JANUARY 22 2024

Formant dynamics in second language speech: Japanese speakers' production of English liquids

Takayuki Nagamine

J. Acoust. Soc. Am. 155, 479–495 (2024)
https://doi.org/10.1121/10.0024351

JASA ARTICLE

Formant dynamics in second language speech: Japanese speakers' production of English liquids

Takayuki Nagamine[a)]
Department of Linguistics and English Language, County South, Lancaster University, Lancaster, LA1 4YL, United Kingdom

ABSTRACT:
This article reports an acoustic study analysing the time-varying spectral properties of word-initial English liquids produced by 31 first-language (L1) Japanese and 14 L1 English speakers. While it is widely accepted that L1 Japanese speakers have difficulty in producing English /l/ and /ɹ/, the temporal characteristics of L2 English liquids are not well-understood, even in light of previous findings that English liquids show dynamic properties. In this study, the distance between the first and second formants (F_2–F_1) and the third formant (F_3) are analysed dynamically over liquid-vowel intervals in three vowel contexts using generalised additive mixed models (GAMMs). The results demonstrate that L1 Japanese speakers produce word-initial English liquids with stronger vocalic coarticulation than L1 English speakers. L1 Japanese speakers may have difficulty in dissociating F_2–F_1 between the liquid and the vowel to a varying degree, depending on the vowel context, which could be related to perceptual factors. This article shows that dynamic information uncovers specific challenges that L1 Japanese speakers have in producing L2 English liquids accurately.
https://doi.org/10.1121/10.0024351

(Received 10 May 2023; revised 4 December 2023; accepted 5 December 2023; published online 22 January 2024)

[Editor: Susanne Fuchs] Pages: 479–495

I. INTRODUCTION

A. Acquisition of English /l/ and /ɹ/ by L1 Japanese speakers

The current study investigates time-varying spectral properties of English liquids produced by first-language (L1) Japanese speakers. Numerous studies have shown that the acquisition of English liquids is particularly challenging for L1 Japanese speakers (e.g., Aoyama *et al.*, 2019; Best and Strange, 1992; Flege *et al.*, 1995; Saito and Munro, 2014; Sheldon and Strange, 1982). They typically perceive English /l/ and /ɹ/ as instances of a single L1 category of Japanese /ɾ/ (e.g., Best and Strange, 1992; Guion *et al.*, 2000). This corresponds to the learning of "similar" phones between L1 and L2 in the Speech Learning model (SLM) (Flege, 1995; Flege and Bohn, 2021) and the single-category (SC) or the category-goodness (CG) assimilation scenarios in the Perceptual Assimilation model of Second Language (L2) Speech Learning (PAM-L2) (Best and Strange, 1992; Best and Tyler, 2007; Hattori and Iverson, 2009), predicting a moderate to substantial difficulty in acquisition of the L2 sounds. SLM posits that perceptual accuracy lays the foundation for accurate L2 speech production because L2 learners develop articulatory rules in the L2 phonetic categories that are established over the course of L2 speech learning (Flege and Bohn, 2021).

The difficulty L1 Japanese speakers face in acquiring English /l/ and /ɹ/ is associated with their sensitivity to the phonetic cues used to distinguish the contrast. The key spectral dimension that contrasts English /l/ and /ɹ/ is the frequency of the third formant (F_3); American English /ɹ/ is associated with a notably low F_3 at 1300 Hz for male speakers and 1800 Hz for female speakers whereas laterals show a high F_3 at approximately 2500–2800 Hz (Espy-Wilson, 1992; Stevens, 2000). The F_2 frequency is associated with the resonance of the vocal tract cavity posterior to the primary constriction for both laterals and rhotics, which are commonly produced with a backed tongue body configuration (Stevens, 2000). Laterals are generally characterised by clear-dark allophony according to syllabic position; "clear" /l/s are often associated with laterals in pre-vocalic, syllable-initial position, and they typically have higher F_2 values and a greater separation between F_2 and F_1 (F_2–F_1) than the post-vocalic "dark" counterpart (Carter and Local, 2007; Recasens, 2012). American English exhibits relatively darker realisations of liquids than British English overall, but syllable-initial laterals in American English are still somewhat "clearer" than syllable-final counterparts (Recasens, 2012). This clear-dark allophony according to the syllable position results from different articulatory configurations, such that the degree of the tongue body retraction is greater for the final laterals than for the initial laterals (Recasens, 2012).

L1 Japanese speakers tend to rely on the less reliable cue of F_2 in their perception of English /l ɹ/ than a more robust cue of F_3 (Iverson *et al.*, 2003; Saito and Munro, 2014). As a result, they tend to produce the distinction along the F_2 dimension instead of learning to make a contrast

[a)]Email: t.nagamine@lancaster.ac.uk

along F_3 (Aoyama *et al.*, 2019; Saito and van Poeteren, 2018). For instance, they produce word-initial English /l/ with a somewhat higher F_2 (approximately 1500 –1800 Hz) than L1 English speakers (approximately 1200–1500 Hz), whereas F_2 frequencies for English /ɹ/ are similar between the two speaker populations (Aoyama *et al.*, 2019; Flege *et al.*, 1995). As for F_3, they produce English /ɹ/ with a relatively high F_3 (2000–2600 Hz) but produce /l/ with F_3 values comparable to L1 English speakers (Aoyama *et al.*, 2019; Flege *et al.*, 1995; Saito and Munro, 2014). Nevertheless, previous research claims that L1 Japanese speakers could learn to use the acoustic cues as L1 English speakers would do, especially for F_1 and F_2; several studies reported similar F_1 values in production of English liquids between L1 Japanese and L1 English speakers (Aoyama *et al.*, 2019; Flege *et al.*, 1995; Saito and Munro, 2014). Saito and Munro (2014) also argue that the use of F_2 is easier for L1 Japanese speakers to acquire than that of F_3 for English /ɹ/ based on findings that L1 Japanese speakers who resided in Canada for longer than 2.5 months produced native-like F_2 values for English /ɹ/ compared to those who had less overseas experience.

The degree of difficulty in L1 Japanese speakers' acquisition of English liquids also varies depending on the vowel context, in which they are better at correctly identifying word-initial English liquids adjacent to front vowels compared to back vowels in perception (Shimizu and Dantsuji, 1983). This might be because L1 Japanese speakers may also perceive English /l/ and /ɹ/ as a sequence of a back vowel and a tap (i.e., [ɯɾ]), possibly due to the vocalic nature of English liquids (Guion *et al.*, 2000). L1 Japanese speakers are more likely to hear a /w/-like percept when perceiving English /l/ and /ɹ/ than L1 English speakers (Best and Strange, 1992; Mochizuki, 1981; Yamada and Tohkura, 1992). These results overall suggest that L1 Japanese speakers are sensitive not only to the phonemic status but also phonetic details of English /l/ and /ɹ/. In particular, Shimizu and Dantsuji (1983) speculate that coarticulatory properties may play a role in explaining the vocalic contextual effects in L1 Japanese speakers' correct identification of English /l/ and /ɹ/.

B. Dynamic analysis of English liquids

Although the errors in segmental realisation in L2 speech are claimed to be rooted in perception, accurate perception does not always entail accurate production (Flege and Bohn, 2021; Sheldon and Strange, 1982). While this does not mean that the role of perceptual accuracy should be discounted, it implies that L2 speech production may be shaped by a combination of factors in addition to perceptual accuracy.

One such possible factor includes the dynamic nature involved in the production of English liquids. Articulation of English liquids requires coordination of multiple articulatory gestures for accurate production (Campbell *et al.*, 2010; Sproat and Fujimura, 1993). English laterals, for instance, involve coordination of tongue tip and dorsum gestures, and the timing and magnitude interact with the syllabic position; a tongue tip gesture precedes a tongue dorsum gesture with a greater magnitude for clear /l/ whereas the two gestures could be timed synchronously for the dark /l/ (Sproat and Fujimura, 1993). English rhotics show similar patterning of gestural timing and magnitude, where labial gestures precede the tongue tip and tongue body gestures (Campbell *et al.*, 2010; Proctor *et al.*, 2019). The dynamic nature of articulation in English liquids suggests that the acoustic characteristics of English liquids are inherently non-static, and it is, therefore, often challenging to select a single point in time that adequately represents liquid quality (Kirkham *et al.*, 2019; Ying *et al.*, 2012).

In addition, acoustic realisations of liquids interact with the neighbouring segments as a result of coarticulation. While coarticulation is often viewed as a consequence of the physiological mechanisms in the transition between segmental targets, some aspects of coarticulation may be language-specific and thus need to be learned (Beristain, 2022; Keating, 1985). Word-initial /ɹ/ in English, for instance, shows lower F_3 values when followed by back vowels compared to other vowel conditions (King and Ferragne, 2020). Similarly, vowel context influences realisations of American English /l/, particularly among word-initial /l/s, such that F_2 values are higher in the /i/ context than in the /a/ context (Recasens, 2012). Coarticulatory effects of liquids could span longer term than the domain of liquid segment itself and provide perceptual basis for listeners to distinguish English /l/ and /ɹ/ (West, 1999a,b).

The findings regarding the dynamic nature of liquid production and liquid-vowel coarticulation may account for the specific difficulties that L1 Japanese speakers have in producing English /l/ and /ɹ/. L1 Japanese speakers tend to substitute English /l/ and /ɹ/ with an alveolar tap or flap [ɾ], a canonical realisation of Japanese /ɾ/ (Riney *et al.*, 2000). Previous articulatory studies show that alveolar taps/flaps show stronger coarticulatory effects with the neighbouring vowels than English laterals and rhotics; while the tongue dorsum gesture is actively involved in the production of English /l/ and /ɹ/, taps and flaps [ɾ] show either less involvement of the tongue dorsum or a "stabilization" tongue dorsum gesture, resulting in stronger coarticulation with the vowel (Morimoto, 2020; Proctor, 2011; Recasens, 1991; Yamane *et al.*, 2015). Furthermore, an x-ray study suggests that L1 Japanese speakers' articulation of English liquids shows greater variability according to the vocalic environment (Zimmermann *et al.*, 1984). In sum, Japanese and English liquids differ in the way they are coarticulated with the vowels, and it can be predicted that L1 Japanese speakers exhibit different liquid-vowel coarticulatory patterns from that of L1 English speakers.

Despite the findings regarding the complexity involved in the production of English liquids, our understanding remains relatively limited regarding the specific mechanism whereby L1 Japanese speakers struggle to produce English /l/ and /ɹ/. This may be because previous research

https://doi.org/10.1121/10.0024351

commonly evaluates liquid quality based on a single-point measurement, in which formant frequencies are measured at one point in time, such as the F_3 minima, the spectral onset, or the spectral release (Aoyama *et al.*, 2019; Flege *et al.*, 1995; Saito and Munro, 2014). Analysis of liquids based on a single measurement, however, inevitably averages out temporal information that may be important for understanding the dynamic characteristics of English liquids.

In the current study, I show that dynamic formant measurement of English liquids allows us to better understand specific challenges that L1 Japanese speakers have in producing English /l/ and /ɹ/. Previous research suggests that (1) L1 Japanese speakers' acquisition of English liquids may be influenced by the phonetic details, such as vowel environments, and (2) English liquids show dynamic characteristics and interactions with the neighbouring vowels. Given these, I hypothesise that L1 Japanese speakers' production of English liquids will exhibit different dynamic acoustical properties compared to L1 English speakers. This study therefore asks what dynamic acoustic properties L1 Japanese speakers would show in their production of English /l ɹ/ compared to L1 English speakers.

I combine static and dynamic analyses of the acoustic properties of English liquids in this study. The static analysis investigates the distance between second and first formants (F_2–F_1) and the third formant (F_3) extracted at the liquid midpoint. The inclusion of this measure allows me to discuss the results in light of previous research in which the single-measurement analysis has been widely used (e.g., Aoyama *et al.*, 2019; Flege *et al.*, 1995; Saito and Munro, 2014; Saito and van Poeteren, 2018). In addition, the time-varying changes in the F_2–F_1 and F_3 values will capture the complex nature of liquid acoustics and the coarticulatory interactions between the liquid and the vowel (Howson and Redford, 2021; Kirkham *et al.*, 2019; Sproat and Fujimura, 1993).

II. METHODS

A. Participants

The data for the current study are obtained from 45 speakers: 31 L1 Japanese learners of English (17 female and 14 male) aged between 18 and 22 years [$M = 19.81$ years, standard deviation (SD) = 1.05] and 14 L1 North American L1 English speakers (11 female and three male) aged between 21 and 43 years ($M = 28.93$ years, $SD = 6.08$).

All of the L1 Japanese speakers were undergraduate university students recruited from two universities in Japan, located near the cities of Nagoya and Kobe, respectively. Their profile is considered to be typical for average Japanese university students who study English as a foreign language; all of them studied English primarily through the school curriculum in either or both primary and secondary schools, and continued it at the tertiary level, with a mean length of English study being 9.31 years ($SD = 2.42$). They did not have an extended stay in an English-speaking country, with the length of overseas experience ranging from none to 4.25 months ($M = 0.77$ months, $SD = 1.35$).

In evaluating L1 Japanese speakers' L2 English proficiency, participants were asked to report their perception on their own oral fluency on a scale of seven, with 1 being "I do not speak English at all." to 7 being "No problems in using English in daily life." This is because there was no common measure available across participants to estimate their English proficiency due to the fact that students have taken different kinds of tests or that first-year students had not yet taken any English language test. Nevertheless, judging from the test scores that some of the participants were able to provide and observations by the researcher who has experience in English language teaching in Japan, their English proficiency is considered to be lower to upper intermediate, which largely agrees with their subjective evaluation of their fluency in English ($M = 3.84$, $SD = 1.10$) (see supplementary material for further details about the participants).[1]

The 14 L1 English speakers identify themselves as fluent L1 speakers of North American English who grew up using English until 13 years of age. Five of them are from Canada and nine are from the United States. They resided in the United Kingdom (UK) at the time of recording; six of them were postgraduate students enrolled at a UK university and the rest worked in companies in the UK. Recruitment of L1 North American English speakers reflects the situation that American English tends to be chosen as a pedagogical model in English language teaching in Japan and therefore it is appropriate for L1 Japanese speakers' production to be compared to that of L1 North American English speakers (Setter and Jenkins, 2005).

B. Data collection

The audio recordings analysed in this study are a subset of data collection for a larger study, in which both articulatory and acoustic data were obtained in a simultaneous high-speed ultrasound-audio recording setting. For this reason, the participants wore an ultrasound headset while recording stimuli for the current study. The participants were recorded in a sound-attenuated booth at universities in the UK for L1 North American English speakers and in a quiet room at universities in Japan for L1 Japanese speakers. In recording some of the L1 Japanese speakers, however, there was minor background fan noise because of the Covid-19 restrictions mandating air ventilation at the time of recording. Acoustic signals were pre-amplified, digitized, and recorded onto a laptop computer via a Sound Devices (Reedsburg, WI) USB-Pre2 audio interface at 44.1 kHz with 16 bit quantisation.

The participants were asked to sit in front of the laptop screen and read the stimuli words in isolation that were displayed one by one orthographically using Articulate Assistant Advanced (AAA) (Edinburgh, UK) software version 220.4.1 (Articulate Instruments, 2022). No carrier phrases were used here because (1) the use of carrier phrases would impose additional difficulty on L1 Japanese speakers,

especially those who were less proficient in English, and (2) the experiment had to be as short as possible due to time constraints in the data collection sessions.

In light of the language mode hypothesis (Grosjean, 2008) that the language setting in an experiment can influence the participants' speech perception and possibly production, the recording sessions for the L1 Japanese speakers were structured as follows. The first half of the experiment, including briefing, equipment setup, and recording of the Japanese words (not presented in this paper), was conducted while I was giving instructions in Japanese. Then, I switched the language of instructions to English and the participant engaged in a short English conversation activity. This included a semi-structured dialogue in which I asked five simple questions to the participants (e.g., "What do you study?," "What do you like the best about the university?," etc.) Finally, the Japanese participants recorded the English words while I gave all the instructions in English. While it would have been theoretically desirable to have someone else who was an L1 English speaker lead the data collection session for English words, it was challenging for reasons of time and room availability given that each session for L1 Japanese speakers took up to 90 min.

The recording session with the L1 North American English speakers did not require such considerations because they recorded English words only. All the procedures were, therefore, conducted in English and each session took up to approximately 60 min. The participants were compensated for their time and participation with the amount of 2000 Japanese yen or 15 British pound stirlings in the form of cash or vouchers commensurate with the regulations at each of the recording venues. The research project has been reviewed and approved by the ethics committees at Lancaster University, Kobe Gakuin University, and Meijo University. Informed consent to take part in the study was obtained in written form from all participants.

C. Materials

Word-initial English /l/ and /ɹ/ were elicited from 16 monosyllabic words (eight minimal pairs), followed by a close front /i/, an open front /æ/, or a close back vowel /u/ (see Table I). The coda consonants were restricted to bilabials /p b m/ or labiodentals /f v/ to minimise the anticipatory coarticulatory effects on the word-initial liquids. All the target words were checked using the Longman Pronunciation Dictionary (Wells, 2008) to ensure that they have the intended vowel environment in American English.

TABLE I. Word list per vowel context.

Vowel context	Words		
/i/	leap / reap	leaf / reef	leave / reeve
/æ/	lap / rap	lamb / ram	lamp / ramp
/u/	lube / rube	loom / room	

D. Segmentation and data processing

Prior to segmentation, audio recordings were low-pass filtered at 11 000 Hz and downsampled to 22 050 Hz. Automatic segmentation was carried out at phoneme level with a Montreal Forced Aligner (MFA) version 2.0.6 (McAuliffe *et al.*, 2017). I then inspected the aligned data visually and manually corrected the segmentation using Praat where necessary (Boersma and Weenink, 2022).

I classified the liquid tokens into two broad categories: approximants and non-approximants, based on the spectrographic representations aided by auditory impressions. This decision reflects the consideration that the L1 Japanese speakers' production of liquids might show a wide range of variations due to the allophonic variation of Japanese /ɾ/ and their articulatory strategies for English /l/ and /ɹ/. Realisations for Japanese /ɾ/ include other types of approximants than English liquids, such as the canonical [ɾ], retroflex flap [ɻ], retroflex lateral approximant [ɭ], and a lateral flap [ɺ] (Akamatsu, 1997; Arai, 2013). They may also use a single strategy or produce a reversed realisation for English /l ɹ/. It could be the case, for instance, that they produce a lateral liquid for both English /l ɹ/. It is also possible that they use [l] for English /ɹ/ and [ɹ] for English /l/. Classifications based on these two broad categories: approximants and non-approximants, therefore, guide me to choose an appropriate type of analysis while maximising the chance of capturing diverse acoustic properties in the L1 and L2 English liquids.

Based on these considerations, I first broadly labelled tokens as approximants if the liquid token in question shows a vowel-like formant structure (Ladefoged and Johnson, 2010). The spectral analysis focuses only on the tokens that are classified here as approximants; it thus excludes 281 non-approximants tokens (e.g., taps or flaps [ɾ]) out of a total of 2914 tokens, leaving 2633 tokens for further processing. The spectrographic examples of an approximant and a non-approximant token are shown in Figs. 1 and 2.

Following this, I segmented the liquid approximant tokens based on the primary cues of a steady state or an approximately steady state of the F_2 and an abrupt change in amplitude in the waveform (Lawson *et al.*, 2011). Laterals and rhotics in English involve various stages, including the transition into the liquid, the steady state, and the transition into the following vowel (Carter and Local, 2007). The current study uses the steady-state portion to define the liquid as in previous studies (Flege *et al.*, 1995; Kirkham, 2017). Although the liquid steady-state is an approximation given the various stages involved in the liquid acoustics mentioned above, this issue can be minimised in the dynamic analysis because it shows holistic time-varying trajectories across the liquid and vowel.

E. Acoustic analysis

This study analyses 2306 liquid tokens for mid-point analysis and 2515 liquid-vowel tokens for dynamic analysis. The detailed breakdown is shown in Table II. The current

https://doi.org/10.1121/10.0024351

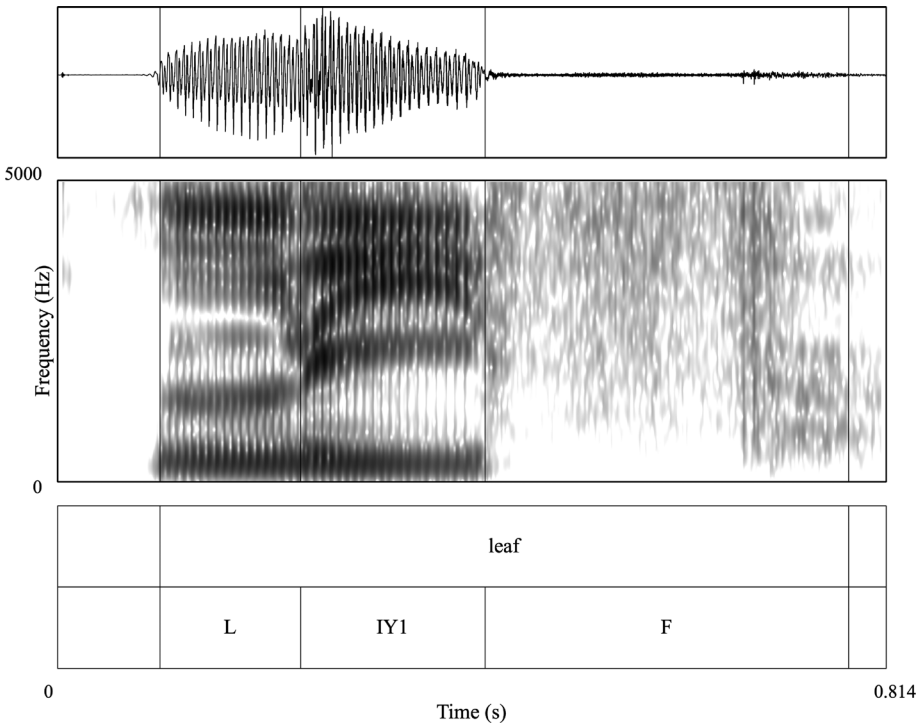

FIG. 1. Example spectrogram of an L1 North American L1 English speaker's production of *leaf*. Labels show phonetic segments in ARPABET, in which "IY1" indicates a stressed high front unrounded vowel /i/.

study compares two acoustic parameters between L1 Japanese and L1 English speakers' production of English liquids: (1) the distance between second (F_2) and first (F_1) formants (F_2–F_1) and (2) the third formant (F_3). F_2–F_1 is used as a measure to evaluate acoustic liquid quality; lower F_2–F_1 values can be related to darker realisations of liquids, resulting from a greater degree of tongue retraction (Howson and Redford, 2021; Sproat and Fujimura, 1993). F_3 is a primary acoustic dimension that distinguishes English /l/ and /ɹ/, and previous research reports robust differences between L1 Japanese and L1 English speakers' production of English liquids.

F_1, F_2, and F_3 values were estimated and extracted with Fast Track, an automatic formant estimation Praat plug-in (Barreda, 2021). Fast Track samples formant frequencies every 2 ms throughout the interval, resulting in smooth trajectories between F_1 and F_3. It then outputs the estimated formant frequencies while aggregating them in a specified

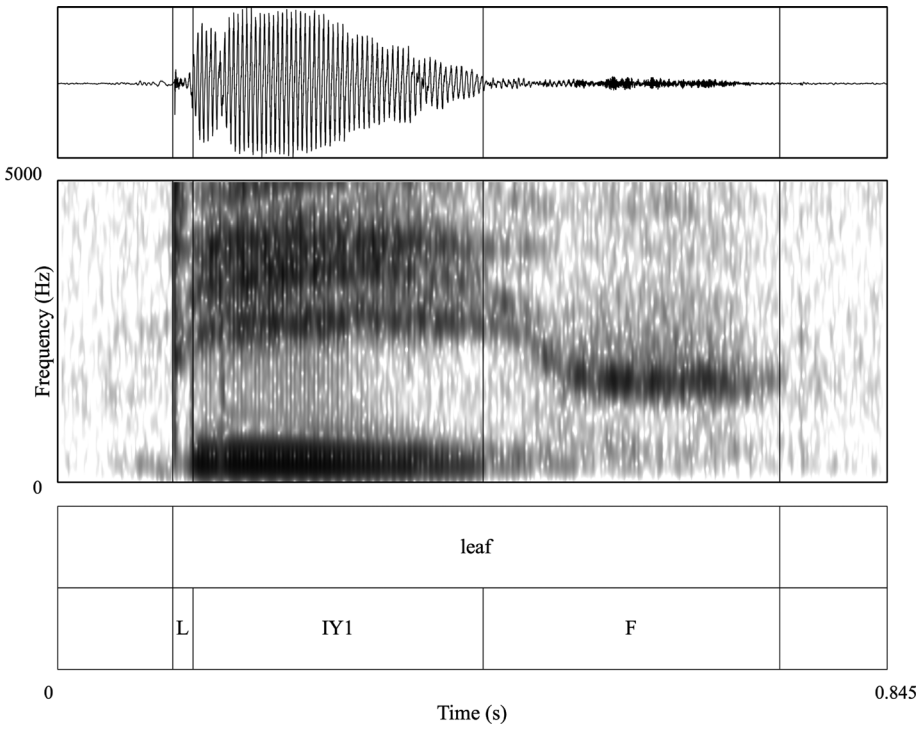

FIG. 2. Example spectrogram of a "definitely a tap" token of *leaf* produced by an L1 Japanese speaker. Labels show phonetic segments in ARPABET, in which "IY1" indicates a stressed high front unrounded vowel /i/.

TABLE II. The number of tokens per vowel context.

Vowel context	/i/	/æ/	/u/
L1 English			
Liquid[a]	155 / 187	188 / 173	119 / 130
Liquid-vowel[b]	177 / 197	199 / 192	130 / 134
L1 Japanese			
Liquid[a]	205 / 246	298 / 286	149 / 170
Liquid-vowel[b]	231 / 284	312 / 310	169 / 180

[a] /l/ tokens on the left; /ɹ/ tokens on the right.
[b] /l/+vowel tokens on the left; /ɹ/+vowel tokens on the right.

number of bins. The current analysis uses 11 data points throughout the liquid-vowel interval for each formant trajectory. The advantage of using Fast Track is that it performs multiple-step formant estimations by adjusting the maximum formant frequency and obtains the "best-winning" analysis based on regression analyses predicting the formant frequency as a function of time (Barreda, 2021). This achieves increased formant estimation accuracy by specifying different formant frequency ranges according to speakers' age and gender.

In the current study, the female and male speakers were analysed separately with different ranges of the upper formant frequency ceiling: between 5000–7000 Hz for female speakers and between 4500–6500 Hz for male speakers. Fast Track then performs 24-step formant estimations with varying upper-frequency ceilings and estimates the formant frequencies at 11 equidistant points during (1) the liquid and (2) the liquid-vowel intervals with a 25 ms window padded before and after the segment. After formant tracking, formant estimation errors can be corrected based on visual inspection of the 24-step analyses. Using this, I visually inspected all the tokens one by one and either improved the formant measurement by nominating a different winning analysis or omitted the tokens when none of the analyses looked reasonable. At this visual inspection stage, 118 tokens out of the 2633 tokens (see Sec. II D) were excluded due to poor formant estimation accuracy.

Finally, Fast Track automatically omits tokens when they are shorter than 30 ms as formant estimation can be challenging for extremely short tokens. As a result, 209 tokens were excluded from the dataset for the static analysis, leaving 2306 tokens for static analysis and 2515 tokens for dynamic analysis. The difference in the number of tokens reflects the greater number of liquid-only tokens being omitted automatically by Fast Track as they were inevitably shorter than liquid-vowel intervals (see supplementary material for the data processing procedure described here).[1]

F. Statistical analysis

All statistical analyses were performed using R version 4.2.2 (R Core Team, 2022) and data visualisation was performed using the *tidyverse* suite (Wickham *et al.*, 2019). Prior to the statistical analysis, the formant values were transformed into Bark scale using the *bark* function in the *emuR* package to allow for cross-speaker comparisons (Jochim *et al.*, 2023).

For the static analysis, Bark-converted F_2–F_1 (Bark F_2–F_1) and F_3 (Bark F_3) at liquid midpoint were modelled using linear mixed-effect models (LME) using the *lme4::lmer* function (Bates *et al.*, 2015). Separate models were constructed for /l/ and /ɹ/, respectively. The fixed effects included (1) the speaker's first language (*L1*: i.e., English vs Japanese), (2) vowel context (*vowel*), and (3) the speaker's gender (*gender*). No interactions were included because initial explorations suggested that the current dataset does not have the statistical power to detect interactions.

Furthermore, an anonymous reviewer suggested classifying the participants into groups according to their English proficiency and including this variable for analysis. Following this, I classified the participants into four groups based on the distribution of their subjective fluency rating scores. L1 Japanese speakers are classified into the *advanced* (rating 5–6, $n = 7$), *intermediate* (rating 4, $n = 14$), and *beginner* (rating 1–3, $n = 10$) groups. L1 English speakers constitute a group on their own (*L1 English*; rating 7, $n = 14$). The L1 English speaker group, however, confounds the *proficiency* variable with the *L1* variable, making the inclusion of the *proficiency* variable problematic. The issue is manifested in the rank-deficient warning for LMEs when both *L1* and *proficiency* are included in the same model, suggesting that two or more variables are not linearly independent from each other. A further analysis using the *caret::findLinearCombos* function shows co-linearity between *L1* and *proficiency* and suggests excluding the level of L1 English speakers from the *proficiency* variable.

For this reason, I perform a separate analysis focussing only on the L1 Japanese speakers' data to investigate the effects of *proficiency* and summarise the results at the end of the static analysis. I have included L1 Japanese speakers only here because inclusion of L1 English speakers might reduce the magnitude of between-group differences among L1 Japanese speakers. The visualisation includes L1 English speakers' data only for the purpose of comparison. I will not explore this extensively as this is not the main focus of the study (see supplementary material for further details of the analysis and results).[1]

The random effect structure for the linear models included by-participant varying slopes and by-participant varying intercepts for vowel contexts and by-word varying intercepts. As a result, the following specification is used for four final models (i.e., models predicting Bark F_2–F_1 and Bark F_3 for /l/ and /ɹ/):

$$\text{lmer}(\text{Bark } F_2\text{–}F_1 \text{ or Bark } F_3 \sim L1 + vowel + gender + (1 \mid word) + (1 + vowel \mid speaker)).$$

The significance of the fixed effects was tested via likelihood ratio testing by comparing the full model and the nested model excluding the fixed effect in question (Winter, 2020). If the full model significantly improved the model fit, I concluded that the main effect significantly influenced the outcome variable. The patterns associated with the vowel

https://doi.org/10.1121/10.0024351

contexts are interpreted via data visualisation for the sake of model simplicity (see supplementary material for additional statistical comparisons).[1]

Second, the dynamic formant analysis used generalised additive mixed models (GAMMs) using the *mgcv::bam* function (Wood, 2017). The non-linear differences between contours can be evaluated in light of *height* and *shape* of the trajectories; the *height* dimension can be modelled via parametric terms, and the *shape* dimension via so-called *smooth terms* that specify the degree of wiggliness of contours (Sóskuthy *et al.*, 2018). Differences between a set of contours can also be directly modelled by incorporating a *reference smooth* (i.e., a contour at the reference level) and the *difference smooth* (i.e., a contour that models the degree of by-group difference of contours) (Sóskuthy, 2017). For more details about GAMMs, please be referred to the existing tutorial papers (e.g., Sóskuthy, 2017; Sóskuthy *et al.*, 2018; Wieling, 2018).

In the current study, I focus on differences in trajectory height and shape between the speaker groups (i.e., English vs Japanese). Separate models were constructed for each combination of the liquid-vowel pairings. Each model predicts the formant values, either Bark F_2–F_1 or Bark F_3, by a parametric term of the speaker's first language and gender, as well as a time-varying reference smooth, a time-varying by-L1 difference smooth, and a time-varying by-gender smooth. It also includes time-by-speaker and time-by-word random smooths.

Note, again, that English proficiency was not included in the GAMMs models together with *L1* as this resulted in inaccurate predictions of the formant trajectories compared to the visualisations of the raw data. Instead, similarly to the linear mixed-effect model analysis, I conducted a separate analysis for the effects of *proficiency* using the L1 Japanese speakers' data only and summarise the relevant results at the end of the dynamic analysis. The choice of including L1 Japanese speakers only reflects the consideration that L1 English speakers' trajectories may be different in both shape and height, which would make it difficult for me to interpret whether statistically significant differences result from speakers' L1 or L1 Japanese speakers' proficiency. This is clear in the visualisations in Figs. 9 and 10, in which L1 English speakers' trajectories are distinct from the three groups of L1 Japanese speakers (see supplementary material for further details).[1]

Residual autocorrelations in the trajectories were corrected using the autoregressive error model (AR model). The autoregressive parameter (rho: ρ) was set as the amount of autocorrelation at lag 1 in the model, estimated using the *start_value_rho* function in the *itsadug* package (van Rij *et al.*, 2020). While this is usually an adequate estimate, the residual autocorrelations were negative in some cases, indicating that a lower value would be optimal (Sóskuthy *et al.*, 2018; Wieling, 2018). In such cases, the new rho value was determined by exploring a range of values and visualising the autocorrelations at lag 1 for each rho value. The final model specification across 12 models (two outcome variables, i.e., Bark F_2–F_1 and Bark F_3) for two liquids (i.e., /l/ and /ɹ/) in three vowel contexts (i.e., /æ/, /i/ and /u/) is

bam(Bark F_2–F_1 or Bark F_3 ~ *L1* + *gender* + s(*time*, bs = "cr") + s(*time*, by = *L1*, bs = "cr") + s(*time*, by = *gender*, bs = "cr") + s(*time*, *speaker*, bs = "fs," xt = "cr," m = 1) + s(*time*, *word*, bs = "fs," xt = "cr," m = 1), method = "ML").

Trajectory height and shape were compared through model comparisons using the *itsadug::compareML* function following the previous research (Kirkham *et al.*, 2019; Sóskuthy, 2017; Sóskuthy *et al.*, 2018) as follows:

(1) I first compared (1) the full model and (2) the nested model excluding the parametric and the smooth terms associated with the speaker's *L1* or *gender*. This allows a comparison of the overall differences associated with these effects in both height and shape between the two contours.
(2) If the above comparison showed a significantly improved model fit of the full model, I then compared (1) the full model and (2) the nested model including the parametric term of *L1* or *gender* but still excluding the by-L1 or by-gender smooth term. This tests whether the two contours differ significantly in shape.

If the full model was still better in the model fit after procedure 2 above, I concluded that both trajectory height and shape were different at a statistically significant level. If the full model improved the model fit for procedure 1 but not for procedure 2, then there was only a difference in trajectory height. Otherwise, I concluded that there was little evidence that the two trajectories are significantly different.

III. RESULTS

A. Liquid static analysis

In this section, I first present the liquid midpoint analysis of F_2–F_1 and F_3 using LMEs in order to investigate the overall trends in liquid quality. The static analysis tests the main effects of *L1*, *vowel*, and *gender* while the liquid-vowel interactions are interpreted via data visualisation. Note that the baseline participant population (i.e., intercept) is the female L1 English speakers in the /æ/ context but the gender is referred to only when the *gender* effect is discussed (see supplementary material for an additional analysis of vowel midpoints).[1]

1. F_2-F_1 midpoint

The model summaries for the F_2–F_1 models are shown in Table III. The lateral F_2–F_1 model predicts that L1 Japanese speakers produce laterals higher at 8.83 Bark than L1 English speakers (6.74 Bark). F_2–F_1 for laterals slightly varies according to the vowel context; F_2–F_1 is the highest in the /i/ context with an averaged F_2–F_1 being at 8.02 Bark, followed by /u/ (7.54 Bark) and /æ/ (6.74 Bark). Male speakers produce laterals with lower F_2–F_1 values at 6.06 Bark.

TABLE III. LME summary: Liquid F_2–F_1 (Bark).

Variable	β	SE	t	$p(\chi^2)$
Lateral /l/				
Intercept	6.74	0.33	20.36	
L1				<0.001
Japanese	1.99	0.38	5.25	
Vowel				<0.001
/i/	1.28	0.16	8.23	
/u/	0.80	0.18	4.50	
Gender				0.072
Male	−0.68	0.36	−1.86	
Rhotic /ɹ/				
Intercept	6.38	0.34	18.53	
L1				<0.001
Japanese	1.86	0.40	4.68	
Vowel				<0.001
/i/	1.15	0.16	7.09	
/u/	0.60	0.13	4.58	
Gender				0.070
Male	−0.75	0.38	−1.97	

The rhotic F_2–F_1 model predicts that L1 English speakers produce rhotics in the /æ/ context at 6.38 Bark and L1 Japanese speakers overall produce 8.24 Bark. It also predicts higher F_2–F_1 overall in the /i/ context (7.53 Bark) and in the /u/ context (6.98 Bark) than in the /æ/ context. Similar to the laterals, male speakers produce rhotics with lower F_2–F_1 values at 5.63 Bark.

Overall, L1 Japanese speakers produce both English /l/ and /ɹ/ with consistently higher F_2–F_1 than L1 English speakers across vowel contexts (Fig. 3), and this is supported by the significant main effect of *L1* for both /l/ [$\chi^2(1) = 17.58$, $p < 0.001$], and /ɹ/ [$\chi^2(1) = 15.68$, $p < 0.001$]. The main effect of *vowel* is also shown to be significant for both /l/ [$\chi^2(2) = 22.74$, $p < 0.001$] and /ɹ/ [$\chi^2(1) = 22.35$, $p < 0.001$]. While male speakers produce liquids with lower F_2–F_1 values than female speakers, this difference was not shown to be statistically significant for either laterals [$\chi^2(1) = 3.23$, $p = 0.073$], or rhotics [$\chi^2(1) = 3.28$, $p = 0.070$].

2. F_3 midpoint

The model summaries for the F_3 models are shown in Table IV. The lateral F_3 model predicts that L1 English speakers produce F_3 at 15.83 Bark for /l/ while L1 Japanese speakers have a slightly lower F_3 at 15.54 Bark. Although model comparisons suggest significant effects of *vowel* for /l/ [$\chi^2(2) = 13.05$, $p = 0.001$], the difference seems to be quite minor; the model predicts 15.65 Bark for /l/ in the /i/ context and 15.44 Bark in the /u/ context. Finally, female speakers produce laterals with higher F_3 values by 1.12 Bark than male speakers overall.

The rhotic F_3 model predicts that L1 English speakers produce 12.17 Bark for /ɹ/ where L1 Japanese speakers produce higher F_3 at 14.05 Bark. Similar to the laterals, slight differences are found for /ɹ/ in the /i/ and /u/ contexts compared to /æ/; the model predicts 12.54 Bark in the /i/ context and 12.21 Bark in the /u/ context. The main effect of *vowel* is also significant here [$\chi^2(2) = 13.78$, $p = 0.001$].

While the main effect of *vowel* influences the F_3 values only slightly for both /l/ and /ɹ/, the effects of *L1* are suggested to be significant for /ɹ/ [$\chi^2(1) = 30.62$, $p < 0.001$] but not for /l/ [$\chi^2(1) = 1.97$, $p = 0.161$]. Figure 4 seems to suggest a bimodal distribution in F_3 (Bark) for L1 English speakers, especially for /l/ in the /i/ and /u/ contexts. This seems to result from gender-related differences, in which male speakers produced liquids with lower F_3 values than female speakers. Indeed, the effects of *gender* are shown to be statistically significant for both laterals [χ^2 (1) = 22.70, $p < 0.001$] and rhotics [χ^2 (1) = 15.87, $p < 0.001$].

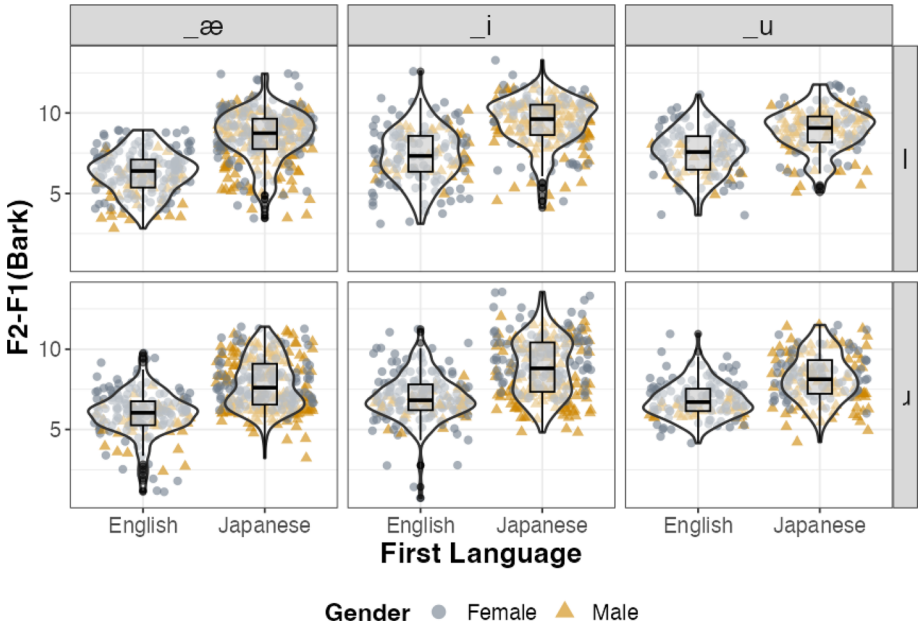

FIG. 3. (Color online) F_2–F_1 (Bark) at liquid midpoint. Each column shows vowel contexts for /l/ (top row) and /ɹ/ (bottom row). Each panel shows distributions of F_2–F_1 (Bark) for L1 English (left) and L1 Japanese (right) speakers. Overlaid is the scatterplot indicating speaker's gender: female (gray circles) and male (yellow triangles) speakers.

https://doi.org/10.1121/10.0024351

TABLE IV. LME summary: Liquid F_3 (Bark).

Variable	β	SE	t	$p(\chi^2)$
Lateral /l/				
Intercept	15.83	0.18	89.35	
L1				0.016
Japanese	−0.29	0.20	−1.44	
Vowel				0.001
/i/	−0.18	0.08	−2.12	
/u/	−0.39	0.08	−4.82	
Gender				<0.001
Male	−1.12	0.19	−5.79	
Rhotic /ɹ/				
Intercept	12.17	0.25	48.56	
L1				<0.001
Japanese	1.88	0.26	7.18	
Vowel				0.001
/i/	0.37	0.08	4.47	
/u/	0.04	0.10	0.41	
Gender				<0.001
Male	−1.15	0.25	−4.53	

3. Effects of L2 proficiency on the midpoint formant measurement

In addition to the main analysis, the effects of *proficiency* are tested for the three groups of L1 Japanese speakers. Grouping is based on their subjective fluency judgement scores: beginner ($n = 10$, rating 1–3), intermediate ($n = 14$, rating 4), and advanced ($n = 7$, rating 5–6). Similarly to the main analysis, separate LME were specified in which Bark F_2–F_1 or Bark F_3 are predicted by fixed effects of *proficiency*, *vowel*, and *gender* with by-item random intercepts and by-speaker random slopes and intercepts for vowels. The results are visualised in Figs. 5 and 6.

The F_2–F_1 models suggested statistically significant effects of *proficiency* on Bark F_2–F_1 for /ɹ/ [$\chi^2(2) = 7.52$, $p = 0.002$], in which the advanced L1 Japanese learners of English produce rhotics with lower F_2–F_1 than those in the beginner and intermediate groups. No statistically significant *proficiency* effects are found for /l/ [$\chi^2(2) = 0.12$, $p = 0.94$]. For Bark F_3, no statistically significant effects of *proficiency* are found for either /l/ [$\chi^2(2) = 0.81$, $p = 0.67$] or /ɹ/ [$\chi^2(2) = 0.057$, $p = 0.97$].

4. Summary: Static analysis

L1 Japanese speakers produce higher F_2–F_1 for both /l/ and /ɹ/ across vowel contexts. F_3 values for /l/ are only slightly lower for L1 Japanese speakers while they produce /ɹ/ with higher F_3 than L1 English speakers across vowel contexts. Male speakers produce liquids with lower F_2–F_1 and F_3 values, and this was particularly the case for F_3. Finally, L1 Japanese speakers in the advanced group produced lower F_2–F_1 than the other groups for /ɹ/.

B. Dynamic analysis

Dynamic analysis in this section now focuses on variation in F_2–F_1 and F_3 trajectories across the liquid-vowel interval using GAMMs. In the visualisation of the liquid-vowel trajectories (Figs. 7 and 8), the liquid portion corresponds roughly to the first third of the interval whereas the vowel corresponds to the second two-thirds. Note the visualisation shows the predictions based on the full models.

1. F_2–F_1 liquid-vowel trajectory

The results of the model comparisons for the F_2–F_1 dynamic analysis are shown in Table V for laterals and Table VI for rhotics. The visualisations are shown in Fig. 7. The model comparisons show that the height and shape of the F_2–F_1 trajectories are significantly different between L1 English and L1 Japanese speakers for both liquids in all vowel contexts. The visualisations of the GAMMs show that

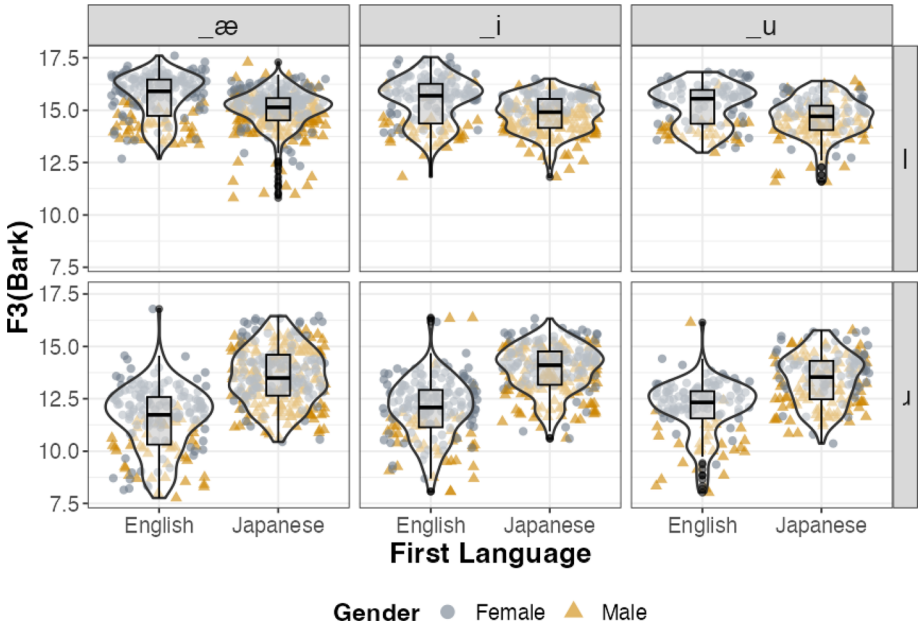

FIG. 4. (Color online) F_3 (Bark) at liquid midpoint. Each column shows vowel contexts for /l/ (top row) and /ɹ/ (bottom row). Each panel shows distributions of F_3 (Bark) for L1 English (left) and L1 Japanese (right) speakers. Overlaid is the scatterplot indicating speaker's gender: female (gray circles) and male (yellow triangles) speakers.

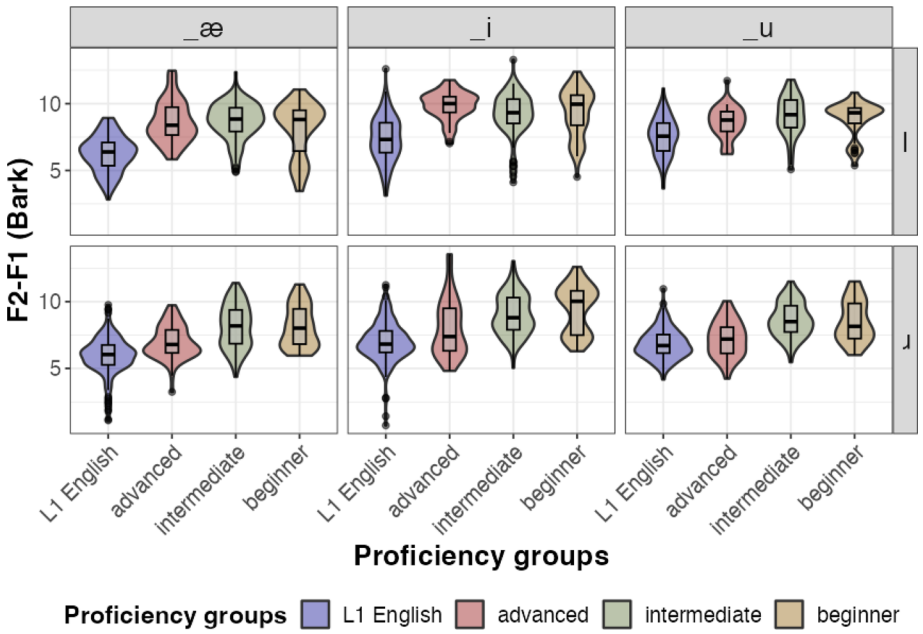

FIG. 5. (Color online) F_2–F_1 (Bark) at liquid midpoint by proficiency groups. Each column shows vowel contexts for /l/ (top row) and /ɹ/ (bottom row). Each panel shows distributions of F_2–F_1 (Bark) for L1 English speakers and three groups of L1 Japanese speakers: advanced, intermediate, and beginner, from left to right.

the trajectories for L1 English and L1 Japanese speakers are similar in the /i/ context (the middle panels in Fig. 7) but look quite different in the /æ/ (left) and /u/ (right) contexts. L1 English speakers follow a similar tendency across the vowel contexts such that they start from lower F_2–F_1 values at the onset of the liquid, showing an increase towards the vowel target and a slight decrease towards the offset of the vowel.

L1 Japanese speakers, on the other hand, show distinct trajectory patterns depending on vowel context. In the /i/ context, their trajectories follow a similar tendency to that of L1 English speakers, but with an earlier rise from the liquid onset towards the vowel resulting in a consistently higher trajectory than L1 English speakers in the first half of the interval. In the /æ/ context, on the other hand, the L1 Japanese speakers show an opposite pattern to L1 English speakers, in which F_2–F_1 values are the highest earlier during the first third of the interval and decrease to the vowel with a small rise towards the end of the interval. Finally, the L1 Japanese speakers' trajectories in the /u/ context show smaller fluctuations than that of L1 English speakers; the trajectory shows almost a linear and monotonic decrease in this vowel context.

Differences associated with *gender* are statistically significant for trajectory height but not for shape for both laterals and rhotics across the vowel contexts. This suggests almost linear differences between female and male speakers' trajectories, in which female speakers show constantly higher trajectories than male speakers, and this is evident in Fig. 7.

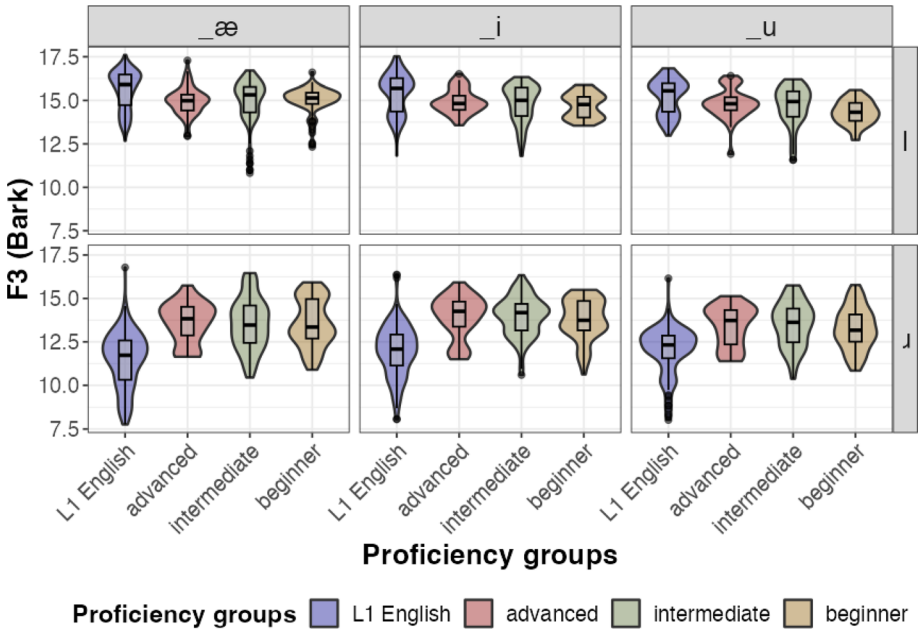

FIG. 6. (Color online) F_3 (Bark) at liquid midpoint by proficiency groups. Each column shows vowel contexts for /l/ (top row) and /ɹ/ (bottom row). Each panel shows distributions of F_2–F_1 (Bark) for L1 English speakers and three groups of L1 Japanese speakers: advanced, intermediate, and beginner, from left to right.

https://doi.org/10.1121/10.0024351

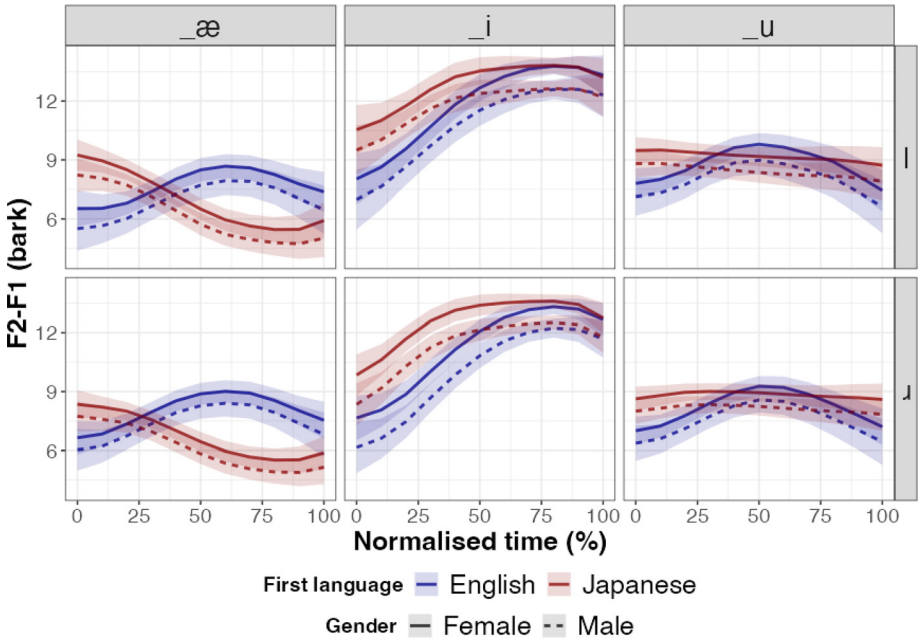

FIG. 7. (Color online) The F_2–F_1 (Bark) trajectories predicted by GAMMs over the liquid-vowel intervals for each liquid (rows) in each vowel context (columns). Each panel shows predictions based on the full model with a mean smooth and 95% confidence interval for L1 English (blue) and L1 Japanese (red) speakers and for female (solid) and male (dashed) speakers.

2. F_3 liquid-vowel trajectory

The model comparisons for F_3 are shown in Table VII for laterals and in Table VIII for rhotics. The visualisations are shown in Fig. 8. The lateral-vowel trajectories (the top half of Fig. 8) show similarities between L1 English and L1 Japanese speakers. The model comparisons suggest that, while the trajectory shape and height are different between L1 English and L1 Japanese speakers in the /i/ context, the trajectories in the /æ/ and /u/ contexts are not statistically significantly different, with the L1 Japanese speakers' trajectories being slightly lower, especially in the first half of the interval.

Even in the lateral-/i/ context where trajectory height and shape are statistically significant, however, a closer look at the GAMMs model specifications and the model comparisons suggest that the difference between the two trajectories is marginal. Neither parametric or smooth terms associated with the *L1* difference were statistically significant in the model summary [$\beta = 0.18$, standard error $(SE) = 0.10$, $t = 1.83$, $p = 0.07$ for the parametric term; $F(6.05) = 1.79$, $p = 0.09$ for the difference smooth]. The model comparison also suggests only a marginal improvement in the Akaike Information Criterion (AIC) values (1561.42 for the full model and 1565.84 for the nested model). Figure 8 also shows that the 95% confidence intervals of two trajectories overlap substantially throughout the liquid-vowel interval.

The /ɹ/-vowel trajectories for F_3, on the other hand, show statistically significant differences in both trajectory

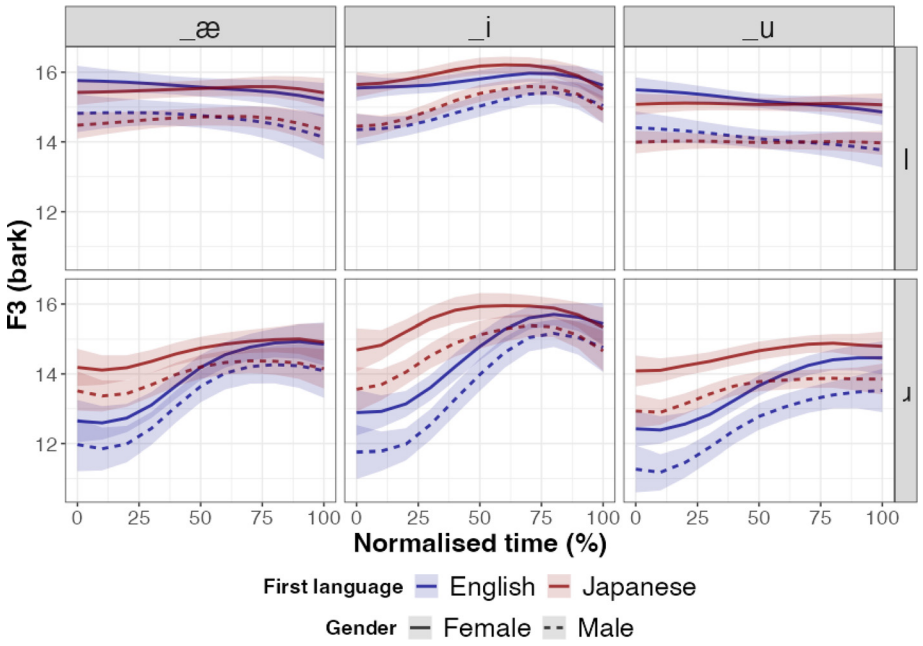

FIG. 8. (Color online)The F3 (Bark) trajectories predicted by GAMMs over the liquid-vowel intervals for each liquid (rows) in each vowel context (columns). Each panel shows predictions based on the full model with a mean smooth and 95% confidence interval for L1 English (blue) and L1 Japanese (red) speakers and for female (solid) and male (dashed) speakers.

TABLE V. Model comparisons for F_2–F_1 GAMMs for laterals.

Comparison	χ^2	*df*	$p(\chi^2)$
/l/: /æ/ context			
Overall: L1	69.88	3	<0.001
Shape: L1	66.63	2	<0.001
Overall: gender	6.67	3	0.004
Shape: gender	1.10	2	0.333
/l/: /i/ context			
Overall: L1	16.54	3	<0.001
Shape: L1	9.68	2	<0.001
Overall: gender	18.91	3	<0.001
Shape: gender	0.34	2	0.712
/l/: /u/ context			
Overall: L1	25.41	3	<0.001
Shape: L1	23.67	2	<0.001
Overall: gender	4.02	3	0.045
Shape: gender	0.07	2	0.929

TABLE VI. Model comparisons for F_2–F_1 GAMMs for rhotics.

Comparison	χ^2	*df*	$p(\chi^2)$
/ɹ/: /æ/ context			
Overall: L1	53.57	3	<0.001
Shape: L1	45.94	2	<0.001
Overall: gender	4.10	3	0.042
Shape: gender	0.06	2	0.938
/ɹ/: /i/ context			
Overall: L1	39.40	3	<0.001
Shape: L1	24.09	2	<0.001
Overall: gender	21.90	3	<0.001
Shape: gender	0.33	2	0.723
/ɹ/: /u/ context			
Overall: L1	21.62	3	<0.001
Shape: L1	17.83	2	<0.001
Overall: gender	4.00	3	0.046
Shape: gender	0.02	2	0.985

TABLE VII. Model comparisons for F_3 GAMMs for laterals.

Comparison	χ^2	*df*	$p(\chi^2)$
/l/: /æ/ context			
Overall: L1	3.12	3	0.100
Shape: L1	—	—	—
Overall: gender	17.57	3	<0.001
Shape: gender	1.22	2	0.295
/l/: /i/ context			
Overall: L1	4.43	3	0.031
Shape: L1	2.53	2	0.080
Overall: gender	33.71	3	<0.001
Shape: gender	5.67	2	0.003
/l/: /u/ context			
Overall: L1	1.81	3	0.306
Shape: L1	—	—	—
Overall: gender	29.91	3	<0.001
Shape: gender	0.00	2	1.000

TABLE VIII. Model comparisons for F_3 GAMMs for rhotics.

Comparison	χ^2	*df*	$p(\chi^2)$
/ɹ/: /æ/ context			
Overall: L1	17.36	3	<0.001
Shape: L1	10.32	2	<0.001
Overall: gender	8.26	3	<0.001
Shape: gender	1.05	2	0.350
/ɹ/: /i/ context			
Overall: L1	43.55	3	<0.001
Shape: L1	26.89	2	<0.001
Overall: gender	22.40	3	<0.001
Shape: gender	3.21	2	0.041
/ɹ/: /u/ context			
Overall: L1	27.42	3	<0.001
Shape: L1	8.31	2	<0.001
Overall: gender	25.96	3	<0.001
Shape: gender	2.87	2	0.057

height and shape in all vowel contexts, although both L1 English and L1 Japanese speakers share a similar trend in the visualisation in Fig. 8. Both groups show lower F_3 values at the liquid onset, which then increase towards the vowel, where L1 English and L1 Japanese speakers' trajectories seem to converge. L1 Japanese speakers' trajectories are overall flatter and higher than that of L1 English speakers across all the vowel contexts.

Finally, similarly to the F_2–F_1 results, the *gender* effect seems to be statistically significant only for the trajectory height. This again suggests that the difference between trajectories for female and male speakers is close to linear (see Fig. 8).

3. Effects of L2 proficiency on formant trajectories

Similar to the static analysis, the effects of L1 Japanese speakers' proficiency have been tested separately from the main analysis. For each liquid-vowel pairing, the models predicts Bark F_2–F_1 or Bark F_3 with parametric terms of *proficiency* and *gender*, a time-varying reference smooth, a time-varying by-proficiency difference smooth, and a time-varying by-gender difference smooth. The random effect is accounted for by time-by-speaker and time-by-word random smooths. The visualisations are shown in Figs. 9 and 10; please note that the predictions shown in these figures are based on the models excluding parametric and smooth terms associated with *gender* because the plots would be too crowded to interpret otherwise.

The analyses for Bark F_2–F_1 suggest a statistically significant effect of *proficiency* on the trajectory height for /ɹ/ in the /u/ context [$\chi^2(6) = 10.24$, $p = 0.002$], in which the F_2–F_1 trajectory for the advanced group is lower than the beginner or intermediate groups. The visualisation in Fig. 9, however, shows that the trajectory shape is quite different between L1 English speakers and the advanced L1 Japanese speakers. For Bark F_3, no statistically significant effects of *proficiency* are found for either /l/ or /ɹ/ for the L1 Japanese speakers.

https://doi.org/10.1121/10.0024351

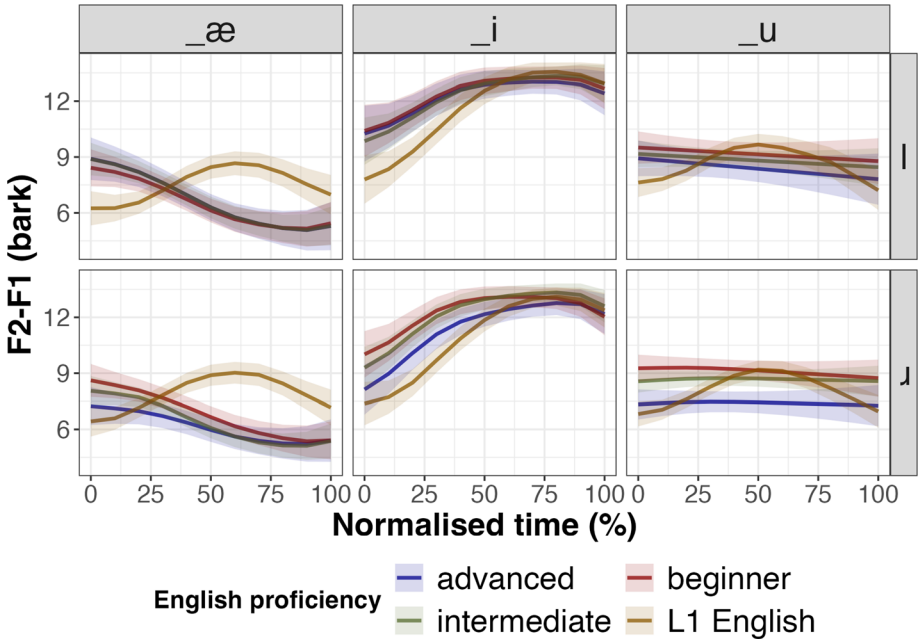

FIG. 9. (Color online) The F_2–F_1 (Bark) trajectories illustrating differences between the different proficiency groups among L1 Japanese speakers predicted by GAMMs over the liquid-vowel intervals for each liquid (rows) in each vowel context (columns). Each panel shows predictions based on the model excluding parametric and smooth terms associated with *gender* for simplicity, with a mean smooth and 95% confidence interval for advanced (blue), intermediate (red), beginner (green) L1 Japanese speakers and L1 English speakers (orange).

4. Summary: Dynamic analysis

The dynamic analysis shows substantial variability in the liquid-vowel realisations between L1 English and L1 Japanese speakers. Shape and height are significantly different for the F_2–F_1 trajectories for both /l/ and /ɹ/, with differences associated not only with the liquid portion corresponding to the first third of the interval but also with the transition patterns into the vowel. The F_3 trajectories for /l/ are largely comparable between L1 English and L1 Japanese speakers with little evidence of statistically significant differences. The F_3 trajectories for /ɹ/, on the other hand, differ substantially in the first half of the interval corresponding to the liquid portion. The effects of *gender* are manifested almost exclusively on the trajectory height, meaning a linear difference between trajectories for female and male speakers. Although advanced L1 Japanese speakers produced the lower F_2–F_1 trajectories in the /ɹ/-/u/ context than the beginner and intermediate groups, the trend is quite different from that of L1 English speakers.

IV. DISCUSSION

A. Spectro-temporal variability in L2 English liquids

The current paper aims to capture time-varying acoustic properties of English liquids produced by L1 English and L1 Japanese speakers. It combines two analyses of F_2–F_1 and F_3: the static analysis at the liquid midpoint and the dynamic analysis over the liquid-vowel interval. The liquid midpoint

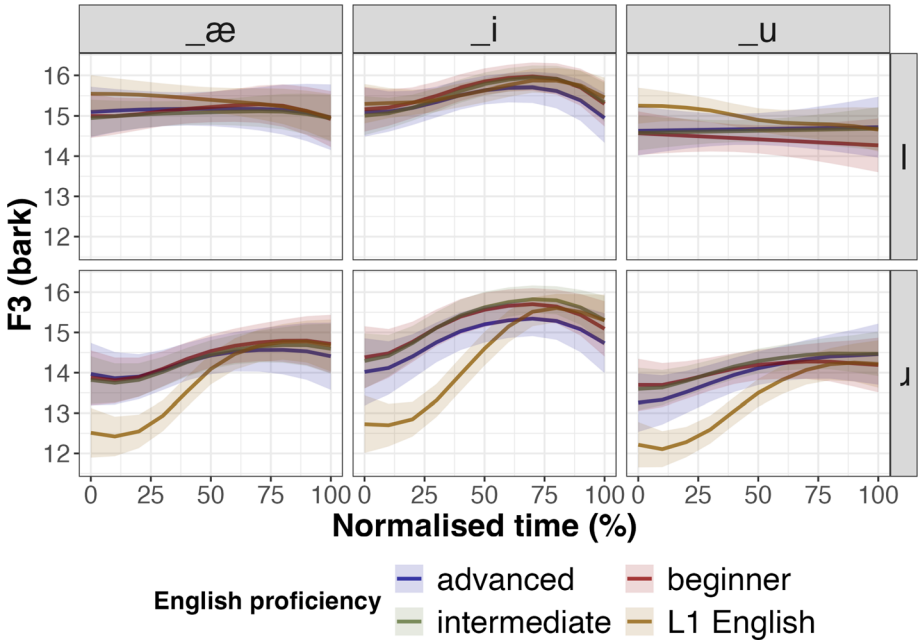

FIG. 10. (Color online) The F3 (Bark) trajectories illustrating differences between the different proficiency groups among L1 Japanese speakers predicted by GAMMs over the liquid-vowel intervals for each liquid (rows) in each vowel context (columns). Each panel shows predictions based on the model excluding parametric and smooth terms associated with *gender* for simplicity, with a mean smooth and 95% confidence interval for advanced (blue), intermediate (red), beginner (green) L1 Japanese speakers and L1 English speakers (orange).

analysis suggests that L1 Japanese speakers constantly produce higher F_2–F_1 for both English /l/ and /ɹ/ and higher F_3 for /ɹ/ than L1 English speakers across vowel contexts. The dynamic analysis, on the other hand, shows that the between-L1 differences are non-linear, highlighting the complexity associated with the production of liquids and liquid-vowel coarticulation.

Comparing the effects of speaker gender and L1 demonstrate the importance of dynamic information in the liquid-vowel sequences. The static analysis shows that male speakers generally produce English liquids with lower F_2–F_1 and F_3 frequencies than female speakers, and the gender difference is statistically significant for F_3. The dynamic analysis further shows clearly that the spectral difference between female and male speakers seems to be linear; GAMMs model comparisons suggest statistically significant differences in trajectory height but not in trajectory shape, and it is quite clear from the visualisations in Figs. 7 and 8 that the differences in trajectories between female and male speakers are (almost) linear.

The dynamic difference associated with speaker L1, on the other hand, draws a much more complicated picture. While the time-varying analysis of F_3 for /l/ indicates little difference between L1 English and L1 Japanese speakers, the F_3 values for /ɹ/ show a clear between-L1 difference in the first half of the interval, indicating differences in acoustic realisations of liquids and the transition into the vowel. Also, the trajectory shape associated with L1 Japanese speakers' /ɹ/ is flatter, resulting in a smaller distinction between /ɹ/ and the vowels. The two language groups slightly differ in the point in time at which F_3 achieves its maximum, such that L1 Japanese speakers seem to achieve the vowel target earlier than L1 English speakers do.

The F_2–F_1 trajectories further highlight the non-linear between-L1 differences in the trajectory (Fig. 7). In particular, L1 Japanese speakers show distinct trajectory patterns across vowel contexts, suggesting that their production of English liquids is subject to greater influence from the following vowels than that of L1 English speakers. The liquid-/i/ trajectories, for example, suggest that L1 Japanese speakers reach the vowel target earlier, given the early onset of the plateau, than L1 English speakers despite a similar trajectory pattern. The linear trend for the liquid-/u/ trajectories also indicates that L1 Japanese speakers do not clearly distinguish the liquid and the vowel on F_2–F_1.

The separate static analyses on the effects of L1 Japanese speakers' English proficiency demonstrated that advanced L1 Japanese-speaking learners of English produced lower F_2–F_1 values for /ɹ/ than the other two groups. Given that L1 English speakers produced lower F_2–F_1 values for /ɹ/ at the liquid midpoint, the findings support the previous claims that English /ɹ/ is easier for L1 Japanese speakers to learn than English /l/ (Aoyama *et al.*, 2004), and that the use of F_2 and F_1 may be easier for them to acquire than that of F_3 (Saito and Munro, 2014). The dynamic analysis in the current study further demonstrates that advanced L1 Japanese speakers' F_2–F_1 trajectory is statistically significantly lower in the /ɹ/-/u/ context than the other two groups. While this could be taken as evidence of the *proficiency* effects, the linear trend of the trajectories across proficiency groups also suggests that even advanced L1 Japanese speakers do not seem to differentiate /ɹ/ and /u/. Fundamentally, this lack of liquid-vowel differentiation might demonstrate a general influence from their L1 (i.e., Japanese). Further research is clearly needed to investigate the effects of L2 proficiency on the formant dynamics by employing more rigorous measures of L2 proficiency, especially given that acoustic profiles of L2 English liquids can be complex (Aoyama *et al.*, 2019).

Overall, the dynamic analysis suggests that L1 Japanese speakers seem to differ not only in acoustic targets of English liquids, as captured in the static analysis, but also in the transition between the liquid and the vowel. The results are in line with the previous findings that the magnitude and timing of spectral changes differ in the production of English liquids by L1 English-speaking children (Howson and Redford, 2021) and by L2 learners of English (Espinal *et al.*, 2020) from that of adult L1 English speakers. These non-linear between-language differences could point to some possible mechanisms whereby L1 Japanese speakers struggle to produce English liquids accurately in light of L2 speech learning.

B. Acquisition of English /l/ and /ɹ/ by L1 Japanese speakers

The overarching question in this study concerns how L1 Japanese speakers differ from L1 English speakers in dynamic acoustic realisations of word-initial English liquids as a function of following vowels. The static analysis suggests that both speaker's L1 and vowel context influence the acoustic realisations of word-initial English /l/ and /ɹ/. The L1 effect is unsurprising, given that it largely agrees with previous findings that L1 Japanese speakers produce both English /l/ and /ɹ/ with higher F_2 and F_3 values than L1 English speakers (Aoyama *et al.*, 2019; Flege *et al.*, 1995; Saito and van Poeteren, 2018). Regarding the vowel effect, the static analysis suggests a general tendency that liquids in the /i/ context are produced with higher F_2–F_1 values than in the /u/ context, whereas the /æ/ context seems to facilitate the lowest F_2–F_1 values for liquids. This could be explained in light of previous findings that the F_2 values in English liquids tend to be higher when preceding a high vowel /i/ than a low vowel /a/ due to different articulatory demands on the tongue dorsum configurations (Recasens, 2012).

The dynamic results demonstrate that L1 Japanese speakers show different patterns of liquid-vowel coarticulatory patterns depending on the following vowel compared to L1 English speakers whose trajectory patterns are consistent across the vowel contexts. The liquid-/u/ trajectories, in particular, suggest that L1 Japanese speakers make a less clear distinction between the liquid and the vowel in the /u/ context. This could corroborate previous perceptual findings that L1 Japanese speakers are more likely to perceive a /w/-like percept when perceiving English /l/ and /ɹ/, resulting in a confusion between English /l ɹ/ and other categories

(e.g., /w/ or [ɯɾ]) and therefore in less success in identifying word-initial liquids in the back vowel context than in the front vowel context (Best and Strange, 1992; Guion *et al.*, 2000; Mochizuki, 1981; Shimizu and Dantsuji, 1983). The data in this study demonstrate that such confusion arising from the vocalic component of English liquids in perception could also be observed in L1 Japanese speakers' production.

Generally, L1 Japanese speakers produce higher F_3 for English /ɹ/ (Aoyama *et al.*, 2019; Flege, 1995; Saito and Munro, 2014). This is apparent in both static and dynamic analyses; in particular, the dynamic analysis for F_3 in Fig. 8 shows that by-group difference largely lies during the liquid portion, suggesting that the difference in F_3 would be attributed to the liquid realisations. Previous research claims that F_2 is an easier acoustic cue for L1 Japanese speakers to acquire (e.g., Saito and Munro, 2014; Saito and van Poeteren, 2018). While variations in F_1 could be negligible between the two speaker populations (e.g., Flege *et al.*, 1995; Saito and Munro, 2014), this claim does not explain well why the F_2–F_1 trajectories, which could derive from variations of F_2, are significantly different both in height and shape between L1 Japanese and L1 English speakers (see Fig. 7). It could therefore be argued that the static analysis only captures a snapshot of acoustical realisations of English liquids, when, in fact, L1 Japanese speakers differ from L1 English speakers in the dynamic spectral characteristics during the liquid-vowel interval.

In addition, an anonymous reviewer suggested a possibility that L1 Japanese speakers might use different dynamic strategies to make a contrast (e.g., through F_2) compared to L1 English speakers. It would, therefore, be worthwhile to investigate how L1 Japanese speakers use dynamic information to make such a phonological contrast, given especially that the Perceptual Assimilation model of L2 Speech Learning (PAM-L2) makes predictions about how L2 speakers assimilate L2 phonological contrasts into their L1 phonology (Best and Tyler, 2007).

Theoretically, the Speech Learning model (SLM) posits that L2 learners store representations of the L2 sounds at the level of the position-sensitive allophones (Flege, 1995; Flege and Bohn, 2021), and previous studies show that L1 Japanese speakers' perception of English /l/ and /ɹ/ is highly subject to the phonetic context and the coarticulatory effects with neighbouring segments (Mochizuki, 1981; Sheldon and Strange, 1982). Taken together, the current results demonstrate that L1 Japanese speakers are influenced by the phonetic details of L2 English liquids, not only in perception but also in production; L1 Japanese speakers show different patterns in the way they dissociate the liquid and vowel clearly, especially in the /u/ context, manifested in their production as different patterns of liquid-vowel coarticulation.

To summarise, the present study shows that the temporal spectral changes during the liquid-vowel intervals are significantly different between L1 English and L1 Japanese speakers along F_2–F_1 for both liquids and F_3 for /ɹ/. The liquid-vowel trajectories of F_2–F_1 in the /i/ and /u/ contexts highlight particularly notable temporal variability in the L1 Japanese speakers' data, suggesting that the liquid-vowel coarticulation could be considered as one of the production properties that L1 Japanese speakers need to acquire in production of English liquids.

V. CONCLUSION

The present study examines the acoustics of L1 Japanese and L1 English speakers' production of word-initial English liquids. The key findings include that L1 Japanese speakers differ in the coarticulatory pattern between the liquid and vowel from L1 English speakers. The dynamic analysis using GAMMs not only generally agrees with the findings from the static analysis but also highlights the robust yet complicated differences between L1 and L2 speech in the formant dynamics. Overall, this study illustrates that the dynamic characteristics are important aspects involved in production of English liquids in the context of L2 speech learning. Directly studying formant dynamics opens discussions around the specific underlying mechanism of L2 speech production under the influence of speakers' L1, and future research will complement the current results using articulatory methods for a better understanding of the factors that may underlie differences in acoustic dynamics shown in this study.

ACKNOWLEDGMENTS

I thank Professor Claire Nance and Dr. Sam Kirkham for their comments and support. Professor Noriko Nakanishi, Profesor Yuri Nishio, and Dr. Bronwen Evans helped me with data collection. The research is financially supported by Graduate Scholarship for Degree-Seeking Students by Japan Student Services Organization (JASSO) and the 2022 Research Grant by the Murata Science Foundation. Data and codes that support the findings of this study are openly available on the Open Science Foundation (OSF) repository at https://osf.io/2phx5/. The author has no conflicts to disclose. This research is approved by ethics committees at Lancaster University, Kobe Gakuin University, and Meijo University. Informed consent was obtained from all participants.

[1]See supplementary material at https://osf.io/2phx5/ for further details about the participants; the data processing procedure; further details of the analysis and results; additional statistical comparisons; and an additional analysis of vowel midpoints.

Akamatsu, T. (**1997**). *Japanese Phonetics: Theory and Practice* (Lincom Europa, München, Newcastle).

Aoyama, K., Flege, J. E., Akahane-Yamada, R., and Yamada, T. (**2019**). "An acoustic analysis of American English liquids by adults and children: Native English speakers and native Japanese speakers of English," J. Acoust. Soc. Am. **146**(4), 2671–2681.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (**2004**). "Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/," J. Phon. **32**(2), 233–250.

Arai, T. (**2013**). "On why Japanese /r/ sounds are difficult for children to acquire," in *Proceedings Interspeech 2013, ISCA*, Lyon, France, pp. 2445–2449.

Articulate Instruments (**2022**). "Articulate Assistant Advanced" (version 220).
Barreda, S. (**2021**). "Fast Track: Fast (nearly) automatic formant-tracking using Praat," Linguistics Vanguard **7**(1), 20200051.
Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**, 1–48.
Beristain, A. M. (**2022**). "The acquisition of acoustic and aerodynamic patterns of coarticulation in second and heritage languages," Ph.D. thesis, University of Illinois Urbana-Champaign, Urbana, IL.
Best, C. T., and Strange, W. (**1992**). "Effects of phonological and phonetic factors on cross-language perception of approximants," J. Phon. **20**(3), 305–330.
Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins Publishing Company, Amsterdam), pp. 13–34.
Boersma, P., and Weenink, D. (**2022**). "Praat: Doing phonetics by computer."
Campbell, F., Gick, B., Wilson, I., and Vatikiotis-Bateson, E. (**2010**). "Spatial and temporal properties of gestures in North American English /r/," Lang. Speech **53**(1), 49–69.
Carter, P., and Local, J. (**2007**). "F2 variation in Newcastle and Leeds English liquid systems," J. Int. Phon. Assoc. **37**(2), 183–199.
Espinal, A., Thompson, A., and Kim, Y. (**2020**). "Acoustic characteristics of American English liquids /ɹ/, /l/, /ɪl/ produced by Korean L2 adults," J. Acoust. Soc. Am. **148**(2), EL179–EL184.
Espy-Wilson, C. Y. (**1992**). "Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English," J. Acoust. Soc. Am. **92**(2), 736–757.
Flege, J. E. (**1995**). "Second language speech learning theory, findings and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD), pp. 233–277.
Flege, J. E., and Bohn, O.-S. (**2021**). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, 1st ed., edited by R. Wayland (Cambridge University Press, Cambridge, UK), pp. 3–83.
Flege, J. E., Takagi, N., and Mann, V. (**1995**). "Japanese adults can learn to produce English /ɹ/ and /l/ accurately," Lang. Speech **38**(1), 25–55.
Grosjean, F. (**2008**). "The bilingual's language mode," in *Studying Bilinguals, Oxford Linguistics* (Oxford University Press, Oxford, New York), pp. 36–66.
Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (**2000**). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," J. Acoust. Soc. Am. **107**(5), 2711–2724.
Hattori, K., and Iverson, P. (**2009**). "English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy," J. Acoust. Soc. Am. **125**(1), 469–479.
Howson, P. J., and Redford, M. A. (**2021**). "The acquisition of articulatory timing for liquids: Evidence from child and adult speech," J. Speech. Lang. Hear. Res. **64**(3), 734–753.
Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**(1), B47–B57.
Jochim, M., Winkelmann, R., Jaensch, K., Cassidy, S., and Harrington, J. (**2023**). "emuR - Main package of the EMU Speech Database Management System," R package version 2.4.2, https://CRAN.R-project.org/package=emuR.
Keating, P. A. (**1985**). "Universal phonetics and the organization of grammars," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic Press, Orlando, FL), pp. 115–132.
King, H., and Ferragne, E. (**2020**). "Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English," J. Phon. **80**, 100978.
Kirkham, S. (**2017**). "Ethnicity and phonetic variation in Sheffield English liquids," J. Int. Phon. Assoc. **47**(1), 17–35.
Kirkham, S., Nance, C., Littlewood, B., Lightfoot, K., and Groarke, E. (**2019**). "Dialect variation in formant dynamics: The acoustics of lateral and vowel sequences in Manchester and Liverpool English," J. Acoust. Soc. Am. **145**(2), 784–794.
Ladefoged, P., and Johnson, K. (**2010**). *A Course in Phonetics, International Edition*, 6th ed. (Wadworth, Boston, MA).
Lawson, E., Stuart-Smith, J., Scobbie, J. M., Yaeger-Dror, M., and Maclagan, M. (**2011**). "Liquids," in *Sociophonetics: A Student's Guide*, edited by M. Di Paolo and M. Yaeger-Dror (Routledge, Abingdon, VA), pp. 72–86.
McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (**2017**). "Montreal Forced Aligner: Trainable text-speech alignment using Kaldi," in *Proceedings Interspeech 2017*, pp. 498–502.
Mochizuki, M. (**1981**). "The identification of /r/ and /l/ in natural and synthesized speech," J. Phon. **9**(3), 283–303.
Morimoto, M. (**2020**). "Geminated liquids in Japanese: A production study," Ph.D. thesis, University of California, Santa Cruz, CA.
Proctor, M. (**2011**). "Towards a gestural characterization of liquids: Evidence from Spanish and Russian," Lab. Phonol. **2**(2), 451–485.
Proctor, M., Walker, R., Smith, C., Szalay, T., Goldstein, L., and Narayanan, S. (**2019**). "Articulatory characterization of English liquid-final rimes," J. Phon. **77**, 100921.
R Core Team (**2022**). "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing.
Recasens, D. (**1991**). "On the production characteristics of apicoalveolar taps and trills," J. Phon. **19**(3-4), 267–280.
Recasens, D. (**2012**). "A cross-language acoustic study of initial and final allophones of /l/," Speech Commun. **54**(3), 368–383.
Riney, T. J., Takada, M., and Ota, M. (**2000**). "Segmentals and global foreign accent: The Japanese flap in EFL," TESOL Quart. **34**(4), 711–737.
Saito, K., and Munro, M. J. (**2014**). "The early phase of /ɹ/ production development in adult Japanese learners of English," Lang. Speech **57**(4), 451–469.
Saito, K., and van Poeteren, K. (**2018**). "The perception–production link revisited: The case of Japanese learners' English /ɹ/ performance," Int. J. Appl. Linguist. **28**(1), 3–17.
Setter, J., and Jenkins, J. (**2005**). "Pronunciation," Lang. Teach. **38**(1), 1–17.
Sheldon, A., and Strange, W. (**1982**). "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," Appl. Psycholinguist. **3**(3), 243–261.
Shimizu, K., and Dantsuji, M. (**1983**). "A study on the perception of /r/ and /l/ in natural and synthetic speech sounds," Stud. Phonologica **17**, 1–14.
Sóskuthy, M. (**2017**). "Generalised additive mixed models dynamic analysis linguistics: A practical introduction," arXiv:1703.05339.
Sóskuthy, M., Foulkes, P., Hughes, V., and Haddican, B. (**2018**). "Changing words and sounds: The roles of different cognitive units in sound change," Top. Cogn. Sci. **10**(4), 787–802.
Sproat, R., and Fujimura, O. (**1993**). "Allophonic variation in English /l/ and its implications for phonetic implementation," J. Phon. **21**(3), 291–311.
Stevens, K. N. (**2000**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).
van Rij, J., Wieling, M., Baayen, R. H., and van Rijn, H. (**2020**). "Itsadug: Interpreting time series and autocorrelated data using GAMMs."
Wells, J. C. (**2008**). *Longman Pronunciation Dictionary*, 3rd ed. (Pearson Education Ltd., Harlow, UK).
West, P. (**1999a**). "The extent of coarticulation of English liquids: An acoustic and articulatory study," in *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS-14)*, edited by J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey (San Francisco, CA), pp. 1901–1904.
West, P. (**1999b**). "Perception of distributed coarticulatory properties of English /l/ and /r/," J. Phon. **27**(4), 405–426.
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (**2019**). "Welcome to the Tidyverse," J. Open Source Softw. **4**(43), 1686.
Wieling, M. (**2018**). "Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English," J. Phon. **70**, 86–116.
Winter, B. (**2020**). *Statistics for Linguists: An Introduction Using R* (Routledge, London, UK).
Wood, S. N. (**2017**). *Generalized Additive Models: An Introduction with R*, 2nd ed. (Chapman and Hall/CRC, New York).

https://doi.org/10.1121/10.0024351

Yamada, R. A., and Tohkura, Y. (**1992**). "The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners," Percept. Psychophys. **52**(4), 376–392.

Yamane, N., Howson, P., and Po-Chun, W. (**2015**). (Grace) "An ultrasound examination of taps in Japanese," in *Proceedings of the 18th International Congress of Phonetic Sciences* Glasgow, UK (August 10–14, 2015), pp. 1–5.

Ying, J., Shaw, J. A., Kroos, C., and Best, C. T. (**2012**). "Relations between acoustic and articulatory measurements of /l/," in *Proceedings of the 14th Australasian International Conference on Speech Science and Technology* (Sydney), pp. 109–112.

Zimmermann, G. N., Price, P., and Ayusawa, T. (**1984**). "The production of English /r/ and /l/ by two Japanese speakers differing in experience with English," J. Phon. **12**(3), 187–193.