Interspeech 2025
17-21 August 2025, Rotterdam, The Netherlands

Lateral channel formation in Australian English /l/: insights from Magnetic Resonance Imaging

Tünde Szalay[1], Michael Proctor[1], Amelia Gully[2], Tharinda Piyadasa[3], Craig Jin[3], David Waddington[3], Naeim Sanaei[4], Sheryl Foster[3,4], Kirrie Ballard[3]

[1]Macquarie University, Australia; [2]University of York, UK; [3]University of Sydney, Australia; [4]Westmead Hospital, Australia

tunde.szalay@mq.edu.au

Abstract

Goals of lateral approximant production are imperfectly understood, partly because of the limitations of most existing data, restricted to the midsagittal plane. To provide more complete information about the configuration of the vocal tract for laterals, /l/-production in three vowel contexts by three Australian English speakers was examined for the first time using a combination of real-time and volumetric Magnetic Resonance Imaging (MRI). Laterals produced by all speakers were characterised by bilateral parasagittal airflow, with some asymmetries in side channel geometries. Central occlusion of the oral airway varied in location, timing, and duration across speakers and vowel contexts, but was consistently associated with reduction in overall acoustic intensity relative to context vowels. These data provide further insights into the complex relationships between articulatory, coarticulatory and acoustic properties of lateral approximants, and their realisation in Australian English.

Index Terms: 3D modelling, intensity, MRI, liquids, dark /l/

1. Introduction

The lateral /l/ is a multigestural segment prototypically produced with a central alveolar closure, dorsal retraction, and lateral channel formation [1, 2], with lateral channel formation being the defining articulatory characteristic of the lateral class [3]. Lateral channels may be formed passively, resulting from tongue elongation caused by simultaneous tongue tip fronting and dorsal retraction, as is typically observed in dark [ɫ] [4]. Thus, as closure is formed, a supralingual pocket of air is trapped behind the midsagittal closure, setting up a side-branch to the main airflow [5–7]. Lateral channels also form in clear [l] articulated with less lingual elongation, where stable timing relations have been observed between the sides and back of the tongue, suggesting active control of lateralisation [8, 9].

Lateralisation, despite being the defining gesture of the class, is imperfectly understood in part due to the limitations of vocal tract imaging methods and to the high volume of inter- and intraspeaker variation [10]. Accurate imaging of lateral channels requires simultaneous mapping of the tongue surface and the soft palate through 3D modelling techniques [10, 11]. In American English, lateral channels were captured by combining magnetic resonance imaging with electropalatography and in Brazilian Portuguese by combining 3D/4D ultrasound with 3D digitised palate impressions [10, 11]. In Australian English (AusE), tongue body elevation relative to tongue blade was estimated through fleshpoint tracking, with lowered blade indicating lateral channel formation [9]. However, fleshpoint tracking cannot capture lateral channels' length or area.

A key acoustic cue to lateral channel formation is anti-resonances, the active cancellation of frequencies in a range determined by the length, area, and relative size of lateral channels [5–7]. For example, a 4.4 cm long lateral channel is predicted to create anti-resonances near or above 2000 Hz [5, 7], whereas a shorter lateral channel of 2.5 cm creates anti-resonances at the higher frequency range of 3000-4000 Hz [6]. Anti-resonances are attributed to unequal lateral channels, but not to lateral channels of equal size [11, 12]. Anti-resonances, caused by active cancellation of frequencies, are difficult to distinguish from a passive lack of resonances caused by lack of enhancement, as both are evidenced by spectral valleys [7]. For example, similar low-energy bands above F3 and F4 were found in Brazilian Portuguese in words with and without laterals (*pala, palha, paia*), suggesting an overall lack of resonance in the region rather than an active cancellation of frequencies caused by lateral channel formation in words containing /l/ [11].

Lateralisation may also be cued by reduced intensity as amplitude is reduced in the higher regions, near the predicted anti-resonances (2000–4000 Hz) [6, 13]. Increased F1 bandwidth caused by the narrowing of the tongue tip constriction further lowers intensity in the lower frequency ranges [6, 13]. Abrupt intensity change is observed impressionistically in lateral acoustics and is used to separate laterals from adjacent vowels during semi-automatic speech segmentation [11, 13–15]. In Brazilian Portuguese, however, the amplitude difference between laterals and adjacent vowels is smaller than the amplitude drop predicted based on the F1 bandwidth, suggesting that the lateral channels boosted rather than dampened intensity [11]. That is, articulatory causes of reduced intensity are imperfectly understood despite their use in segmentation and the early comparisons of laterals to quiet vowels [16].

To further examine articulatory gestures underlying reduced intensity in Australian English (AusE), we captured lateral channels through 3D vocal tract modelling using volumetric Magnetic Resonance Imaging (MRI), constriction formation through real time Magnetic Resonance Imaging (rtMRI), and intensity drop in time-aligned acoustic data. Our aims were to examine (1) lateral channels using 3D vocal tract modelling for the first time in AusE; (2) midsagittal articulatory- and acoustic variation in /l/ in three vowel contexts; and (3) links between lateralisation, coronal closure, and reduced intensity.

2. Methods

Data were collected as part of a larger project examining speech motor control development in AusE-speaking adolescents and young adults [15].

2.1. Participants, stimulus, and procedure

Three young adult native speakers of AusE (Speaker 141: male, 22 years, Speaker 144: female, 22 years, and Speaker 152: fe-

10.21437/Interspeech.2025-1065

male, 21 years) produced sustained and intervocalic laterals in a series of speech tasks recorded outside and inside an MRI scanner. Intervocalic laterals were elicited in the non-words *ili, ala, olo* between three corner vowels: high front /iː/, low /ɐː/, and the highest back vowel of AusE, /oː/ [17]. Participants were cued visually using a progress bar to sustain /l/ for 7.6 s and /VlV/ sequences for 1.5 s.

Each token was recorded once in a quiet room with a Glottal Enterprises EG2-PCX2 digital speech recorder to familiarise the participant with the experimental materials. Sustained lateral productions were later recorded twice during a volumetric MRI scan and intervocalic laterals five times during an rtMRI scan. A total of 3 (participants) × 3 (1 pre-scan + 2 volumetric scan) = 9 sustained /l/ tokens and 3 (participants) × 3 (vowel contexts) × 6 (1 pre-scan + 5 rtMRI) - 3 (one missing rtMRI repetition) = 51 /VlV/ tokens were included in the analysis.

2.2. Data acquisition

MRI data were acquired at Westmead Hospital (Sydney, Australia), on a Siemens Magnetom Prisma 3T scanner with a 64-channel head/neck receiver array coil following the protocol in [15]. The speaker's upper airway was imaged while lying supine. Data were acquired from an 8 mm slice aligned with the mid-sagittal plane, over a 280×280 mm field of view, using a 2D RF-spoiled, radially-encoded FLASH sequence [18].

3D configuration of the vocal tract during sustained lateral production was captured using volumetric imaging of the upper airway. Data were acquired using a T1-weighted fast 3D gradient-echo sequence, with a spatial resolution of 160×160×32 px over a 256×256×64 mm field of view centred on the pharynx: a voxel resolution of 1.6×1.6×2.0 mm.

Audio was recorded concurrently in-scanner at 16 kHz using an Opto-acoustics FOMRI-III ceramic noise-canceling microphone designed for MRI environments [19]. rtMRI data were reconstructed in Matlab into midsagittal videos with a pixel resolution of 0.83 mm2, encoded as 72 frames per second MP4 files. Audio and video were time-aligned automatically during postprocessing using a synchronisation pulse between the microphone and the rtMRI scanner recorded during data collection.

2.3. Sustained /l/ analysis

Vocal tract boundaries were segmented in volumetric images of the upper airway during sustained /l/ production using the semi-automatic Snake-segmentation feature of ITK-SNAP (Fig. 3) [20]. Snake-segmentation is a user-guided active contour segmentation method during which segmentation is initiated at user-defined points while the remaining areas are generated automatically based on image contrast [20]. Acoustic characteristics of sustained /l/ were not analysed due to low signal-to-noise ratio in volumetric MRI recordings.

2.4. Intervocalic /l/ analysis

rtMRI videos and time-aligned in-scanner audio recordings were analysed using a Matlab-based custom graphical interface [21]. Image frames were identified corresponding to articulatory target postures for pre- and post-lateral vowels, and lateral coronal closure and release (Fig. 1) [15]. Vowel targets were located at the centre frame of the stable articulatory position associated with each segment (Fig. 1, bottom L, R). Lateral coronal closure was located at the first frame after any observable gap between the tongue tip (TT) and alveolar ridge (Fig. 1, bottom centre L). Lateral coronal release was located at the first frame when a gap between the TT and alveolar ridge was first observed after closure (Fig. 1, bottom centre R). Coronal closure was achieved in all but one out of 51 tokens. In the token with no alveolar closure, coronal gestural onset was identified in the frame when the highest TT position was reached and offset when the TT started moving away from the highest position.

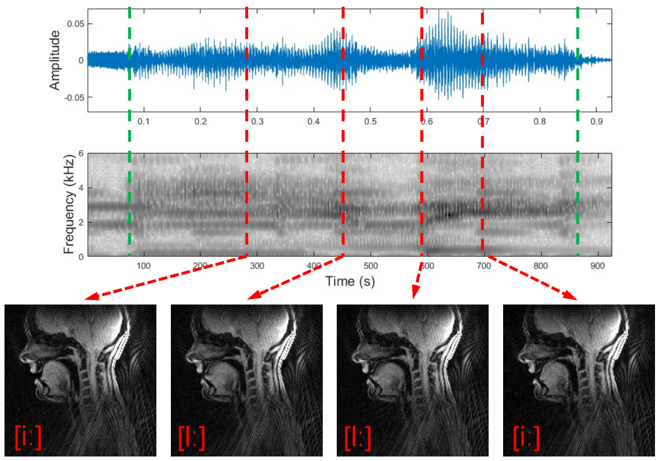


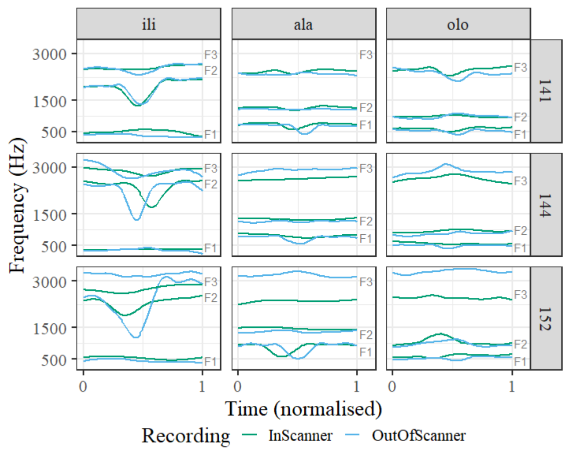
Figure 1: *Lateral dynamics in the front vowel context.* Waveform and spectrogram of noise-cancelled in-scanner recording of /ili/ (green lines: acoustic onset and offset), time-aligned with MRI frames captured at (L to R, red lines): pre-/l/ vowel target, /l/ onset, /l/ offset, post-/l/ vowel target.

Audio recordings were force-aligned using MAUS to locate segment boundaries which were then hand-corrected (Fig. 1, green line) [22–24]. Formant frequencies were estimated automatically in Praat every 20 ms over a 64 ms Gaussian analysis window with 75% overlap, 50 dB dynamic range, and a pre-emphasis filter increasing spectral slope above 100 Hz by 6 dB/octave [25]. Five formants were tracked up to 4.5 kHz ceiling for tokens with lower F2 values typically found in the male speaker's back vowel, and up to 5.5 kHz for higher F2 values, typically found in the female speakers' front vowels [26]. Formant values were then corrected manually [26]. At each timepoint where formants were estimated, intensity values were estimated with a pitch floor of 100 Hz. Values were first squared, then convolved with a Gaussian analysis window (Kaiser-20; sidelobes below -190 dB), yielding an analysis window with an effective duration of 3.2 / 100 Hz. Formant trajectories and intensity contours were smoothed with Generalised Additive Models (GAM) for visualisation purposes (Fig. 2). GAMs were fitted separately for each participant and vowel context to smooth across repetitions using the geom_smooth() function with GAM method and the default 0.95 confidence interval from the R *ggplot2* library [27, 28].

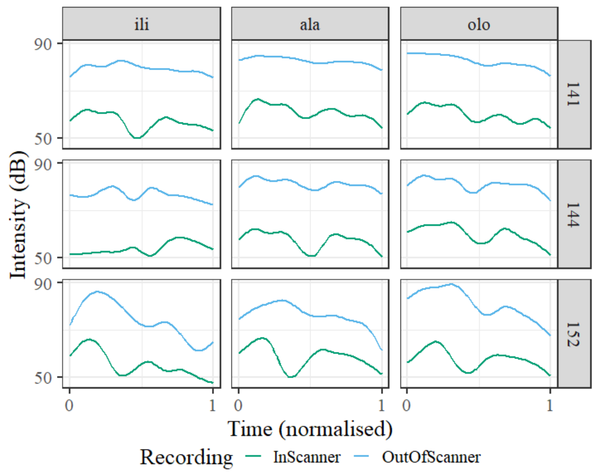
2.5. Data validation

To validate acoustic measurements of in-scanner recordings, in-scanner and out-of-scanner measures were compared (Fig. 2). Formant values were comparable between in-scanner and out-of-scanner recordings of /VlV/ sequences with the exception of Speaker 152's F3 estimated in in-scanner recordings. The reduced F3 reliability is attributed to scanner noise (Fig. 2a).

Intensity was consistently higher in out-of-scanner recordings compared to in-scanner recordings due to scanner noise



(a) *Formant trajectories*



(b) *Intensity trajectories*

Figure 2: *Formant (top) and intensity (bottom) trajectories.* Five in-scanner repetitions of (L to R): /iːliː/, /ɐːlɐː/, /oːloː/ per speaker (rows). Trajectories for /VlV/ sequences (thin lines) smoothed using GAMs (thick lines, 95% CI not shown).

Fig. 2b). Intensity varied between the first and the second vowel depending on which vowel was stressed by the participants, as well as within the vowels due to participants extending each segment, causing intensity variation within the vowel.

3. Results and Discussion

3.1. Sustained /l/: Vocal tract models

3D vocal tract models capturing airflow in sustained lateral production revealed that three speakers of AusE produced sustained /l/ with a central occlusion and two lateral channels (Fig. 3). Lateral channel size varied between speakers with Speaker 144 exhibiting narrower channels (Fig. 3b). Interspeaker variation in lateral channel size in AusE /l/ is consistent with findings on interspeaker variation in American English, in which four speakers produced lateral channels with areas varying from 0.05 cm^2 to 0.5 cm^2 [10].

Speaker 141 exhibited asymmetric channels with narrow right and wider left channel (Fig. 3a), consistent with observations in American and Australian English. In American English, one out of four speakers exhibited asymmetric lateral channel formation, a similar ratio to the current dataset [10]. In AusE, six out of six speakers consistently demonstrated asymmetric lateral channel formation, such that right-handed speakers produced right-dominant lateral channels and left-handed speakers produced left-dominant lateral channels [9].

3.2. /l/ in three vowel contexts

Articulatory-acoustic characteristics of intervocalic /l/ exhibited considerable coarticulatory variation between vowel contexts (Figs. 4–5). Midsagittal images revealed that coronal place of articulation varied in anteriority with vowel frontness: dental-alveolar in /iːliː/, alveolar in /ɐːlɐː/, and post-alveolar in /oːloː/ (Fig. 4). Tongue dorsum articulation varied in frontness and height according to vowel context and between speakers. Tongue dorsum showed coarticulatory fronting in /iːliː/ for two speakers (141, 144), and coarticulatory resistance for one (152). Tongue dorsum showed coarticulatory lowering for two speakers in /ɐːlɐː/ (141 and 152), and resistance for one (144). Tongue dorsum was backed in /oːloː/ for all speakers, with Speaker 141 backing the dorsum in the back vowel context only (Fig. 4).

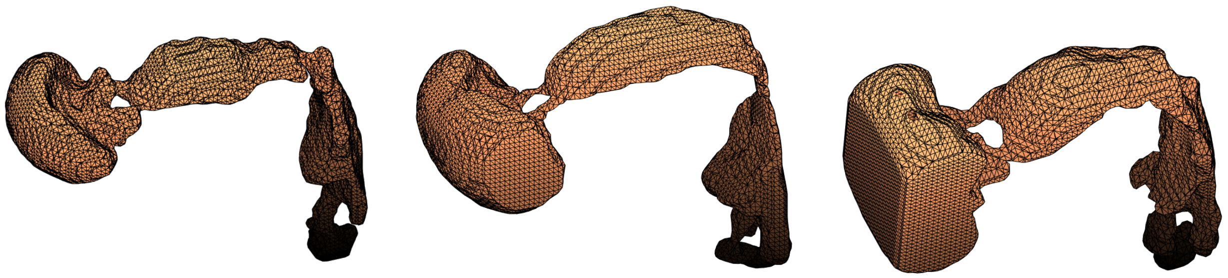
Formant characteristics showed the coarticulatory variation expected, based on the articulatory data. F1 showed an inverse relationship with tongue body- and vowel height and F2 correlated with tongue body- and vowel frontness (Figs. 4-5). F2 appeared to be impacted more strongly by the high front vowel compared to the other vowels, despite the inconsistent tongue body fronting across speakers in this context (Figs. 4–5).

The dataset is consistent with active as well as with passive lateral channel formation, suggesting inter- and intra-speaker variation. Evidence for coarticulatory fronting demonstrated by Speakers 141 and 144 is at odds with consistent tongue elongation, so it appears unlikely that lateral channels are formed passively in front vowel contexts for all speakers, consistent with active lateralisation [9,15]. Simultaneous tongue tip closure and dorsum backing observed in the back vowel context, as well as 144's low and 152's front vowel context may lead to passive lateral channel formation [4].

3.3. Intensity, lateralisation, and coronal constriction

Lateral intensity was reduced relative to adjacent vowels for all speakers and vowel contexts; however, inter- and intraspeaker variation was present (Figs. 2b, 5). Intensity reduction was larger relative to /ɐː/ and /oː/ than to /iː/, potentially due to the overall lower intensity of the high vowel. Speaker 152 produced larger intensity reduction than speakers 141 and 144, while Speaker 141 produced noticeably asymmetric channels (Figs. 3, 5). Asymmetric lateral channels can contribute anti-resonances to the spectrum, and thus reduce intensity in the higher frequency ranges, whereas anti-resonances created by symmetric lateral channels at the same frequency range cancel each other out without lowering intensity [12, 13]. However, no clear link was evident between lateral channel asymmetry and intensity reduction in this dataset. Insights into lateral channel asymmetry offered by the current dataset is limited by the error associated with segmentation of anatomical and airway features in volumetric data at this level of resolution.

Intensity reduction may also be attributed to the loss of energy in lower frequency ranges as the coronal constriction narrows [13]. However, the coronal constriction in this dataset showed considerable inter- and intra-speaker variation and was imperfectly aligned with the intensity dip (Fig. 5). Only Speaker 144 achieved tongue tip closure consistently prior to reaching the intensity minima (Fig. 5). In contrast, Speakers



(a) *Speaker 141* (b) *Speaker 144* (c) *Speaker 152*

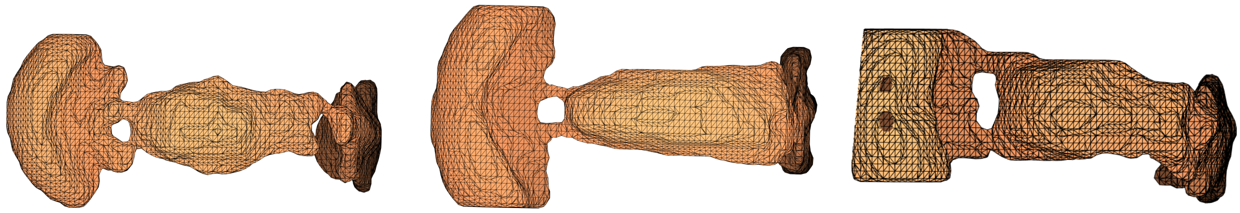
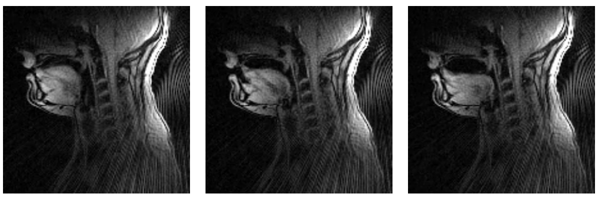


Figure 3: *Three dimensional vocal tract configuration showing airflow during sustained /lː/ for three speakers (L-R: 141, 144, 152).* Tract volume viewed from L. superior anterior perspective (top) and superior perspective (bottom). Right lateral channel toward the top and left toward the bottom of the figure. Central occlusion marked by lack of airflow. Anterior part of volume extends beyond lips.



(a) */iːliː/ (Rep. 1)* (b) */ɐːlɐː/ (Rep. 1)* (c) */oːloː/ (Rep. 1)*

Figure 4: *Coarticulatory impact of corner vowels on midsagittal /l/ articulation.* Speaker 141, L-to-R: /iːliː/, /ɐːlɐː/, /oːloː/.

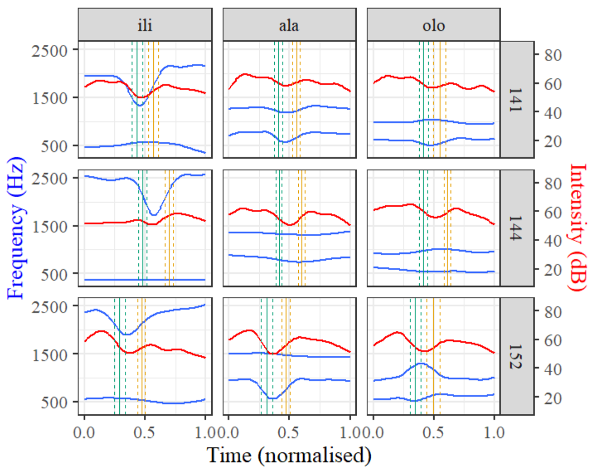


Figure 5: *Formant trajectories (left y-axis, blue) and intensity contours (right y-axis, red), aligned with coronal closure onset (mean and sd.vertical solid and dashed green lines) and offset (mean and sd.vertical solid and dashed orange lines).* Five in-scanner repetitions of (L to R): /iːliː/, /ɐːlɐː/, /oːloː/ per speaker (rows). Formant trajectories and intensity contours (thin lines) smoothed using GAMs (thick lines, 95% CI not shown).

141 and 152 achieved tip closure at the minima, suggesting delayed tip gesture relative to the acoustic onset of /l/ (Fig. 5). Imperfect alignment of tongue tip constriction and reduced intensity observed in this dataset is consistent with the intensity dip cueing lateralisation rather than tip closure.

In many varieties of English, including AusE, /l/ exhibits positional allophony [17]. Clear onset [l] is short, articulated with a simultaneous tongue tip and dorsum gesture, with the latter susceptible to anticipatory coarticulation [29–31]. Dark coda [ɫ] is long, articulated with a dorsum gesture preceding the tip gesture, with the dorsum gesture demonstrating coarticulatory resistance [29–31]. Laterals produced by Speaker 152 had a delayed tip gesture and stable dorsum backing across vowel contexts, potentially consistent with this speaker producing intervocalic dark [ɫ]. Intervocalic [ɫ] is attributed to the participant being prompted to sustain /VlV/ nonwords for 1.5 s., resulting in longer, and thus darker lateral production.

4. Conclusion

The multimodal dataset presented here provided new insight on lateral channel formation. Volumetric MRI allowed for 3D modeling of sustained /l/, revealing lateral channel formation for the first time in AusE. Consistent with other varieties of English, lateral channels showed interspeaker variation with one speaker producing asymmetric channels. Midsagittal articulatory data revealed inconsistent tongue elongation and potential variation between clear and dark /l/ in intervocalic position. Laterals produced in the front vowel context were less likely to be articulated with tongue elongation, thus providing evidence for active lateralisation, whereas laterals produced with the dark allophone were more likely to be elongated, providing evidence for passive lateral channel formation. Both clear and dark laterals were characterised by intensity reduction potentially aligned with lateral channel formation rather than tongue tip closure, allowing for the use of intensity drop as a cue to segmentation. Future work will examine intensity reduction at different frequency ranges to disentangle intensity reduction attributed to lateral channel formation and tongue tip closure.

5. Acknowledgment

This work was supported by Australian Research Council Discovery Grant DP220102933 and approved by the Sydney Children's Hospitals Network Human Research Ethics Committee (approval number: 2022/ETH00752).

6. References

[1] S. Giles and K. Moll, "Cinefluorographic study of selected allophones of English /l/," *Phonetica*, vol. 31, no. 3-4, p. 206–227, 1975.

[2] M. Stone and A. Lundberg, "Three-dimensional tongue surface shapes of English consonants and vowels," *The Journal of the Acoustical Society of America*, vol. 99, no. 6, p. 3728–3737, 1996.

[3] P. Ladefoged and I. Maddieson, *The sounds of the world's languages*. Oxford, UK; Cambridge, Mass.: Blackwell, 1996.

[4] C. P. Browman and L. M. Goldstein, "Gestural syllable position effects in American English," in *Producing speech: contemporary issues*, F. Bell-Berti and J. L. Raphael, Eds. New York: AIP Press, 1995.

[5] G. Fant, *Acoustic Theory of Speech Production*. The Hague: Mouton, 1960.

[6] K. N. Stevens, *Acoustic Phonetics*. Cambridge, Mass; London: MIT Press, 1998.

[7] K. Johnson, *Acoustic and auditory phonetics*. Blackwell, 2003.

[8] M. Proctor, "Towards a gestural characterisation of liquids: Evidence form Spanish and Russian," *Laboratory Phonology*, vol. 2, p. 451–485, 2011.

[9] J. Ying, J. A. Shaw, C. Carignan, M. Proctor, D. Derrick, and C. T. Best, "Evidence for active control of tongue lateralization in Australian English /l/," *Journal of Phonetics*, vol. 86, p. 101039, 2021.

[10] S. S. Narayanan, A. A. Alwan, and K. Haker, "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. part I. The laterals," *The Journal of the Acoustical Society of America*, vol. 101, no. 2, p. 1064–1077, 1997.

[11] S. Charles and S. M. Lulich, "Articulatory-acoustic relations in the production of alveolar and palatal lateral sounds in Brazilian Portuguese," *The journal of the Acoustical Society of America*, vol. 145, no. 6, pp. 3269–3288, 2019.

[12] Z. Zhang and C. Y. Espy-Wilson, "A vocal-tract model of American English /l," *The Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1274–1280, 2004.

[13] M. Tabain, A. Butcher, G. Breen, and R. Beare, "An acoustic study of multiple lateral consonants in three Central Australian languages," *The Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. 361–372, 2016.

[14] S. Charles, "Articulation and acoustics of lateral speech sounds," Ph.D. dissertation, Indiana University, 2022.

[15] T. Szalay, M. Proctor, A. Gully, T. Piyadasa, C. Jin, D. Waddington, N. Sanaei, Y. Yue, S. Foster, and K. Ballard, "Lateral articulation across vowel contexts: insights from a Magnetic Resonance Imaging case study," in *Proceedings of the 19th Australasian International Conference on Speech Science and Technology*, 2024.

[16] M. Joos, *Acoustic phonetics*. JSTOR, 1948, vol. 24, no. 2.

[17] F. Cox and J. Fletcher, *Australian English pronunciation and transcription*. Cambridge University Press, 2017.

[18] A. J. Kennerley, D. A. Mitchell, A. Sebald, and I. Watson, "Real-time magnetic resonance imaging: mechanics of oral and facial function," *British Journal of Oral and Maxillofacial Surgery*, vol. 60, no. 5, pp. 596–603, 2022.

[19] Optoacoustics Ltd., "FOMRI-II version 2.2," 2007.

[20] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[21] M. I. Proctor, D. Bone, and S. S. Narayanan, "Rapid semi-automatic segmentation of real-time Magnetic Resonance Images for parametric vocal tract analysis," in *Interspeech*, Makuhari, 26-30 Sept. 2010, pp. 1576–1579.

[22] F. Schiel, "Automatic Phonetic Transcription of Non-Prompted Speech," in *Proc. 14th Intl. Congress of Phonetic Sciences*, J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey, Eds., San Francisco, CA, USA, 1999, p. 607–610.

[23] ——, "A statistical model for predicting pronunciation." in *ICPhS*, 2015.

[24] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, p. 326–347, 2017.

[25] P. Boersma and D. Weenink, "Praat 6.4.05." 2024. [Online]. Available: http://www.fon.hum.uva.nl/praat/

[26] T. Szalay, T. Benders, F. Cox, and M. Proctor, "Vowel merger in Australian English lateral-final rimes: /æo-æ/," in *Proc. 18th Australasian Intl. Conf. on Speech Science and Technology*, R. Billington, Ed., 2022.

[27] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2024. [Online]. Available: https://www.R-project.org/

[28] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. [Online]. Available: https://ggplot2.tidyverse.org

[29] R. Sproat and O. Fujimura, "Allophonic variation in English /l/ and its implications for phonetic implementation," *Journal of phonetics*, vol. 21, no. 3, pp. 291–311, 1993.

[30] M. Proctor and R. Walker, "Articulatory bases of sonority in English liquids," *The sonority controversy*, vol. 18, p. 289, 2012.

[31] M. Proctor, R. Walker, C. Smith, T. Szalay, L. Goldstein, and S. Narayanan, "Articulatory characterization of English liquid-final rimes," *Journal of Phonetics*, vol. 77, p. 100921, 2019.