Lateral articulation across vowel contexts: insights from a Magnetic Resonance Imaging case study

Tünde Szalay[1], Michael Proctor[1], Amelia Gully[2], Tharinda Piyadasa[3], Craig Jin[3]
David Waddington[4], Naeim Sanaei[5], Yaoyao Yue[3], Sheryl Foster[4,5], Kirrie Ballard[6]

[1]Department of Linguistics, Macquarie University, Australia
[2]Department of Language and Linguistic Science, University of York, UK
[3]School of Electrical and Computer Engineering, University of Sydney, Australia
[4]Image X Institute, University of Sydney, Australia
[5]Radiology Department, Westmead Hospital, Australia
[6]Discipline of Speech Pathology, University of Sydney, Australia

tunde.szalay@mq.edu.au

Abstract

The goals of lateral production are complex and imperfectly understood, partly because of the limitations of existing data. Structural Magnetic Resonance Imaging provides rich information about details of lateral production not available using other methods. /l/-articulation by a British English speaker was examined in three vowel contexts using real-time and volumetric Magnetic Resonance Imaging. Onset of lateralisation was characterized acoustically by decreased intensity and development of anti-formants, independent of the degree of tongue dorsum retraction and lingual elongation in different coarticulatory contexts. These patterns suggest that, for this speaker, active lateral channel formation is a primary goal of clear-/l/ production.

Index Terms: liquids, approximants, articulatory-acoustic relationships, coarticulation, goals of production, rtMRI

1. Introduction

English /l/ is a multigestural segment prototypically produced with a central alveolar closure, dorsal retraction and lowering, and lateral channel formation [1, 2]. Lateral approximants are characterised by complex intergestural and articulatory-acoustic relationships [3, 4, 5]. Lateral channels may form passively when the tongue is elongated through simultaneous tongue tip fronting to achieve alveolar closure and dorsal retraction, as is typically observed in dark [ɫ] [6]. 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 that there may be active control of lateralisation [7, 8]. Many details of lateral production are still not well understood, in part due to the limitations of methods used to study the configuration of the vocal tract.

Acoustic data offer important insights into lateral production, as /l/ typically shows three distinct formants below 5 kHz [3, 5, 9]. The low F1 ($\sim$250–500 Hz) is associated with a Helmholtz resonance between the relatively large back cavity volume and the oral constriction space [3, 5]. F1 increases when the oral constriction is reduced, contributing to the higher F1 in dark [ɫ] produced with a weakened coronal contact [3, 10]. F2 ($\sim$1.2–1.5 kHz) is associated with the back cavity, such that retracting or raising the tongue dorsum increases back cavity length and lowers F2 in dark [ɫ] [3]. Lateralized airflow can give rise to spectral zeros whose properties depend on the length and asymmetry of the channels; anti-resonances >3 kHz can result when a pocket of air above the tongue forms a side branch to the primary lateralized airway [9, 11]. Anti-resonances raise the 3rd formant, and a high F3 well separated from F2 is one of the defining acoustic features of lateral approximants [5, 11, 12].

Lateral production has been studied using sustained /l/ [3, 4] and in specific vowel contexts [5, 10], yet dorsal posture – a key gesture affecting tongue elongation and F2 – varies with vowel-context, assuming a similar articulatory target to that of adjacent vowels [13]. In American English, onset [l] is coarticulated more strongly with the vowel than coda [ɫ] [13], predicting considerable coarticulatory F2 variation; however, Catalan /lˠ/ is articulated with a lower dorsum between low vowels, and is produced with a relatively stable F2 across vowel contexts [14].

To further examine these relationships, we analyzed time-aligned articulatory and acoustic data in a single speaker study of Standard Southern British English (SSBE) /l/ produced in three vowel contexts, using real-time (rtMRI) and volumetric magnetic resonance imaging (MRI). Our aims are to (1) identify articulatory and acoustic /l/ targets; (2) describe the coarticulatory influences of vowel context on /l/; and (3) link the articulatory changes caused by vowel context to acoustic changes.

2. Methods

Data were collected during the pilot phase of a larger project examining development of speech motor control in adolescents. An adult female L1 speaker of SSBE (Author 3) produced intervocalic laterals in a series of speech tasks recorded out of and inside an MRI scanner. Laterals were elicited between three corner vowels: high front /iː/, low /ɑː/, and high back /uː/. Each token was recorded once in a quiet room with a Glottal Enterprises EG2-PCX2 digital speech recorder to familiarize the participant with the experimental materials. The same utterances were later recorded three times during a rtMRI scan, and additionally as sustained lateral productions during a volumetric MRI scan. A total of 3 (vowel contexts) $\times$ (1 pre-scan + 3 rtMRI) + 1 (volumetric MRI) = 13 laterals were included in the analysis.

2.1. Data acquisition

MRI data were acquired at Westmead Hospital (Sydney, New South Wales), on a Siemens Magnetom Prisma 3T scanner with a 64-channel head/neck receiver array coil. 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 [15]. Audio was recorded concurrently in-scanner at 16 kHz using an Opto-acoustics FOMRI-III ceramic noise-canceling microphone designed for MRI environments [16]. rtMRI data were reconstructed in Matlab into midsagittal videos with a pixel resolution of $0.83\ \text{mm}^2$, encoded as 72 frames per second MP4 files. Audio and video were time-aligned during postprocessing and video reconstruction based on visual inspection of the audio signal and the video frames.

3D configuration of the vocal tract during sustained (7.6 s) 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 $160 \times 160 \times 32$ px over a $256 \times 256 \times 64$ mm field of view centred on the pharynx: a voxel resolution of $1.6 \times 1.6 \times 2.0$ mm.

2.2. Phonetic data analysis

rtMRI videos and time-aligned in-scanner audio recordings were analyzed using a Matlab-based custom graphical interface. Image frames were identified corresponding to articulatory target postures for pre- and post-lateral vowels, and lateral coronal closure, target, and coronal release (Figs. 1, 3, 6).

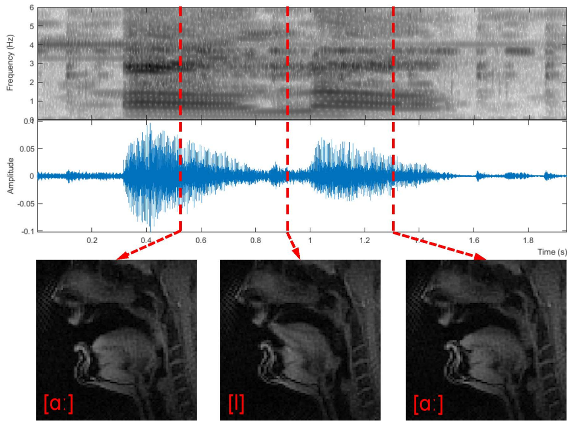
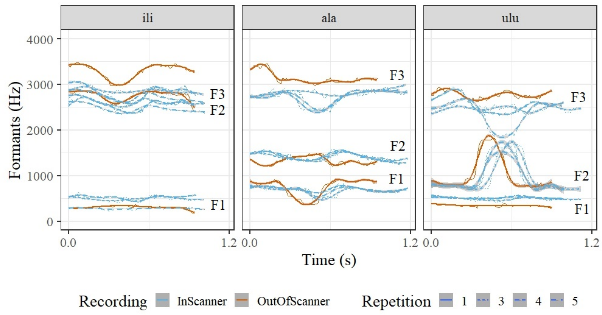


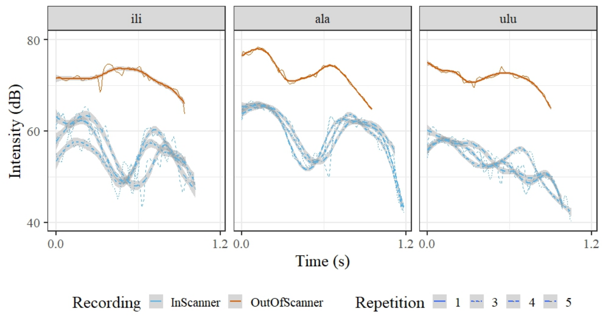
Figure 1: *Intervocalic lateral production, low vowel context.* Spectrogram and waveform of noise-cancelled in-scanner recording of /ɑlɑ/, time-aligned with rtMRI frames captured at vowel and lateral lingual target postures.

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. 6, bottom centre L). The lateral target was located at the centre frame of the interval over which contact was maintained between the TT and alveolar ridge (Fig. 1, bottom centre). Lateral coronal release was located at the first frame when a gap between the TT and alveolar ridge was first observed after closure (Fig. 6, bottom centre R). Coronal closure was achieved in every token, showing that none of the laterals were vocalised. Lingual target postures were identified in all tokens despite the motion blur in some frames, as slow and hyperarticulated speech yielded visible sustained lingual targets.

Audio recordings were force-aligned using MAUS to locate segment boundaries which were then hand-corrected [17, 18, 19]. Formant trajectories were estimated automatically and corrected manually in Praat [20]. Formant frequencies were estimated every 10 ms over a 40 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. Five formants were tracked up to a 5.5 kHz ceiling for tokens with higher F2 values, and up to 5 kHz for lower F2 values, then corrected manually [21]. At each timepoint where formants were estimated, intensity values were estimated with a pitch floor of 100 Hz. Characteristic formant trajectories and intensity contours for laterals in each vowel context were generated by fitting a Generalised Additive Model (GAM) to the set of time series for each experimental item (Fig. 2).



(a) *F1, F2, and F3 trajectories*



(b) *Intensity contours*

Figure 2: *Formant trajectories (a) and intensity contours (b).* Out-of-scanner (red) and 3 in-scanner (blue) repetitions of (L-to-R): [iːliː], [ɑːlɑː], [uːluː]. Individual repetition timeseries (thin lines) fitted with GAMs (thick lines + grey s.d.).

2.3. Acoustic data validation

Formant and intensity values estimated in in-scanner recordings were validated against acoustic measures estimated in out-of-scanner recordings. Formants generally tracked poorly in in-scanner recordings due to scanner noise and signal processing involved in noise-reduction. F1 and F2 estimates from in- and out-of-scanner recordings aligned most closely, but F3 estimates were consistently lower for in-scanner recordings, compared to out-of-scanner equivalents, and showed larger discrepancies between repetitions (Fig. 2a). Manual correction of 3rd formant trajectories in these data was determined to be too unreliable, so F3 was not analysed for in-scanner recordings.

Utterance durations were consistently longer for in-scanner recordings, primarily due to lengthening of pre-lateral vowels, but the same general patterns can be observed as the speaker hyperarticulated during the rtMRI scan (Fig. 2). Vowel and lateral intensity was consistently higher in out-of-scanner recordings compared to in-scanner recordings (Fig. 2b). Intervocalic laterals showed an intensity dip relative to the initial vowel in all tokens other than out-of-scanner /ili/ (Fig. 2b), which is attributed to speech variation, rather than vowel or recording environment effects. Laterals, however, only showed lower intensity relative to the final vowel in [ili] and /ɑlɑ/, but not in /ulu/, due to /i/ and /ɑ/ having higher intensity than /u/ (Fig. 2b).

3. Results and Discussion

Complete midsagittal occlusion in the dental-alveolar region was observed in all lateral tokens (Fig. 3); no vocalized /l/ was produced in these data. Achieving tongue tip contact in intervocalic /l/ is consistent with the lack of undershoot in this position in American- and British English [1, 22]. Extended /l/ duration could also contribute to achieving alveolar closure by providing enough time for the tongue tip to reach its target [23, 24].

(a) [iːliː] (Rep. 4) (b) [ɑːlɑː] (Rep. 4) (c) [uːluː] (Rep. 4)

Figure 3: *Midsagittal /l/ articulation at TT target in three vowel contexts.* L-to-R: [iːliː], [ɑːlɑː], [uːluː].

Volumetric image data show the 3D vocal tract configurations used to produce laterals in each vowel context, including key details of /l/ articulation beyond the midsagittal plane (Fig. 4). The central occlusion formed by the TT against the alveolar ridge can be seen anterior to the mid-oral cavity. On either side of the occlusion, narrow lateral channels connect the mid-oral airway to the anterior part of the vocal tract formed by the sublingual cavity. The precise geometry of the lateral channels cannot be determined from this volume because these regions will also include some dentition (teeth do not image in MRI); yet, the data reveal two largely symmetrical lateral channels and a complete central alveolar occlusion.

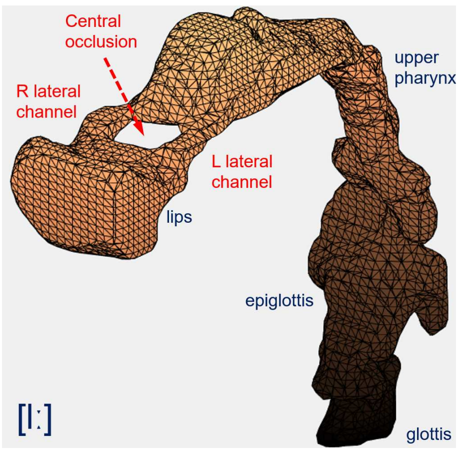


Figure 4: *Three dimensional vocal tract configuration during sustained [lː].* Tract volume viewed from L. superior anterior perspective. Anterior part of volume extends beyond lips.

3.1. Acoustic characterization of intervocalic laterals

Laterals were elicited in intervocalic environments, where they were acoustically delineated by a drop in intensity (Fig. 2b), formant transitions specific to the vowel context (Fig. 7), and appearance of anti-formants. Intensity drop cued lateral onset and offset in the [iː] and [ɑː] contexts and lateral onset in the [uː] context (Fig. 2b). A spectrogram generated from the pre-scan recording of [ɑːlɑː] (Fig. 5) reveals a prominent antiresonance centred at 3.7 kHz throughout the lateral interval (690 to 920 ms). In Fant's model, both reduced intensity and antiresonance would arise from a side branch of length ∼23 mm [9, 25, 26]. Although the precise length of the supralingual air pocket cannot be determined because of uncertainties associated with dentition, the frequencies of the main antiformants observed in these spectra are broadly consistent with Fant's acoustic model applied to the vocal tract configurations revealed by the imaging data. Anti-resonances in a similar region were observed in American English /l/, while intensity drop characterises Turkish and Brazilian Portuguese laterals [4, 27].

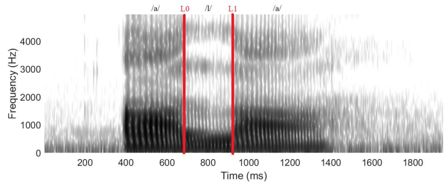


Figure 5: *Spectrogram of [ɑːlɑː] (out-of-scanner): 6 ms Kaiser windows, 2 ms overlap, 1024 pt FFT.* L0: beginning of lateral; L1: end of lateral. Primary anti-formant centred at 3.7 kHz.

3.2. Vocalic influences on lateral production

Imaging data reveal large coarticulatory influences of vowel context on lateral production. Coronal place of articulation varies in anteriority with vowel frontness: dental-alveolar for [iːliː] (Fig. 3a) and alveolar for [ɑːlɑː] (Fig. 3b), both produced with an apical TT gesture. In [uːluː], the midsagittal constriction occurs at a more retracted post-alveolar target through sub-laminal TT closure and a more retroflexed coronal gesture (Fig. 3c). The dorsum is raised and fronted in the high-front vowel context (Fig. 3a), lowered in the low vowel context

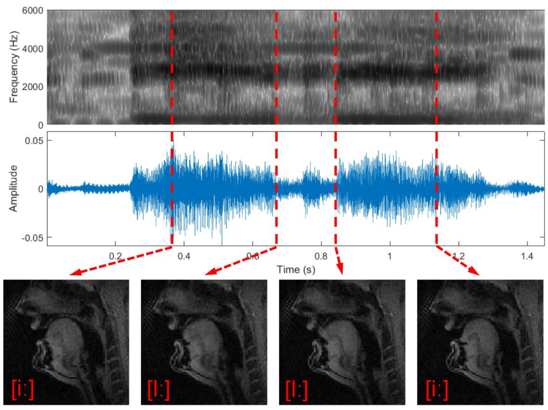


Figure 6: *Lateral dynamics in the front vowel context.* Spectrogram and waveform of noise-cancelled in-scanner recording of /ili/, time-aligned with MRI frames captured at (L to R): pre-/l/ vowel target, /l/ onset, /l/ target, post-/l/ vowel target.

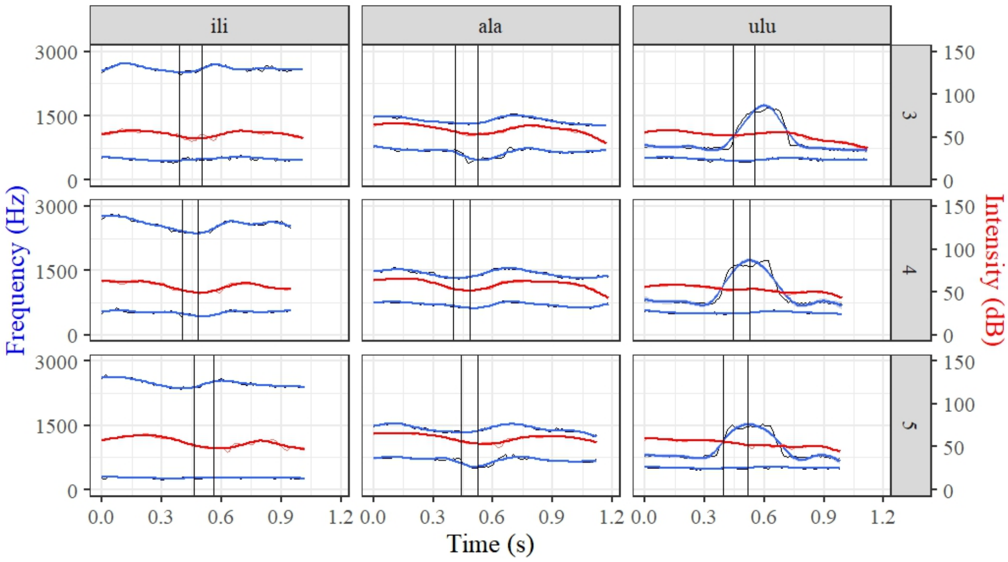


Figure 7: *Formant trajectories (left y-axis, blue) and intensity contours (right y-axis, red), aligned with onset and offset of coronal closure (vertical black lines).* Three in-scanner repetitions of (L to R): /ili/, /ala/, /ulu/. Top to bottom: Repetitions 1 to 3. Formant trajectories and intensity contours (thin lines) smoothed using GAMs (thick lines, std. dev. in grey).

(Fig. 3b), and high and back in the back vowel context (Fig. 3c).

F1 in intervocalic laterals ranged from 350 Hz in high vowel contexts to 750 Hz between low vowels (Fig. 7). F1 trajectories are relatively stable throughout [iːliː] and [uːluː], consistent with the stability in tongue height observed in the corresponding image sequences of these utterances (Fig. 6). F1 was higher throughout [ɑːlɑː] utterances, lowering in the transition from the initial vowel to the lateral, then rising to a peak of ∼750 Hz at lateral TT release (Fig. 7) – higher than F1 values previously reported for sustained laterals (350–450 Hz) [3]. The raised F1 in [ɑːlɑː] is consistent with the pervasive tongue body lowering observed in the corresponding midsagittal image sequences of laterals in low vowel contexts (Fig. 1).

A wide range of F2 frequencies are seen in these data. Overall, F2 trajectories show the expected correlations with tongue frontness. In non-front vowel contexts, lateral F2 ranged between 1450–1550 Hz. F2 was most stable throughout [ɑːlɑː] utterances, lowering slightly before TT closure and peaking after TT release, following a similar trajectory to F1 (Fig. 7). [iːliː] was also characterized by F2 lowering in the pre-lateral vowel, however, F2 did not reach the same target in the lateral, remaining much higher (>2 kHz) than in the other vowel contexts (Fig. 7). Between back vowels, F2 rose sharply to peak at ∼1500 Hz at the point of TT closure, then relowering after TT release to the F2∼750 Hz of context [uː] (Fig. 7). Variation in lateral formants is consistent with lateral formants not being sufficient for distinguishing laterals from other segments in SSBE, similarly to other languages (e.g., Turkish, Brazilian Portuguese, Central Australian languages) [28, 27].

These formant trajectories are consistent with the coarticulatory patterns observed in imaging data. Lateral F2 is affected more strongly by adjacent /iː/, compared to /ɑː/ and /uː/. Midsagittal images reveal that the tongue body is more advanced throughout [iːliː] compared to the other vowel contexts, and does not retract as much at the lateral target (Figs. 1, 6). Stronger coarticulatory influences of front vowels, and palatals more generally, have also been demonstrated in Catalan and other languages [29]. As a result of this coarticulation, this speaker's laterals are not consistently produced with an elongated tongue, so it appears unlikely that lateral channels are formed passively in front vowel contexts [6]. These data – although limited – lend more support for models proposing active lateral channel formation in /l/ production [7, 8]. The same patterns of production might also arise if tongue blade width were an active parameter of control [30], allowing side channels to form around a narrowed TT central constriction.

4. Conclusions and future research

The dataset demonstrates the value of multi-modal data in the phonetic characterization of complex segments. Understanding dynamic patterns of articulation beyond the midsagittal plane and their acoustic consequences is particularly important for lateral approximants. This speaker consistently produced hyperarticulated intervocalic laterals with central TT closure and formation of lateral channels, characterized acoustically by antiformant(s) and reduced intensity relative to vowels. Tongue body anteriority during lateral production and formant trajectories – especially F2 – were strongly influenced by vowel context, and may provide less consistent cues to lateralization. Inconsistent tongue body retraction across vowel contexts suggests that active lateral channel formation, rather than lingual elongation, is a primary goal of /l/ production for this speaker.

The data are limited in scope, as only a small number of lateral exemplars from a single speaker of SSBE have been analyzed, and the speech is hyperarticulated due to the nature of the task and the unusual environment in which it was produced. More detailed analysis of the geometry and dynamics of lateral channel formation in different phonological environments is required to better understand how lateralization is achieved, and how /l/ can be characterized in articulatory and acoustic domains. Dynamic imaging in the coronal plane and modelling of dentition in MRI data will help inform these issues. Robust tracking of F3 in in-scanner recordings will be important to better characterize the acoustic dynamics of lateral production.

5. Acknowledgements

This work was supported by Australian Research Council Discovery Grant DP220102933.

6. References

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

[2] M. Stone, A. Faber, L. J. Raphael, and T. H. Shawker, "Cross-sectional tongue shape and linguopalatal contact patterns in [s],[ʃ], and [l]," *Journal of Phonetics*, vol. 20, no. 2, pp. 253–270, 1992.

[3] 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, pp. 1064–1077, 1997.

[4] X. Zhou, C. Y. Espy-Wilson, M. Tiede, and S. Boyce, "An MRI-based articulatory and acoustic study of lateral sound in American English," in *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 4182–4185.

[5] 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.

[6] C. P. Browman and L. M. Goldstein, "Gestural syllable position effects in American English," in *Producing speech: contemporary issues (for Katherine Safford Harris)*, F. Bell-Berti and L. J. Raphael, Eds. New York: AIP Press, 1995, pp. 19–34.

[7] M. Proctor, "Towards a gestural characterization of liquids: Evidence from Spanish and Russian," *Laboratory Phonology*, vol. 2, no. 2, pp. 451–485, 2011.

[8] 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.

[9] G. Fant, *Acoustic theory of speech production, with calculations based on X-ray studies of Russian articulations*. s'Gravenhage: Mouton, 1960.

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

[11] K. N. Stevens, *Acoustic Phonetics*. Cambridge, Mass; London: MIT Press, 2000.

[12] C. Y. Espy-Wilson, "Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English," *The Journal of the Acoustical Society of America*, vol. 92, no. 2, pp. 736–757, 1992.

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

[14] D. Recasens, M. D. Pallarès, and J. Fontdevila, "Co-articulatory variability and articulatory–acoustic correlations for consonants," *International Journal of Language & Communication Disorders*, vol. 30, no. 2, pp. 203–213, 1995.

[15] 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.

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

[17] 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.

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

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

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

[21] 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.

[22] J. M. Scobbie and M. Pouplier, "The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final english/l," *Journal of Phonetics*, vol. 38, no. 2, pp. 240–259, 2010.

[23] E. Lawson and J. Stuart-Smith, "The effects of syllable and sentential position on the timing of lingual gestures in /l/ and /r/," in *International Congress of Phonetic Sciences (ICPhS 2019)*, Melbourne, Australia, 2019, pp. 547–551.

[24] P. Strycharczuk, J. M. Scobbie, A. Przewozny, C. Viollain, and S. Navarro, "Gestural delay and gestural reduction. articulatory variation in/l/-vocalisation in southern british english," *The corpus phonology of English: Multifocal analyses of variation*, pp. 9–29, 2020.

[25] 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.

[26] K. Johnson, *Acoustic and Auditory Phonetics*, ser. Acoustic and Auditory Phonetics. Malden, MA: Wiley-Blackwell, 2011.

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

[28] 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.

[29] D. Recasens and C. Rodríguez, "A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound," *Journal of Phonetics*, vol. 59, pp. 58–75, 2016.

[30] C. P. Browman and L. M. Goldstein, "Articulatory gestures as phonological units," *Phonology*, vol. 6, no. 2, pp. 201–251, 1989.