THE /el-/æl/ MERGER IN AUSTRALIAN ENGLISH: ACOUSTIC AND ARTICULATORY INSIGHTS

Chloé Diskin[1], Deborah Loakes[1,2], Rosey Billington[1,2], Hywel Stoakes[3], Simón Gonzalez[2,4], Sam Kirkham[5]

[1]School of Languages and Linguistics, The University of Melbourne, Australia [2]ARC Centre of Excellence for the Dynamics of Language, Australia [3]The University of Auckland, New Zealand [4]The Australian National University, Australia [5]Lancaster University, United Kingdom

chloe.diskin@unimelb.edu.au; dloakes@unimelb.edu.au; rbil@unimelb.edu.au; h.stoakes@auckland.ac.nz; simon.gonzalez@anu.edu.au; s.kirkham@lancaster.ac.uk

ABSTRACT

This paper investigates a merger-in-progress of /e/-/æ/ in prelateral contexts for speakers of Australian English in Victoria. Twelve participants (7F, 5M) were recorded producing a wordlist resulting in acoustic and concurrent articulatory data via stabilised mid-sagittal ultrasound tongue imaging. Focusing on a subset of the data comprising short front vowels /ɪ, e, æ/ in /hVt/ and /hVl/ contexts, findings show that there are robust acoustic differences between /e/ and /æ/ preceding /t/, as anticipated. However, individual differences emerge for /e/ and /æ/ preceding /l/, with highly gradient production patterns across the speakers, ranging from speakers who exhibit merger behaviour to those who maintain categorical distinctions. The evidence for merging behaviour across speakers is similar, but does not map directly, across both the acoustic and articulatory data, and illustrates the value of incorporating a range of data types in investigating a merger-in-progress.

Keywords: Australian English, /el/-/æl/ merger, acoustics, articulation, ultrasound

1. INTRODUCTION

A merger of /e/ and /æ/ in prelateral contexts has been reported among speakers of Australian English (AusE) in southern parts of Victoria [4, 18, 20, 21] (as well as in other English varieties [8, 14, 27]). In locations where it occurs, it appears to be completely entrenched for some community members, but still in progress for others (e.g. [19] for L1 Australian Aboriginal English speakers vs. mainstream AusE speakers in the Warrnambool region). Since the merger is not found consistently among all speakers in southern Victoria, it is more accurately described as a merger-in-progress (see e.g. [13]). Previous work adopting an acoustic phonetic methodology to determine the extent of this merger in production shows that merged speakers tend to acoustically merge /el/ with /æl/ [4, 21], while non-merging speakers have a very low and retracted /e/ before /l/, but still keep the vowels distinct [18, 21].

There have also been investigations into the perception of this merger by AusE speaker-listeners via vowel categorisation tasks. These show varying degrees of merger in perception, with some listeners answering at random when they are faced with /el/ and /æl/, and others maintaining a distinction [7, 18]. While not directly aligned, perception and production behaviour in relation to this merger tend to be linked, with speakers merging in production more likely to do so in perception [18]. Merging in perception has also been shown to be highly regionally specific [20].

The /el/-/æl/ merger occurs in the context of rapid diachronic change in the short front vowel system of AusE in production [5] and perception [22], and is thought to occur precisely because of highly variable input listeners receive for the /e/-/æ/ contrast, even in locations where the merger does not occur [20, 21]. The combination of rapidly changing vowel qualities and a particularly dark /l/ provides favourable conditions for merger [21], with highly coarticulated vowels in prelateral contexts masking cues to a contrast and possibly being the instigator for change (see [23, 12] on listener-induced sound changes; and [21] which links this with the /el/-/ æl/ merger).

To date there have been no investigations into this merger in articulation – an aspect that is of pertinence due to the incipient status of the phenomenon. Other research into mergers shows that articulatory studies allow investigations of aspects of speaker behaviour that cannot be accessed via acoustic analysis alone [6, 16]. Overall, there is limited articulatory research on AusE; while there are some investigations of vowel articulation [2, 25, 28], and lateral articulation [17, 32], there are no articulatory investigations of vowel production in prelateral contexts.

1.1. Aims of the study

We present findings from an exploratory study of the /el/-/æl/ merger-in-progress among 12 native AusE speakers (7F, 7M, average age 31; *SD*=5). All were born and raised in Melbourne, with the exception of one born and raised on the Mornington Peninsula (~90 km from Melbourne). The present study is part of a wider project with a range of participants and data types. The guiding research question is: To what extent do AusE speakers in Victoria show evidence for this merger in articulation and how does this relate to accompanying acoustic cues?

2. METHOD

2.1. Materials and procedures

We report on acoustic and articulatory data collected via laboratory-based recording of a wordlist with simultaneous ultrasound tongue imaging. The full wordlist contained 54 items, but here we focus on the subset pertinent to our current research question, containing short front vowels /ɪ, e, æ/ in /hVt/ and /hVl/ contexts: *hit*, *het*, *hat* and *hill*, *hell*, *Hal*. Stimuli were presented to participants orthographically using Articulate Assistant Advanced (AAA) software [1], which was also used to capture the data. An experimenter advanced through the wordlist items and the participant produced three repetitions of the target item in isolation. The items were not randomised. The total number of tokens in this subset of the data was 216. Participants differed in production speed for individual tokens (mean word duration 478ms, *SD*=112).

Acoustic data were captured using a Røde NT-3 hypercardioid condenser microphone at a sampling rate of 44.1kHz with 16-bit depth. Mid-sagittal ultrasound tongue images were generated using a Mindray DP6600 ultrasound machine and a 65EC10EA microconvex transducer with 120° field of view, set to a frequency of 6.5MHz and imaging depth of 9.7cm. The probe was stabilised using an Articulate Instruments Probe Stabilisation headset worn by participants [26]. Data were captured at a frame rate of ~30 frames per second. Concurrent with the tone produced at the beginning of recording for each item, a flash was imposed on the ultrasound image to enable audio-video synchronisation [31].

2.2. Data processing and analysis

2.2.1. Acoustic data

Automatic segmentation of the speech signal was performed via WebMAUS [15], using the AusE language model. Segment boundaries in the output TextGrid files were checked and manually corrected in Praat [3], with reference to wideband spectrograms and corresponding waveforms. The maximum F2 value between 25% and 75% of each segmented vowel was then automatically identified and added in a point tier using a Praat script. Given that the acoustic boundaries between vowels and coda laterals are difficult to precisely identify, and are influenced by the range of lateral articulations (e.g. vocalised or velarised /l/), this point was chosen as it is likely to be further removed from the lower F2 values associated with the lateral coda in the /hVl/ items, while also avoiding formant perturbations in cases of initial breathiness and creak preceding the coda in the /hVt/ items. In most cases, the annotated point was close to the start of the chosen range (on average 33%, *SD*=13%). A hierarchical database was constructed using the EMU Speech Database Management System [29], including tiers for the word, segments, and max F2. F1 and F2 characteristics at the labelled max F2 were queried and analysed using the emuR package in R [24, 30]. Values were examined for all vowels in the tokens, but the primary interest in relation to our research question is whether individual speakers show acoustic differences between /e/ and /æ/ preceding /l/ as well as preceding /t/.

2.2.2. Ultrasound data

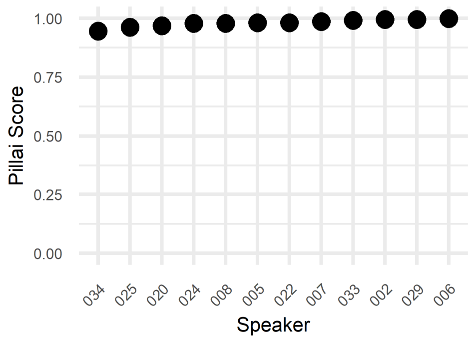
Ultrasound images were de-interlaced, resulting in a frame rate of ~60 frames per second. Video and audio were synchronised in AAA using the recorded tone and the flash imposed on the video as alignment reference points. Using edge-detection within AAA, splines were automatically fitted to sections of the image corresponding to the surface of the tongue, for frames across each produced token. Fitted splines were hand-corrected where required. The video frame nearest to the labelled max F2 between 25%-75% of each vowel was identified. Spline data for these frames was extracted, to allow for exploration of lingual configuration at this point in the vowel in relation to spectral characteristics at the same point. Spline data were visualised using the UVA Shiny app [9] in which tongue contours are plotted using XY Cartesian coordinates. Differences between contours are therefore based on height and advancement in the Cartesian space. We examined whether there were differences in overall tongue shape for the vowels in both *het-hat* and *hell-Hal* for individual speakers, and to what extent this corresponds to acoustic difference, or lack of difference, between the vowels for each speaker, particularly in the pre-lateral context.

3. RESULTS

3.1. Acoustic results

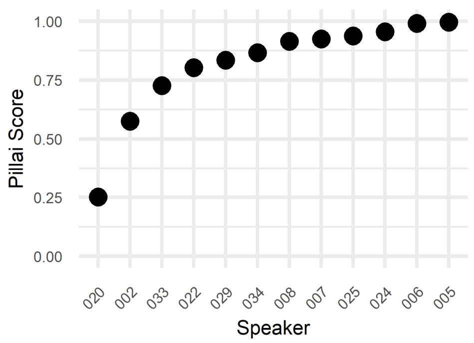
For all participants, F1 and F2 values for /e/ and /æ/ preceding /t/ are indicative of the robust acoustic distinction we expect to find for this pair of vowels in this consonantal context. For all participants, F1 values are higher for /æ/ ($\bar{x}$ 957Hz, σ 144Hz) compared to /e/ ($\bar{x}$ 666Hz, σ 129Hz), and F2 values are lower for /æ/ ($\bar{x}$ 1712Hz, σ 182Hz) compared to /e/ ($\bar{x}$ 2098Hz, σ 171Hz), as can be seen in the example plots in Figs. 3-5. Fig. 1 shows Pillai scores derived from these measures, following e.g. [10, 11], as an indicator of the extent of acoustic overlap between vowels; values approaching 1 indicate a high level of distinctiveness and values closer to 0 indicate greater overlap in the F1/F2 space. As can be seen, all participants have high Pillai scores, within a narrow range (0.95 to 1.0).

Figure 1: Pillai scores for all participants, comparing /e/ and /æ/ preceding /t/.



Patterns are somewhat different when comparing F1 and F2 values for /e/ and /æ/ preceding /l/. While there is still a tendency towards higher F1 values for /æ/ ($\bar{x}$ 933Hz, σ 125Hz) than for /e/ ($\bar{x}$ 776Hz, σ 141Hz) in this context, there are only slightly lower F2 values for /æ/ ($\bar{x}$ 1698Hz, σ 130Hz) compared to /e/ ($\bar{x}$ 1727Hz, σ 182Hz), and there is more variable production behaviour across participants.

Figure 2: Pillai scores for all participants, comparing /e/ and /æ/ preceding /l/.



Pillai scores based on pre-lateral tokens are given in Fig. 2. Some participants have values similarly high to those in Fig. 1; e.g. participant 005 has the highest score, 1.0, and as shown in Fig. 3, this corresponds to tokens of pre-lateral /e/ and /æ/ which are distinct from one another in the F1/F2 space. In comparison, participant 020 has the lowest score, 0.25, and as shown in Fig. 4, this corresponds to an overlap in the distribution of pre-lateral /e/ and /æ/ tokens. Participant 033, whose score is an intermediate 0.73, shows similar but not quite overlapping distributions for /e/ and /æ/ (Fig. 5.)

Figure 3: F1/F2 for /ɪ, e, æ/ before /l, t/ (participant 005).

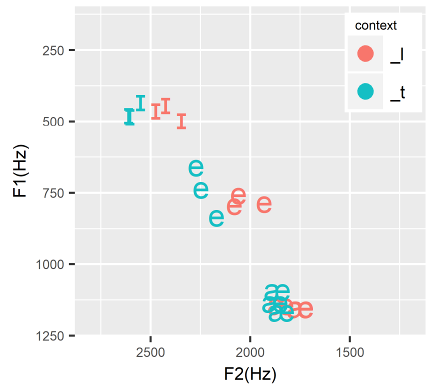


Figure 4: F1/F2 for /ɪ, e, æ/ before /l, t/ (participant 020).

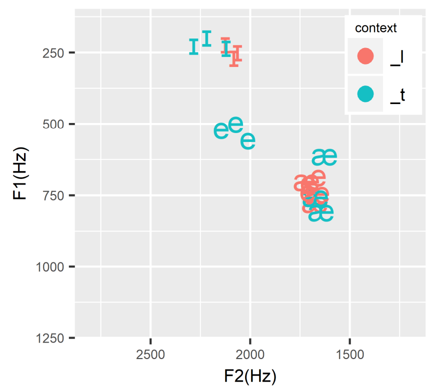
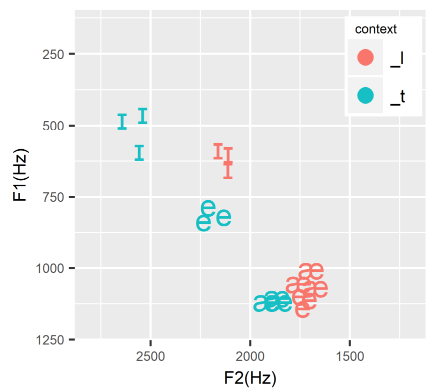


Figure 5: F1/F2 for /ɪ, e, æ/ before /l, t/ (participant 033).



3.2. Articulatory results

Midsagittal tongue splines show that all participants exhibit distinct lingual configurations for /e/ and /æ/ preceding /t/. The splines for /e/ are indicative of a more anterior tongue root position and raised tongue dorsum than for /æ/, as seen in the left panels in Figs. 6-8. This differs from splines for these vowels preceding /l/, which show more varied patterns across participants, but also points towards a degree of gestural convergence. For participant 005, whose Pillai score of 1.0 suggests they do not tend to merge these vowels pre-laterally, splines show similarities in the anterior constriction but there is still a difference in the tongue root, as in the right panel in Fig. 6. For participant 020 (Pillai score 0.25), there is overlap between the splines for each vowel, as seen in Fig. 7. For participants with Pillai scores between these, any visible differences between the splines for pre-lateral tokens are generally smaller than those observed preceding /t/, and may be apparent in only a small part of the imaged area, as seen for participant 033 in Fig. 8. Where there are differences, the nature of these varies across participants. Several (e.g. 002, 007, 025, 034) show splines with a high degree of overlap, similar to participant 020 (Fig. 7).

Figure 6: Midsagittal tongue splines for /e/ and /æ/ before /t/ (left) and before /l/ (right), for participant 005.

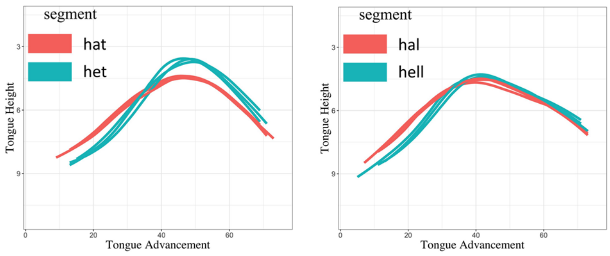


Figure 7: Midsagittal tongue splines for /e/ and /æ/ before /t/ (left) and before /l/ (right), for participant 020.

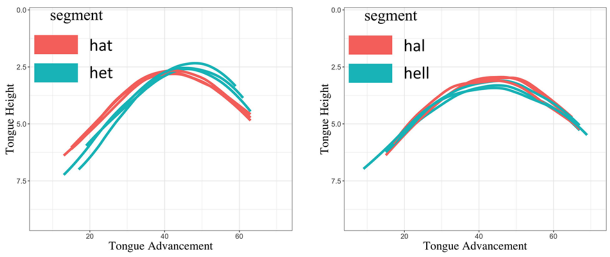
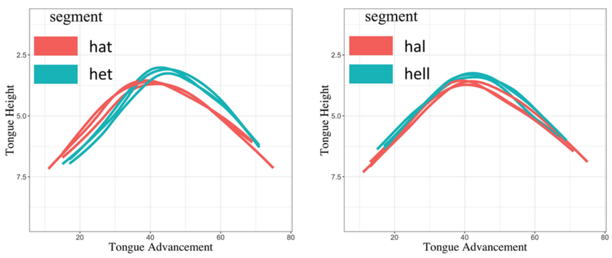


Figure 8: Midsagittal tongue splines for /e/ and /æ/ before /t/ (left) and before /l/ (right), for participant 033.



4. CONCLUDING DISCUSSION

Results suggest that, as expected, all speakers maintain a categorical distinction between /e/ and /æ/ in the non-lateral context, as evidenced by the lack of acoustic overlap and distinct lingual configurations for each vowel. Production patterns for /e/ and /æ/ preceding /l/ are more varied; for three participants, Pillai scores are in the 0.95 to 1.0 range that was observed for all participants in the non-lateral context, and scores remain high for others. However, there are indications that for some participants there is acoustic overlap between /e/ and /æ/ pre-laterally. Interestingly, this appears to be less widespread than found in earlier acoustic studies with participants from southern Victoria (e.g. [18, 20]). While the acoustic findings point towards evidence of merger behaviour among these participants, patterns are highly gradient. This is not unexpected given that the pre-lateral merger has been described as a sound change in progress and is not fully entrenched across different populations in southern Victoria [21]. Articulatory results similarly show varied production patterns for pre-lateral /e/ and /æ/, but also a tendency towards less distinct tongue splines for /e/ and /æ/ preceding /l/ than preceding /t/. The participants who exhibit the most acoustically distinct pre-lateral /e/ and /æ/ (e.g. 005) show reasonably distinct tongue splines; whereas those with the least acoustically distinct tokens (e.g. 020) have overlapping tongue splines. However, there is a range of possibilities in between; some speakers with high Pillai scores (i.e. less evidence of merging behaviour) have tongue splines which largely overlap (007, 025, 024).

Overall, acoustic and articulatory patterns align in various ways in the present data, illustrating some of the complexities associated with phonetic investigations of sound changes underway. A wider range of methodological approaches may reveal more consistent correlations between the acoustics and articulation, which may not necessarily be evident based on static mid-sagittal tongue splines alone. Objective quantification measures of contour difference are being tested for inclusion in future work. The articulatory findings also highlight the need for further research examining e.g. the lateral channel [32] and lip rounding, to more comprehensively understand the articulation of coda-lateral rhymes in AusE. Dynamic analyses will also be crucial in understanding production behaviour in pre-lateral contexts (including variability in the production of /l/ as either velarised or vocalised), and will be explored in ongoing research. To better assess the degree and spread of the /el/-/æl/ merger, there is also a need for more data, and data types, across AusE speakers from Victoria. Participants in this study undertook additional tasks, including a studio-recorded sociolinguistic interview and wordlist, and a perception task (forced choice vowel categorisation). The present study has elucidated previously undescribed aspects of speaker behaviour in AusE, in conditions which are highly variable. Future work will triangulate acoustic, articulatory and perceptual patterns for pre-lateral /e/ and /æ/ and the mechanisms underpinning the merger-in-progress.

5. REFERENCES

[1] Articulate Instruments Ltd. 2012. *Articulate Assistant Advanced user guide: Version 2.14*. Edinburgh: Articulate Instruments Ltd.

[2] Blackwood Ximenes, A., Shaw, J. A., Carignan, C. 2017. Comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from American and Australian English. *J. Acoustical Society of America* 142, 363–377.

[3] Boersma, P., Weenink, D. 2018. Praat: Doing phonetics by computer [Version 6.0.40]. Available: http://www.praat.org/.

[4] Cox, F., Palethorpe, S. 2004. The border effect: Vowel differences across the NSW – Victorian border. *Proc. ALS 2003* Newcastle, 1-27.

[5] Cox, F., Palethorpe, S. 2008. Reversal of short front vowel raising in Australian English. *Proc. Interspeech 2008* Brisbane, 342-345.

[6] De Decker, P. M., Nycz, J. R. 2012. Are tense [æ]s really tense? The mapping between articulation and acoustics. *Lingua* 122, 810-821.

[7] Diskin, C., Loakes, D., Clothier, J. 2018. Varietal differences in categorisation of /ɪ e æ/: A case study of Irish and Australian English listeners in Melbourne. *Proc. SST2018* Sydney, 105-108.

[8] Finn, P. 2004. Cape Flats English: Phonology. In Burridge, K., Kortmann, B. (eds), *A Handbook of Varieties of English*, vol. 1: Phonology. Berlin: Mouton de Gruyter, 964-984.

[9] Gonzalez, S. 2018. UVA. Available: https://osf.io/z5hyj/

[10] Hall-Lew, L. 2010. Improved representation of variance in measures of vowel merger. *Proc. Meetings on Acoustics* 9, paper 060002.

[11] Hall-Lew, L., Nycz, J. R. 2013. Best practices in measuring vowel merger. *Proc. Meetings on Acoustics* 20, paper 060008.

[12] Harrington, J., Kleber, F., Reubold, U., Schiel, F., Stevens, M. 2018. Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in Cognitive Science* 10, 1-21.

[13] Hay, J., Warren, P., Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *J. Phonetics* 34, 458-484.

[14] Ingram, J., Mühlhäusler, P. 2004. Norfolk Island Pitcairn English: Phonetics and Phonology. In Burridge, K., Kortmann, B. (eds), *A Handbook of Varieties of English*, vol. 1: Phonology. Berlin: Mouton de Gruyter, 780-804.

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

[16] Lawson, E., Scobbie, J., Stuart-Smith, J. 2013. Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation. *J. Phonetics* 41, 198-210.

[17] Lin, S., Palethorpe, S., Cox, F. 2012. An ultrasound exploration of Australian English /CVl/ words. *Proc. SST2012* Sydney, 105-108.

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

[19] Loakes, D., Fletcher, J., Hajek, J., Clothier, J., Volchok, B. 2016. Short vowels in L1 Aboriginal English spoken in Western Victoria. *Proc. SST2016* Sydney, 33-36.

[20] Loakes, D., Hajek, J., Clothier, J., Fletcher, J. 2014. Identifying /el/-/æl/: A comparison between two regional Australian towns. *Proc. SST2014* Canterbury, 41-44.

[21] 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, 29-49.

[22] Mannell, R. 2004. Perceptual vowel space for Australian English lax vowels: 1988 and 2004. *Proc. SST2004* Sydney, 221-226.

[23] Ohala, J. J. 1993. Sound change as nature's speech perception experiment. *Speech Communication* 13, 155-161.

[24] R Core Team. 2018. R: A language and environment for statistical computing [Version 3.4.2]. Available: https://www.r-project.org.

[25] Ratko, L., Proctor, M., Cox, F., Veld, S. 2016. Preliminary investigations into the Australian English articulatory vowel space. *Proc. SST 2016* Sydney, 117-120.

[26] Scobbie, J. M., Wrench, A. A., van der Linden, M. 2008. Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement *Proc. 8th Intl. Seminar on Speech Production* Strasbourg, 373–376.

[27] Thomas, B., Hay, J. 2006. A pleasant malady: The Ellen/Allan merger in New Zealand English. *Te Reo* 48, 69-93.

[28] Watson, C. I., Harrington, J., Palethorpe, S. 1998. A kinematic analysis of New Zealand and Australian English vowel spaces. *Proc. 5th Intl. Conference on Spoken Language Processing* Sydney, 2363-2366.

[29] Winkelmann, R., Harrington, J., Jänsch, K. 2017. EMU-SDMS: Advanced speech database management in R. *Computer Speech & Language* 45, 392-410.

[30] Winkelmann, R., Jänsch, K., Cassidy, S., Harrington, J. 2018. emuR: Main package of the EMU Speech Database Management System [Version 1.0.0]. Available: https://cran.r-project.org.

[31] Wrench, A. A., Scobbie, J. M. 2008. High-speed cineloop ultrasound vs. video ultrasound tongue imaging: comparison of front and back lingual gesture location and relative timing. *Proc. 8th Intl. Seminar on Speech Production* Strasbourg, 57-60.

[32] Ying, J., Carignan, C., Shaw, J. A., Proctor, M., Derrick, D., Best, C. 2017. Temporal dynamics of lateral channel formation in /l/: 3D EMA data from Australian English. *Proc. Interspeech 2017*, 2978-2982.