a sharp fall between Steps 3 and 4. This creates a curve which resembles more closely the classic s-curve of categorical perception. In contrast, for the two IndE listener groups, the category crossover occurs later and more gradually. For baseline Ind-IndE listeners, the crossover does not occur until Step 7, and even then, it barely goes lower than 0.50, suggesting no clear, unambiguous perception of the TRAP vowel. For the Aus-IndE listeners, the crossover occurs a little earlier, at Step 6. This suggests a stronger overall bias towards perceiving *pet* (i.e., the DRESS vowel) among the IndE groups and a less categorical response to step changes in the continuum.

Despite the general trend of a later crossover for IndE listeners overall, as well as a higher relative bias towards the DRESS vowel, the results indicate clear differences in the categorization behavior between the Ind-IndE and Aus-IndE listener groups (significant effect of STEP and GROUP interaction for Aus-IndE listeners: $\beta = 0.31$, SE = 0.12, $z = 2.60$, $p = 0.009$; OR = 1.36, 95% CI [1.08, 1.72]). Firstly, at all steps except Step 1, the Aus-IndE group shows a weaker response bias towards perceiving DRESS, relative to the Ind-IndE

group (the purple trajectory is lower than the blue trajectory). Secondly, a substantial proportion of responses (approximately 45%) for the Ind-IndE group remained as *pet* even at Step 7, indicating that this step is almost entirely ambiguous for them, compared to approximately 31% of responses at Step 7 for Aus-IndE listeners. Further, pairwise comparisons of *pet* responses show a sharper decrease across steps for Aus-IndE listeners than for Ind-IndE listeners, starting at step = 3.875 (Aus-IndE: $z = 2.614$, $p = 0.009$; Ind-IndE: $z = 1.800$, $p = 0.070$). In other words, although both Indian groups pattern distinctly from AusE listeners, the Aus-IndE listeners' behavior resembles more closely that of AusE listeners, particularly in their propensity to categorize the TRAP vowel in this context. This pattern suggests a modest degree of perceptual accommodation among Aus-IndE to D2 for the non-lateral context, likely due to exposure to the ambient variety. Figure 1b illustrates the probability of responses along the continuum for the *set-sat* pair. As with *pet-pat*, regression modeling results revealed significant main effects of STEP ($\beta = -1.37$, $SE = 0.12$, $z = -11.424$, $p < 0.001$) and GROUP (Type II Wald $\chi^2(2) = 15.79$, $p < 0.001$), and a significant interaction between STEP and GROUP (Type II Wald $\chi^2(2) = 23.97$, $p < 0.001$). A summary of the model results for *set-sat*, including coefficients and their significance tests, can be viewed in Table A5 (Appendix C). As depicted in Figure 1b, at almost every step (Steps 3–7), both IndE listener groups (blue and purple trajectories) have more DRESS responses than the AusE group (red trajectory), indicating that both IndE listener groups have a bias towards hearing DRESS for this lexical pair (similarly to *pet-pat*, albeit not quite as consistently). However, compared to the AusE group, the Aus-IndE listeners displayed less perceptual bias towards *set* overall ($\beta = 0.62$, $SE = 0.13$, $p < 0.001$; OR = 1.85, 95% CI [1.43, 2.39], before or after the acoustic midpoint (i.e., in Steps 5 and 6) than Ind-IndE listeners ($\beta = 0.33$, $SE = 0.14$, $p = 0.017$; OR = 1.39, 95% CI [1.06, 1.82]). In other words, as with *pet-pat*, their behavior showed some accommodation towards that of the AusE group.

With regard to crossover, the AusE listeners again exhibited an earlier crossover than the two IndE listener groups, with the majority of AusE listeners categorizing the stimuli as *sat* by Step 5, i.e., a little later than they did for *pet-pat*. By Step 7, AusE listeners almost unanimously categorized the stimuli as *sat*, which also mirrors the almost unanimous categorization of the Step 1 stimulus as *set*. The more closely unanimous perception of the endpoints of the continuum for *set-sat* are possibly an artifact of the onset fricative in this pair which (compared with the voiceless stop closure internal for the *pet-pat* pair) could play a critical role in providing early acoustic evidence (through coarticulatory spectral cues). In contrast, while some responses from Aus-IndE listeners were recorded at Step 5, more generally, both IndE cohorts showed a relatively high probability of *set* responses than *sat* responses at Step 5 and still showed many *set* responses even by Step 7. This is particularly true for the Aus-IndE listeners, with approximately 22% *set* responses at Step 7, suggesting a tendency to perceive the more open vowel stimulus as *set*. This pattern suggests a potential adaptation in their perceptual responses as a result of exposure to Australian English in their new sociolinguistic environment.

In summary, for the non-lateral context, both Indian groups show a distinct perceptual response from that of AusE listeners, with a greater bias towards hearing DRESS at all points of the continuum from Step 3 onwards, and a clear reluctance to perceive TRAP until the very end of the continuum. The response curves show greater perceptual ambiguity for the Indian listeners, especially towards the TRAP end of the continuum, as evidenced by a less steep response curve. However, at almost all points, the Aus-IndE response curve is somewhat closer to the AusE response curve, suggesting that a degree of perceptual accommodation has occurred, with a mapping of acoustic stimuli to perceptual categories that have shifted to some extent towards that of AusE mapping.

5.1.3. Pre-Lateral Contexts

Figure 2a illustrates the probability of responses to stimuli along the seven-step continuum for the *hell–Hal* pair. All three groups show a dominant percept of the DRESS vowel for Steps 1–4, but a weaker and more inconsistent percept of the TRAP vowel for Steps 6–7. A generalized linear mixed-effects model revealed a significant main effect of STEP ($\beta = -1.07$, SE = 0.10, $z = -10.762$, $p < 0.001$), indicating that listeners' categorization responses changed systematically across the acoustic continuum. A significant main effect of GROUP also emerged (Type II Wald $\chi^2(2) = 7.127$, $p = 0.028$), suggesting overall differences in perceptual patterns across the three groups. A significant STEP * GROUP interaction (Type II Wald $\chi^2(2) = 20.05$, $p < 0.001$) showed that the effect of acoustic variation along the continuum differed across the cohorts. However, the patterns are more nuanced for this word pair. A summary of the model results for *hell–Hal*, including coefficients and their significance tests, can be viewed in Table A6 (Appendix C).

As shown in Figure 2a, and unlike the non-lateral context where both IndE groups showed more DRESS responses than the AusE group, for the lexical pair *hell–Hal*, only the Ind-IndE group shows (marginally) more DRESS responses than the AusE group. Indeed, with this lexical pair, Aus-IndE listeners exhibit similar patterns to the AusE group, at least for Steps 1–3, with both showing a slightly lower rate of *hell* (i.e., DRESS) responses compared with the Ind-IndE listener group, particularly at Steps 2 and 3. From Step 4 onwards, however, the behavior of the Aus-IndE group patterns even less closely with AusE than that of the Ind-IndE group. The latter reaches the perceptual crossover threshold around Step 5, although IndE listeners continue to be less certain in categorizing stimuli as *Hal*. The Aus-IndE group finally identifies the stimulus at Step 6 mostly as *Hal*, following a very sharp shift in responses after Step 5. However, by Step 7, this trend reverses, and the Aus-IndE group appears to be answering at random, with the rate of *hell* responses increasing (to approximately 46%, up from approximately 38% at Step 6). In comparison with the AusE group, the significant STEP * GROUP interaction for their cohort further highlights their distinct perceptual trajectory across the continuum ($\beta = 0.43$, SE = 0.12, $z = 3.636$, $p < 0.001$). In other words, the Aus-IndE group responds more variably and with greater uncertainty to the TRAP end of the stimuli continuum than does the Ind-IndE group ($\beta = 0.03$, SE = 0.12, $z = 0.277$, $p = 0.782$), suggesting anything but accommodation towards Australian English, and diverging quite markedly from the findings in the non-lateral context. With regard to a possible lexical effect, the item 'Hal', a given name for males in Australia (albeit less commonly for younger people), may not be familiar to IndE listeners, but this does not provide an explanation for why there is greater uncertainty for Aus-IndE listeners than for Ind-IndE listeners.

A less variable, but somewhat similar, picture emerges for another pre-lateral pair *shell–shall*. Unlike *hell–Hal*, this pair is lexically balanced, as the words at the two ends of the continuum are familiar and frequent in both Australian English and Indian English. As such, it arguably provides a clearer illustration of lexical perception based more purely on acoustic differences, without the confound of large discrepancies in lexical frequency. As illustrated in Figure 2b, responses cluster more tightly across the groups than those for *hell–Hal*, with the crossover threshold occurring around Step 4 for the three groups. Indeed, logistic regression results revealed a significant main effect of STEP ($\beta = -1.37$, SE = 0.12, $z = -10.999$, $p < 0.001$), indicating systematic variation in vowel response across the continuum.

As Figure 2b clearly shows, the slight bias towards a TRAP response in the AusE group (relative to both IndE groups) that was observed throughout the 7-step continua for the non-lateral context (i.e., *pet–pat* and *set–sat*), as well as in Steps 5–7 for the pre-lateral *hell–Hal* context, is clearly *not* present for *shell–shall*. On the contrary, the AusE group

shows a higher preference (relative to the IndE groups) towards DRESS for Steps 1–4. Indeed, while there was no main effect of GROUP (Type II Wald $\chi^2(2) = 2.36, p = 0.3$), the interaction between STEP and GROUP was significant (Type II Wald $\chi^2(2) = 14.978, p < 0.0001$), pointing to group-specific perceptual boundaries for the two IndE cohorts relative to the AusE cohort (Aus-IndE: $\beta = 0.53, SE = 0.14, z = 3.808, p < 0.001$; Ind-IndE: ($\beta = 0.32, SE = 0.14, z = 2.292, p < 0.02$). Both IndE-speaking groups show less certainty in selecting *shell* responses at Steps 1–3. For instance, the probability of hearing *shell* at Step 1 was 94% for AusE listeners, compared to 80–81% for Ind-IndE and Aus-IndE listeners. A summary of the model results for *shell–shall*, including coefficients and their significance tests, can be viewed in Table A7 (Appendix C).

One possible explanation is that the two IndE groups anticipated a phonetically higher DRESS vowel than the stimuli heard in the lateral context, and that is due to differences in the lateral between D1 and D2. Even though the AusE speaker who provided the recordings for the stimuli was not a ‘merger’ herself, she is likely to show allophonic variation pre-laterally (where the lateral is dark); it is precisely such a conditioned variation that is likely to have motivated and led to the TRAP–DRESS merger in speakers for whom the merger has been documented. The fact that the two IndE groups did not appear to anticipate this, and therefore did not account for it in their perception, would suggest that coarticulatory effects on a pre-lateral vowel in IndE (where the lateral is light; Wells, 1982; Gargesh, 2006; Shaktawat, 2024) differ from those in AusE (where the lateral is dark). This effect may not have been evident for *hell–Hal* due to a possible negative bias against selecting *Hal*. In other words, perceptual response behavior appears to be mediated by an interplay of lexical effects, L1 acoustic mapping, and knowledge of subtle, sub-phonemic cues that may span several segments.

The high lexicality and the availability of acoustic cues into the coda of this pair should, arguably, promote good conditions for greater certainty in the perceptual response. Overall, all three groups show a high likelihood (over 75% of responses) of selecting *shall* by Steps 5–7 and a fairly robust s-curve in the identification response. However, when viewed over the whole trajectory, there is a notable difference in behavior between the two IndE groups. The similarity in response patterns observed between Ind-IndE and Aus-IndE cohorts at Steps 1–4 shifts at Step 5, with the perceptual behavior of the two IndE cohorts beginning to diverge. Aus-IndE listeners show reduced certainty in identifying *shall* at Steps 5–7 (approximately 81–84%) in comparison to Ind-IndE listeners, whose likelihood of selecting *shall* responses is 94% (Step 7), which is even higher than that of the AusE cohort.

After initial similarities, this divergence of the Aus-IndE group from the Ind-IndE group in the later steps is somewhat odd, especially given the unexpected direction. Our expectations are that exposure to the Australian English variety will lead to a subtle recalibration of perceptual boundaries for this vowel contrast regardless of phonetic context, i.e., whether it be in the pre-lateral or non-lateral environment. However, unlike in the non-lateral context where perceptual responses are *more* similar to that of AusE speakers, in the pre-lateral context, such exposure has led to a notable *divergence* from AusE responses. In the lexical pairs considered so far, Aus-IndE responses show the least confidence in identifying TRAP at the TRAP end of the continuum in the pre-lateral context. Somewhat paradoxically, this pushes the pattern of their responses even further away from that of the AusE group than that of the Ind-IndE group.

We see traces, albeit weaker, of this apparent anomaly for the *pell–pal* pair as well (see a summary of the model results in Table A8 (Appendix C)). Firstly, as with all lexical pairs, the analysis revealed a significant main effect of STEP ($\beta = -1.41, SE = 0.13, z = -11.071, p < 0.001$), strongly indicating clear sensitivity to the acoustic continuum (OR = 0.24, 95%

CI [0.19, 0.31]), as evident in Figure 3a. However, as already illustrated for *hell–Hal* and *shell–shall*, the pre-lateral context is distinct from the non-lateral context. Firstly, the apparent slight bias towards TRAP for the AusE listeners (relative to the IndE groups), observed in the non-lateral context *throughout the continuum*, is not evident here in the pre-lateral context. Secondly, and as illustrated in Figure 3a, responses across the three cohorts were tightly clustered—particularly at Steps 1–3 and 6–7—with a sharp decline from *pell* to *pal* occurring between Steps 3 and 4.

Although more tightly clustered than for the non-lateral context, the response patterns in the figure show a lower probability of selecting *pell* at Steps 4–7 for AusE listeners in comparison with both IndE listener groups, i.e., an earlier perception of TRAP in the continuum. This is particularly so when compared with the Aus-IndE group, which shows a lesser propensity to perceive TRAP. Indeed, this was shown to be statistically significant in that the interaction between STEP and GROUP further revealed differences for the Aus-IndE group ($\beta = 0.48$, SE = 0.14, $z = 3.304$, $p < 0.001$), indicating that the Aus-IndE listeners' point of categorization across the continuum of steps was distinct from that of the AusE group. In particular, as with *Hell–Hal* and *shell–shall*, Aus-IndE responses show the least confidence in identifying TRAP at the TRAP end of the continuum in the pre-lateral context, pushing Aus-IndE responses even further away from AusE responses than those of Ind-IndE listeners.

These results suggest that while listeners across cohorts broadly followed a similar response pattern along the acoustic continuum, there were group-specific differences in the precise location of category boundaries, that is, in which step the perception shifts, particularly for the Aus-IndE group. What is perhaps surprising, however, is that the divergent response behavior of the Aus-IndE group when compared with the Ind-IndE group is not uniformly a movement towards the behavior of the AusE group. It would appear that the pre-lateral context in particular presents a more 'difficult' set of contrasts, perceptually, for the Aus-IndE group.

Figure 3b shows the results for one of the pairs of nonce words, *skell–skall*, i.e., in a paradigm where there can be no lexicality bias. Once again, we see that the apparent slight bias towards TRAP in AusE listeners (relative to the IndE groups), observed in the non-lateral context, is not evident here either. Regression analyses revealed a significant main effect of STEP (*skell–skall*: $\beta = -0.89$, SE = 0.04, $z = -20.525$, $p < 0.001$) but, unlike with the lexical pairs, no main effect of GROUP (*skell–skall*: Type II Wald $\chi^2(2) = 0.82$, $p = 0.7$), with the final (best fit) model excluding a STEP * GROUP interaction. For all three cohorts, there is a relatively steep identification curve, with a perceptual threshold for all three falling at Step 4 (no pairwise group contrasts were significant at Step 4) and—unlike any of the lexical pair paradigms—notably tight clustering of responses at both DRESS and TRAP ends of the continuum. A summary of the model results for *skell–skall*, including coefficients and their significance tests, can be viewed in Table A9 (Appendix C).

5.1.4. Summary of Vowel Categorization Results

To summarize, the analysis of responses across the stimuli continuum and the three cohorts revealed distinct behavioral patterns between the Aus-IndE and the two IndE groups. Broadly speaking, in the non-lateral context, both IndE groups showed stronger bias towards perceiving the DRESS vowel across the continuum compared to the AusE group. As suggested above, this pattern could be due to the TRAP vowel being lower in both L1 Indian languages and Indian English than in Australian English, leading to an earlier crossover from /e/ to /æ/. However, the perceptual responses (See Figure 1a,b) clearly show Aus-IndE listeners' responses to be somewhat closer (compared to Ind-IndE responses) to those of AusE listeners, indicating an effect of exposure to the variants of the TRAP vowel

in this variety (in the context of vowel merger in progress). Further, the pre-lateral context also revealed differences between Aus-IndE and Ind-IndE perceptual responses but of a more complex nature. In this context, somewhat counter-intuitively, the IndE subgroup closest to AusE was the IndE ‘base’ group, i.e., listeners in India, with little to no prior experience of AusE. It was suggested that while all three groups would have ‘knowledge’ of universal coarticulatory effects of dark laterals on preceding vowels, and could therefore compensate for this, the Aus-IndE group’s particularly divergent response was likely mediated by exposure to and experience with a singularly wide pool of possible variants of DRESS and multiple variants of /l/ in Australian English, with the latter combined with Indian English variants. In contrast, Indian English(es) are characterized by a light /l/ in all syllable positions, a feature not only commonly retained among first-generation Indian migrants but also potentially more phonetically exaggerated (Shaktawat, 2024). Thus, while exposure to the Australian English vowel variants may have ‘benefitted’ Aus-IndE listeners in the non-lateral context, it appears to have incurred perceptual costs in the pre-lateral context.

5.2. Effects of LOR and Gender for Aus-IndE Listeners

To address the second research question—whether the length of exposure to Australian English predicts Aus-IndE listeners’ perceptual responses—we conducted a set of conditional inference tree analyses for each stimulus pair. Specifically, we focused on the predictive variables of LOR and GENDER. Gender is of relevance in this set of analyses for two reasons: (a) gender has been included in the final model for the pairs *set–sat*, *hell–Hal*, *shell–shall*, and *pell–pal* in the response results (See Table A3 in Appendix C); (b) previous research on D2 acquisition has reported gender differences in vowel categorization in non-lateral contexts for Irish English listeners in Australia (Diskin-Holdaway et al., 2024). While conditional inference trees were fitted for each stimulus pair, among the pairs, LOR emerged as a contributing factor for the non-lateral pair *set–sat* but not for *pet–pat* and did not contribute to response patterns for any of the pairs with a lateral coda. GENDER emerged as a significant contributing factor only for *set–sat* and *pet–pat* (non-lateral contexts). This section presents selected conditional inference trees (Tagliamonte & Baayen, 2012; Levshina, 2015) for non-lateral (*set–sat*) and lateral contexts (*shell–shall*, *skell–skall*).

With regard to interpretation of the tree diagrams, each oval (node) represents a variable that significantly contributes to the data split ($p < 0.05$), along with its associated p -value. The ‘branches’ illustrate the levels of these variables, indicating how the data is divided. At the bottom of each branch, the bar plots (referred to as ‘leaves’) show the proportion of listeners who selected DRESS over TRAP in each node. The number of observations for each end node (‘bin’) is shown in parenthesis above the bar plots (Levshina, 2015). Prior to reporting and interpreting the results, we illustrate how to interpret conditional inference trees by examining Figure 4 which represents *set–sat*. The first branching, as shown in the topmost oval of Figure 4, is based on STEP. We see that this split occurs between Step 4 and Step 5, dividing the responses in relation to the acoustic midpoint for the contrast, as indicated by the branches labeled ≤ 4 and > 4 . Ovals 2 and 5 correspond to further branching between different steps in the perceptual task. The branching to the left of the figure, at oval 2, illustrates differences in responses between Steps 1–3 and Step 4, as indicated by the branches ≤ 3 and > 3 . The branching to the right, at oval 5, separates the observations into two branches labeled as > 5 (with 156 of *sat* observations at node 11) and ≤ 5 (leading to oval 6). A subsequent split at oval 6 shows branching by gender (M vs. F), with responses for male listeners grouped under node 7 (48 observations of *set*) and responses for female listeners followed by another split at oval 8, based on LOR. For female

listeners, differences in responses depend on LOR, as indicated by the branches labeled ≤ 6.92 years and > 6.92 years.

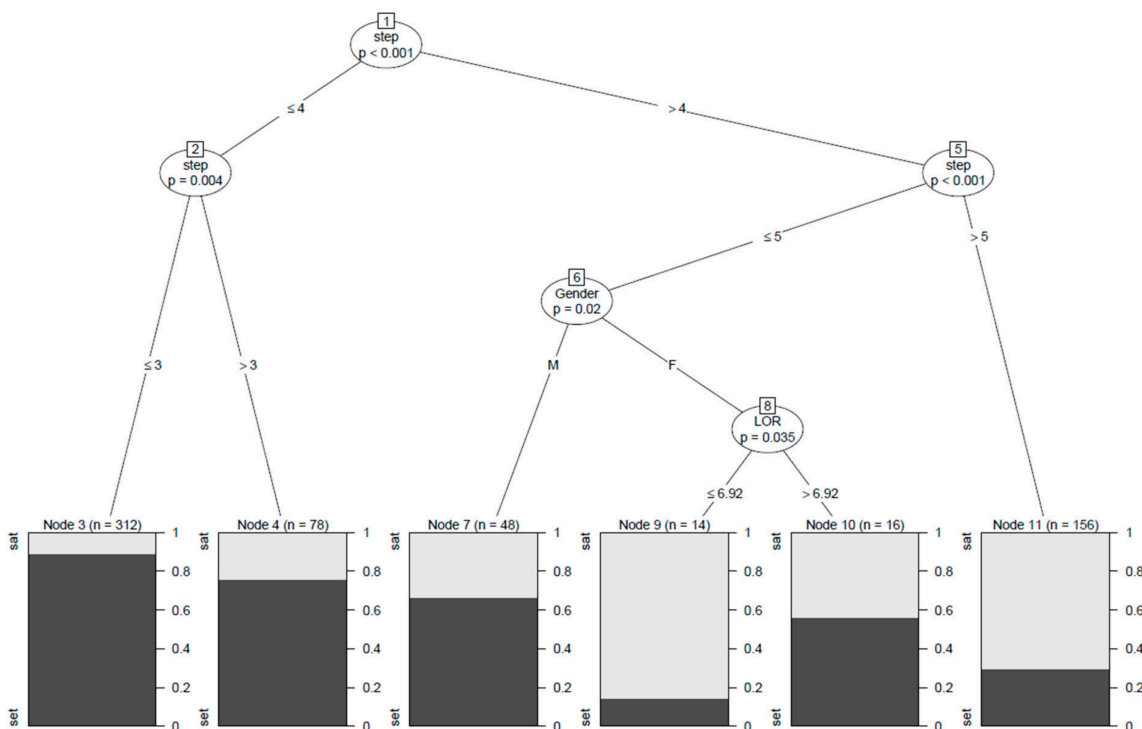


Figure 4. Conditional inference tree for *set*–*sat* continuum by Aus-IndE listeners, showing significant effect of LOR and gender.

Based on the conditional inference tree for *set*–*sat* (Figure 4), we see that for this non-lateral context, the main branching occurs between Steps 4 and 5. Node 6 separates responses based on gender, with males producing more *pet* responses at or before Step 5 (M). A further split at node 8 for female listeners (F) separates females with fewer than 6.92 years’ residence in Australia, who produce more *TRAP* responses (14 observations, node 9) than females with more than 6.92 years’ residence (16 observations, node 10). The fact that LOR is only significant at this point in the continuum suggests that the increase in *TRAP* responses for this female sub-group is quite sudden (i.e., generating a particularly steep identification curve). Cross-referencing with Figure 1b (purple graph), we see that Step 5 is the point of greatest ambiguity for the Aus-IndE cohort; and from the inference tree we now see that the male participants’ response is more squarely *DRESS*, while the female response is varied. Female listeners with a shorter LOR have more *TRAP* responses; the longer-LOR females are more squarely ambiguous, i.e., close to the 50% mark. Upon closer examination of the demographic data for the Aus-IndE females, we see that all but one female listener are students in their twenties and have an LOR ≤ 6.92 . We will return to these results and provide a possible explanation in the Discussion of the paper (Section 6). Interestingly, while the overall branching pattern for the second pair in the non-lateral context, *pet*–*pat* (not illustrated here), was similar to *set*–*sat*, with *GENDER* playing a role in conditioning responses and females being more likely to respond with *pat* at Step > 4 as compared to males. LOR had no effect on the responses of female listeners.

Turning to the pre-lateral context illustrated in Figures 5 and 6 for pairs *shell*–*shall* and *skell*–*skall*, we observe no effects of LOR or *GENDER*, with *STEP* being the only strong predictor for listener responses. This pattern, i.e., *STEP* as the only significant factor contributing to listener responses, was observed for all stimulus pairs in the pre-lateral condition, suggesting that any influence from coarticulatory cues of the lateral is neutral to these

factors. However, the structure of the conditional inference trees for real and nonce word pairs reveals notable differences. For the real-word pair *shell–shall*, responses were more evenly distributed, resulting in a more balanced, S-shaped categorization curve. In contrast, the nonce-word pair *skell–skall* showed a distinct shift in curve shape between Step 3 and Step 4, as indicated by the branches ≤ 4 and > 4 , where listeners displayed complete ambiguity—categorizing the stimulus as DRESS or TRAP with equal probability.

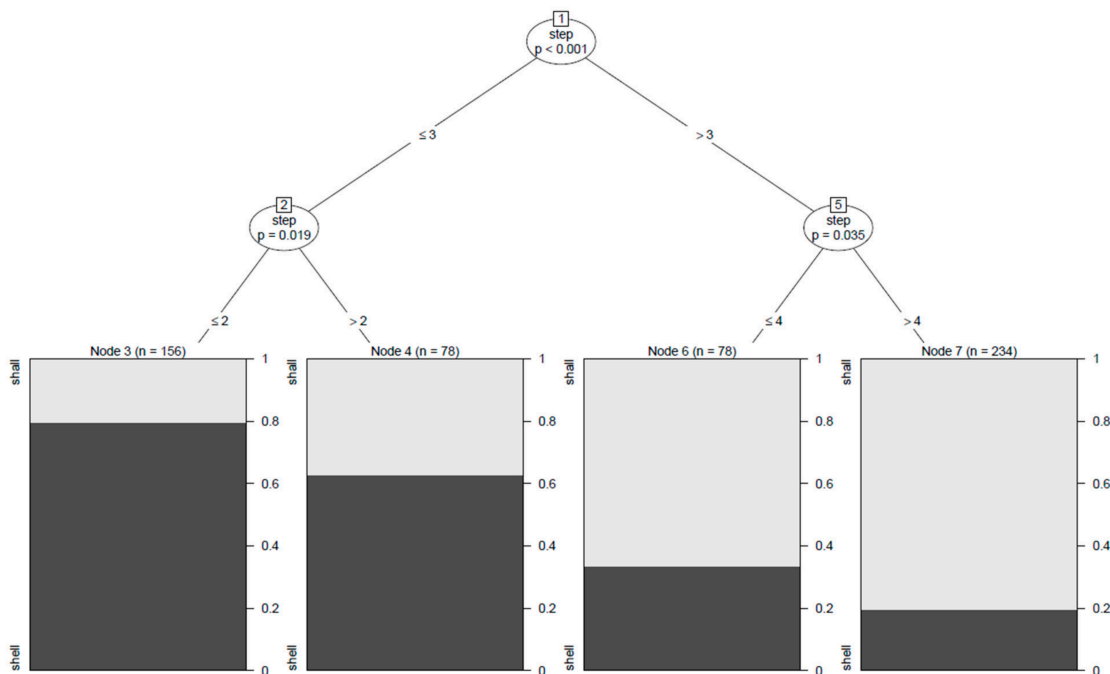


Figure 5. Conditional inference tree for *shell–shall* continuum by Aus-IndE listeners.

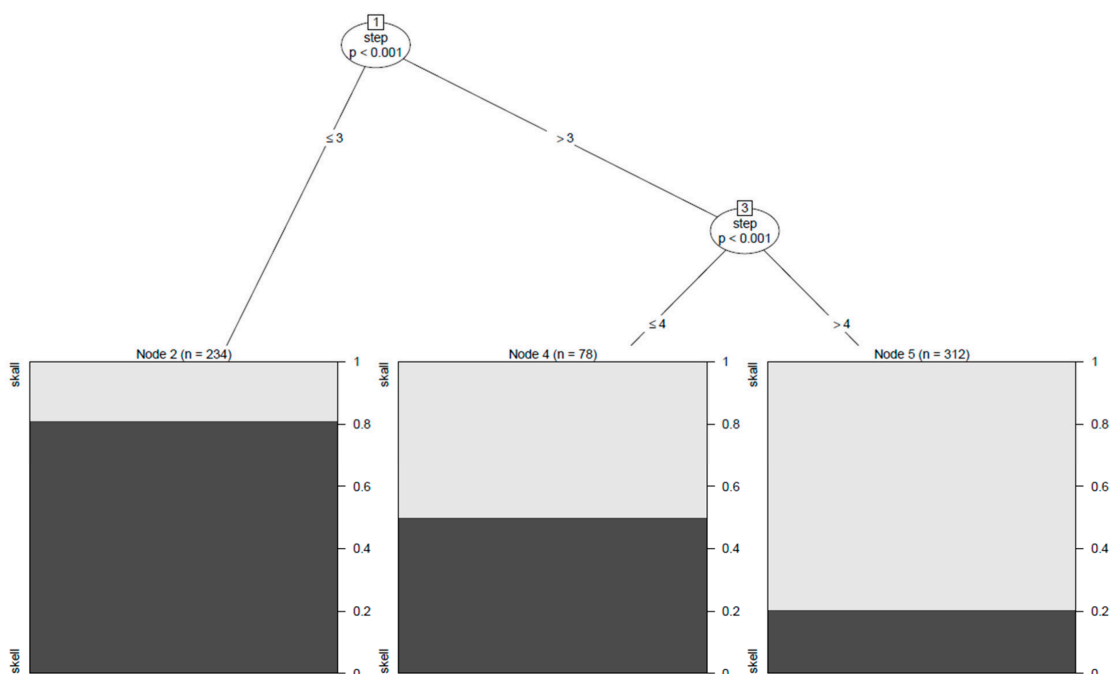


Figure 6. Conditional inference tree for *skell–skall* continuum by Aus-IndE listeners.

6. Discussion

In this paper, we investigated the perception of Australian English lax vowels /e æ/ in three groups of listeners: Australian English, Indian English, and Australian Indian En-

glish (first-generation migrants to Australia). The aim of the study was to analyze the perceptual behavior of first-generation Indian English-speaking migrants in Melbourne to assess the extent and nature of contact-induced variation and potential change, i.e., whether, and to what degree, the Indian migrants had adapted to their new linguistic environment in Australia (i.e., to Australian English, their D2) with respect to perception. In investigating this, it was also necessary to investigate the perceptual behavior of “baseline” Australian English and Indian English listeners. We additionally investigated a range of extralinguistic variables, specifically the influence of length of residence in the new environment, and gender. The phonetic context of the target vowels /e æ/ was varied (before a lateral vs. before a non-lateral) to additionally explore the impact of current allophonic variability (and partial merger) in the D2 being acquired.

Firstly, considering the overall results, the Aus-IndE listeners in the non-lateral condition were observed to apparently have perceptual responses *in-between those of the AusE and the baseline comparison involving Ind-IndE listeners*, suggesting a degree of perceptual adaptation. On first consideration, this result appears straightforward as expected, perhaps deceptively so: given their cumulative linguistic experience both prior to and following migration, the Aus-IndE cohort has been exposed to a wider overall pool of variants for both DRESS and TRAP, arguably affording them greater perceptual flexibility. This aligns with findings mentioned earlier by Clopper and Walker (2017) that listeners with exposure to multiple varieties have wider boundaries for phoneme identification (also see Loakes et al., 2024b for Aboriginal English listeners in Australia). It also supports observations (Munro et al., 1999; Ziliak, 2012; Nycz, 2018) that, in both the loss (or shift away) of first dialect features and the acquisition of second dialect features, resulting forms are often intermediate between the two dialects. We note that evidence of such intermediate forms is not necessarily evidence of an *ongoing* process of accommodation *per se* in the sense of listeners *gradually moving their perceptual targets towards* those of AusE. It may just be that they end up with a hybrid version at the center of a wider pool that encompasses the reach of both IndE and AusE variants. Longitudinal studies could potentially cast light on whether D2 listeners move their representation or merely widen its ‘catchment area’.

However, in the pre-lateral condition, the findings paint a rather more complex picture. In this condition, and somewhat paradoxically, the Aus-IndE listeners appear to move *away* from the response behavior of AusE listeners, rather than towards it, particularly by showing a substantial degree of perceptual confusion toward the TRAP endpoint of the continuum for *hell–Hal*, and also—albeit to a lesser extent—for *shell–shall* and *pell–pal*. Furthermore, this apparent greater confusion on the part of Aus-IndE speakers was not the case with *skell–skall*, i.e., where nonce word pairs were used. In this context, all three listener groups responded relatively similarly. Thus, the apparent move *away* from the D2 target, in the pre-lateral context, only occurs in the context of a real lexical choice.

These seemingly paradoxical findings suggest that adaptation to the new linguistic environment entails more than a simple shift in phonetic categories at the segmental level, and as such demands a more nuanced analysis. While speculative, one possible explanation emerges from the following observation: Aus-IndE listeners have not just simply been exposed to a wider pool of variants for the phonetic categories of the DRESS and TRAP vowels; they have also acquired some knowledge beyond the spatial-acoustic coordinates of segmental targets. Such knowledge includes *structural* properties of D2, i.e., its regulation of intersegmental coordination, knowledge of contextually conditioned allophonic variation (the influence of Australian English coda laterals being dark), speaker-variable DRESS lowering and possible merger before the lateral. The addition of D2 structural knowledge renders their linguistic competence more complex, thereby posing a greater challenge for the mapping of acoustic input onto perceptual targets. It is not simply a case of their hav-

ing a wider pool of acoustic variants in their bank of episodic experience, but that these variants are associated with particular contexts and, furthermore, varyingly so (since this feature in Mainstream Australian English, for this region at least, is itself in flux).

The theory of greater cognitive load could explain what appears to be a paradoxically regressive effect in this context, i.e., more exposure leading to greater uncertainty of mapping. Such an effect is somewhat reminiscent of the U-shaped development often observed in developmental phonology and L1 acquisition more generally (e.g., Bernhardt & Stemberger, 1998), whereby greater exposure brings knowledge of more complexity and thereby increases the challenge. Similarly, exposure to a D2 brings exposure to new variants, which at the simplest level expands the listener's perceptual repertoire; however, it also brings knowledge of multiple possible cross-dialectal mappings between the acoustic input and lexical representations, making the task harder.

Furthermore, what matters is *reference to a specific linguistic system*, which—critically here—also includes the lexicon. The findings with regard to the nonce pair further support this. In this case, where no single cohort has any pre-existing knowledge or experience of either option, responses show no cross-cohort differences. There are no statistically significant differences between cohorts, either in terms of end of continuum response behavior or along the response curve. The nonce items were word-like, and so we can reasonably assume that listeners carried out the task as at least a pseudolinguistic one. However, none of the participant cohorts had any pre-existing pool of experienced variants for these hypothetical lexical items on which to map the acoustic stimuli, and thus all three cohorts could only base their perceptual response on the same process of 'raw' acoustic discrimination. Taken together, these patterns reflect a process of perceptual identification and representation that is deeply embedded in actual produced, and therefore experienced, language. Our findings lend support to models of speech perception that suppose phonetically rich lexical representations, with lexical access mediated through episodic memory of realized productions (e.g., Goldinger, 1996; Hay & Foulkes, 2017). This would suggest that D2 acquisition advances through lexically mediated shifts in perceptual targets, rather than at the level of the segment.

Finally, the conditional inference tree analysis offers initial insight into whether, and if so how, length of exposure to Australian English may shape perceptual responses among Aus-IndE listeners. However, the emerging picture suggests nuanced differences warranting tentative interpretations. Gender emerged as a significant predictor of response for the pair *set-sat* (but not *pet-pat*) and was also mediated by LOR, with female listeners showing divergent response patterns based on their time in Australia. Notably, Aus-IndE females with a shorter LOR (≤ 6.92 years) exhibited more Australian-like perceptual behavior when listening to continuum *set-sat*, with a shift toward TRAP responses at Step 5, while longer-residing females showed more uncertainty (i.e., were less like Aus-IndE listeners). As noted in the preceding section, the demographic profile for the shorter LOR female group is primarily younger students in tertiary education with the majority in their late twenties or early thirties. Their perceptual behavior could quite plausibly have been shaped by their (current or very recent) experiences in tertiary education settings, where they were likely to have increased exposure to and interaction with native Australian English speakers. Importantly, the response curve alone (i.e., Figure 1b) would not be able to capture these sub-group differences, highlighting the value of decision tree models in uncovering socially conditioned variation.

Nevertheless, the relationship between LOR and perceptual behavior is unlikely to be linear or solely driven by exposure. As Diskin-Holdaway et al. (2024) found in their study of Irish English listeners in Australia, shorter LOR was associated with more AusE-like responses for the *hell-Hal* pair, but for *het-hat*, females with LOR of 6+ years had more

DRESS responses. Similarly, in the present study, LOR had no predictive value in the pre-lateral context, where STEP was the only significant factor. Taken together, these findings underscore the complexity of perceptual adaptation in contact settings and suggest that its trajectory is shaped by a combination of factors—including but not limited to LOR and gender—and that other sociolinguistic factors, such as listener linguistic experience with L2/D2 (Flege & Bohn, 2021) and social network (See Lonergan, 2013 and Diskin-Holdaway et al., 2024) need to be considered.

At this point, we would like to draw on some preliminary findings from the production data referred to in Sections 4.2 and 4.3. As part of the production task, real words containing KIT, DRESS AND TRAP vowels in lateral and non-lateral contexts were elicited from the same speakers (Payne & Maxwell, in preparation) included in the present study. Production results provide further evidence for linguistic adaptation to Australian English among the Aus-IndE speaker group while also presenting a nuanced picture. First, there is no effect of phonetic context for the Ind-IndE baseline group, with the F1/F2 vowel spaces of female speakers characterized by a greater separation between DRESS and TRAP in terms of vowel height in both conditions (non-lateral and pre-lateral). In contrast, for the AusE speaker cohort, on the whole, the pre-lateral context leads to the lowering of DRESS towards TRAP (characterized by less separation between those vowels compared to the non-lateral context), i.e., there is clear phonetic conditioning in the pre-lateral context, affecting the acoustic difference in this vowel pair. Further, production results show DRESS lowering in the pre-lateral context for Aus-IndE speakers as well, with the F1 values closer to those in the AusE group than to the Ind-IndE group. However, the overall picture is one of nuanced adjustment by the Aus-IndE speakers, in that while F1 values are seen to have shifted towards the AusE cohort, the F2 values of DRESS indicate a significantly fronter vowel compared to that of the AusE cohort and more similar to the F2 values in the productions of Indian English speakers in India. This suggests that Indian English speakers in Australia are more “attuned” to the salient feature of vowel height when adapting their speech behavior and produce a lower vowel than the Ind-IndE cohort, while maintaining Indian English-like vowel frontness.

Of note, and again echoing the perceptual results for the Aus-IndE group, gender exerts a significant effect on vowel production, albeit in a different way and one which is mitigated by age and not LOR. In production, younger Aus-IndE females show stronger adaptation than males to Australian English in pre-lateral contexts. Vowel lowering (i.e., greater movement towards merger) is also more evident for female speakers in the AusE cohort, whose DRESS–TRAP patterns show significantly tighter F1 clustering in the pre-lateral context than their male counterparts.

7. Conclusions

In conclusion, we have confirmed the findings of earlier studies reporting perceptual adjustment to a new linguistic environment, with evidence for intermediary patterns of behavior between D1 and D2, at least in some contexts. Our findings, however, point to a more complex process of adjustment which involves the acquisition not just of a wider pool of phonetic variants, and thus changes in the perceptual target, but also of structural knowledge, regulating the mapping between acoustic input and the lexical representations. It would appear that acquiring such knowledge results—at least initially and in this group—in an apparent regression in D2 acquisition. Such effects are seen to flatten out across cohorts for nonce-word stimuli, for which listeners have no prior experience, suggesting that D2 acquisition is mediated via shifts in or amplification of phonetically rich lexical representations, and a recalibration of stimulus-to-representation mappings, rather than at the level of the segment. An intriguing question that arises from these findings is

which dialect does the Aus-IndE cohort actually have? Is this D1, D2, or a new sub-variety of one or the other? Alternatively, is it to be characterized as transitional between one variety and another and not yet stable? The findings of the present study suggest a “hybrid” model of linguistic adaptation, which incorporates the features of D1 and D2 but is also characterized by sociolinguistic variation. Further work, including combined and more detailed analyses of the perception and production data, as well as an investigation of the acoustic properties of the coda lateral among the Aus-IndE speaker cohort in comparison to the ‘baseline’ Ind-IndE and AusE speakers, will seek to answer these questions.

Author Contributions: Conceptualization, O.M. and E.P.; methodology, O.M., E.P. and D.L.; formal analysis, O.M., E.P. and M.S.; investigation, O.M. and E.P.; writing—original draft preparation, O.M., E.P. and D.L.; writing—review and editing, O.M., E.P. and D.L.; visualization, M.S. and O.M.; project administration, O.M. and E.P.; funding acquisition, O.M. and E.P. All authors have read and agreed to the published version of the manuscript.

Funding: The project was supported by the ARC Centre of Excellence for the Dynamics of Language: CE140100041 and the Leverhulme Trust, International Academic Fellowship: IAF-2020-013 ‘Indian English on the move: language contact and change in new urban diasporas’.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Faculty of Arts Human Ethics Advisory Group (HEAG), University of Melbourne 2057345.1 on 27 July 2020.

Informed Consent Statement: Participants’ consent was obtained in two ways: 1. Verbal consent, at the start of a scheduled Zoom/WhatsApp meeting. They were also given an opportunity to read the Plain Language Statement (sent beforehand by email) and ask questions; 2. Online consent, a tick box was embedded at the start of the listening task and the online questionnaire, and participants could choose whether they consent or do not consent.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study. Requests to access the datasets should be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AusE	Australian English
IndE	Indian English
Ind-IndE	Indian English listener group in India
Aus-IndE	Indian English listener group in Australia

Appendix A

Table A1. Formant values of input recordings for the listening task stimuli.

Word Pair	Contrast	Pre-Lateral	Vowel 1		Vowel 2	
			F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)
<i>shell–shall</i>	[e]-[æ]	yes	727	2183	959	1869
<i>pell–pal</i>	[e]-[æ]	yes	784	2228	1025	1939
<i>hell–Hal</i>	[e]-[æ]	yes	784	2224	1023	1877
<i>pet–pat</i>	[e]-[æ]	no	761	2411	999	1987
<i>set–sat</i>	[e]-[æ]	no	719	2384	989	1954
<i>skell–skall</i>	[e]-[æ]	yes	772	2216	977	1978

Appendix B

Table A2. Formant values for each step by stimulus pair after resynthesis.

Stimulus Pair	F1	F2
<i>shell–shall</i>		
Step 1	727	2183
Step 2	753	2170
Step 3	775	2222
Step 4	814	2192
Step 5	938	2044
Step 6	957	1945
Step 7	959	1869
<i>hell–Hal</i>		
Step 1	784	2224
Step 2	797	2178
Step 3	856	2106
Step 4	919	2040
Step 5	976	1999
Step 6	1001	1910
Step 7	1023	1877
<i>pell–pal</i>		
Step 1	784	2228
Step 2	808	2177
Step 3	876	2144
Step 4	951	2089
Step 5	1006	2121
Step 6	1020	1994
Step 7	1025	1939
<i>pet–pat</i>		
Step 1	761	2411
Step 2	778	2310
Step 3	820	2204
Step 4	895	2131
Step 5	941	2051
Step 6	964	1998
Step 7	999	1987
<i>set–sat</i>		
Step 1	719	2384
Step 2	768	2330
Step 3	849	2283
Step 4	937	2211
Step 5	955	2149
Step 6	978	2075
Step 7	989	1954

Table A2. *Cont.*

Stimulus Pair	F1	F2
<i>skell–skall</i>		
Step 1	772	2216
Step 2	809	2170
Step 3	839	2116
Step 4	883	2073
Step 5	920	2045
Step 6	932	2026
Step 7	977	1978

Appendix C

Table A3. Model fit parameters for generalized linear mixed-effects models for each continuum (word pair).

Continuum	Final Model	Log Lik	AIC	BIC	LRT (vs. Baseline)	C-Value
<i>pet–pat</i>	<i>response ~ step * group + (1 listener)</i>	−759	1533	1572	$\chi^2(5) = 621, p < 0.001$	0.835
<i>set–sat</i>	<i>response ~ step * group + gender + (1 listener)</i>	−717	1451	14,950	$\chi^2(6) = 880, p < 0.001$	0.872
<i>hell–Hal</i>	<i>response ~ step * group + gender + (1 listener)</i>	−715	1445	1489	$\chi^2(6) = 629, p < 0.001$	0.840
<i>shell–shall</i>	<i>response ~ step * group + gender + (1 listener)</i>	−715	1446	1489	$\chi^2(6) = 780, p < 0.001$	0.843
<i>pell–pal</i>	<i>response ~ step * group + gender + (1 listener)</i>	−680	1376	1420	$\chi^2(6) = 924, p < 0.001$	0.884
<i>skell–skall</i>	<i>response ~ step + group + (1 listener)</i>	−916	1842	1869	$\chi^2(3) = 710, p < 0.001$	0.810

Table A4. Coefficients and their significance tests for the final model for *pet–pat*.

Predictor	β	SE	OR	95% CI	z	p-Value
(Intercept)	4.44	0.47	84.43	[33.41, 213.34]	9.379	<0.001
Step	−1.09	0.10	0.34	[0.28, 0.41]	−10.985	<0.001
Cohort (Aus-IndE vs. AusE)	0.07	0.61	1.08	[0.32, 3.57]	0.121	0.904
Cohort (Ind-IndE vs. AusE)	1.41	0.65	4.10	[1.16, 14.53]	2.184	0.029
Step × Cohort (Aus-IndE)	0.31	0.12	1.36	[1.08, 1.72]	2.614	0.009
Step × Cohort (Ind-IndE)	0.22	0.12	1.25	[0.98, 1.59]	1.810	0.070

Table A5. Coefficients and their significance tests for the final model for *set–sat*.

Predictor	β	SE	OR	95% CI	z	p-Value
(Intercept)	6.09	0.59	443.61	[139.24, 1413.28]	10.310	<0.001
Step	−1.37	0.12	0.25	[0.20, 0.32]	−11.424	<0.001
Cohort (Aus-IndE vs. AusE)	−1.91	0.65	0.15	[0.04, 0.53]	−2.921	0.003
Cohort (Ind-IndE vs. AusE)	0.11	0.69	1.12	[0.29, 4.36]	0.166	0.868
Gender	−0.50	0.27	0.60	[0.36, 1.02]	−1.889	0.059
Step × Cohort (Aus-IndE)	0.62	0.13	1.85	[1.43, 2.40]	4.678	<0.001
Step × Cohort (Ind-IndE)	0.33	0.14	1.39	[1.06, 1.82]	2.395	0.017

Table A6. Coefficients and their significance tests for the final model for *hell–Hal*.

Predictor	β	SE	OR	95% CI	z	p-Value
(Intercept)	5.20	0.54	180.93	[62.18, 526.42]	9.540	<0.001
Step	−1.07	0.10	0.34	[0.28, 0.42]	−10.762	<0.001
Cohort (Aus-IndE vs. AusE)	−1.15	0.65	0.32	[0.09, 1.13]	−1.771	0.077
Cohort (Ind-IndE)	0.39	0.66	1.48	[0.40, 5.44]	0.595	0.552
Gender	−0.15	0.25	0.86	[0.52, 1.41]	−0.593	0.553
Step × Cohort (Aus-IndE)	0.43	0.12	1.54	[1.22, 1.94]	3.636	<0.001
Step × Cohort (Ind-IndE)	0.03	0.12	1.03	[0.81, 1.32]	0.277	0.782

Table A7. Coefficients and their significance tests for the final model for *shell–shall*.

Predictor	β	SE	OR	95% CI	z	p-Value
(Intercept)	5.34	0.61	208.47	[63.62, 683.15]	8.818	<0.001
Step	−1.37	0.12	0.26	[0.20, 0.33]	−10.999	<0.001
Cohort (Aus-IndE vs. AusE)	−2.45	0.68	0.09	[0.02, 0.33]	−3.597	<0.001
Cohort (Ind-IndE vs. AusE)	−1.90	0.67	0.15	[0.04, 0.55]	−2.841	0.004
Gender	0.18	0.31	1.20	[0.65, 2.19]	0.580	0.562
Step × Cohort (Aus-IndE)	0.53	0.14	1.70	[1.30, 2.24]	3.808	<0.001
Step × Cohort (Ind-IndE)	0.32	0.14	1.38	[1.05, 1.82]	2.292	0.022

Table A8. Coefficients and their significance tests for the final model for *pell–pal*.

Predictor	β	SE	OR	95% CI	z	p-Value
(Intercept)	6.31	0.40	552.17	[252.25, 1208.70]	15.796	<0.001
Step	−1.24	0.06	0.29	[0.26, 0.32]	−21.326	<0.001
Cohort (Aus-IndE vs. AusE)	−0.74	0.37	0.48	[0.23, 0.98]	−2.019	0.043
Cohort (Ind-IndE vs. AusE)	−0.54	0.35	0.58	[0.29, 1.15]	−1.558	0.119

Table A9. Coefficients and their significance tests for the final model for *skell–skall*.

Predictor	β	SE	OR	95% CI	z	p-Value
(Intercept)	3.69	0.31	39.96	[21.93, 72.82]	12.046	<0.001
Step	−0.89	0.04	0.41	[0.38, 0.45]	−20.525	<0.001
Cohort (Aus-IndE vs. AusE)	−0.21	0.33	0.81	[0.43, 1.55]	−0.637	0.524
Cohort (Ind-IndE vs. AusE)	−0.28	0.31	0.76	[0.41, 1.40]	−0.894	0.371

Notes

- Throughout the paper, we predominantly rely on Wells’s (1982) lexical sets. In the instances of describing Mainstream Australian English, we adopt the transcription system for vowel phonemes by Harrington et al. (1997), where the DRESS vowel is represented as /e/ and the TRAP vowel as /æ/. The HCE system is now a widely accepted representation of the average pronunciation by speakers of Mainstream Australian English (Cox et al., 2024).
- In this paper, we present a subset of the results as part of a larger project. The remainder of the target words (14 pairs), together with the speech recordings elicited from the same participants, are part of the ongoing analyses.
- Loakes et al. (2024a) was also modeled on the method in Harrington et al. (2008), Kleber et al. (2011), and Kendall and Fridland (2012).

References

ABS. (2021). *People in Australia who were born in India. 2021 Census country of birth QuickStats*. Australian Bureau of Statistics. Available online: https://www.abs.gov.au/census/find-census-data/quickstats/2021/7103_AUS (accessed on 3 June 2025).

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

- Bernhardt, B. M., & Stemberger, J. (1998). *Handbook of phonological development: From a nonlinear constraints-based perspective*. Academic Press.
- Billington, R. (2011). Location, location, location! Regional characteristics and national patterns of change in the vowels of Melbourne adolescents. *Australian Journal of Linguistics*, 31(3), 275–303. [CrossRef]
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9–10), 341–345.
- Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, 8, e9414. [CrossRef]
- Clopper, C. (2014). Sound change in the individual: Effects of exposure on cross-dialect speech processing. *Laboratory Phonology*, 5(1), 69–90. [CrossRef]
- Clopper, C. (2021). Perception of dialect variation. In J. Pardo, L. Nygaard, R. Remez, & D. Pisoni (Eds.), *The handbook of speech perception* (pp. 333–364). Wiley-Blackwell.
- Clopper, C., & Walker, A. (2017). Effects of lexical competition and dialect exposure on phonological priming. *Language and Speech*, 60(1), 85–109. [CrossRef]
- Coats, S., Diskin-Holdaway, C., & Debbie, L. (2025). *Regional distribution of the /eɪ/-/æɪ/ Merger in Australian English*. Proceedings of the 12th workshop on NLP for similar languages, varieties and dialects, Abu Dhabi, United Arab Emirates, January 19 (pp. 147–156).
- Cox, F. (1999). Vowel change in Australian English. *Phonetica*, 56, 1–27. [CrossRef]
- Cox, F., & Fletcher, J. (2017). *Australian English pronunciation and transcription* (2nd ed.). Cambridge University Press.
- Cox, F., & Palethorpe, S. (2004). The border effect: Vowel differences across the NSW–Victorian border. In C. Marovsky (Ed.), *Proceedings of the 2003 conference of the Australian Linguistic Society* (pp. 1–27). School of Language and Media, University of Newcastle.
- Cox, F., & Palethorpe, S. (2008). Reversal of short front vowel raising in Australian English. In J. Fletcher, D. Loakes, R. Göcke, D. Burnham, & M. Wagner (Eds.), *Proceedings of Interspeech 2008 Incorporating SST 2008, Brisbane, Australia, September 22–26* (p. 34245). ISCA.
- Cox, F., & Palethorpe, S. (2019). Vowel variation in a standard context across four major Australian cities. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th international congress of phonetic sciences, Melbourne, Australia, August 5–9* (p. 577–581). Australasian Speech Science and Technology Association Inc., and International Phonetic Association.
- Cox, F., Penney, J., & Palethorpe, S. (2024). Australian English monophthong change across 50 years: Static versus dynamic measures. *Languages*, 9(3), 99. [CrossRef]
- Diskin-Holdaway, C., Loakes, D., & Clothier, J. (2024). Categorisation of short front lax vowels by Irish and Chinese migrants in Melbourne: Variability in cross-language and cross-dialect processing in a dialect-familiar context. *Phonetica*, 81(1), 1–41. [CrossRef]
- Domange, R. (2020). Variation and change in the short vowels of Delhi English. *Language Variation and Change*, 32(1), 49–76. [CrossRef]
- Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2007). The perception of phonemic contrasts in a non-native dialect. *JASA Express Letters*, 121, EL131–EL136. [CrossRef]
- Elvin, J., Williams, D., & Escudero, P. (2016). Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *Journal of the Acoustical Society of America*, 140(1), 576–581. [CrossRef]
- Evans, B. G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America*, 115(1), 352–361. [CrossRef]
- Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, 121(6), 3814–3826. [CrossRef] [PubMed]
- Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage. Available online: <https://www.johnfox.ca/Companion/> (accessed on 15 April 2025).
- Gargesh, R. (2006). South Asian Englishes. In B. B. Kachru, Y. Kachru, & C. L. Nelson (Eds.), *The handbook of world Englishes*. Wiley. [CrossRef]
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166–1183. [CrossRef] [PubMed]
- Grama, J., Travis, C., & González, S. (2019). Initiation, progression and conditioning of the short-front vowel shift in Australian English. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th international congress of phonetic sciences, Melbourne, Australia, August 5–9* (pp. 1769–1773). Australasian Speech Science and Technology Association Inc., and International Phonetic Association.
- Harrington, J., Cox, F., & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics*, 17(2), 155–184. [CrossRef]
- Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America*, 123(5), 2825–2835. [CrossRef]
- Hay, J., & Foulkes, P. (2017). The evolution of medial/t/over real and remembered time. *Language*, 92(2), 298–330. [CrossRef]

- Hickey, R. (Ed.). (2004). *Dialects of English and their transportation* (pp. 33–58). Cambridge University Press.
- Hosmer, D. W., & Lemeshow, S. (2000). Introduction to the logistic regression model. In W. A. Shewhart, S. S. Wilks, D. W. Hosmer, & S. Lemeshow (Eds.), *Applied logistic regression* (pp. 1–30). John Wiley & Sons, Inc. [CrossRef]
- Kachru, B. B. (1983). Models for new Englishes. In J. Cobarrubias, & J. A. Fishman (Eds.), *Progress in language planning: International perspectives* (pp. 145–170). De Gruyter Mouton.
- Kawahara, H. (2006). STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. *Acoustical Science and Technology*, 27(6), 349–353. [CrossRef]
- Kendall, T., & Fridland, V. (2012). Variation in perception and production of mid-front vowels in the U.S. southern vowel shift. *Journal of Phonetics*, 40(2), 289–306. [CrossRef]
- Kleber, F., Harrington, J., & Reubold, U. (2011). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, 55(3), 383–405. [CrossRef] [PubMed]
- Kunkel, S., Passoni, E., & de Leeuw, E. (2023). Perceptual discrimination of phonemic contrasts in Quebec French: Exposure to Quebec French does not improve perception in hexagonal French native speakers living in Quebec. *Languages*, 8(3), 193. [CrossRef]
- Ladefoged, P., & Johnson, K. (2015). *A course in phonetics* (7th ed.). Wadsworth.
- Lenth, R. (2023). *emmeans: Estimated marginal means, aka least-squares means* (R package version 1.8.5). Available online: <https://CRAN.R-project.org/package=emmeans> (accessed on 17 November 2022).
- Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins.
- Levshina, N. (2021). Conditional inference trees and random forests. In M. Paquot, & T. Gries (Eds.), *Practical handbook of corpus linguistics* (pp. 611–643). Springer.
- Loakes, D., Clothier, J., Hajek, J., & Fletcher, J. (2014). An investigation of the /e/–/æ/ merger in Australian English: A pilot study on production and perception in South-West Victoria. *Australian Journal of Linguistics*, 34(4), 436–452. [CrossRef]
- Loakes, D., Clothier, J., Hajek, J., & Fletcher, J. (2024a). Sociophonetic variation in vowel categorization of Australian English. *Language and Speech*, 67(3), 870–906. [CrossRef]
- Loakes, D., Escudero, P., Clothier, J., & Hajek, J. (2019). Tracking vowel categorisation behaviour longitudinally: A study across three x three-year increments (2012, 2015, 2018). In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International congress of phonetic sciences, Melbourne, Australia, August 5–9* (pp. 2787–2891). Australasian Speech Science and Technology Association Inc., and International Phonetic Association.
- Loakes, D., Fletcher, J., & Clothier, J. (2024b). One place, two speech communities: Differing responses to sound change in Mainstream and Aboriginal Australian English in a small rural town. In F. Kleber, & T. Rathcke (Eds.), *Speech dynamics: Synchronic variation and diachronic change* (pp. 117–144). Chapter 4. De Gruyter Mouton.
- Loakes, D., & Gregory, A. (2024). Acoustic analysis of vowels in Australian Aboriginal English spoken in Victoria. *Languages*, 9(9), 299. [CrossRef]
- Loakes, D., Hajek, J., & Fletcher, J. (2017). Can you t[æ]ll I'm from M[æ]lbourne? An overview of the DRESS and TRAP vowels before /L/ as a regional accent marker in Australian English. *English World-Wide*, 38(1), 29–49. [CrossRef]
- Lonergan, J. (2013). *An acoustic and perceptual study of Dublin English phonology* [Unpublished Ph.D. thesis]. University College Dublin.
- Mannell, R. (2004). Perceptual vowel space for Australian English lax vowels: 1998 and 2004. In S. Cassidy, F. Cox, R. Mannell, & S. Palethorpe (Eds.), *Proceedings of 10th Australian international conference on speech science and technology* (pp. 221–226). ASSTA.
- Mauk, M. D., & Buonomano, D. V. (2004). The neural basis of temporal processing. *Annual Review of Neuroscience*, 27, 307–340. [CrossRef]
- Maxwell, O., Diskin-Holdaway, C., & Loakes, D. (2023). Attitudes towards Indian English among young urban professionals in Hyderabad, India. *World Englishes*, 42, 272–291. [CrossRef]
- Maxwell, O., & Fletcher, J. (2009). Acoustic and durational properties of Indian English vowels. *World Englishes*, 28(1), 52–70. [CrossRef]
- Maxwell, O., & Payne, E. (2018). Pitch accent types and tonal alignment of the accentual rise in Indian English(es). In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of speech prosody 2018, Poznan, Poland, June 13–16* (pp. 942–946). ISCA.
- Maxwell, O., & Payne, E. (2023). Investigating (rhythm) variation in Indian English: An integrated approach. In R. Fuchs (Ed.), *Speech rhythm in learner and second language varieties of English* (pp. 17–57). Springer.
- Maxwell, O., Payne, E., & Billington, R. (2018). Homogeneity vs. heterogeneity in Indian English: Investigating influences of L1 on f0 range. *Interspeech*, 2018, 2191–2195. [CrossRef]
- Munro, M. J., Derwing, T., & Flege, J. (1999). Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics*, 27(4), 385–403. [CrossRef]
- Nycz, J. (2013). Changing words or changing rules? Second dialect acquisition and phonological representation. *Journal of Pragmatics*, 52, 49–62. [CrossRef]

- Nycz, J. (2018). Stylistic variation among mobile speakers: Using old and new regional variables to construct complex place identity. *Language Variation and Change*, 30(2), 175–202. [CrossRef]
- Payne, E., & Maxwell, O. (2018). Durational variability as a marker of prosodic structure in Indian English(es). In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of speech prosody 2018, Poznan, Poland, June 13–16* (pp. 749–753). ISCA.
- Payne, E., Maxwell, O., Fuchs, R., & Wang, Y. Payne, E., Maxwell, O., Fuchs, R., & Wang, Y. (2023). Lexical stress perception in Indian Englishes. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th international congress of phonetic sciences* (pp. 2890–2894). Guarant International.
- Payne, E., Maxwell, O., & Volchok, B. (2019). Tense lax contrasts in Indian English vowels: Transfer effects from L1 Telugu at the phonetics phonology interface. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia, August 5–9* (pp. 1079–1083). Australasian Speech Science and Technology Association Inc., and International Phonetic Association.
- Pearce, J. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2, 10. [CrossRef]
- R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.0). R Foundation for Statistical Computing.
- Satyanath, S., & Sharma, R. (2016). The growth of English in Delhi: New perspectives in a multilingual setting. In J. N. Singh, A. Kantara, & D. Cserző (Eds.), *Downscaling culture: Revisiting intercultural communication* (pp. 192–227). Cambridge Scholars.
- Schmidt, P., Diskin-Holdaway, C., & Loakes, D. (2021). New insights into/el-/æ/merging in Australian English. *Australian Journal of Linguistics*, 41(1), 66–95. [CrossRef]
- Sethi, J. (1980). Word accent in educated Punjabi speakers' English. *Bulletin of the Central Institute of English*, 16(2), 35–48.
- Shaktawat, D. (2024). The effect of Indian contact and Glaswegian contact on the phonetic backward transfer of Glaswegian English (L2) on Hindi and Indian English (L1). *Languages*, 9(4), 118. [CrossRef]
- Sharma, D. (2005). Dialect stabilization and speaker awareness in non-native varieties of English. *Journal of Sociolinguistics*, 9(2), 194–224. [CrossRef]
- Siegel, J. (2010). *Second dialect acquisition*. Cambridge University Press.
- Sirsa, H., & Redford, M. (2013). The effects of native language on Indian English sounds and timing patterns. *Journal of Phonetics*, 41, 393–406. [CrossRef]
- Tagliamonte, S., & Baayen, H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135–178. [CrossRef]
- Trudgill, P., & Hannah, J. (2008). *International English: A guide to the varieties of standard English* (5th ed.). Hodder Education.
- Wells, J. (1982). *Accents of English*. Cambridge University Press.
- Wiltshire, C. R. (2005). The 'Indian English' of Tibeto-Burman language speakers. *English World-Wide: A Journal of Varieties of English*, 26(3), 275–300. [CrossRef]
- Wiltshire, C. R. (2020). *Uniformity and variability in the Indian English accent*. Cambridge University Press.
- Ziliak, Z. (2012). *The relationship between perception and production in adult acquisition of a new dialect's phonetic system* [Doctoral dissertation, University of Florida].

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.