

Copyright (2020) Acoustical Society of America. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the Acoustical Society of America.

The following article appeared in JASA Express Letters and may be found at <https://asa.scitation.org/doi/full/10.1121/10.0001758>. This article is made available here with permission from the publisher.

Acoustic characteristics of American English liquids /ɹ/, /l/, /ɹl/ produced by Korean L2 adults

Alexis Espinal, Austin Thompson, and Yunjung Kim

Citation: *The Journal of the Acoustical Society of America* **148**, EL179 (2020); doi: 10.1121/10.0001758

View online: <https://doi.org/10.1121/10.0001758>

View Table of Contents: <https://asa.scitation.org/toc/jas/148/2>

Published by the *Acoustical Society of America*

ARTICLES YOU MAY BE INTERESTED IN

[Cross-generational acoustic comparisons of tonal Kyungsang Korean stops](#)

The Journal of the Acoustical Society of America **148**, EL172 (2020); <https://doi.org/10.1121/10.0001757>

[Elicitation context does not drive F0 lowering following voiced stops: Evidence from French and Italian](#)

The Journal of the Acoustical Society of America **148**, EL147 (2020); <https://doi.org/10.1121/10.0001698>

[F0-related head movement in blind versus sighted speakers](#)

The Journal of the Acoustical Society of America **148**, EL190 (2020); <https://doi.org/10.1121/10.0001353>

[Superiority of blind over sighted listeners in voice recognition](#)

The Journal of the Acoustical Society of America **148**, EL208 (2020); <https://doi.org/10.1121/10.0001804>

[A typology of laterals in twelve English dialects](#)

The Journal of the Acoustical Society of America **148**, EL72 (2020); <https://doi.org/10.1121/10.0001587>

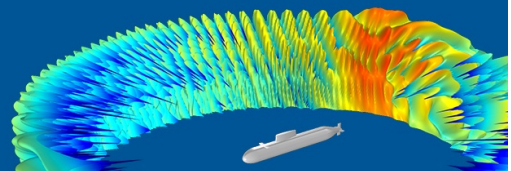
[Indicators of anterior–posterior phase difference in glottal opening measured from natural production of vowels](#)

The Journal of the Acoustical Society of America **148**, EL141 (2020); <https://doi.org/10.1121/10.0001722>

COMSOL Day
Acoustics

A free, online event where you can attend
multiphysics simulation sessions, ask
COMSOL staff your questions, and more

JOIN US MAY 25 »



Acoustic characteristics of American English liquids /ɹ/, /l/, /ɹl/ produced by Korean L2 adults

Alexis Espinal, Austin Thompson, and Yunjung Kim^{a)}

Department of Communication Sciences and Disorders, Louisiana State University, 85 Hatcher Hall,
Field House Drive, Baton Rouge, Louisiana 70803, USA
jespi13@lsu.edu, atho184@lsu.edu, ykim6@lsu.edu

Abstract: This study examined the acoustic characteristics of American English liquids /ɹ/, /l/, and /ɹl/ produced by 14 adult learners of English (L2) and 13 native speakers of English. Several temporal and spectral measures were examined, including a novel measure to describe the relative timing of the maximum constriction during liquid production. The results indicated that L2 speakers rely more on duration contrasts to distinguish the three liquids than spectral contrasts. Reduced spectral differences among the liquids in L2 speakers are discussed concerning the influence of the native language of L2 speakers.

© 2020 Acoustical Society of America

[Editor: Brad H. Story]

Pages: EL179–EL184

Received: 3 May 2020 Accepted: 29 July 2020 Published Online: 20 August 2020

1. Introduction

Despite the well-known challenge in mastering American English liquids /ɹ/ and /l/ among second language (L2) learners, attention has been focused on a limited number of languages (e.g., Japanese, Cantonese) and perceptual distinctions between the two sounds (Aoyama *et al.*, 2004; MacKain *et al.*, 1981; Strange and Dittmann, 1984). The difficulties with the liquid sounds have been generally accounted for by the different phonetic/phonemic inventories and phonological positions of the sounds between the first language (L1) and L2 (Flege *et al.*, 1995).

The current study sought to examine the acoustic characteristics of American English liquids produced by adult L2 speakers whose native language was Korean. Like Japanese and Cantonese, Korean has only one liquid phoneme, which is transcribed as [l], and in intervocalic contexts, as [r]. An example of different realization of Korean liquid across different positions is the word pair, *pulli* [pul.li] “disadvantage” vs *puli* [pu.ri] “beak.” While English allows both /ɹ/ and /l/ to occur in various phonological positions such as word-initial and -final, as well as in consonant clusters, positions for the single liquid are very limited in Korean. For example, English loan words containing obstruent-liquid clusters are typically re-syllabified by vowel epenthesis because Korean does not permit consonant clusters. Further, the articulatory gestures used to produce American English liquids differ depending on the phonetic context. For example, /ɹ/ in prevocalic and intervocalic positions are created with lip rounding, while /ɹ/ in the word-final position is not (Delattre and Freeman, 1968). As such, Korean speakers have difficulty perceiving and producing the contrast between /ɹ/ and /l/ in the same way as native English speakers (Borden *et al.*, 1983; Ingram and Park, 1998). The Speech Learning Model (Flege, 1995; Flege *et al.*, 2002) suggests that Korean speakers may have difficulty perceiving and producing American English liquids due to categorical assimilation. That is, Korean speakers may consider American English liquids as allophones of Korean liquids, rather than creating a separate category for American English liquids.

In addition to the two frequently studied liquids, /ɹ/ and /l/, the combination, /ɹl/ (as in *Carl*), was included in the study because Koreans who are learning English as a second language frequently complain about it being even more challenging than a singleton liquid. To the best of our knowledge, this is the first investigation of English /ɹl/ production among L2 speakers. We limited our analysis to word-final liquids in consideration of the effects of position within words on acoustic characteristics of liquids (Idemaru and Holt, 2011). In addition, liquid accuracy in L2 speakers is known to vary across the position of the sounds. For example, word-final liquids produced by Korean learners of English are less accurate than in other positions when perceptually judged by native speakers of English (Jun, 2003).

^{a)}Current address: School of Communication Science and Disorders, Florida State University, 201 W. Bloxham St. Warren Building, Rm 509 Tallahassee, FL 32301

To characterize liquids produced by native speakers of Korean learning American English as an L2, the current study selected a set of acoustic measures based on the critical role of formant frequencies, particularly the third formant (F3), in perception and production of English liquids (Dalston, 1975; Espy-Wilson, 1992; Saito and Lyster, 2012). For example, Saito and Lyster (2012) reported that listeners perceive /ɹ/ when its F3 dips below 2000 Hz during the constriction interval and as /l/ when its F3 exceeds 2400 Hz. Albeit limited, acoustic studies also reported liquids produced by Korean L2 speakers to be different from native English speakers with respect to F2 and F3 values. Park and Jang (2016a) found significantly higher F2 and lower F3 for /l/ produced by Korean speakers than /l/ produced by native English speakers. Therefore, in this study, we included three spectral measures: (1) relative timing of maximal constriction, (2) frequency difference between F3 and F2 at the maximal constriction, and (3) Euclidean distance between /ɹ/ and /l/ on the F2-F3 plane (see Sec. 2 for details). One temporal measure, liquid duration, was also included.

We explored a new measure, relative timing of maximum constriction, as an attempt to examine the temporal organization of the formant trajectories during liquid production in L2 speakers. This measure was determined by locating the point at which F2 and F3 were closest during the liquid on a normalized time scale, given the potential differences in liquid duration across speakers and stimuli. Therefore, the temporal location of the closest distance between F2 and F3 was expressed in percentage, 0% meaning the onset and 100% meaning the offset of the vocalic nuclei containing the target liquid. This approach was taken to capture dynamic interactions between F2 and F3 during a relatively long period of liquid sounds, instead of using a conventional method in which F2 and F3 values are obtained at a fixed, single time point. In addition, the frequency difference at this maximum constriction was obtained.

To this aim, three questions were posed: (1) do the L1 and L2 speaker groups show differences in word-final /ɹ/, /l/, and /ɹl/ (Speaker Group Effects), (2) do the three phonetic segments differ in the four acoustic measures described above (Sound Effects), and (3) does the degree of difference between L1 and L2 vary across phonetic stimuli (Interaction between Speaker Group and Sound)?

2. Methodology

2.1 Participants, speech tasks, and recording procedure

The participants included a total of 27 speakers, 13 native English speakers (3 men, 10 women), and 14 native speakers of Korean (7 men, 7 women). L1 speakers ranged in age from 19 to 24 years ($M = 23.31$, $SD = 9.38$) and were native speakers of Southern European American English dialect. L2 speakers ranged in age from 23 to 58 years ($M = 42.00$, $SD = 11.87$) and in time of residence in the United States from 3 weeks to 32 years ($M = 12.80$, $SD = 11.89$). The average age of arrival to the United States was 29 years ($SD = 6.95$), with a range from 23 to 51 years. All L2 speakers reported acquiring English throughout their education, with a mean age of 11.29 years ($SD = 2.61$). All participants reported no history of communication problems and signed consent forms. The Louisiana State University Institutional Review Board approved the study.

Participants were asked to read *The Caterpillar* passage in a comfortable, conversational mode (Patel et al., 2013). The passage includes four words ending with /ɹl/ (*Caterpillar*, *coaster*, *car*, *roller*), two words ending with /l/ (*well*, *tall*), but no words ending with /ɹl/. Therefore, the participants were additionally asked to read five times the following sentence, *Carl got a croaking frog*. Connected speech rather than words in isolation was employed in the study to elicit the natural production of the target sounds. Participants were blind to the specific purpose of the study focusing on the liquid production.

Audio recording was obtained using an electromagnetic articulography system, Wave (NDI, Canada), in a sound-attenuating booth, with a sampling rate of 20 kHz and 16-bit quantization. An AKG C1000S microphone positioned approximately 30 cm from the speaker was used to record the speech stimuli. Kinematic data were simultaneously collected as part of a larger study but are not reported in the current study. To minimize the interference of sensors with natural articulatory movements, participants were provided with 10 min of adaptation before recordings (Dromey and Hunter, 2016).

2.2 Acoustic analysis

Acoustic data were segmented using the spectrographic view in the Time-Frequency Analysis Software Program for 32-bit Windows (TF32; Milenkovic, 2004). The formant extraction was based on linear predictive coding (LPC) analysis, and manual correction was performed when errors of automatic extraction were found. Data were analyzed using a custom R script (R Core Team, 2020).

As previously addressed, the following four acoustic measures were obtained: (1) duration (ms) of vocalic nuclei containing the target liquid in consideration of the difficulty of segmenting liquids from the surrounding transition (Carter, 2003; Dalston, 1975), (2) relative timing (%) of maximum constriction within the vocalic nucleus, (3) the frequency difference (Hz)

between F2 and F3 at the maximum constriction, and (4) the Euclidean distance (Hz) in the F2-F3 plane between /ɹ/ and /l/. In cases that the maximum constriction of the target liquid lasted for more than a sampling period (0.05 ms of the study), the relative timing of maximum constriction was coded to compute as the midpoint of the constriction [similar to the method of Recasens and Espinosa (2009)]. Raw durations of vocalic nuclei were reported in the study except for the relative timing of maximum constriction measure.

The current study computed the Euclidean distance between /ɹ/ and /l/ at two time-points, the temporal midpoint of the entire vocalic nucleus, and the time point at which maximum constriction occurred. This comparison was made as an attempt to examine if the maximum constriction point would provide more sensitive measurement point for possible L1-L2 differences given that L2 studies have frequently used the temporal midpoint to extract formant frequencies for liquids (e.g., Park and Jang, 2016a). However, several studies have shown acoustic and articulatory characteristics of liquids using different time points, such as the point of maximum linguopalatal constriction (Recasens and Espinosa, 2009) and the midpoint of the steady state (Carter and Local, 2007).

2.3 Statistical analysis

Prior to analysis, frequencies were normalized using the Bark Scale (Smith and Abel, 1999) to partially account for gender and age differences across speakers. The research questions were examined using two statistical procedures. First, a two-way multivariate analysis of variance (MANOVA)¹ was used to investigate the main effects of speaker group and sound, as well as the interaction between speaker group and sound. For this analysis, the independent variables were speaker group and sound, and the dependent variables were (1) Duration, (2) Max F3-F2 Constriction, and (3) Relative Timing of Max Constriction. The second procedure, a one-way MANOVA, was used to investigate the speaker group effects for the two Euclidean distance measures. For this analysis, speaker group was the sole independent variable and the dependent variables were (1) the Euclidean distance between /ɹ/ and /l/ measured from the temporal midpoint and (2) the Euclidean distance between /ɹ/ and /l/ measured from the max constriction.

For both statistical procedures, sex was submitted as a covariate to control for physiological differences (i.e., MANCOVA). However, after the initial analysis, sex was not significant for any of the dependent variables, suggesting that sex has a nonsignificant relationship on the target measures (For the two-way MANCOVA analysis, sex was nonsignificant, $F[3, 415] = 0.007$, $p = 0.999$; for the one-way MANCOVA analysis, sex was nonsignificant, $F[2, 23] = 0.003$, $p = 0.997$). Subsequently, sex was removed as a covariate from the analyses.

3. Results

The reliability of acoustic measurements was checked using Pearson Product Moment Correlation coefficients. For intra-measurer reliability, approximately 40% of the data (5 L1 and 5 L2 speakers) were randomly selected and re-measured four months after the initial measurement. For inter-measurer reliability, the second measurer randomly selected 40% of the data (5 L1 and 5 L2), blinded to the original measurement. The correlation coefficients indicated strong agreement for both intra-measurer and inter-measurer reliability; $r = 0.98$ and 0.95 , respectively.

Table 1 summarizes the descriptive results for each speaker group and measure. Statistical results revealed significant group effects on Duration, Euclidean distance between /ɹ/

Table 1. Means and standard deviations for each speaker group and measure.

Measure	L1			L2		
	/ɹ/	/l/	/ɹl/	/ɹ/	/l/	/ɹl/
Duration (ms)	215.06 (58.87)	196.77 (48.62)	259.31 (45.17)	291.12 (93.07)	288.56 (92.11)	332.13 (69.11)
Max F3-F2 Constriction (in bark)	0.47 (0.35)	1.15 (0.41)	0.74 (0.16)	0.74 (0.34)	0.88 (0.41)	0.89 (0.35)
Relative Timing of Max Constriction (%)	75.93 (30.97)	41.79 (46.64)	35.18 (21.83)	46.88 (38.54)	61.11 (34.99)	65.78 (35.06)
Euclidean distance between /ɹ/ and /l/ at the temporal midpoint (in Bark)		1.48 (0.47)			1.05 (0.49)	
Euclidean distance between /ɹ/ and /l/ at the max constriction (in Bark)		3.00 (2.25)			1.26 (0.52)	

and /l/ at the temporal midpoint, and Euclidean distance between /ɪ/ and /l/ at the max constriction. Prior to time normalization, the liquids were longer for L2 compared with L1 [$F(1, 418) = 90.28, p < 0.001$]. Euclidean distance was smaller for L2 speakers compared to L1 for both measurement points, temporal midpoint and maximum constriction point. However, reduced Euclidean distance was more apparent when the distance was measured at the maximum constriction point [$F(1, 25) = 6.83, p = 0.01$] than temporal midpoint [$F(1, 25) = 5.41, p < 0.02$]. No group effects were found for Max F3-F2 Constriction and Relative Timing of Max Constriction. However, the interpretation of these nonsignificant group effects is complicated by the presence of a significant speaker group and sound interaction (reported below).

In addition, significant main effects of liquid sounds were found for the three measures, Duration [$F(2, 418) = 17.23, p < 0.001$], Max F3-F2 Constriction [$F(2, 418) = 38.20, p < 0.001$], and Relative Timing of Max Constriction [$F(2, 418) = 5.00, p = 0.007$]. Again, the interpretation of the significant main effects of sound for Max F3-F2 Constriction and Relative Timing of Max Constriction is complicated by the significant speaker group and sound interaction (reported below).

Last, the two-way MANOVA results showed a significant interaction between speaker group and sound for Max F3-F2 Constriction [$F(2, 418) = 13.42, p < 0.001$] and Relative Timing of Max Constriction [$F(2, 418) = 35.43, p < 0.001$], but not for Duration. While the durational pattern among the three stimuli was the same between the two speaker groups, the timing and degree of constriction showed a different pattern between the two speaker groups (Fig. 1).

As seen in the middle panel of Fig. 1, the Relative Timing of the Max Constriction in the production of /ɪ/ occurred later for L1 speakers compared to the L2 speakers [$F(1, 241) = 41.73, p < 0.001$]. For /ɪl/, the Relative Timing of the Max Constriction occurred earlier for L1 speakers compared to L2 speakers [$F(1, 125) = 38.28, p < 0.001$]. For /l/, there was no significant difference in the Relative Timing of the Max Constriction between the two groups [$F(1, 52) = 0.667, p = 0.418$].

The right panel of Fig. 1 shows the *post hoc* comparisons for Max F3-F2 Constriction. For /ɪ/, the Max F3-F2 Constriction was smaller for the L1 speakers compared to the L2 speakers, indicating a tighter constriction [$F(1, 241) = 14.82, p < 0.001$]. For /l/, the Max F3-F2 Constriction was larger for the L1 speakers compared to the L2 speakers, indicating a wider constriction [$F(1, 52) = 7.43, p = 0.009$]. The two speaking groups produced statistically similar degrees of constriction for /ɪl/ [$F(1, 125) = 0.761, p = 0.385$].

4. Discussion

The findings of the study provide acoustic data on liquid production between native speakers of American English and L2 speakers with a Korean language background. The main findings can be summarized as follows. First, Korean L2 speakers showed overall lengthened durations compared to L1 speakers, while maintaining the durational differences among liquids as in L1 speakers. Second, however, spectral distinctiveness among the three liquids in L2 was either reduced or different compared to L1 speakers. Specifically, the spectral distinction between /ɪ/ and /l/, measured by Euclidean distance on the F2-F3 plane, was reduced in L2 speakers. In addition, the location and degree of maximal constriction across the three liquids were different between L1 and L2 speakers.

4.1 Temporal vs spectral characteristics of L2 speakers

L2 speakers produced longer liquids consistently across the three liquid stimuli than L1 speakers, possibly reflecting “cautious” production of these challenging sounds to achieve better accuracy.

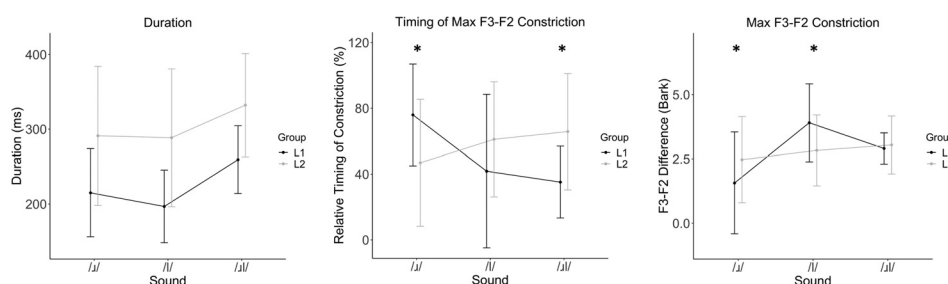


Fig. 1. Line graphs displaying interactions between speaker groups and liquid sounds for Duration, Relative Timing of Max Constriction, and Max F3-F2 Constriction. Asterisks indicate significant *post hoc* pairwise comparison results. *Post hoc* comparisons were only examined for the measures with significant speaker group and sound interactions (i.e., Relative Timing of Max Constriction and Max F3-F2 Constriction). Duration was found to have significant group and sound effects. However, there was no significant group and sound interaction for Duration.

However, albeit lengthened, the durational pattern across the three liquids, from shortest to longest, /l/-ɹ/-ɹl/, did not differ between the speaker groups. This finding may imply that L2 speakers rely more on durational contrasts rather than spectral contrasts to produce acoustic distinctions among the three liquids, because spectral distinction among liquids, such as the Euclidean distance between /ɹ/ and /l/ on F2-F3 planes, was significantly reduced for L2 compared to L1. In other words, native Korean speakers may be more successful in creating durational modifications than spectral modifications required for the English liquids. In general, this finding is in line with prior work reporting L2 speaker's overall preference of durational differences to spectral differences in perception and production of English vowels [e.g., perception: Yazawa *et al.* (2019); production: Kim *et al.* (2018)]. However, contrary to our finding, Park and Jang (2016a) found that both Korean and native English speakers did not utilize duration to differentiate /l/ and /ɹ/ productions. The inconsistent findings may be explained, in part, by the methodological differences between the two studies. Besides different speech stimuli (words in a carrier phrase vs passage reading), Park and Jang (2016a) measured normalized duration, while the current study did not.

As a methodological issue, and to confirm that the extraction of formant frequencies at the point where F2 and F3 were maximally close provided a sensitive measure of L1-L2 differences, we measured Euclidean distances between /ɹ/ and /l/ at two temporal points, the midpoint and the closest point between F2 and F3 trajectories (i.e., the maximum constriction). At both measurement points, the L2 group showed significantly reduced spectral distances between /ɹ/ and /l/ compared to the L1 group. However, the difference was more apparent when measured at the maximum constriction point. The reduced Euclidean distance between the two liquids in L2 speakers is largely accounted for by two contributors, high F3 for /ɹ/ and high F2 for /l/. Despite the critical role of F3 in distinguishing the two English liquids both in perception and production, in the latter case, a significant dip for /ɹ/ (Dalston, 1975; Saito and Lyster, 2012), F3 was significantly higher for the L2 group compared to the L1 group. In addition, F2 for /l/ was also higher for L2 compared to L1. High F2 for /l/ may reflect an influence of the native language of the L2 speakers. In Korean, /l/ is the only liquid that can occur at the word-final position, and it has higher F2 than English /l/ based on acoustic studies for each language (Dalston, 1975; Espy-Wilson, 1992; Kim and Lotto, 2004).

Two constriction-related measures were included in the study, and the two speaker groups showed significantly different patterns across the liquids. First, the degree of vocal tract constriction, acoustically indexed by the F3-F2 difference at the point of minimum difference, was ordered from /ɹ/-ɹl/-l/, smallest to largest, in L1 speakers. This pattern is consistent with the requirement of a strong dip in F3 to produce /ɹ/. However, L2 speakers exhibited significantly less constriction for /ɹ/ and significantly more constriction for /l/ compared to L1 speakers, which led to similar degrees of constriction across the phonetic stimuli (Fig. 1, right panel).

The other constriction measure was the temporal location of the maximum constriction, as an index of temporal organization of the constriction within vocalic nuclei containing liquids. This exploratory measure was selected in consideration of previous work, most of which has focused on one fixed, slice-in-time analysis window (frequently temporal midpoint) to capture formant frequency characteristics. The results show an opposing pattern between the two speaker groups. In contrast, the L1 group placed the constriction in the order of /ɹl/-l/-ɹ/, the L2 group placed in the order of /ɹ/-l/-ɹl/ (Fig. 1, middle panel). For L1 speakers, the dip for F3 happens toward the offset of /ɹ/, which moves to earlier for the combination sequence /ɹl/. On the contrary, the constriction occurred significantly earlier for /ɹ/ and later for /ɹl/ in L2 speakers compared to L1 speakers. We argue that these new measures provide a useful method to identify the spectral characteristics of liquids produced by L2 speakers.

4.2 Conclusion and limitations

The study showed acoustic differences in liquid production between two speaker groups, L1 and L2, which may serve as a source of perceptual sound errors and degree of foreign accent. In general, the findings imply that durational contrasts seem more accessible to L2 speakers than spectral contrasts to distinguish the three liquid sounds included in the study. This conclusion is based on the finding that Korean speakers did not produce adequate degrees of constriction for /ɹ/, /l/, or /ɹl/ with the timing observed in native speakers of American English (specifically a Southern European American English dialect), whereas they showed the durational differences across the liquids as in L1 speakers. Heavy use of durational differences is consistent with a perceptual study, which reported that duration is used by Korean L2 speakers to identify /ɹ/ and /l/, but not by native speakers of English (Park and Jang, 2016b).

With the increasing number of native speakers of a language other than English living in the United States (US), the exposure to accented speech is growing; so is the need to develop

effective, data-driven accent neutralization approaches. The findings may serve as empirical data for such effort, by emphasizing spectral differences between the two speaker groups.

References and links

¹MANOVA, the multivariate extension of the analysis of variance design, was used to avoid inflating the familywise error rate that occurs when conducting multiple univariate tests on the same data.

- 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.
- Borden, G., Gerber, A., and Milsark, G. (1983). "Production and perception of the /l/-/l/ contrast in Korean adults learning English," *Lang. Learn.* **33**(4), 499–526.
- Carter, P. (2003). "Extrinsic phonetic interpretation: Spectral variation in English liquids," in *Phonetic Interpretation: Papers in Laboratory Phonology VI* (Cambridge University Press, Cambridge), pp. 237–252.
- Carter, P., and Local, J. (2007). "F2 variation in Newcastle and Leeds English liquid systems," *J. Int. Phon. Assoc.* **37**(2), 183–199.
- Dalston, R. M. (1975). "Acoustic characteristics of English /w,r,l/ spoken correctly by young children and adults," *J. Acoust. Soc. Am.* **57**(2), 462–469.
- Delattre, P., and Freeman, D. C. (1968). "A dialect study of American r's by x-ray motion picture," *Linguistics* **6**(44), 29–68.
- Dromey, C., and Hunter, E. (2016). "Speech adaptation to kinematic recording sensors," *J. Acoust. Soc. Am.* **140**(4), 3220–3221.
- 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., Schirru, C., and MacKay, I. R. A. (2002). "Interaction between the native and second language phonetic subsystems," *Speech Commun.* **40**(4), 467–491.
- Flege, J. E., Takagi, N., and Mann, V. (1995). "Japanese adults can learn to produce English /r/ and /l/ accurately," *Lang Speech* **38**, 25–55.
- Idemaru, K., and Holt, L. L. (2011). "Production and perception of English /l/ and /r/ by 4-, 5-, and 8-year-old children," paper presented at the *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong.
- Ingram, J. C. L., and Park, S.-G. (1998). "Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners," *J. Acoust. Soc. Am.* **103**(2), 1161–1174.
- Jun, E. (2003). "The perception of English /l/ and /r/ by Korean speakers," *Stud. Phon. Phon. Morphol.* **9**(2), 471–484.
- Kim, D., Clayards, M., and Goad, H. (2018). "A longitudinal study of individual differences in the acquisition of new vowel contrasts," *J. Phon.* **67**, 1–20.
- Kim, M., and Lotto, A. J. (2004). "Acoustic measurements of Korean approximants," *Korean Lang. Am.* **9**, 70–77, see www.jstor.org/stable/42922859.
- MacKain, K. S., Best, C. T., and Strange, W. (1981). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," *Appl. Psycholing.* **2**(4), 369–390.
- Milenkovic, P. (2004). "TF32," University of Wisconsin-Madison, <http://userpages.chorus.net/cspeech/> (Last viewed 1/28/2020).
- Park, S., and Jang, T.-Y. (2016a). "Acoustic characteristics of English liquids produced by Korean learners of English," *Stud. Phon. Phonol. Morphol.* **22**(2), 289–315.
- Park, S., and Jang, T.-Y. (2016b). "Perception of English intervocalic liquids by Korean learners of English," *Lang. Ling.* **71**, 53–78.
- Patel, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., Ramage, L., Tyler, S., and Russell, S. (2013). "The caterpillar: A novel reading passage for assessment of motor speech disorders," *Am. J. Speech Lang. Pathol.* **22**(1), 1–9.
- R Core Team (2020). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/> (Last viewed 12/11/2019).
- Recasens, D., and Espinosa, A. (2009). "An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan," *J. Acoust. Soc. Am.* **125**(4), 2288–2298.
- Saito, K., and Lyster, R. (2012). "Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /l/ by Japanese learners of English," *Lang. Learn.* **62**(2), 595–633.
- Smith, J. O., and Abel, J. S. (1999). "Bark and ERB bilinear transforms," *IEEE Trans. Speech Audio Proc.* **7**(6), 697–708.
- Strange, W., and Dittmann, S. (1984). "Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English," *Percept. Psychophys.* **36**(2), 131–145.
- Yazawa, K., Whang, J., Kondo, M., and Escudero, P. (2019). "Language-dependent cue weighting: An investigation of perception modes in L2 learning," *Second Language Research* (published online).