Characterizing Rhotic Articulation in Australian English using Ultrasound

Michael Proctor, Jae-Hyun Kim, Joshua Penney, Louise Ratko, Felicity Cox

Department of Linguistics, Macquarie University, Australia
michael.proctor@mq.edu.au

Abstract

Rhotic approximants in English are produced with considerable articulatory variability across speakers and contexts. Although this variability has been examined in North American, British and New Zealand varieties, it has not been documented using instrumental methods in Australian English (AusE). We therefore examined rhotic approximants produced by six speakers of AusE in nine vowel contexts using ultrasound tongue imaging. Three broad patterns of tongue shaping were observed, and speakers differed in the type and degree of vocalic influence on rhotic posture. Implications for phonological characterization of AusE /ɹ/ are discussed, along with methodological considerations for ultrasound studies.

Index Terms: rhotics, ultrasound tongue imaging, Australian English, liquids, articulation

1. Introduction

The phonetic and phonological characterization of rhotic approximants remains an ongoing topic of research, and the complexity and variability of this class of sounds presents many challenges.

Acoustically, many rhotic approximants are characterized by a lowered third formant in some environments [1, 2], but F3 trajectories are not a robust universal perceptual correlate of rhoticity, and acoustic properties of /ɹ/ exhibit complex variability across speakers and environments [3]. Furthermore, the relationships between acoustic properties and rhotic articulation are not well understood [4, 5]

American English /ɹ/ is produced with variable tongue configurations, which have been described using taxonomies that vary in complexity describing two [6], six [7], and eight [8] different categories of rhotic articulation. The most common defining characteristics proposed to differentiate these variant /ɹ/ types in American English – and similar rhotic variation in other Englishes – include 'retroflex' vs. 'bunched' [9], or 'tip-up' vs. 'tip-down' [10]. The distinction is complicated by individual speaker variation, sound change, and inconsistent percepts of rhotic type. In an ultrasound study of 27 speakers of American English, Mielke et al. 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" [11, p.101].

Further insights have been provided into English /ɹ/ through analysis of the gestures involved in production. Articulatory studies of North American Englishes using MRI [12, 13], X-ray microbeam [14, 15], ultrasound and video [16, 15, 17] have revealed that /ɹ/ involves coordination of two lingual gestures, with an additional labial gesture observed in onset environments [18]. Australian English (AusE) is a non-rhotic variety [19], so although the properties of rhotics occurring in coda environments are not relevant, the phonetic similarities of AusE onset /ɹ/ suggests that it involves similar goals of production as those of other English varieties.

Heyne et al. [20] investigated rhotic articulation in New Zealand English (NZE), which like AusE, is a non-rhotic variety. Sixty-two speakers produced 13 words containing /ɹ/ in different phonological environments, and tongue shapes were categorized according to the four main patterns described by Delattre and Freeman [7]. 25 speakers consistently produced tip-down variants, 12 consistently produced tip-up rhotics, and 25 speakers produced /ɹ/ with variable tongue shapes; tip-up allophones were more commonly observed in back vowel contexts, and tip-down before high front vowels.

To date, no systematic study of AusE /ɹ/ articulation has been conducted; however, a pilot ultrasound study of rhotics in six speakers from Sydney [21] reveals patterns of production similar to the NZE study [20]. Four broad tongue shapes were observed in word-initial /ɹ/ produced before three different vowel qualities /iː-ɐː-oː/, but further details of production were not analyzed.

Liquid consonants, /ɹ/ and /l/, are typically amongst the most difficult English sounds to master for both first [22] and second language learners, and studies of phonological acquisition in monolingual AusE speaking children in Sydney show that lateral approximants are acquired relatively late [23]. These developmental trajectories suggest that Australian English /ɹ/ may also be characterized by gestural complexity requiring fine control of articulators [24], the details of which are not yet understood. More data is needed to understand the goals of production of AusE /ɹ/ and the articulatory variability it exhibits across speakers and environments.

1.1. Aims

The aim of this study is to investigate lingual articulation of /ɹ/ in AusE, using ultrasound tongue imaging. To the best of our knowledge, this is the first instrumental study to examine AusE rhotic production in a wide range of vowel contexts, allowing us to offer initial insights into patterns of /ɹ/ production. A secondary aim of this study is to establish robust methods and experimental materials for ongoing studies of English rhotics in Australia.

2. Methods

Six female adult monolingual speakers of AusE participated in the study (Mean age = 23.2 years, SD = 2.9, range: 19–27). All participants were born and raised in Australia and had at least one parent who was also born and raised in Australia. All participants had completed their primary and secondary education in Australia. Data were acquired as part of an ongoing study of AusE rhotics. Participants were undergraduate students at Macquarie University, and received course credit for their participation. The ethical aspects of this study have been approved by the Human Research Ethics Committee of the institution affiliated with the authors.

2.1. Experimental Materials

Rhotics were elicited in word-initial position before nine different vowels distributed across three broad places of articulation (Table 1). Each target word was elicited in the carrier phrase *'It's a ___'* to ensure that the tongue body was in a neutral position for schwa prior to the rhotic production, enabling observation of both retraction and raising for the following onset rhotic. Each individual item was presented orthographically on a monitor in a pseudo-random order and read aloud by the participant in a self-paced recording session divided into three blocks. During the recording session, participants also produced rhotics in a range of other contexts, which are not analysed here. The elicitation was monitored by the experimenter so that trials compromised by mispronunciations, atypical prosody or noise interference could be re-recorded immediately. A total of 9 (items) × 3 (repetitions) × 6 (participants) = 162 trials were recorded for analysis.

Table 1: *Stimuli used to elicit Australian English /ɹ/ in word-initial position before nine different vowels.*

ITEM	TARGET	V PLACE
'It's a reef'	/ɹiːf/	High Front
'It's a rip'	/ɹɪp/	High Front
'It's a ref'	/ɹep/	High Front
'It's a rap'	/ɹæp/	Low
'It's a raft'	/ɹɐːft/	Low
'It's a rough'	/ɹɐf/	Low
'It's a raw'	/ɹoː/	Back
'It's a rook'	/ɹʊk/	Back
'It's a rob'	/ɹɔb/	Back

2.2. Data acquisition

Lingual articulation was tracked in the midsagittal plane using ultrasound tongue imaging. Data were elicited and recorded using the Articulate Assistant Advanced (AAA) software Version 220.2.0 [25] at Macquarie University. A microconvex probe (2-4 MHz, 20 mm radius) was located beneath the participant's chin, held in place using an Articulate Instruments Aluminium Probe Stabilisation Headset [26].

Ultrasound video data were acquired at a temporal resolution of 55 to 60 f.p.s. with the probe frequency at 3 MHz, depth at 120mm, the focus depth at 96 mm and with a 83.2° field of view. Speech audio was recorded concurrently at a sampling rate of 22,050 Hz using a RØDE NTG1 condenser shotgun microphone located approximately 30 cm in front of the participant, offset 15°, connected to a Focusrite Scarlett Solo 3rd Generation preamplifier.

Hard palates were located using water swallow trials and by fitting a convex hull to the region of maximum lingual excursion observed during obstruent production [27]. Palates were manually traced using the fan spline in the AAA system. Midsagittal palate traces were exported as a 12-point set of cartesian coordinate pairs defined with respect to a fiducial line located immediately above the ultrasound probe.

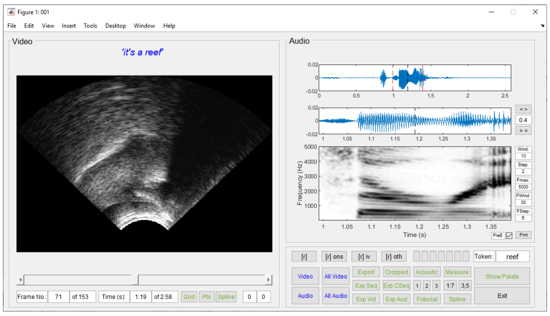

Figure 1: ***Ultrasound Inspection and Analysis**. Matlab-based tool used for inspecting and analysing time-aligned ultrasound video and audio data.*

Video data were exported from the AAA system in uncompressed AVI format (RGB24 encoding) with a spatial resolution of 400 × 300 px over a 188 × 141 mm field of view (1 px = 0.47 mm). Companion audio recordings were exported in uncompressed 16 bit mono WAV format.

2.3. Phonetic data analysis

Ultrasound video and companion audio recordings were inspected using a custom Matlab-based graphical user interface facilitating frame-by frame navigation of exported AAA data with time-aligned audio (Fig. 1).

Rhotic targets were identified by inspecting ultrasound video sequences in the interval following the schwa. Two dimensions of rhotic articulation were tracked: (i) raising/advancement of the coronal part of the tongue towards the palate or alveolar ridge, and (ii) retraction of the posterior part of the tongue dorsum observable in the region anterior to the hyoid shadow. The rhotic target was identified as the frame in which the coronal part of the tongue achieved maximal excursion away from the initial lingual posture. Where the coronal part of the tongue maintained a maximally raised/advanced posture for multiple frames, the central frame was chosen.

Three phonetically trained analysts independently identified the rhotic target frame in each utterance in the experimental corpus. 486 target frame numbers were recorded (3 analysts × 162 trials) and compared. Analysts agreed on rhotic target frames for most trials, and in no case did target frame numbers differ by more than 4 across analysts; in these cases, ultrasound video data were re-examined to check rhotic targets.

2.4. Ultrasound image analysis

Image frames at each rhotic target were exported in uncompressed JPEG format. The three frames corresponding to rhotic targets for each repetition in each vowel context were combined into a single image by computing the mean intensity of each pixel in the 400 × 300 matrix. 27 images were generated for each participant, illustrating mean tongue posture for each rhotic in each vowel context. Images were further combined to illustrate rhotic tongue posture at each broad place of articulation of context vowels by generating mean images from the set of all target frames for [iː-ɪ-e] (front), [æ-ɐː-ɐ] (low), and [oː-ʊ-ɔ] (back) vowel contexts. Palate traces were superimposed on images for reference.

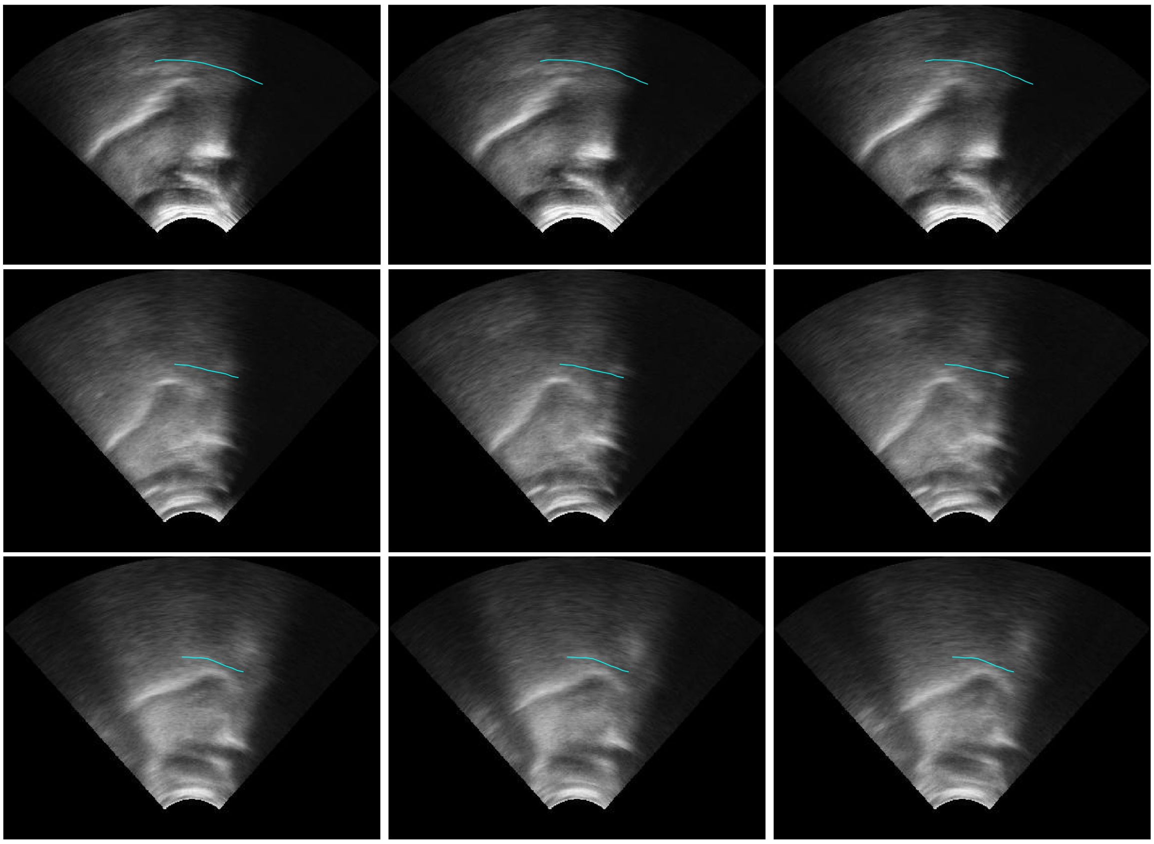

Figure 2: ***Mean tongue posture at rhotic targets***. *Mean image calculated from target frame in three repetitions of three words. L-to-R: back vowel contexts [oː-ʊ-ɔ], low vowel contexts [æ-ɐː-ɐ], front vowel contexts [iː-ɪ-e]. Top row: Participant W003; Middle row: W006; Bottom row: W007. Blue line: palate trace. (Front of mouth: right side of image).*

3. Results

Two lingual gestures were observed in all rhotics produced by all speakers in the experimental corpus: dorsal retraction toward a mid- to low-pharyngeal target, and coronal raising toward a target in the alveolar-palatal region. For all speakers, the two gestures were largely synchronous; precise intergestural timings have not yet been quantified. Participants differed in (i) the place of articulation of the coronal gesture, (ii) the tongue shape at rhotic target, and (iii), the degree and type of influence of vowel context on target rhotic posture.

Mean midsagittal images capturing mid-consonantal tongue postures for word-initial /ɹ/ produced by each participant in high front, low, and back vowels contexts are illustrated in Figures 2 and 3. Individual differences in coronal articulation can be observed: more retracted toward a palatal target for Speakers W003, W006, W009 and W014, and more anterior for W007, who uses a post-alveolar coronal gesture in all contexts. Speaker W010 articulates /ɹ/ with an alveolar coronal gesture in back vowel contexts, a more retracted palatal constriction in front vowel contexts, and a more distributed laminar coronal gesture intermediate between these two places in low vowel contexts.

Rhotics produced by all six speakers were characterized by a 'saddle' – some degree of concavity in the mid-lingual region – in at least some vowel contexts. For Speakers W003, W006, W007 and W009, all rhotics were produced with a mid-lingual saddle, regardless of vowel context. Speaker W014 articulates /ɹ/ with a more bunched tongue posture showing minimal dorsal concavity. Speaker W010 produces rhotics before low vowels (*'rap'–'raft'–'rough'*) with a globally convex tongue shape, but other realizations show saddles at different parts of the tongue, depending on context: the concavity appears at a similar mid-lingual location to other speakers before front *'reef'–'rip'–'ref'*, but at a more anterior part of the tongue before back vowels *'raw'–'rook'–'ref'* (Fig. 3, middle row).

4. Discussion

Consistent lingual postures during rhotic production were observed in all vowel contexts by five of the six speakers in our study – tongue shapes that broadly correspond to the 'tip-down' configuration classified as 'Type 4' by Delattre and Freeman [7]. The predominance of this tongue shape is consistent with previous findings for NZE [20] and North American Englishes [11], in which 'bunched' variants were also the most common rhotic allophones. In contrast to these previous studies, none of the AusE speakers consistently and exclusively produced 'tip-up' variants; however, six speakers is too a small sample from which to draw further conclusions at this stage.

Furthermore, closer inspection of even these initial data suggests that categorical classification of tongue postures into binary taxonomies misrepresents the complexity of articulation involved. Rhotics produced by W009 in non-front vowel contexts, for example, demonstrate some evidence of retroflexion towards a palatal target (Fig. 3, top row), which is not well

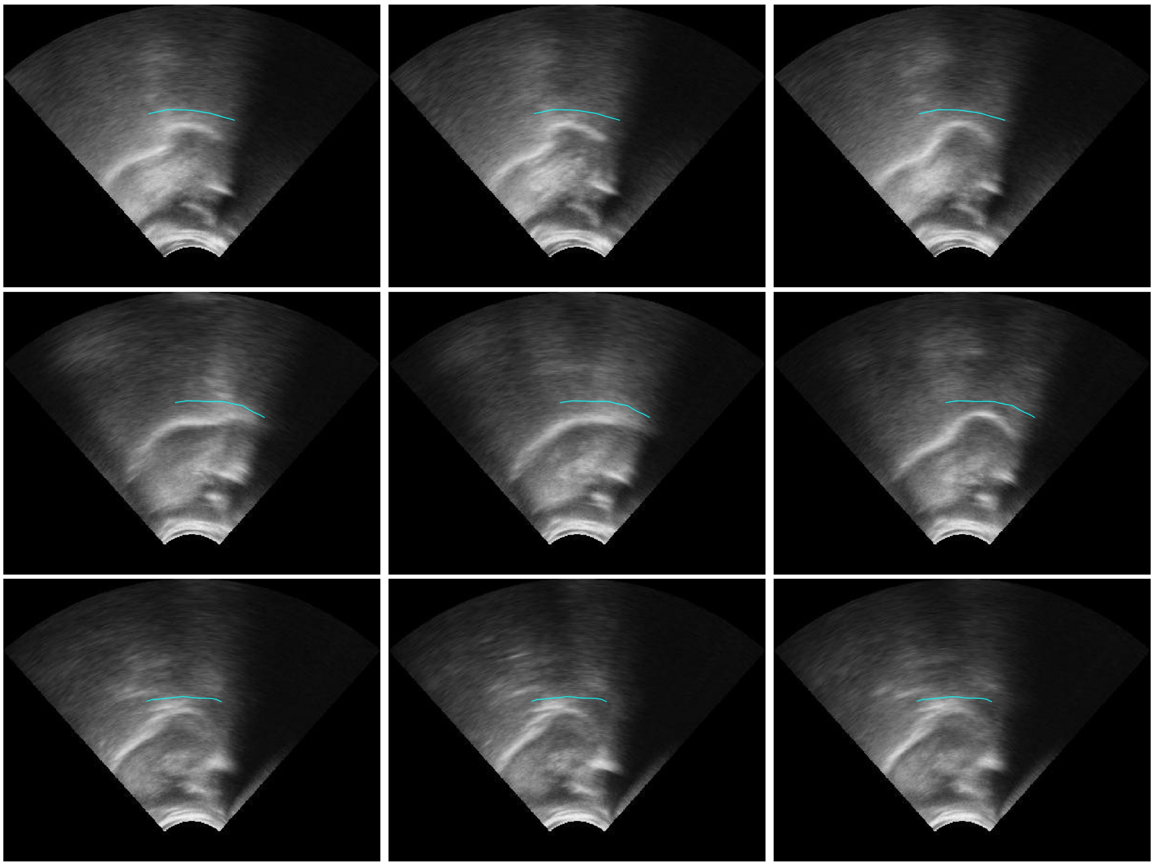

Figure 3: ***Mean tongue posture at rhotic targets***. *Mean image calculated from target frame in three repetitions of three words. L-to-R: back vowel contexts [oː-ʊ-ɔ], low vowel contexts [æ-ɐː-ɐ], front vowel contexts [iː-ɪ-e]. Top row: Participant W009; Middle row: W010; Bottom row: W014. Blue line: palate trace. (Front of mouth: right side of image).*

characterized as 'bunched', nor as either 'tip-up' or 'tip-down'. More systematic investigation of tongue shaping in these and other speakers of Australian English is required to better represent articulation in the midsagittal plane, and how this relates to goals of production [28, 29, 30]. Key to this will be dynamic analysis of tongue shaping, including quantification of intergestural timing and examination of how the coordination of lingual gestures shapes the tongue over time in different vowel contexts.

4.1. Methodological considerations

In this study, data were acquired using the AAA system [25] and exported for inspection and analysis using a custom Matlab-based tool (Fig. 1). This allowed for greater flexibility in analysis, but requires robust synchronization of exported audio and video data. While this is usually feasible, platform-independent methods for audio-video synchronization during and after ultrasound recordings are important for validation and maximal flexibility in data analysis and interpretation (e.g. [31]).

The carrier phrase *'It's a ___'* worked well for elicitation of onset /ɹ/ in an environment where key lingual gestures could be observed before a range of vowels. It was important to monitor speakers during data collection to ensure that they produced the sentences with a pre-rhotic schwa, as one speaker (W007) produced a diphthongized [æɪ] article in some trials. While this is unlikely to have influenced target tongue postures significantly, it will have affected gestural timing and coarticulation.

The aluminium headset [26] was effective in stabilizing the ultrasound probe, but partly due to the weight (∼0.8 Kg), most speakers experienced some discomfort over the duration of a 40 minute experiment. More ergonomic options (e.g. [32]) may facilitate improved user comfort. Some displacement/rotation of the probe may have affected consistency of the imaging plane and therefore reliability of the data for some speakers, and inconsistency in the relative location of hyoid and mandible shadows across speakers makes direct comparison of place of articulation difficult. No bite plate was used to calibrate ultrasound data with anatomical landmarks [33], so images in Figures 2–3 should be interpreted with caution.

5. Conclusions

This study is a first step towards the systematic investigation of Australian English /ɹ/ production using instrumental methods. Initial investigation of lingual articulation in six speakers reveals two coordinated lingual gestures: a mid pharyngeal tongue body gesture, and a coronal gesture realized at a speaker-specific place of articulation. Consistent tongue shaping was observed across vowel contexts for five speakers, in a posture characterized by some degree of mid-lingual concavity. These data further demonstrate the utility of midsagittal ultrasound tongue imaging as a method for characterizing general goals of rhotic approximants, and for revealing individual speaker variation.

6. Acknowledgements

This project was supported by Macquarie University Research Acceleration Scheme (MQRAS0001053) "Visualising complex speech sounds: a study of rhotic consonants – evidence from ultrasound of the tongue". We would like to thank Khan Milligan who assisted with data collection.

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] 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.

[4] 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.

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

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] B. Gick, "Articulatory correlates of ambisyllabicity in English glides and liquids," in *Papers in laboratory phonology VI: Phonetic interpretation*, J. Local, R. Ogden, and R. A. M. Temple, Eds. Cambridge: CUP, 2003, pp. 222–236.

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

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

[21] 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.

[22] K. Crowe and S. McLeod, "Children's English consonant acquisition in the United States: A review," *American J. Speech-Language Pathology*, vol. 29, no. 4, pp. 2155–2169, 2020.

[23] S. Lin and K. Demuth, "Children's acquisition of English onset and coda /l/: Articulatory evidence," *JSLHR*, vol. 58, no. 1, pp. 13–27, 2015.

[24] P. J. Howson and M. A. Redford, "The acquisition of articulatory timing for liquids: Evidence from child and adult speech," *JSLHR*, vol. 64, no. 3, pp. 734–753, 2021.

[25] Articulate Instruments, *Articulate Assistant Advanced User Guide: Version 220.2.0.*

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

[27] M. A. Epstein and M. Stone, "The tongue stops here: Ultrasound imaging of the palate," *JASA*, vol. 118, no. 4, pp. 2128–2131, 2005.

[28] I. Stavness, B. Gick, D. Derrick, and S. Fels, "Biomechanical modeling of English /r/ variants," *JASA*, vol. 131, no. 5, pp. EL355–EL360, 2012.

[29] S. Stolar and B. Gick, "An index for quantifying tongue curvature," *Canadian Acoustics*, vol. 41, no. 1, 2013.

[30] K. M. Dawson, M. K. Tiede, and D. Whalen, "Methods for quantifying tongue shape and complexity using ultrasound imaging," *Clinical Linguistics & Phonetics*, vol. 30, no. 3-5, pp. 328–344, 2016.

[31] A. Eshky, J. Cleland, M. S. Ribeiro, E. Sugden, K. Richmond, and S. Renalds, "Automatic audiovisual synchronisation for ultrasound tongue imaging," *Speech Communication*, vol. 132, pp. 83–85, 2021.

[32] L. Spreafico, M. Pucher, and A. Matosova, "Ultrafit: A speaker-friendly headset for ultrasound recordings in speech science," in *Proc. Interspeech*, 2018, pp. 1917–1520.

[33] L. Ménard, J. Aubin, M. Thibeault, and G. Richard, "Measuring tongue shapes and positions with ultrasound imaging: A validation experiment using an articulatory model," *Folia Phoniatrica et Logopaedica*, vol. 64, no. 2, pp. 64–72, 2012.