**Proceedings Paper:**

Proctor, Michael, Szalay, Tuende, Piyadasa, Tharinda et al. (6 more authors) (2025) Rhotic Articulation in Australian English:Insights from MRI. In: Proceedings of Interspeech 2025. Interspeech 2025, 17-21 Aug 2025 , NLD

# Rhotic Articulation in Australian English: Insights from MRI

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

[1]Macquarie University, Australia
[2]University of York, United Kingdom
[3]University of Sydney, Australia
[4]Westmead Hospital, Australia

michael.proctor@mq.edu.au

## Abstract

English rhotics are realized with rich allophony across speakers, contexts and varieties, but Australian English /ɹ/ has not previously been examined in detail. Rhotic approximants produced in three vowel contexts by four speakers of Australian English were captured using real-time and volumetric structural magnetic resonance imaging. /ɹ/ was articulated with bunched tongue postures by two speakers and more apical configurations by two speakers, but all rhotics were characterized by three coordinated gestures: tongue tip, tongue body and labial constrictions. These data shed new light on the complex goals of production of rhotic approximants beyond the midsagittal plane, and their realization and extent of variation in Australian English.

**Index Terms**: rhotic approximant, Australian English, speech production, MRI, liquid consonant

## 1. Introduction

Rhotic approximants have been the focus of a large body of research because of their great diversity, incompletely understood complexity, and special salience as speech sounds. The complexity and variability of this class of consonants presents many challenges for phonetic and phonological characterization.

Acoustically, a lowered third formant is a canonical characteristic of English /ɹ/ in many phonological environments [1, 2], but rhotic approximants show great phonetic variability across speakers and environments [3, 4], and the complex relationships between rhotic articulation and its acoustic consequences are imperfectly understood [5, 6].

American English /ɹ/ is produced with remarkable variability among speakers, and a focus of much research into rhotics has been characterizing the range of tongue configurations involved. Different taxonomies have been proposed to capture the main patterns of articulation shared across speakers, involving up to eight different categories [7, 8, 9]. Classifications have distinguished 'retroflex' vs. 'bunched', and 'tip-up' vs. 'tip-down' /ɹ/ in North American varieties and other Englishes with similar rhotic approximants [10, 11, 12].

Characterization of /ɹ/ allophony is complicated by many factors that interact with individual speaker variation, including sound change, register, prosodic and contextual influences, and inconsistent percepts of rhotic type. For example, Mielke et al.'s ultrasound study of 27 speakers of American English found that "two speakers used only retroflex /ɹ/, sixteen use only bunched /ɹ/, and nine use both /ɹ/ types, with idiosyncratic allophonic distributions". Furthermore, "these allophony patterns are covert, because the difference between bunched and retroflex /ɹ/ is not readily perceived by listeners" [13, p.101]. Another ultrasound study of 62 speakers of New Zealand English found that 25 speakers consistently produced tip-down variants, 12 consistently produced tip-up rhotics, and 25 speakers produced /ɹ/ with variable tongue shapes [12].

Important insights into properties of English /ɹ/ have been obtained through analysis of its consistent gestures. North American English rhotics have been studied using MRI [14, 15, 10, 16], X-ray microbeam [17, 18], ultrasound and video [19, 20, 13]. These and other studies have revealed that /ɹ/ is typically produced through the coordination of two lingual gestures, and an additional labial gesture in onset environments [21]. It remains to be seen whether Australian English /ɹ/ is characterized by the same patterns of gestural coordination. Australian English (AusE) is a non-rhotic variety [22], so for most speakers /ɹ/ does not occur in codas, but the phonetic similarities in onset and intervocalic environments suggest that AusE /ɹ/ involves similar goals of production to rhotics in other English varieties.

AusE /ɹ/ has previously been investigated using ultrasound tongue imaging (UTI). A study of six speakers from Sydney [23] found similar patterns of /ɹ/ production as in NZE [12]: four broad tongue shapes were observed word-initially before /iː-ɐː-oː/. UTI was also used to examine midsagittal articulation of onset /ɹ/ by six speakers from Sydney before the full range of AusE vowel contexts [24]. Three broad patterns of tongue shaping were observed, and although speakers differed in the type and degree of vocalic influence on rhotic posture, the data revealed two coordinated lingual gestures across all speakers and contexts: a mid pharyngeal tongue body gesture, and a coronal gesture realized at a speaker-specific place of articulation. These studies have provided an initial articulatory phonetic characterization of Australian English /ɹ/, but more data are needed to understand the goals of production and how the tongue is shaped by the constituent gestures within and beyond the midsagittal plane of the vocal tract.

The goal of this study is to begin to examine AusE /ɹ/ articulation in new detail using structural magnetic resonance imaging, to more comprehensively characterize this consonant phonetically and to shed more light on its goals of production. To the best of our knowledge, this is the first study of Australian English rhotics using MRI. By combining real-time and volumetric MRI data, we aim to provide richer insights into the articulation of these complex liquid consonants and individual speaker variation.

## 2. Methods

Data were collected as part of an ongoing project examining speech production in Australian English. Four adult native speakers (2 female, 2 male) of Australian English produced rhotic approximants in a series of speech tasks recorded out of and inside an MRI scanner. Female participants F21 and F22 were 21 and 22 years, and male participants M21 and M22 were 21 and 22 years old respectively. All participants were raised in Australia, completed their secondary education in Sydney, and had at least one parent who was born and raised in Australia.

### 2.1. Experimental Materials

Rhotics were elicited between three corner vowels: high front /iː/, low /ɐː/, and back /oː/, and in isolation as a sustained continuant [ɹːː]. Each token was recorded twice in a quiet room with a Glottal Enterprises EG2-PCX2 digital speech recorder to familiarize the participant with the experimental materials. Elicitation items were presented orthographically on a monitor and read aloud by the participant. Trials compromised by mispronunciations or atypical prosody were re-recorded. The same utterances were later recorded five times during a rtMRI scan, and the sustained rhotics twice more during a volumetric MRI scan. A total of 3 (vowel contexts) × (2 pre-scan + 5 rtMRI) + 4 (2 pre-scan + 2 volumetric MRI) = 25 rhotics were recorded for analysis from each participant.

### 2.2. Data acquisition

MR imaging was conducted at Westmead Hospital (Sydney, Australia), on a Siemens Magnetom Prisma 3T scanner with a 64-channel head/neck receiver array coil. Speakers lay supine in the scanner bore, reading elication materials presented on a screen visible through a mirror. Image data were acquired from an 8 mm slice aligned with the mid-sagittal plane of the upper airway, over a 280 × 280 mm field of view, using a 2D RF-spoiled, radially-encoded FLASH sequence [25].

Audio was recorded concurrently in-scanner at 16 kHz using an Opto-acoustics FOMRI-III ceramic noise-canceling microphone designed for MRI environments [26]. Speech audio and rtMRI data were time-aligned during postprocessing using synchronization signals saved with each in-scanner recording, and reconstructed into videos with a pixel resolution of 0.83 $mm^2$, encoded as 72 frames per second MP4 files.

3D configuration of the vocal tract during rhotic production was captured using volumetric imaging. Participants sustained [ɹːː] for 7.6 s, following timed prompts presented on the elicitation screen. Data were acquired over a 256 × 256 × 64 mm field of view centred on the pharynx, using a T1-weighted fast 3D gradient-echo sequence with a spatial resolution of 160 × 160 × 32 px, to resolve the structures of the entire upper airway with a voxel resolution of 1.6 × 1.6 × 2.0 mm.

### 2.3. Phonetic analysis

rtMRI videos and time-aligned in-scanner audio recordings were analyzed using a Matlab-based custom graphical interface [27]. Image frames corresponding to articulatory landmarks were located with reference to the time-aligned audio signal and spectrogram (Fig. 1).

Volumetric MRI data were analyzed using ITK-SNAP [28]. Lingual tissue boundaries were segmented iteratively across sagittal, coronal, and axial orientations to identify tongue configurations used by each speaker while articulating sustained rhotic approximants. The entire tongue mass was first segmented from the airway and surrounding tissue to validate the anatomical structures. To more effectively illustrate lingual articulation and allow comparison between speakers (Fig. 7), the inferior part of each tongue was then truncated by bisecting the volume with an oblique axial plane connecting the lowest point of the sublingual cavity to the junction between the tongue root and base of the epiglottis.

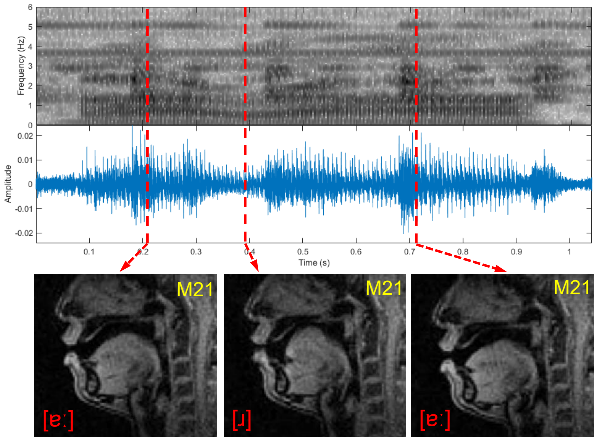


Figure 1: ***Identification of articulatory landmarks in real-time MRI data.*** *Top: audio and spectrogram of in-scanner speech recording (Speaker M21, utterance [ɐːɹɐː]); Bottom: time-aligned image frames extracted from rtMRI video.*

## 3. Results

### 3.1. Intervocalic Rhotic Production

Rhotic production in a low vowel context [ɐːɹɐː] is illustrated in Figs. 2–3. For each speaker, three frames captured using the real-time MRI sequence are shown, representing midsagittal tongue postures at (i) the pre-rhotic vowel target; (ii) the acoustic mid point of the rhotic; and (iii) the post-rhotic vowel target. Each speaker produced the [ɐːɹɐː] utterance five times with similar patterns of articulation; the images shown are taken from the first repetition.

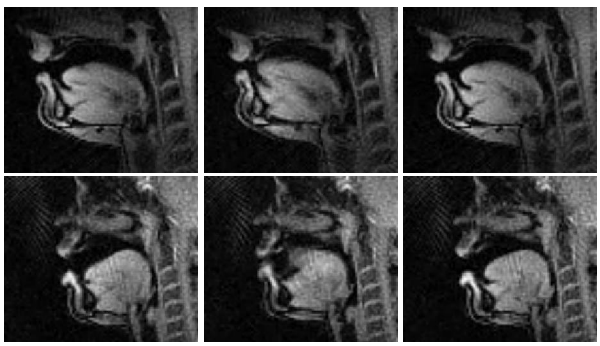

Figure 2: ***Rhotic production in a low vowel context* [ɐːɹɐː], *Speakers M22 (top row) and F21 (bottom row)***. *Left: pre-rhotic vowel posture [ɐː]; Centre: target rhotic posture [ɹ]; Right: post-rhotic vowel target [ɐː].*

In the [ɐː_ɐː] context, Speakers M22 and F21 articulate /ɹ/ with an apical coronal gesture. The tongue tip (TT) constriction location target is alveolar for Speaker M22, and post-alveolar for Speaker F21 (Fig. 2, centre column). In the same vocalic environment, Speakers M21 and F22 articulate /ɹ/ with a laminal coronal gesture at a palatal constriction location. The tongue shape for Speaker M21 at the rhotic target is characterized by a mid-lingual saddle, while Speaker F22 shows a globally bunched midsagittal posture for /ɹ/, resembling that of a high central vowel (Fig. 3, centre column).

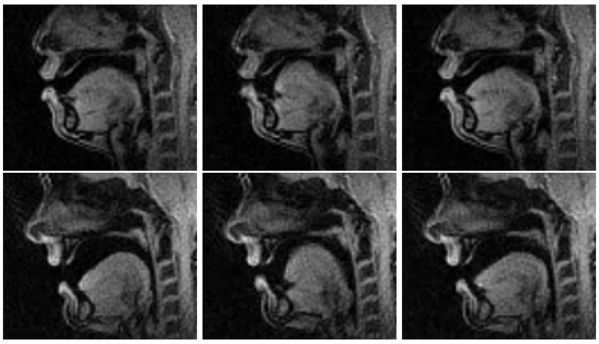

Figure 3: ***Rhotic production in a low vowel context* [ɐːɹɐː], *Speakers M21 (top row) and F22 (bottom row)***. *Left: pre-rhotic vowel posture [ɐː]; Centre: target rhotic posture [ɹ]; Right: post-rhotic vowel target [ɐː].*

Midsagittal [ɹ] articulation in a back vowel context is illustrated in Fig. 4. Three frames are shown for each speaker, revealing midsagittal lingual configurations at (i) initial [oː], (ii) [ɹ], and (iii) final [oː] targets. Speakers M22, F21 and F22 produced the [oːɹoː] utterance five times with similar patterns of articulation; images taken from the first repetition are shown here. Speaker M21 produced the final intervocalic rhotic with a different context vowel quality [ʉːɹʉː] which cannot be compared directly with back vowel contexts of the other three speakers.

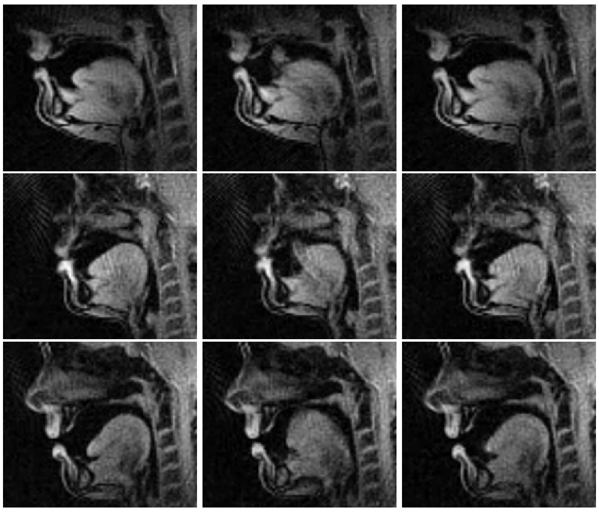

Figure 4: ***Rhotic production in a back vowel context* [oːɹoː], *Speakers M22 (top row), F21 (middle row) and F22 (bottom row)***. *Left: pre-rhotic vowel posture [oː]; Centre: target rhotic posture [ɹ]; Right: post-rhotic vowel target [oː].*

Rhotic articulation in the high-front vowel context is illustrated in Figs. 5–6. Three frames captured during production of the sequence [iːɹiː] are shown for each speaker, at (i) initial [iː] target; (ii) the acoustic mid point of [ɹ]; and (iii) the post-rhotic [iː] target. Each speaker produced the [iːɹiː] utterance five times with similar patterns of articulation; the images shown are taken from the first repetition.

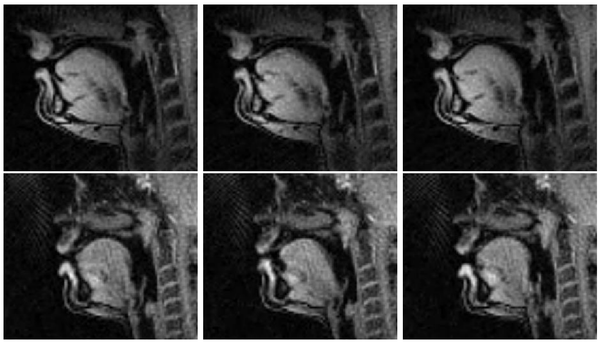

Figure 5: ***Rhotic production in a high-front vowel context* [iːɹiː], *Speakers M22 (top row) and F21 (bottom row)***. *Left: pre-rhotic vowel posture [iː]; Centre: target rhotic posture [ɹ]; Right: post-rhotic vowel target [iː].*

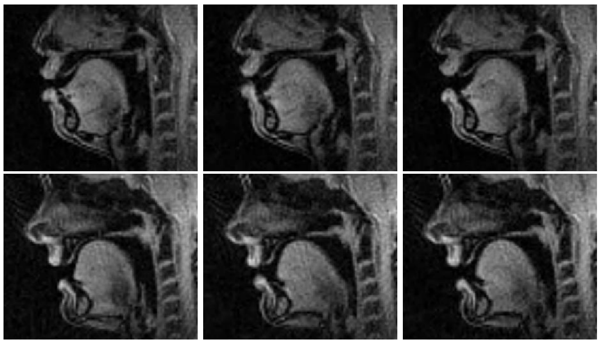

Figure 6: ***Rhotic production in a high-front vowel context* [iːɹiː], *Speakers M21 (top row) and F22 (bottom row)***. *Left: pre-rhotic vowel posture [iː]; Centre: target rhotic posture [ɹ]; Right: post-rhotic vowel target [iː].*

### 3.2. Sustained Rhotic Production

More detailed information about articulation of Australian English /ɹ/ is provided by the volumetric MRI data captured from the same speakers. Fig. 7 shows three dimensional lingual configurations adpoted by each speaker during sustained [ɹː] production. The superior surface of each tongue is shown from a viewpoint located above and behind the centre of the tongue, beyond the left cheek. The entire volume of each speaker's tongue was segmented to identify the complete lingual structure, but to allow clearer visualization and comparison, only the top part of each tongue is illustrated in Fig. 7.

## 4. Discussion

These data reveal two broad patterns of tongue shaping in Australian English /ɹː/, most clearly differentiated in the low vowel context. Rhotics are generally produced by Speakers M22 and F21 with a more clearly separated coronal articulatory component (Fig. 2), while Speakers M21 and F22 show a more globally bunched tongue posture at the rhotic target (Fig. 3). These patterns of individual speaker variation are compatible with previous typologies of /ɹː/ allophony [7, 10, 11], and the influence of vowel context in these data is broadly consistent with previous findings for New Zealand English that tip-up allophones are more commonly observed in back vowel contexts, and tip-down before high front vowels [12]. Beyond these general pat-

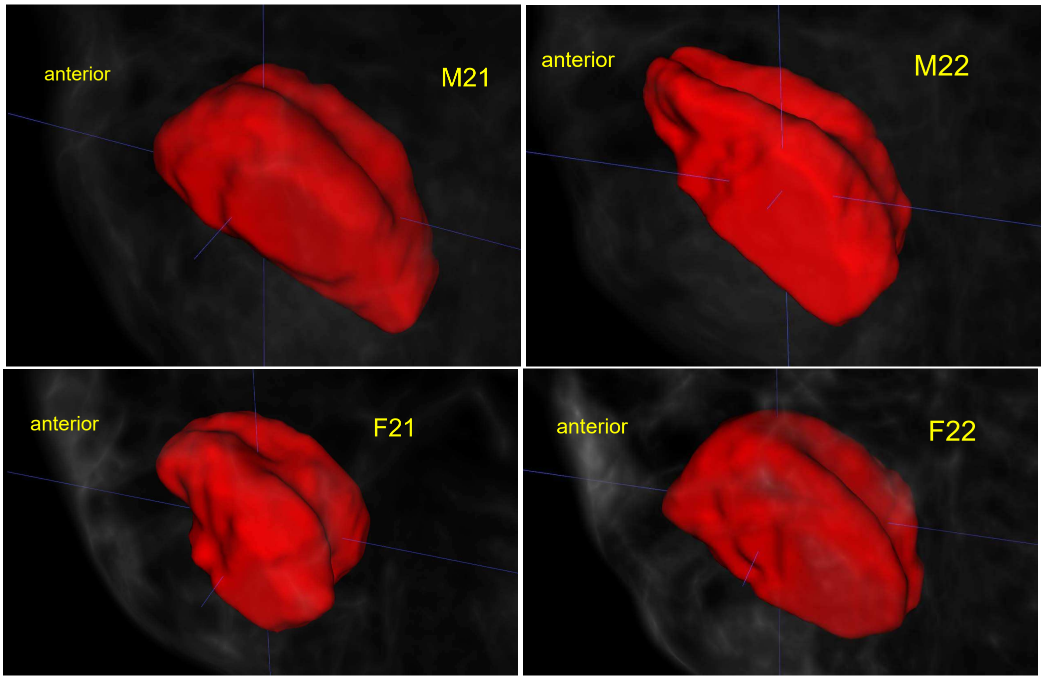


Figure 7: *Lingual configurations during sustained rhotic production. Top row: Speaker M21 (L) and M22 (R); Bottom row: Speaker F21 (L) and F22 (R). Superior tongue mass depicted (above oblique axial plane bisecting sublingual cavity and tongue root), visualized from left superior posterior viewpoint. Tongue tip oriented towards left of each image.*

terns, it is difficult to describe the full range of variation and details of production in terms of global tongue configuration. A better understanding of the goals of rhotic production can be obtained by comparing patterns of articulation across different vowel contexts to gain insights into gestural constituency and coordination in Australian English /ɹː/.

### 4.1. Gestural Characterization of AusE /ɹː/

All rhotics produced in low and back vowel contexts by these speakers demonstrate advancement and/or raising of some part of the front of the tongue towards some part of the anterior oral cavity. The type of coronal articulation and the location of the target varies: Speakers M22 and F21 use a more apical coronal gesture with the anterior tip of the tongue approximating the alveolar ridge or post-alveolar region (Figs. 2, 4), while Speakers M21 and F22 form a laminal constriction with the tongue body at a palatal target (Figs. 3, 4). Regardless of these differences, rhotics produced by all speakers involve a coronal approximation gesture.

After the tongue body is raised and advanced towards the palatal target of the initial vowel in [iːɹiː] sequences, retraction of the tongue body/root towards the rear pharyngeal wall can be observed before the tongue advances again during articulation of the final [iː] (Figs. 5 & 6). These patterns demonstrate that Australian English /ɹː/ involves a tongue body gesture with a pharyngeal target. The constriction location of the tongue body gesture appears be speaker-specific, but some degree of retraction of the tongue dorsum was observed in rhotics produced by all four speakers in [iːɹiː] utterances in these data.

Rhotics produced in unrounded vowel contexts [iːɹiː] and [ɐːɹɐː] by these speakers also showed labial approximation in the intervocalic interval (Figs. 2–6), revealing evidence of a labial gesture for Australian English /ɹː/. For Speakers M21 and M22, consistent labialization was only observed in [iːɹiː] and [ɐːɹɐː] sequences, but rhotics produced by Speaker F21 in the back vowel context were also characterized by a narrower labial constriction between the rounded vowels, consistent with /ɹː/ involving an independent labial approximation gesture, rather than simply coarticulated labial protrusion (Fig. 4).

### 4.2. Future Directions

Additional data from more speakers are needed to better understand the type and scope of individual speaker variation in /ɹː/ articulation in Australian English. Improved temporal resolution of real-time MRI data would allow closer analysis of intergestural timing across a wider range of phonological contexts to shed more light on the way that tongue tip, tongue body and labial gestures are coordinated in Australian English /ɹː/, and how these coordinative patterns are influenced by wider coarticulatory and prosodic factors. Acoustic analysis of recorded speech and simulations based on MRI-derived vocal tract models are planned, to help validate the articulatory characterization of /ɹː/ and inform our understanding of the acoustic consequences of different aspects of articulation within and across speakers.

## 5. Conclusions

These data demonstrate how image data combined from different MRI modalities can provide new insights into details of rhotic articulation in under-described language varieties. Australian English /ɹ/, as in other English varities, is produced with considerable inter-speaker variability; nevertheless, each of the rhotic allophones captured here can be characterized as a complex articulatory structure involving coordination of tongue tip, tongue body and labial gestures.

## 6. Acknowledgements

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

## 7. References

[1] R. M. Dalston, "Acoustic characteristics of English /w, r, l/ spoken correctly by young children and adults," *JASA*, vol. 57, no. 2, pp. 462–469, 1975.

[2] O. Fujimura and D. Erickson, "Acoustic phonetics," in *The Handbook of Phonetic Sciences*, W. Hardcastle and J. Laver, Eds. Oxford: Blackwell, 1997, pp. 65–115.

[3] R. E. Hagiwara, "Acoustic realizations of american/r/as produced by women and men," vol. WPP, No. 90, 1995.

[4] B. Heselwood and L. Plug, "The Role of F2 and F3 in the Perception of Rhoticity: Evidence from Listening Experiments," in *Proc. ICPhS*, 2011, pp. 867–870.

[5] C. Y. Espy-Wilson, S. E. Boyce, M. Jackson, S. Narayanan, and A. Alwan, "Acoustic modeling of American English /r/," *JASA*, vol. 108, no. 1, pp. 343–356, 2000.

[6] P. J. Howson and P. J. Monahan, "Perceptual motivation for rhotics as a class," *Speech Communication*, vol. 115, pp. 15–28, 2019.

[7] K. Sebregts, R. van Hout, and H. Van de Velde, "Sociophonetics and rhotics," in *The Routledge Handbook of Sociophonetics*. Routledge, 2023, pp. 195–213.

[8] P. Delattre and D. C. Freeman, "A dialect study of American r's by x-ray motion picture," *Linguistics*, vol. 44, pp. 29–68, 1968.

[9] M. K. Tiede, S. E. Boyce, C. K. Holland, and K. A. Choe, "A new taxonomy of American English /r/ using MRI and ultrasound," *JASA*, vol. 115, no. 5, pp. 2633–2634, 2004.

[10] X. Zhou, C. Y. Espy-Wilson, S. Boyce, M. Tiede, C. Holland, and A. Choe, "A magnetic resonance imaging-based articulatory and acoustic study of 'retroflex' and 'bunched' American English /r/," *JASA*, vol. 123, no. 6, pp. 4466–4481, 2008.

[11] E. Lawson, J. M. Scobbie, and J. Stuart-Smith, "The social stratification of tongue shape for postvocalic /r/ in Scottish English," *J. Sociolinguistics*, vol. 15, no. 2, pp. 256–268, 2011.

[12] M. Heyne, X. Wang, D. Derrick, K. Dorreen, and K. Watson, "The articulation of /ɹ/ in New Zealand English," *JIPA*, vol. 50, no. 3, pp. 366–388, 2020.

[13] J. Mielke, A. Baker, and D. Archangeli, "Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /r/," *Language*, vol. 92, no. 1, pp. 101–140, 2016.

[14] A. Alwan, S. Narayanan, and K. Haker, "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics," *JASA*, vol. 101, no. 2, pp. 1078–1089, 1997.

[15] B. Gick, A. M. Kang, and D. H. Whalen, "MRI evidence for commonality in the post-oral articulations of English vowels and liquids," *Journal of Phonetics*, vol. 30, no. 3, pp. 357–371, 2002.

[16] M. Proctor and R. Walker, "Articulatory bases of English liquids," in *The Sonority Controversy*, S. Parker, Ed. Berlin: De Gruyter, 2012, vol. 18, pp. 285–312.

[17] B. Gick, K. Iskarous, D. H. Whalen, and L. M. Goldstein, "Constraints on variations in the production of English /r/," in *6th Intl. Seminar on Speech Production*, Sydney, 2003, pp. 73–78.

[18] K. Iskarous, "The Articulation of the Palatal Gesture in American English [r]," in *Proc. 7th Intl. Seminar on Speech Production*, Ubatuba, 2006.

[19] B. Gick and F. Campbell, "Intergestural timing in English /r/," in *Proc. 15th Intl. Congress of Phonetic Sciences*, Universitat Autònoma de Barcelona. Barcelona: Proc. XVth Intl. Congress of Phonetic Sciences, 2003, pp. 1911–1914.

[20] B. Gick, F. Campbell, S. Oh, and L. Tamburri-Watt, "Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids," *J. Phon*, vol. 34, no. 1, pp. 49–72, 2006.

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

[22] F. Cox and S. Palethorpe, "Australian English," *JIPA*, vol. 37, no. 03, pp. 341–350, 2007.

[23] F. Cox, L. Ratko, J.-H. Kim, J. Penney, and M. Proctor, "Investigating rhotic production by Australian English speakers using ultrasound imaging," in *Australian Linguistic Society Annual Conf.*, University of Sydney, 29 Nov - 01 Dec. 2023.

[24] M. Proctor, J.-H. Kim, J. Penney, L. Ratko, and F. Cox, "Characterizing rhotic articulation in Australian English using ultrasound," in *Australasian Intl. Conf. on Speech Science and Technology (19th SST)*. Australasian Speech Science and Technology Association, 2024, pp. 82–86.

[25] 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.

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

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

[28] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel 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.