Regional Distribution of the /el/-/æl/ Merger in Australian English

Steven Coats[1], **Chloé Diskin-Holdaway**[2], **Debbie Loakes**[2]

[1]English, Faculty of Humanities, University of Oulu, Finland
[2]School of Languages and Linguistics, The University of Melbourne, Australia

Correspondence: steven.coats@oulu.fi

Abstract

Prelateral merger of /e/ and /æ/ (where words like *celery* and *salary* are both pronounced with [æ] in the first syllable) is a salient acoustic feature of speech from Melbourne and the state of Victoria in Australia, but little is known about its presence in other parts of the country. In this study, automated methods of data collection, forced alignment, and formant extraction are used to analyze the regional distribution of the vowel merger within all of Australia, in 4.3 million vowel tokens from naturalistic speech in 252 locations. The extent of the merger is quantified using the difference in Bhattacharyya's distance scores based on phonetic context, and the regional distribution is assessed using spatial autocorrelation. The principal findings are that the merger is most prominent in Victoria, especially southern Victoria, and least prominent in Sydney and New South Wales. We also find preliminary indications that it may be present in other parts of the country.

1 Introduction

The past 20 years have seen an increased interest in the analysis of regional phonetic variation in Australian English. Prelateral merger of /e/ and /æ/ (where words like *celery* and *salary* are both pronounced with [æ] in the first syllable) is a salient feature of the speech of southern Victoria (VIC), particularly in the city of Melbourne, and has been researched in a number of studies (see e.g. Schmidt et al., 2021; Loakes et al., 2017), including some more recent work on perception of the merger (Diskin-Holdaway et al., 2024; Loakes et al., 2024a,b). In locations where it does occur, it is reported to be completely entrenched for some speakers, but still in progress or almost absent for others (Diskin et al., 2019a; Loakes et al., 2024b). This vowel merger is important because (1) it is one of the few documented features that appears to distinguish the accent of (southern) Victorians from the accent of speakers from other states; and (2), due to its absence among certain speakers, including in VIC, it is unclear whether this represents a true sound change. Most empirical studies of the phenomenon have utilized relatively small datasets of word list recordings from few locations, and little is known about the presence of the merger in other parts of the country.

In recent years, the rise of automated methods of acoustic analysis and the availability of vast amounts of naturalistic speech data have opened up new opportunities for (socio-)phonetic research. Automated formant analysis of naturalistic speech (e.g., Brand et al., 2021; Coto-Solano et al., 2021; Renwick and Stanley, 2020) has been made possible by tools for vowel and formant extraction (e.g., Reddy and Stanford, 2015; Rosenfelder et al., 2015), which are increasingly incorporated into data extraction and processing pipelines (e.g., Coats, 2023; Méli et al., 2023), allowing researchers to work with large samples of real-world, "ecologically valid" speech. Although word list data offers a valuable point of comparison, formant values derived from these contexts may not fully align with those obtained from more natural speech, which, although it exhibits variability due to phonological, lexical, and syntactic influences, as well as various situational and social factors, is generally more representative of everyday communication than is data collected in controlled settings (Liberman, 2019), in addition to providing results that can be more statistically robust and generalizable.

This paper demonstrates the feasibility of working with a recent large naturalistic speech dataset from Australia (Coats, 2024a,b) to investigate prelateral merger of /e/ and /æ/. In addition, the study provides an overview of the phenomenon as it occurs across the whole country, providing further evidence for regional phonetic variation for Australian English, a variety that although long considered regionally homogeneous, has "begun to

exhibit more widespread social and regional variation than has previously been acknowledged" (Cox and Fletcher, 2017, p. 20).

2 Previous Work

Realization of /el/ as [æl] in Melbourne/Victoria was first observed at the end of the 1980s (see, e.g., Bradley, 2008), yet phonetic research was not carried out until Cox and Palethorpe (2004) recorded teenage girls in three towns in New South Wales (NSW) and in Wangaratta, VIC. That study found that /e/ before /l/ was lowered and retracted among the VIC as compared to the NSW speakers, and effectively realized as [æ]. However, no such phenomenon was observed among either of the groups when /e/ occurred before the consonant /d/: in those cases, it was pronounced as /e/, suggesting that the merger was exclusively in prelateral contexts. Since then, studies into production and perception have shown a high degree of variability in speaker-listener behavior, with research showing a complete merger of /el/-/æl/ for certain speakers in Melbourne/VIC, while others exhibit a broader range of phonetic behavior. For example, the merger seems to be more common for middle-aged and older speakers (Diskin et al., 2019b; Schmidt et al., 2021), but this is also dependent on the community (Loakes et al., 2024b). Additionally, older speakers might "hypocorrect" and produce /æl/ as [el] (Loakes et al., 2011; Schmidt et al., 2021).

In a previous study of variable merger behavior, Diskin et al. (2019b) analyzed the speech of 12 Melbourne speakers in their thirties reading words containing the short front vowels /ɪ, e, æ/. Prelateral merger behavior of /e/ and /æ/ was found for 9 speakers, but there were individual differences in both their acoustics and their articulation, which was also measured via ultrasound tongue imaging. Diskin et al. (2019a) extended the dataset from (Diskin et al., 2019b) to compare wordlist and naturalistic speech, and again found individual variation, where some speakers had the merger only in the wordlist, but not in their naturalistic speech, whereas for other speakers, it was only in their naturalistic speech and not in their wordlist. Schmidt et al. (2021) examined 628 reading list tokens from 13 older speakers aged 51-80 from Ocean Grove, VIC. They found no merger of /e/ and /æ/ before the /d/ consonant, but significant merger in prelateral pairs such as *palate* and *pellet*. In one of the few studies outside of VIC, and the only known study in Queensland (QLD), Gregory (2019) found evidence for the merger for some speakers in a study of word list recordings of 17 speakers from Northern QLD.

Perception studies (investigating whether people hear the /el/ and /æl/ as the same or different) have shown further support for the merger in Melbourne/Victoria, especially in the state's southernmost locations, and with lexical frequency playing a small but crucial role in some of the differences between older and younger listeners. For example, younger listeners are biased toward hearing the first name *Mel* when presented with a choice between *Mel* and *Mal* because of an increase in popularity (and thus frequency) of the name *Mel* over time (Loakes et al., 2024a,b).

Based on the prior research, which has primarily centered on VIC speakers and word list data, we propose two research questions which guide our paper:

1. Is the merger of /el/-/æl present across all states of Australia, or only in VIC?

2. How does the merger pattern in a large-scale corpus of naturalistic speech, compared to the small samples of controlled word list data that has dominated previous research?

3 Data and Methods

3.1 Vowel Extraction

The starting point for the project was data from CoANZSE Audio comprising short excerpts of transcripts and audio content from 38,786 videos uploaded to YouTube channels of Australian councils (for details, see Coats 2024b). For each of the 404 Australian CoANZSE locations, 20-word audio segments were aligned with the corresponding Automatic Speech Recognition (ASR) textual content, using the Montreal Forced Aligner (McAuliffe et al. 2017) and its default English acoustic model and dictionary (v3.0.0). This model was trained using audio data from ten datasets, including the Common Voice English v8.0 dataset, which contains 50,285 sentences spoken by Australian speakers (Ardila et al., 2020). Additionally, the `adapt` functionality of the Montreal Forced Aligner, which tunes the acoustic model based on the Gaussian Mixture Model means of the data to be aligned, was employed.

Formant values for /e/ and /æ/, based upon the pronunciations of the Montreal Forced Aligner English Dictionary v.3.0.0, were then extracted at the

midpoints of the targeted vowel segments using Parselmouth-Praat (Jadoul et al., 2018), a Python interface for Praat (Boersma and Weenink, 2024), with an automatic time step based on the duration of the sound file, five formants, and a maximum formant frequency of 5,500 Hz. A window length of 0.025 seconds and pre-emphasis above 50 Hz were applied, and the F1 and F2 values, along with their bandwidths, were retrieved.

Forced alignment and vowel extraction returned 9,264,705 vowel tokens (3,826,298 for /e/ and 5,438,407 for /æ/), which were then filtered and labeled for context as prelateral or non-prelateral. A process of filtering removed tokens in unstressed syllables, as determined by the CMU pronunciation dictionary (Weide et al., 1998); common English stopwords were excluded using a list from NLTK (Bird et al., 2009). Phonetic context was determined by the phone labels from the CMU dictionary. Locations with at least 20 tokens in each context (/æl/, /æC/, /el/, and /eC/, where C represents any non-lateral consonant) were retained for analysis. After filtering, 4,297,259 vowel tokens from 252 locations remained for the ensuing analysis.

Table 1 shows the number of tokens for each state/territory-level location and each context.

Loc.	Context	count	Loc.	Context	count
ACT	/æl/	548	SA	/æl/	10,456
	/æC/	11,308		/æC/	240,279
	/el/	1,232		/el/	22,726
	/eC/	11,917		/eC/	269,945
NSW	/æl/	20,105	TAS	/æl/	4,178
	/æC/	465,825		/æC/	89,067
	/el/	46,508		/el/	8,815
	/eC/	531,894		/eC/	94,512
NT	/æl/	85	VIC	/æl/	29,097
	/æC/	1,346		/æC/	625,318
	/el/	163		/el/	69,308
	/eC/	1,590		/eC/	683,640
QLD	/æl/	13,875	WA	/æl/	5,233
	/æC/	332,394		/æC/	133,116
	/el/	28,041		/el/	14,016
	/eC/	375,405		/eC/	155,317

Table 1: Vowel and dataset counts across Australian states and territories (ACT=Australian Capital Territory; NSW=New South Wales; NT=Northern Territory; QLD=Queensland; SA=South Australia; TAS=Tasmania; VIC=Victoria; WA=Western Australia). /C/ stands for any consonant other than /l/.

For plotting formant values (Fig. 3), we used a z-scaled version of Nearey's transformation, a speaker-extrinsic method, applied to each formant and each vowel token. The Nearey transformation for a formant F is given by:

$$F_{\text{nearey}} = \log(F) - \log(\text{central frequency})$$

where central frequency is the geometric mean of the formant values across all tokens.

$$\text{central frequency} = \exp\left(\frac{1}{N}\sum_{i=1}^{N}\log(F)\right)$$

F_{neary} scores were then converted to a z-score.

3.2 Vowel Overlap Measure

The Bhattacharyya coefficient BC between two probability distributions P and Q is defined as

$$BC = \left(\int \sqrt{P(x)\cdot Q(x)}\,dx\right)$$

To quantify the extent of vowel overlap, we used Bhattacharyya distance, which is the negative logarithm of the Bhattacharyya coefficient (Bhattacharyya, 1943), a measure which has been proposed as an alternative to Pillai's trace metric (Pillai, 1955), and has been employed in previous work in phonetics (Warren, 2018). Like Pillai's trace, Bhattacharyya's distance can be employed to characterize the overlap of two distributions of F1 and F2 values. However, while the MANOVA model that generates Pillai's trace assumes multivariate normality (Johnson, 2015), Bhattacharyya distance can be applied to non-normally distributed data and is generally more robust to differences in sample size (see Stanley and Sneller, 2023). This makes Bhattacharyya distance a versatile choice for comparing vowel distributions under varying sample conditions, especially when additional covariates in a MANOVA analysis are not required, as is the case in the present study.

Bhattacharyya's distance was calculated for /æ/ and /e/, using all the tokens recorded in each location, for both prelateral and for non-prelateral contexts. Like Pillai's trace, a value of zero indicates complete overlap for two distributions (in this study, complete merger of /æ/ and /e/), while larger values indicate the underlying vowels are more distinct in F1/F2 space.

After confirming that the Bhattacharyya distance for prelateral and non-prelateral contexts was significantly different (mean Bhattacharyya before /l/ = 0.173, mean Bhattacharyya in other contexts = 0.431, $t = -31.037$, $p < 0.001$), for each location in the dataset, we subtracted the Bhattacharyya

distance value for the prelateral context from the value for the non-prelateral context. Lower Bhattacharyya distances for the /el/-/æl/ context indicates greater overlap or merger. Higher Bhattacharyya distances for the /eC/-/æC/ context indicate less overlap. This **Bhattacharyya difference** measure thus characterizes the extent to which the prelateral context results in different realizations of these vowels, compared to non-prelateral contexts. Positive difference values indicate that the vowels are more merged in prelateral context than in non-prelateral context.

3.3 Spatial Analysis

Spatial autocorrelation, a method proven to be effective for analyzing language data, including vowel formants (Grieve et al., 2011, 2013), was applied in this study. Two spatial autocorrelation metrics were used: **Moran's I**, which assesses all locations in a dataset and provides a summary measure of the overall spatial correlation (Moran, 1950), and the **Getis-Ord local** G_i^* statistic (Getis and Ord, 1992; Ord and Getis, 1995; Getis, 2010), which identifies spatial clusters by comparing the values at each location to those of its neighboring locations in the context of the entire dataset.

Both statistics rely on a **spatial weights matrix** W, which quantifies the influence of nearby measurements on a given location's values. Neighbors can be assigned binary weights based on a distance threshold, or weights can be calculated as a function of distance or other criteria. In this study, an inverse distance spatial weights matrix was used: for locations within a specified minimum threshold distance, the weight for location j relative to location i was defined as $w_{ij} = \frac{1}{d_{ij}}$. The spatial autocorrelation analysis was conducted using PySAL (Rey and Anselin, 2010).

Moran's I takes values between -1 and 1, where positive values indicate clustering of similar values, negative values suggest even dispersion, and a value of zero signifies a random distribution. The G_i^* statistic is computed for each location in the dataset and does not have a fixed range. A positive G_i^* value means the sum of the values at a specific location and its neighbors is greater than what would be expected based on the global distribution, while a negative G_i^* suggests the sum is lower than expected.

The significance of Moran's I can be computed using a normal approximation of the distribution of the statistic under the null hypothesis of no spatial autocorrelation, or, for values that are not normally distributed, with randomized permutations. G_i^* significance is mostly calculated using a z-score. For detecting clusters of high values, $z \geq 1.645$ is significant at $p = 0.05$. To detect both high and low clusters, a two-tailed test with $|z| \geq 1.96$ is used at $p = 0.05$. Essentially, G_i^* can be viewed as a localized indicator of spatial clustering, aggregating local values and comparing them to a global average.

4 Results

Overall, the Bhattacharyya difference at the 252 locations had a mean value of 0.258, with a standard deviation of 0.132; the range of values was -0.469 to 0.529. The distribution of values is depicted in Fig. 1.

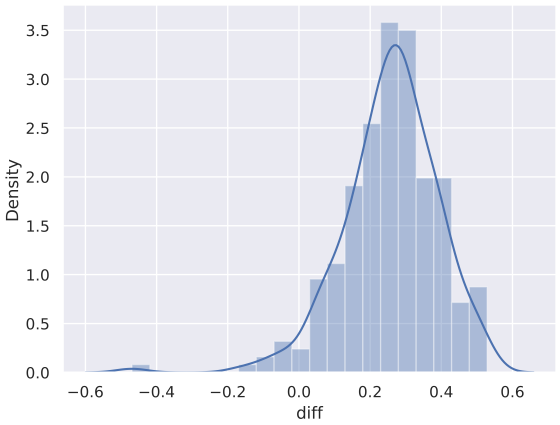

Figure 1: Distribution of Bhattacharyya difference values for 252 locations

Difference is highest for VIC, followed by WA, the ACT, QLD, TAS, SA, NSW, and the NT (Fig. 2).

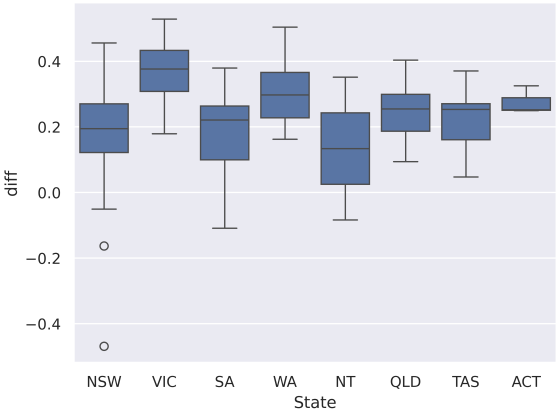

Figure 2: Bhattacharyya difference by state/territory

Victorian speakers tend to have substantially lowered (more [æ]-like) vowels for /e/ in prelateral position. Fig. 3 shows a subset of the data: eight of the most frequent words of 5 characters or fewer, with the targeted prelateral and non-prelateral contexts, in the NSW and VIC subcorpora. For each word, the location corresponds to the Nearey-transformed and z-scaled mean formant values for the targeted phone, and the subscript indicates the number of extracted tokens of that word. As can be seen, for prelateral contexts (on the left-hand side), VIC speakers have substantially lower /e/ values than do NSW speakers in the words *dealt, held, sell, else, tell, help*, and *well*. For *value*, on the other hand, /æ/ is much lower for NSW-located tokens, suggesting that it remains distinct from /e/ in prelateral contexts for these speakers. For non-prelateral contexts (on the right-hand side), mean Nearey-z values for frequent words are quite close for VIC and NSW tokens, and no clear regional tendency prevails.

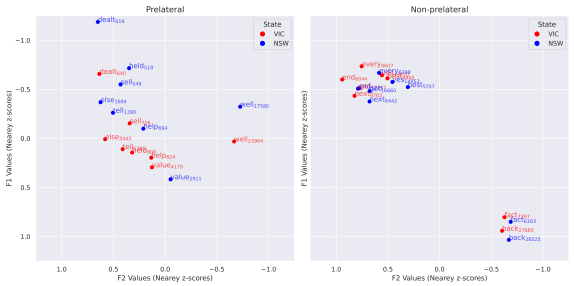

Figure 3: Mean locations of most frequent words, prelateral context (left) and non-prelateral context(right), Nearey-z-score-transformed F1/F2 values

To investigate the possibility that the merger is affected by word frequency, we correlated frequency with F1, with F2, and with the Euclidean F1/F2 distance. This was done for the 9,838 word types in the dataset (4,297,259 word tokens) as well as for all combinations of vowel and context. No correlations resulted in an $r \geq |0.07|$ or a significant p-value.

4.1 Regional Distribution

The largest Bhattcharyya difference values were found for four councils in the Melbourne metropolitan area: Maroondah City Council, City of Stonnington, City of Whittlesea, and Glen Eira City Council, ranging from 0.506 to 0.529. The lowest difference value, -0.469, was found for Narrabri Shire in northern NSW. This value is an outlier, as the locations with the lowest values otherwise had difference scores in the range of 0 to -0.16.

Large difference values were also found for data from councils in WA, including Armadale, Kalamunda, Kwinana, and Joondalup, in the Perth area, which registered difference values ranging from 0.394 to 0.504. In QLD, the highest values were found for Redland City, in the Brisbane area (0.404), Balonne Shire (0.396), Cairns (0.385), and Banana Shire (0.383). Values for SA were mixed, with relatively high difference values found for a few councils in the Adelaide area (0.379 for West Torrens and 0.315 for Charles Sturt), but low values for others (-0.109 for Yorke Peninsula Council and 0.055 for Mount Barker District Council). Tasmanian difference values were also mixed, ranging from 0.047 for Circular Head to 0.371 for the city of Launceston. In the ACT, the three sampled councils showed middling difference values from 0.249 to 0.325.

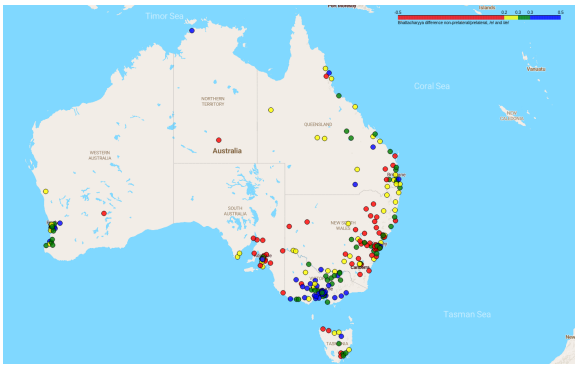

Figure 4: Bhattacharyya Difference values

Fig. 4 depicts the raw Bhattacharyya difference values for the 252 sampled locations, with colors indicating quantiles.[1]

Moran's I, calculated on the basis of the Bhattacharyya difference for all locations, was found to have a value of 0.235 for this dataset. Due to the non-normality of the data, a p-value was calculated using 999 random permutations of the underlying difference values, resulting in a p-value of 0.001. Thus, the difference in Bhattacharyya distance values for non-prelateral and prelateral contexts for the vowels /e/ and /æ/ in this dataset can be considered to be moderately clustered.

[1]The images in Fig. 4 and Fig. 5 are screenshots of interactive maps that can be found at https://stcoats.github.io/AU_Bhatt_map.html and https://stcoats.github.io/AU_Bhatt_Gi_map_v2.html.

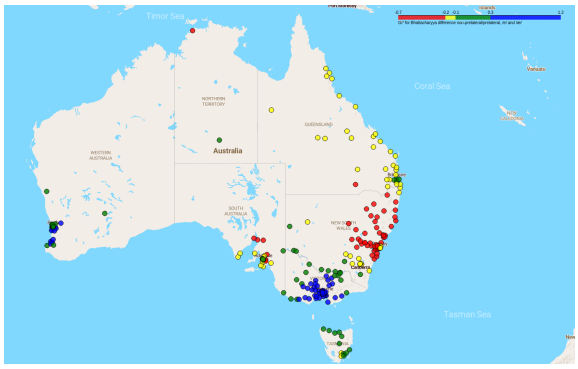

Figure 5: G_i^* values for Bhattacharyya Difference

Fig. 5 shows G_i^* values, calculated on the basis of the Bhattacharyya difference, at the 252 locations where the merger was analyzed. As can be seen, the difference is largest in Melbourne and neighboring VIC localities, and is smallest in NSW, especially in Sydney and environs and the Central Coast. Values are also high in WA in the Perth metropolitan area and in adjacent councils.

5 Discussion

Spatial distribution of Bhattacharyya difference values provides preliminary confirmation that the /el/-/æl/ merger is primarily a Victorian phenomenon, and particularly in southern Victoria. The regional pattern is evident in the mapped raw values (Fig. 4), and becomes clearer when the difference value for each location is converted to a G_i^* statistic (Fig. 5).

Raw difference values in the 252 sampled locations are heterogeneous: although the highest values are found in Melbourne, and the merger is evident to a lesser degree in other parts of VIC, consistent with previous research (Loakes et al., 2024a), some high values can also be found in, for example, WA, QLD, and TAS. The lowest values are found in NSW and SA, and locations with low values can be found throughout Australia.

This heterogeneity likely reflects variability at several levels: Firstly, in terms of the sample size and demographic characteristics of the recorded tokens at each location, secondly, in terms of the audio quality for the recordings, which vary between channels and also among the different videos uploaded by a single channel, and finally, in terms of the presence or absence of the merger for individual speakers. This last point aligns with research findings in fine-grained phonetic studies and perception studies, as noted in the introduction.

Despite the inherent variability in the data, the large sample size tends to reduce the impact of this variability on the analysis. According to the Central Limit Theorem, the sample mean approaches the population mean as sample size increases, leading to more reliable aggregate characteristics. As a result, a geographical pattern emerges more clearly in the difference value map in Fig. 4, and especially in the spatial autocorrelation map in Fig. 5, even if some of underlying data points contain errors (see Section 5.1, below).

One unexpected finding is that the WA localities sampled in the corpus exhibit relatively high Bhattacharyya difference values (and thus also high G_i^* values), in some cases almost as high as those in the greater Melbourne area. Although it is possible that the merger is present among some WA speakers, it has not previously been noted in the literature (or remarked upon as a salient feature of Perth speech), as far as we know. Docherty et al. (2018, Fig. 4) note that the distance between mean values for the /e/ and /æ/ vowels in conversational speech from Perth is smaller than when read aloud, which is typical for English varieties, but they do not mention presence of /el/-[æl]. While our results for WA may reflect the presence of Melburnians who have relocated to Perth, given the large number of sampled videos in 28 different WA locations, this possibility seems unlikely for all individuals. Further investigation of these WA data are also warranted, and will form the basis of a future study.

The prospect that the vowel merger may be mediated by word frequency, an idea which has been proposed in several studies of historical vowel shifts (Bybee, 2002; Pierrehumbert, 2001; Hay et al., 2015), is not corroborated in this data. While we find no evidence for broad-based frequency effects, a more fine-grained analysis of particular locations or lexical items may reveal frequency associations.

5.1 Caveats

A number of important caveats must be taken into account concerning the underlying data and the measurement of formant values. First, CoANZSE transcripts are generated by ASR, and contain errors. The nature of the merger under consideration is such that in some cases, phonological contrasts are eliminated, making it difficult for an ASR algorithm to determine the correct lexical item.[2] In

[2] An example can be found in a transcript from the Horsham Rural City Council, Victoria, entitled *Dental Health Tips for Families*, in which a speaker is transcribed as having said ... *so even harder objects like your carrot sticks and* ***salary****...*. This excerpt can be listened to on CoANZSE Audio at `https://tinyurl.com/mtv2adp3`.

addition, ASR errors may result in false positives (i.e., the targeted phonological context being incorrectly identified, for example if a speaker said *until* but the ASR transcribed *and tell*) as well as false negatives (i.e., the targeted phonological context being missed, e.g. *bill* instead of *bell*). Nevertheless, the number of words for which /e/ can be substituted with /æ/ and result in a legitimate lexical item is small, compared to the overall number of word types containing these phones. Furthermore, the word error rate of CoANZSE data has been calculated to be 0.14 (Coats, 2024c). Given that these errors are distributed across multiple categories for any phone (e.g. a word containing /e/ could be mistranscribed as containing /ɪ/, but also with /i/, /eɪ/, etc.), the "noisiness" of the data is unlikely to result in vowel extraction errors that would systematically shift the results, especially given the sheer size of the sample, at almost 4.3 million tokens. For a few word pairs, the merger may actually be underrepresented in this data due to ASR errors: searching the CoANZSE Audio website reveals, in addition to hits where the ASR has mistranscribed *celery* as *salary*, several instances of *watching tally*.

Another caveat concerns formant values. Transformation of formant values to ensure comparability is a common procedure in phonetic analysis (see, e.g., Adank et al., 2004, Fabricius et al., 2009, Flynn, 2011, Kendall and Thomas, 2010), but because transcript data from YouTube is not diarized (i.e., there are no indications of changes in speaker turn), normalization at speaker level to account for sex-associated differences in vocal tract length was not possible. Instead, we used a scaled Nearey transformation. As Thomas and Kendall note, Nearey's method, a version of which was used for vowel normalization for the data presented in the *Atlas of North American English* (Labov et al., 2005) is "best only when a study has an exceptionally high subject count" (Thomas and Kendall, 2007), a condition which is likely for this data, although the exact number of speakers is unknown.

Despite this, corpus-phonetic analysis of large datasets without speaker labels is relatively uncharted territory, and the most suitable technique for vowel formant normalization for such data remains to be determined. One possibility for this and similar data would be to automatically diarize and induce speaker sex/gender labels, using pyannote for diarization (Bredin, 2023; Plaquet and Bredin, 2023) and wav2vec2-large-xlsr-53-gender-recognition-librispeech (Ferreira, 2024) for speaker gender identification. Future work with this data may undertake these steps.

A third caveat concerns the identities of the persons speaking in the sampled videos: Although it is reasonable to assume that most members of local councils in Australia are resident in or near the locations of those councils, this cannot be guaranteed. As for their residence histories, they are not known. Mobility is a fact of Australian life, and while disqualifying speakers on the basis of prior residence history may be a valid methodological step in studies concerned with the historical evolution and spread of a particular regional language feature, in this study, we have not considered the diachronic development of prelateral merger of /e/ and /æ/.

6 Summary and Future Outlook

This study has considered prelateral merger of /e/ and /æ/ in a large dataset of geolocated naturalistic speech. We used Bhattacharrya diference, a measure of overlap for multidimensional distributions, to characterize the F1 and F2 values for the two vowels in prelateral and non-prelateral contexts. We find that the merger is most evident in southern VIC and Melbourne, largely confirming previous findings based mostly on word- and reading-list data, but it can also be identified in other state/territory locations, including WA.

While this study demonstrates the feasibility of using large, naturalistic speech datasets for phonetic analysis, the results are to be interpreted with caution due to the inherent heterogeneity of the underlying data. Several possibilities for further investigation of the merger using this data present themselves, including 1) Semi-automatic (or manual) annotation of a curated subset of the data in order to investigate the interaction of the merger with demographic parameters; 2) A focus on particular phonological contexts and/or lexical items; 3) A focus on particular discourse content (for example, is the merger more evident when topics pertaining to Melbourne are under discussion in the council meetings that comprise the majority of the underlying data?); and 4) A focus on specific locations or regions which exhibit variability in this data but which have not previously been considered as exhibiting the merger, most notably Perth, but also TAS, as well as QLD, where the merger has already been remarked upon in previous studies. In addition, future work could also explore regional

differentiation in other vowel contrasts. One example is the prenasal raising of /æ/ (where words like *hand* sound like [he:nd]), which is known to vary along various sociophonetic dimensions such as gender, level of linguistic diversity in the community and age (Penney et al., 2023; Gregory, 2019), but has not yet been investigated from the perspective of regional variation.

Finally, we propose that continued work with this data may help to bridge the "sociophonetic gap" by integrating small-scale analysis of carefully collected word-list tokens with large-scale studies of naturalistic speech. As pointed out by Docherty et al. (2018, p. 786), "the deployment of socially marked phonetic features in speech performance is [...] considered to be fundamentally driven by an individual's construction and expression of identity". Naturalistic speech datasets, such as the one used in this study, could potentially contribute to our understanding of how complex configurations of situational contexts and sociostylistic factors shape particular phonetic realizations – provided they have been carefully filtered and annotated for discourse contexts and personal identity parameters. Future work along these lines, we hope, will be able not only to shed light on the /el/-/æl/ merger in Australia more generally, but also to explore whether this merger may be moving from being below the level of consciousness in Melbourne/VIC (Loakes et al., 2017) to a potential indexical marker of Melbourne/VIC identity.

References

Patti Adank, Roel Smits, and Roeland van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, 116(5):3099–3107.

Rosana Ardila, Megan Branson, Kelly Davis, Michael Henretty, Michael Kohler, Josh Meyer, Reuben Morais, Lindsay Saunders, Francis M. Tyers, and Gregor Weber. 2020. Common Voice: A massively-multilingual speech corpus. *Preprint*, arXiv:1912.06670.

Anil Bhattacharyya. 1943. On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society*, 35:99–110.

Steven Bird, Edward Loper, and Ewan Klein. 2009. *Natural Language Processing with Python*. O'Reilly Media Inc.

Paul Boersma and David Weenink. 2024. Praat: Doing phonetics by computer [Computer program]. Version 6.3.10.

David Bradley. 2008. Regional characteristics of Australian English: Phonology. In Bernd Kortmann, Edgar W. Schneider, and Kate Burridge, editors, *A Handbook of Varieties of English, Volume 1: Phonology. Part 3: The Pacific and Australasia*, pages 111–123. De Gruyter Mouton, Berlin, New York.

James Brand, Jen Hay, Lynn Clark, Kevin Watson, and Márton Sóskuthy. 2021. Systematic co-variation of monophthongs across speakers of New Zealand English. *Journal of Phonetics*, 88.

Hervé Bredin. 2023. Pyannote.audio 2.1 speaker diarization pipeline: Principle, benchmark, and recipe. In *INTERSPEECH 2023*, pages 1983–1987.

Joan Bybee. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14(3):261–290.

Steven Coats. 2023. A pipeline for the large-scale acoustic analysis of streamed content. In *Proceedings of the 10th International Conference on CMC and Social Media Corpora for the Humanities (CMC-Corpora 2023)*, pages 51–54.

Steven Coats. 2024a. Building a searchable online corpus of Australian and New Zealand aligned speech. *Australian Journal of Linguistics*, 0(0):1–17.

Steven Coats. 2024b. CoANZSE Audio: Creation of an online corpus for linguistic and phonetic analysis of Australian and New Zealand Englishes. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 3407–3412.

Steven Coats. 2024c. Noisy data: Using automatic speech recognition transcripts for linguistic research. In Steven Coats and Veronika Laippala, editors, *Linguistics across disciplinary borders: The march of data*, page 17–39. Bloomsbury Academic, London.

Rolando Coto-Solano, James N. Stanford, and Sravana K. Reddy. 2021. Advances in completely automated vowel analysis for sociophonetics: Using end-to-end speech recognition systems With DARLA. *Frontiers in Artificial Intelligence*, 4.

Felicity Cox and Janet Fletcher. 2017. *Australian English Pronunciation and Transcription*, 2 edition. Cambridge University Press.

Felicity Cox and Sallyanne Palethorpe. 2004. The border effect: Vowel differences across the NSW/Victorian border. In *Proc. 2003 Conference, Australian Linguistics Society*.

Chloé Diskin, Deborah Loakes, Rosey Billington, Simón Gonzalez, Ben Volchok, and Josh Clothier. 2019a. Sociophonetic variability in the /el/-/æl/

merger in Australian (Melbourne) English: Comparing wordlist and conversational data. Poster presented at NWAV48, Eugene, Oregon.

Chloé Diskin, Deborah Loakes, Rosey Billington, Hywel Stoakes, Simón Gonzalez, and Sam Kirkham. 2019b. The /el-/æl/ merger in Australian English: Acoustic and articulatory insights. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, pages 1764–1768.

Chloé Diskin-Holdaway, Debbie Loakes, and Josh Clothier. 2024. Variability in cross-language and cross-dialect perception. How Irish and Chinese migrants process Australian English vowels. *Phonetica*, 81(1):1–41.

Gerard Docherty, Paul Foulkes, Simon Gonzalez, and Nathaniel Mitchell. 2018. Missed connections at the junction of sociolinguistics and speech processing. *Topics in Cognitive Science*, 10(4):759–774.

Anne H. Fabricius, Dominic Watt, and Daniel Ezra Johnson. 2009. A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. *Language Variation and Change*, 21(3):413–435.

Alef Iury Siqueira Ferreira. 2024. wav2vec2-large-xlsr-53-gender-recognition-librispeech. Accessed: 2024-10-31.

Nicholas Flynn. 2011. Comparing vowel formant normalisation procedures. *York Papers in Linguistics Series*, 2(11):1–28.

Arthur Getis. 2010. Spatial autocorrelation. In Manfred M. Fischer and Arthur Getis, editors, *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, pages 255–278. Springer, Berlin, Heidelberg.

Arthur Getis and J. K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3):189–206.

Adele Gregory. 2019. The [æ]nds of the earth: an investigation of the DRESS and TRAP vowels in Northern Queensland. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, pages 1754–1758.

Jack Grieve, Dirk Speelman, and Dirk Geeraerts. 2011. A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change*, 23(2):193–221.

Jack Grieve, Dirk Speelman, and Dirk Geeraerts. 2013. A multivariate spatial analysis of vowel formants in American English. *Journal of Linguistic Geography*, 1(1):31–51.

Jennifer B. Hay, Janet B. Pierrehumbert, Abby J. Walker, and Patrick LaShell. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition*, 139:83–91.

Yannick Jadoul, Bill Thompson, and Bart de Boer. 2018. Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, 71:1–15.

Daniel Ezra Johnson. 2015. Quantifying vowel overlap with Bhattacharyya's affinity. *New Ways of Analyzing Variation (NWAV44), Toronto*.

Tyler Kendall and Erik R. Thomas. 2010. Vowels: Vowel manipulation, normalization, and plotting in R [R library].

William Labov, Sharon Ash, and Charles Boberg. 2005. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. De Gruyter Mouton, Berlin • New York.

Mark Y. Liberman. 2019. Corpus phonetics. *Annual Review of Linguistics*, 5:91–107.

Debbie Loakes, Josh Clothier, John Hajek, and Janet Fletcher. 2024a. Sociophonetic variation in vowel categorization of Australian English. *Language and Speech*, 67(3):870–906.

Debbie Loakes, Janet Fletcher, and Josh Clothier. 2024b. One place, two speech communities: Differing responses to sound change in Mainstream and Aboriginal Australian English in a small rural town. In Felicitas Kleber and Tamara Rathcke, editors, *Speech dynamics: Synchronic variation and diachronic change*, pages 117–144. De Gruyter Mouton, Berlin, Boston.

Deborah Loakes, John Hajek, and Janet Fletcher. 2011. /el/-/æl/ transposition in Australian English: Hypercorrection or a competing sound change? In *Proceedings of the 17th International Congress of Phonetic Sciences*.

Deborah Loakes, John Hajek, and Janet Fletcher. 2017. Can you t[æ]ll I'm from M[æ]lbourne? *English World-Wide*, 38(1):29–49.

Adrien Méli, Steven Coats, and Nicolas Ballier. 2023. Methods for phonetic scraping of YouTube videos. In *6th International Conference on Natural Language and Speech Processing (ICNLSP 2023)*, volume 6, pages 244–249.

P. A. P. Moran. 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23.

J. K. Ord and Arthur Getis. 1995. Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, 27(4):286–306.

Joshua Penney, Felicity Cox, and Sallyanne Palethorpe. 2023. Variation in pre-nasal raising of trap in australian english. *The Journal of the Acoustical Society of America*, 154($4_{s}upplement$) : $A334 - -A334$.

Janet B. Pierrehumbert. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Joan L. Bybee and Paul J. Hopper, editors, *Frequency and the Emergence of Linguistic Structure*, pages 137–158. John Benjamins, Amsterdam.

K. C. S. Pillai. 1955. Some new test criteria in multivariate analysis. *The Annals of Mathematical Statistics*, 26(1):117 – 121.

Alexis Plaquet and Hervé Bredin. 2023. Powerset multi-class cross entropy loss for neural speaker diarization. In *INTERSPEECH 2023*, pages 3222–3226.

Sravana Reddy and James N. Stanford. 2015. Toward completely automated vowel extraction: Introducing DARLA. *Linguistics Vanguard*, 1(1):15–28.

Margaret E. L. Renwick and Joseph A. Stanley. 2020. Modeling dynamic trajectories of front vowels in the American South. *The Journal of the Acoustical Society of America*, 147(1):579–595.

Sergio J. Rey and Luc Anselin. 2010. PySAL: A Python library of spatial analytical methods. In Manfred M. Fischer and Arthur Getis, editors, *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, pages 175–193. Springer, Berlin, Heidelberg.

Ingrid Rosenfelder, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hilary Prichard, and Jiahong Yuan. 2015. FAVE 1.1.3.

Penelope Schmidt, Chloé Diskin-Holdaway, and Debbie Loakes. 2021. New insights into /el/-/æl/ merging in Australian English. *Australian Journal of Linguistics*, 41(1):66–95.

Joseph A. Stanley and Betsy Sneller. 2023. Sample size matters in calculating Pillai scores. *The Journal of the Acoustical Society of America*, 153(1):54–67.

Erik R. Thomas and Tyler Kendall. 2007. NORM: The vowel normalization and plotting suite. Accessed: 2024-10-29.

Paul Warren. 2018. Quality and quantity in New Zealand English vowel contrasts. *Journal of the International Phonetic Association*, 48(3):305–330.

Robert Weide et al. 1998. The Carnegie Mellon Pronouncing Dictionary. Release 0.7b.