말소리와 음성과학

Phonetics and Speech Sciences

Check for updates

Comparing English and Korean speakers' word-final /rl/ clusters using dynamic time warping

Hyesun Cho*

Department of Education, Graduate School of Education, Dankook University, Yongin, Korea

Abstract

The English word-final /rl/ cluster poses a particular problem for Korean learners of English because it is the sequence of two sounds, /r/ and /l/, which are not contrastive in Korean. This study compared the similarity distances between English and Korean speakers' /rl/ productions using the dynamic time warping (DTW) algorithm. The words with /rl/ (*pearl*, *world*) and without /rl/ (*bird*, *word*) were recorded by four English speakers and four Korean speakers, and compared pairwise. The F2-F1 trajectories, the acoustic correlate of velarized /l/, and F3 trajectories, the acoustic correlate of /r/, were examined. Formant analysis showed that English speakers lowered F2-F1 values toward the end of a word, unlike Korean speakers, suggesting the absence of /l/ in Korean speakers. In contrast, there was no significant difference in F3 values. Mixed-effects regression analyses of the DTW distances revealed that Korean speakers produced /r/ similarly to English speakers but failed to produce the velarized /l/ in /rl/ clusters.

Keywords: English rhotic, English lateral, dynamic time warping, formant, velarized l, dark l

1. Introduction

1.1. The /rl/ Sequence

The Korean language does not have the phonemic contrast between /r/ and /l/ whereas English does. Due to a lack of the phonemic contrast, Korean learners of English have difficulties in distinguishing English /r/ and /l/ sounds in perception and production (Borden et al., 1983; Jang, 2005; Kim & Rhee, 2019). When the two sounds are hard to distinguish from each other, the production of the sequence of the two, /rl/, becomes even harder. In addition, Korean phonology does not allow coda clusters to surface [*Complex Coda>>Max(C)] unlike English. Due to the difficulties in production,

perception, and phonological constraints, the English word-final /rl/ cluster is often simplified as one liquid sound by Korean learners of English. As a result, the English words *world* [wɜːɹld] and *word* [wɜːɹd] are hard for Korean speakers to differentiate in their perception and production. Against this backdrop, the present study examines the production of the word-final /rl/ clusters by English and Korean speakers with a novel approach, the dynamic time warping (DTW) algorithm (Giorgino, 2009).

Previous research has focused more on Korean speakers' production of English /r/ and /l/ separately, instead of the cluster of the two (Ingram & Park, 1998; Kang, 1999; Kim & Rhee, 2019; Park & Jang, 2016; Sohn & Lim, 2020). Korean speakers'

difficulties involve both perceptual and production aspects of /r/ and /l/. They are hard to distinguish in perception because they are not contrastive sounds in Korean. L2 phones that correspond to a single category in L1 are hard to differentiate (Flege, 1987). In production, Korean /l/ is not velarized in word-final position (Iverson & Sohn, 1994), so Korean has articulatory difficulties in the production of English dark /l/.

The difficulties are closely related to the phonological context where they appear. According to the corpus-based study of Kim & Rhee (2019), in word-final position, a single coda /r/ was more accurately pronounced than a single coda /l/ (_car_ vs. _all_). In English, /l/ is velarized in word-final position (Gimson, 1989; Ladefoged & Johnson, 2011), which is difficult for Korean learners to produce (Jang, 2005).

Few studies focused on the production of the /rl/ cluster. Kwon (2010) examined the production of the /rl/ sequences in English words _girl_ and _early_ by Korean and English speakers. The comparison of formant transition slopes shows that the slope was significantly different in Korean vs. English speakers in all three formants (F1, F2, F3) in the production of _girl_. On the other hand, for _early_, F2 does not show a significant difference between Korean and English speakers, where the /l/ is a syllable onset.

1.2. Acoustic Correlates of /r/ and /l/

The production of American /r/ is highly variable depending on context and speaker, ranging between the two extreme ends: bunched or retroflexed /r/ (Delattre & Freeman, 1968; Guenther et al., 1999; Hwang, 2021; Zhou et al., 2008). Despite large variations in articulation, the acoustic characteristic is relatively simple, a "deep dip" in F3 (Espy-Wilson, 1992; Feng, 2020; Guenther et al., 1999; Idemaru & Holt, 2013; Johnson, 2003; Ladefoged & Maddieson, 1996). The low F3 for /r/ is due to the resonance from the anterior cavities formed by the palatal constriction (Stevens, 1998) and lip rounding (Delattre & Freeman, 1968). The acoustic stability of F3 is maintained by the tradeoff relationships between the length of the front cavity and the length and size of the palatal cavity (Guenther et al., 1999:13).

The alveolar lateral approximant /l/ has two allophones, light [l] in syllable onset position and dark [ɫ] in syllable coda position in English (Carr, 2020; Gimson, 1989; Ladefoged & Johnson, 2011). In the articulation of dark or velarized /l/, the front of the tongue is pushed down while maintaining the alveolar contact, and at the same time, the back of the tongue is raised toward the dorsum. F1 increases as the tongue body is lowered, while F2 decreases as the tongue-dorsum retracts. As a result, F2-F1 values decrease for dark /l/ (Ladefoged & Johnson, 2011; Sproat & Fujimura, 1993; Stevens, 1998; Sohn & Lim, 2020). Darkness of /l/, or the degree of velarization, is gradient, depending on boundary strength, so /l/ is more velarized in word-final position than in intervocalic or post-boundary position (Sohn & Lim, 2020).

To summarize, the acoustic correlate of /r/ is F3 and the acoustic correlate of dark /l/ is F2-F1. Thus, in the present study, these two acoustic measures are collected to examine the production of /rl/ sequences.

1.3. The Dynamic Time Warping (DTW) Algorithm

The DTW algorithm is an algorithm to compare time-series data that change over time (Giorgino, 2009; Sakoe & Chiba, 1978). The algorithm is widely used in various fields, such as speech recognition, econometrics, and other general time-series mining. Acoustic measures, such as F0 and formant frequencies, are also time-series data, so DTW can be used to compare two sets of acoustic data. The DTW algorithm compares a pair of temporal data of different lengths by stretching or compressing them to make them match each other as much as possible. The algorithm finds the best matching points of two sounds in a way that minimizes the summed cost of differences. The matching of two sounds is not based on the same-interval time-step but based on the series of the points whose summed differences are the smallest, or the optimal path. Therefore, the DTW algorithm can deal with two sets of data of different lengths. The differences computed from the optimal path are summed. The total difference or the remaining cumulative distance is the phonetic distance of two compared sounds, expressed in a single numerical value. The greater the distance, the more different the two sounds are.

The DTW algorithm was used in Cho et al. (2021) to measure the phonetic distance between two words. They measured phonetic distances in the Mel-frequency cepstral coefficients (MFCC) values. In the present study, I use the DTW algorithm for formant comparison (F2-F1 and F3, the acoustic correlates of /r/ and /l/ mentioned in Section 1.2.). The DTW algorithm has been applied to formant data in the literature (Boujnah et al., 2021; Kasuya et al., 1994; Regier, 2014).

The DTW algorithm is appropriate for the study of spoken words because they have different durations when produced by different speakers and even when produced by the same speakers. Moreover, it can capture the dynamic characteristics of formants that change over time. Most previous research has depended on some selected subsets of formant measures, such as formant values at static time points, formant differences, or formant transition slopes between two time-points (e.g., Kwon, 2010; Park & Jang, 2016; Sohn & Lim, 2020). These measures may not fully capture the characteristics of formant trajectories that change over time. The DTW method compares the formant contours as a whole, finding the optimal matching points between two data sets.

2. Research Methods

2.1. Recording

2.1.1. Subjects

The speakers were 4 native speakers of American English (E1, E2, E3, E4) and 4 Korean speakers (K1, K2, K3, K4). The American speakers had rhotic accents (three midwestern, one western) and they were in their late 20's to late 30's. The Korean speakers were in their 20's with high-intermediate English proficiency (TOEIC scores 795−900). All the subjects were compensated for their participation.

2.1.2. Speech materials

The words containing the /rl/ sequence, _pearl_ and _world_, were compared with _bird_ and _word_ that have the same place of articulation in word-initial position but a sequence /rd/ in word-final position. The words were embedded in a carrier sentence, "I said __ today."

Table 1. Speech materials

Word	IPA	POA (initial)	Final sounds
(a) *pearl*	[pɜːɹl]	lab	rl
bird	[bɜːɹd]	lab	rd
(b) *world*	[wɜːɹld]	lab-dor	rld
word	[wɜːɹd]	lab-dor	rd

IPA, International Phonetic Alphabet; POA, place of articulation.

Table 1 shows the target words. The place of articulation of the word-initial segments was controlled in each pair, (a) (labial) and (b) (labial-dorsal). In each pair, final sounds were different, /rl/ vs. /rd/. *World* has an additional /d/, but it was included for comparison with *word*, which is minimally different by /l/.

2.1.3. Recording procedures

The Korean speakers were recorded in a sound-attenuated recording studio in a university. The sampling frequency was 44.1 kHz, with a sample size of 16 bits. The English speakers were recorded online using an online recorder, Vocaroo (http://vocaroo.com), due to the pandemic. They were asked to read the sentences in a quiet room. The subjects were given a list of sentences in randomized order and asked to read the list three times in a quiet room.

2.2. Analysis Procedures

2.2.1. Pairwise comparison

For each speaker, there were 12 tokens (4 words×3 repetitions). The words recorded by English speakers were all pairwise compared with those recorded by Korean speakers [(12 tokens by an English speaker×12 tokens by a Korean speaker)×4 Korean speakers=576 pairs]. The pairs included all possible combinations between English (E1−E4) speakers and Korean speakers (K1−K4) (e.g. E1-K1, E1-K2, E1-K3, E1-K4 for speaker E1). The words recorded by English speakers were also compared by those recorded by the same group of English speakers [(12 tokens by an English speaker×12 tokens by another English speaker)×4 English speakers= 576 pairs]. For example, English speaker E1's tokens were pairwise compared with the tokens by English speaker E1, E2, E3, E4 in turn. From here, pairs with identical sounds [same tokens by the same speaker, 48 in total (4 words×3 repetitions×4 speakers)] were excluded. Thus there were a total of 1,104 pairs. The DTW distances were computed for each of these pairs. The differences are compared in terms of (i) formants and (ii) similarity distances measured by DTW. The details are explained in the next two sections.

2.2.2. Formant measures

Formants (F1, F2, F3) were automatically collected using a Praat script. The formant values were collected at three points, the beginning, middle, and end of the voiced interval (the vowel and

/r(l)/) in each word. Some speakers optionally released word-final obstruent /d/ (*bird*, *world*, *word*). The optional closure and release for the final /d/ were excluded in the analysis.

2.2.3. Similarity distance using dynamic time warping

The DTW algorithm was used to measure the similarity between the words produced by Korean and English speakers. The codes for DTW were created in Python in Google Colaboratory.[1] The *dtw-python library* (Giorgino, 2009) was used to implement the DTW algorithm. The *praat* functions in the package *parselmouth* (Boersma & Weenink, 2020; Jadoul et al., 2018) were used to automatically extract formant values in the Python environment,[2] and the *os.walk* function in the package *os* was used to loop through the sound files in Google Drive.

F2-F1 and F3 values were collected for each word from the midpoint to the end of the voiced portion (vowel and /r(l)/) of the word, i.e., the second half of the rime excluding /d/. The first half of the rime was excluded to avoid the effect of the onset consonant. It is difficult to reliably segment vowels and /r/ because vowels are rhoticized before /r/ (Allen, 1979; Celce-Murcia et al., 2010:217; Chung & Pollock, 2014). The acoustic correlates of r-colored vowels and /r/ are both F3. Thus, instead of segmenting the vowel and following /r/, the second half of the rime excluding /d/ was used for similarity analysis.

The F2-F1 and F3 trajectories were compared for each pair of words using DTW. The distance in each pair is expressed by a single numerical number (DTW distance). As explained in Section 2.2.1., there were 1,104 pairs, so a total of 1,104 DTW distance values for F2-F1 and a total of 1,104 DTW distance values for F3 were obtained. Formant values and DTW distances were visualized and analyzed using R (version 4.1.2) (R Core Team, 2021).

2.3. Hypotheses

Firstly, it is expected that DTW distances between the words produced by English speakers and the words produced by Korean speakers (between-language distances) will be greater than those between the words produced by English speakers (within-language distances). Secondly, assuming the /rl/ sequence is hard for Korean speakers to pronounce than /rd/, it is expected that the DTW distance between Korean and English speakers will be greater in /rl/ clusters than in /rd/ clusters.

3. Results

3.1. Formants

Figures 1 through 4 show the changes of the formants at three points (beginning, middle, end) in words produced by English and Korean speakers. In all figures, there is a deep dip in F3 values in the middle of the word in both English and Korean speakers, which is an acoustic characteristic for /r/ (Ladefoged & Maddieson, 1996). On the other hand, F2 shows different patterns depending on speaker

---

[1] The codes were written based on the codes "Calculating similarities between time-series signals using a dynamic time warping algorithm" provided by Sunghye Cho at the winter workshop by the Korean Society of Speech Sciences (December 18, 2021). For the present study, the modifications were made to collect formants instead of MFCC and to loop through Google Drive folders and compare the sound files in them.

[2] http://blog.syntheticspeech.de/2021/03/10/how-to-extract-formant-tracks-with-praat-and-python/ (Adaptation from Feinberg (2022) with permission of Center For Open Science.

L1. In Figure 1, whereas English speakers' F2 does not much change from the mid to the end of the word (with wide variation), Korean speakers' F2 substantially increases toward the end of the word. F2 is lowered by a dorsal constriction that is required syllable-final dark /l/ (Johnson, 2003:163), so this means that Korean speakers did not have a dorsal constriction for dark /l/. In brief, the trajectories in Figure 1 show that Korean speakers had the articulatory gestures for /r/ but not for dark /l/ in the production of *pearl*.

On the other hand, in *bird* (Figure 2), both F2 and F3 gradually increase from the midpoint to the end of the word, in both English and Korean speakers. We can see that the difference between F1 and F2 (F2-F1) becomes greater toward the end of the word in both groups, which is expected due to the absence of /l/ in *bird*.
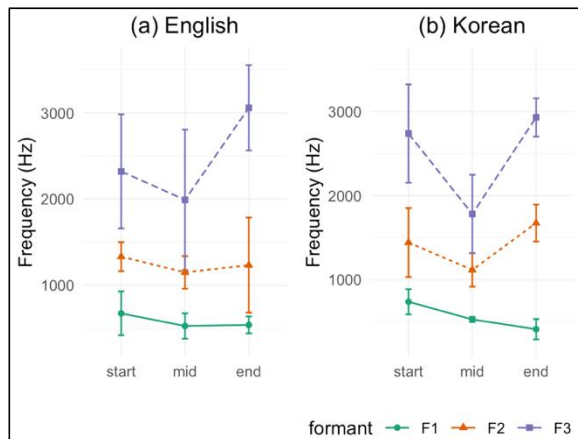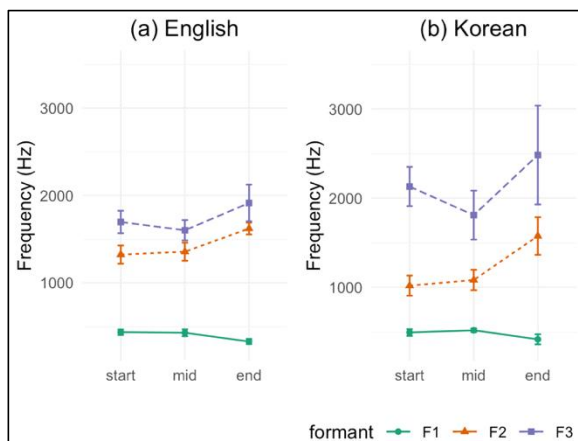

Figure 1. Formant changes in *pearl* [pɜːɹl].

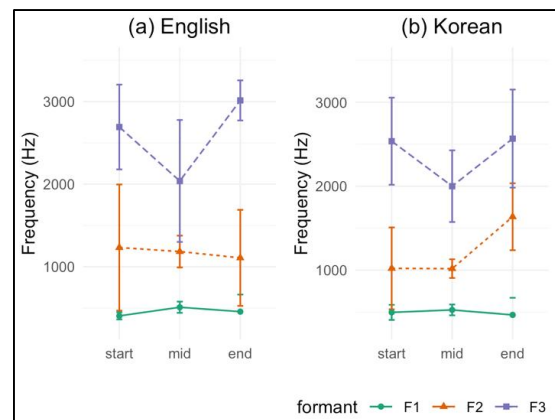**Figure 2.** Formant changes in *bird* [bɜːɹd].

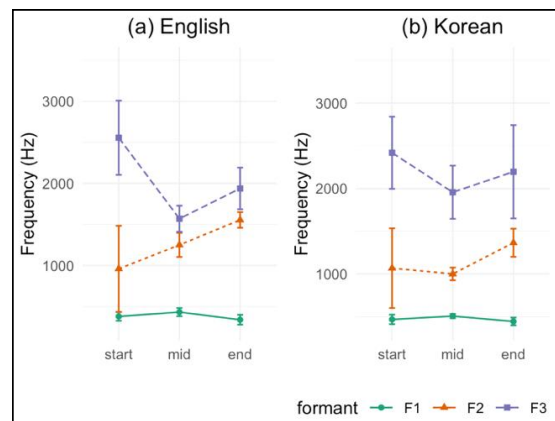**Figure 3.** Formant changes in *world* [wɜːɹld].

**Figure 4.** Formant changes in *word* [wɜːɹd].

Similar patterns are found in *world* vs. *word* in Figures 3 and 4. In *world* (Figure 3a) by English speakers, F3 is lowered and then increases, suggesting the production of /r/. F2 slightly decreases toward the end of the word, and F2 and F1 get closer to each other (i.e., smaller F2-F1), suggesting the production of [ɫ]. In contrast, in the same word produced by Korean speakers (Figure 3b), there is a dip in F3, suggesting the presence of /r/, but the difference between F2 and F1 becomes greater toward the end of the word, suggesting the absence of [ɫ].

On the other hand, in Figure 4 for *word*, the patterns of F3 and F2-F1 are similar to each other in English and Korean speakers. There is a dip in F3 for the articulation of /r/, and the difference between F2 and F1 increases toward the end of the word in both speaker groups.

In summary, the words *world* and *word* are distinguished by F2 by English speakers, comparing Figures 3a and 4a. However, the two words are not differentiated by Korean speakers. The formant trajectories are not much different in Figures 3b and 4b. The same applies to Figures 1 and 2.
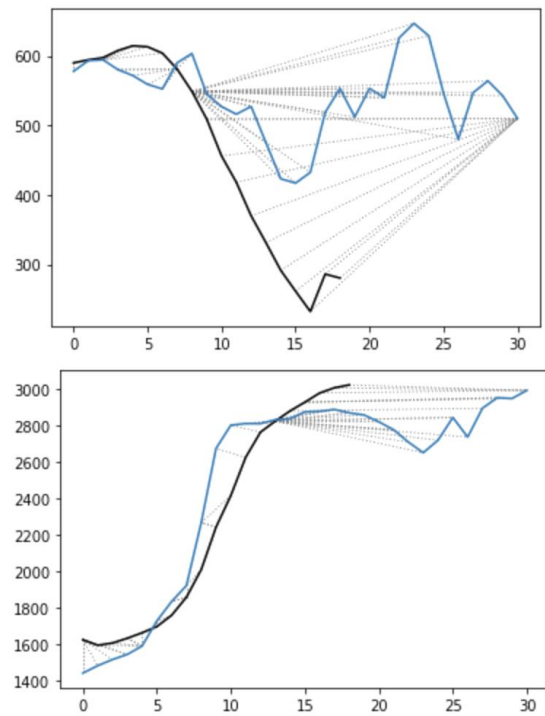
3.2. Dynamic Time Warping

Figures 5 and 6 each show examples of two trajectories aligned by the DTW algorithm. The solid lines are the trajectories of formant values over time. The dotted straight lines connect two points in each trajectory, which are the optimal path that minimizes the distance between the trajectories.
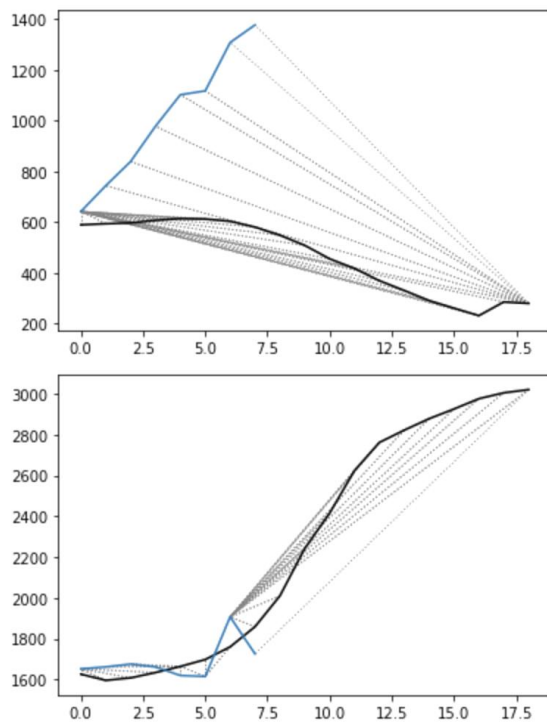
Figure 5. Alignment example using DTW, comparing English (E3, black) and Korean (K4, blue) speakers for the second half of *world*. Top: F2-F1; DTW distance, 330. Bottom: F3; DTW distance, 391. DTW, dynamic time warping.

Figure 5 shows an example of how the formant trajectories of English and Korean speakers are aligned (black: English, blue: Korean). The plot at the top shows the alignment between F2-F1 values, which decrease over time for the English speaker but increase for the Korean speaker. Given that dark /l/ has low F2-F1, this shows that the English speaker produced dark /l/ toward the end of the word but the Korean speaker did not. The plot at the bottom shows the alignment of F3 trajectories between the Korean and English speakers. Both trajectories tend to increase toward the end. It is also noticeable that the DTW algorithm can compare the words with different durations as shown here (Korean: about 7.5 ms, English: about 17.5 ms). The DTW distances are 330 in the top panel and 391 in the bottom panel, which means that in the illustrated example, F3 trajectories are more different from each other than F2-F1 trajectories.

Figure 6. Alignment example using DTW, comparing two English speakers (E3, black; E4, blue) for the second half of *world*. Top: F2-F1; DTW distance, 56. Bottom: F3; DTW distance, 47. DTW, dynamic time warping.

On the other hand, English speakers are more similar to each other in F2-F1 as well as in F3. Figure 6 illustrates the alignment of F2-F1 (top) and F3 trajectories (bottom) of two tokens of *world* each produced by English speakers (E3: black, E4: blue). In the top panel, the black line is the F2-F1 trajectory of speaker E3, and the blue line is that of speaker E4. Though E3 and E4 have individual differences, both lines show decreasing trends. Individual variations are also observed in F2 values in Figure 3a, so individual variations are expected in F2-F1 values as well. E3's trajectory drastically decreases over time. On the other hand, E4's trajectory has perturbations, but a descending trend can be found, looking at the beginning and endpoints of the trajectory.

The plot at the bottom shows the alignment of F3 trajectories for the two English speakers, which are similar to each other. The DTW distances are 56 in the top panel and 47 in the bottom panel, which means that in the illustrated example, F2-F1 trajectories are more different each other than F3 trajectories. Compared with examples in Figure 5, the DTW distances between English speakers are much lower than those between English and Korean speakers despite perturbations in formant trajectories. Thus, we can see that the DTW algorithm can capture overall trends despite local perturbations.

3.3. Statistical Analysis

Mixed-effects linear regression models were fitted to the data with DTW distances as a dependent variable, formants (F2-F1, F3 separately), and the speaker group that is compared with the English speakers (L1 English or Korean) as fixed effects. Random intercepts for two speaker groups separately were included.
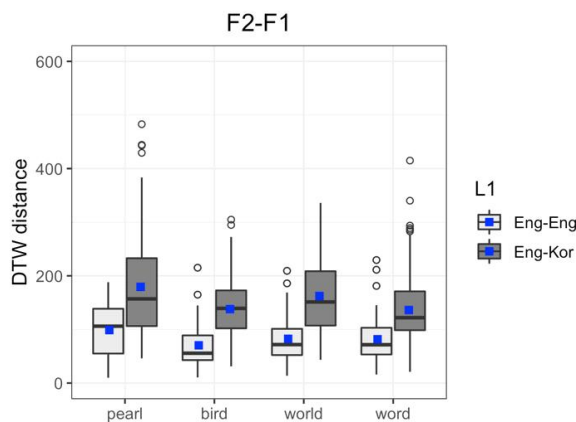
Figure 7. DTW distances in F2-F1. DTW, dynamic time warping.

Figure 8. DTW distances of F3. DTW, dynamic time warping.

Figure 7 shows the DTW distances in F2-F1 for the within-language pairs (English-English) (light-grey boxes) and the between-language pairs (English-Korean) (dark-grey boxes). As expected, the between-language distances are greater than the within-language distances. This means that the F2-F1 values are more similar to each other among English speakers when compared with Korean speakers. It is also noticeable that in the English-Korean pairs, the DTW distances of *pearl*, *world* are greater than those of *bird*, *word*. That is, the Korean speakers are more different from the English speakers in the production of the /rl/ cluster than the /rd/ cluster.

Table 2. Mixed-effects regression of dynamic time warping distances depending on F2-F1

	B	*β*	*SE*	*df*	*t*-value	*p*-value
Intercept	76.08		8.33	9.89	9.13	<.0001
L1:KOR	61.06	.43	9.64	8.05	6.33	<.001
Coda:rl	14.64	.10	5.14	1,091	2.84	<.01
L1:KOR× Coda:rl	19.05	.12	7.13	1,091	2.67	<.01

A mixed-effects linear regression analysis shows that the differences are all significant (Table 2). In the production of the /rl/ and /rd/ clusters, Korean speakers are significantly different from English speakers [$t(8.05)=6.33$, $p<.001$]. The positive coefficient value ($B=61.06$) indicates that the DTW distance is greater in English-Korean pairs than in English-English pairs. The /rl/ and /rd/ clusters were also significantly different in L2-L1 values [$t(1,091)=2.84$, $p<.01$], due to the presence and absence of dark /l/. The positive coefficient value ($B=14.64$) indicates that the DTW distance is greater when the coda is /rl/ than /rd/. This indicates that the production of /rl/ varies more than /rd/ in English as well as Korean speakers. The interaction of L1 and coda was also significant [$t(1,091)=2.67$, $p<.01$], with a positive coefficient ($B=19.05$). This means that the DTW distance is greater in English-Korean pairs and when the coda is /rl/. That is, Korean speakers produced /rl/ words differently from English speakers, and the difference is significantly greater compared with /rd/ words. The standardized coefficient ($β$) shows that the language effect is greater than the coda effect. That is, the between-language distances are greater than the differences due to coda.
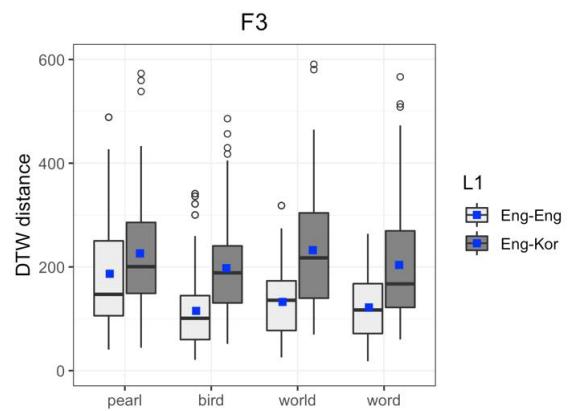
Figure 8 shows the DTW distances in F3 for the within-language pairs (English-English) (light-grey boxes) and the between-language pairs (English-Korean) (dark-grey boxes). For each word, the distance is greater when the compared speaker's L1 is Korean. It can also be seen that the DTW distances of the /rl/ cluster are greater than the /rd/ cluster. That is, for English-Korean pairs (dark-grey), the DTW distances are greater in *pearl* than in *bird*, and in *world* than in *word*. *Pearl* and *bird* look also different in English-English pairs (light-grey), so there seem great deviations among English speakers with the word *pearl*.

Table 3. Mixed-effects regression of dynamic time warping distances depending on F3

	B	*β*	*SE*	*df*	*t*-value	*p*-value
Intercept	118.62		20.07	8.49	5.91	<.001
L1:KOR	82.14	.36	26.19	6.71	3.14	<.05
Coda:rl	45.10	.20	8.71	1,090	5.18	<.0001
L1:KOR× Coda:rl	−3.26	−.01	12.05	1,090	−.27	.787

The result of mixed-effects linear regression is shown in Table 3. The between-language differences were significant [($t(6.71)=3.14$, $p<.05$]. This means that English speakers are more similar to each other than English vs. Korean speakers in F3. The positive coefficient value ($B=82.14$) indicates that the DTW distance is greater in Korean-English pairs than English-English pairs. The coda effect (/rl/ or /rd/) is also significant [$t(1,090)=5.8$, $p<.0001$]. The positive coefficient value ($B=45.10$) indicates that the DTW distance is greater for /rl/ than /rd/. However, the interaction term suggests that the F3 for the /rl/ cluster is not particularly different in the English-Korean pairs [$t(1,090)=−.27$, $p=.787$]. This indicates that both speaker groups have similar F3 trajectories, suggesting that the Korean speakers' /r/ production was similar to the English speakers' /r/ production at least for these words.[3]

4. Discussion and Conclusion

The present study examines the phonetic realization of word-final /rl/ sequences by English vs. Korean speakers. The word-final /rl/ cluster poses a challenge for Korean speakers because it is a sequence of two sounds that are not contrastive in Korean. In this paper we looked at two acoustic correlates of the /rl/ sequence: F2-F1 for the presence of dark /l/ and F3 for the presence of

rhoticity for /r/. The DTW distances show clear differences between English and Korean speakers' /rl/ production. A significant language-dependent difference is found with F2-F1 values. Whereas English speakers lower F2-F1 values toward the end of the word, Korean speakers do not. On the other hand, English and Korean speakers do not show significant differences in F3 trajectories. This means that when producing word-final /rl/, the Korean speakers produced /r/ relatively similar to the English speakers, but they did not produce the velarized /l/.

Unlike the previous research on Korean speakers' /r/ and /l/ production, the present study explores the use of the DTW algorithm for the direct comparison of formant trajectories. Despite temporal variations and segmental perturbations, the results indicate where the differences lie. Considering the mixed-effects results of the DTW distances, the Korean speakers produced /r/ in /rl/ clusters similarly to English /r/ but failed to produce dark /l/ in /rl/ clusters.

Note that the Korean language also has a phoneme /l/ which has the allophones [ɾ] in intervocalic position and [l] in word-final position. So in Korean, English coda /rl/ is transcribed as /l/ in loanwords, but phonetically, the Korean /l/ is not velarized. Rather, the Korean /l/ is closer to English /r/ than to English dark /l/ in terms of articulation. Korean /l/ is retroflex by some speakers, and Korean speakers' production of English /r/ is also retroflex (Hwang, 2021:368). This is further supported by error analyses in the previous studies. In coda position, the error rate of English /l/ by Korean speakers is high because dark /l/ is hard to pronounce (Park, 2004). Similarly, Kim & Rhee (2019:61) showed that single coda /r/ shows a lower error rate than single coda /l/ (69% vs. 58%).

Although the present study has limitations due to the small number of speakers and target words, it offers some initial observations. Further research is needed with more speakers and target words with various phonological environments. For example, the /rl/ clusters across syllable boundary [e.g. *early* [ɜːɹ.li] as in Kwon (2010)] can be examined using DTW. One may hypothesize that Korean speakers' *earl* ([ɜːɫ]) will be more different from English speakers' *earl* whereas Korean and English speakers' production *early* will be relatively more similar to each other compared to *earl*, because of the absence of dark /l/ in *early*. We can also test with different preceding vowels, such as [ɑ] (*Karl*). We may expect similar results if the production difficulties are due to differences in the phonemic system in L1 and L2, rather than coarticulation with preceding vowels.

The pedagogical implication is that in the production of coda /rl/, Korean learners of English should be guided to consciously make gestures for /l/ at the end of the word and not to omit it. The instruction should start with making it clear that /r/ and /l/ are two contrastive sounds in English, and both sounds in /rl/ should be separately articulated.

References

Allen, G. D. (1979, December). Transcription of the American /r/.

Current issues in the phonetic sciences: Proceedings of the IPS-77 Congress (pp. 1019-1025). Miami Beach, FL.

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer (version 6.1.35) [Computer program]. Retrieved from http://www.praat.org/

Borden, G., Gerber, A., & Milsark, G. (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning*, *33*(4), 499-526.

Boujnah, S., Sun, X., Marshall, D., Rosin, P. L., & Ammari, M. L. (2021). A novel approach for speaker recognition in degraded conditions. In N. Derbel, & O. Kanoun (Eds.), *Advanced methods for human biometrics* (pp. 139-146). Cham, Switzerland: Springer.

Carr, P. (2020). *English phonetics and phonology: An introduction* (3rd ed.). Hoboken, NJ: Wiley-Blackwell.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *English pronunciation: A course book and reference guide* (2nd ed.). Cambridge, UK: Cambridge University Press.

Cho, S., Nevler, N., Parjane, N., Cieri, C., Liberman, M., Grossman, M., & Cousins, K. A. Q. (2021). Automated analysis of digitized letter fluency data. *Frontiers in Psychology*, *12*, 654214.

Chung, H., & Pollock, K. E. (2014). Acoustic characteristics of adults' rhotic monophthongs and diphthongs. *Communication Sciences & Disorders*, *19*(1), 113-119.

Delattre, P., & Freeman, D. C. (1968). A dialect study of American R's by X-ray motion picture. *Linguistics: An Interdisciplinary Journal of the Language Sciences*, *6*(44), 29-68.

Espy-Wilson, C. Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English. *The Journal of the Acoustical Society of America*, *92*(2), 736-757.

Feinberg, D. R. (2022). *Parselmouth Praat scripts in Python*. Retrieved from https://doi.org/10.17605/OSF.IO/6DWR3

Feng, Z. (2020). Effects of identification and pronunciation training methods on L2 speech perception and production: Training adult Japanese speakers to perceive and produce English /r/-/l/. *Studies in Applied Linguistics & TESOL*, *20*(2), 57-83.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47-65.

Gimson, A. C. (1989). *An introduction to the pronunciation of English* (4th ed.). London, UK: Edward Arnold.

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: The dtw package. *Journal of Statistical Software*, *31*(7), 1-24.

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *The Journal of the Acoustical Society of America*, *105*(5), 2854-2865.

Hwang, Y. (2021). Articulatory characteristics of word-final English /ɹ/ produced by Korean learners of American English. *Studies in Phonetics, Phonology and Morphology*, *27*(2), 353-371.

Idemaru, K., & Holt, L. L. (2013). The developmental trajectory of

---

3 An anonymous reviewer suggested examining the effects of the final /d/ in *world*, which can be optionally released. English speakers released the final /d/ in 5 tokens (5/12=42%) and Korean speakers released it in 6 tokens (6/12=50%). For *world*, a mixed-effects regression analysis was conducted with DTW distance as a dependent variable, language, final /d/ release, and their interaction as independent variables, and speaker random intercepts. The results show that the effect of absence or presence of final /d/ release was not significant [t(597)=−1.02, p=.311 for F2-F1, t(95.6)=.047, p=.96 for F3].

children's perception and production of English /r/-/l/. *Journal of the Acoustical Society of America*, *133*(6), 4232-4246.

Ingram, J. C. L., & Park, S. G. (1998). Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America*, *103*(2), 1161-1174.

Iverson, G. K., & Sohn, H. S. (1994). Liquid representation in Korean. In Y. K. Kim-Renaud (Ed.), *Theoretical issues in Korean linguistics* (pp. 79-100). Stanford, CA: Stanford University Centre for the Study of Language and Information.

Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1-15.

Jang, T. (2005). Construction of an English speech database for Korean learners of English. *Language and Linguistics*, *35*, 293-310.

Johnson, K. (2003). *Acoustic and auditory phonetics*. Malden, MA: Blackwell.

Kang, H. S. (1999). Production and perception of English /r/ and /l/ by Korean learners of English: An experimental study. *Speech Sciences*, *6*, 7-24.

Kasuya, H., Tan, X., & Yang, C. S. (1994, September). Voice source and vocal tract characteristics associated with speaker individuality. *Proceedings of the 3rd International Conference on Spoken Language Processing* (*ICSLP 94*) (pp. 1459-1462). Yokohama, Japan.

Kim, R. E., & Rhee, S. C. (2019). A study on English liquids in the rated L2 English speech corpus of Korean learners. *Korean Journal of English Language and Linguistics*, *19*(1), 53-75.

Kwon, J. S. (2010). *An experimental study on the English sonorant cluster /rl/ produced by native speakers and young Korean EFL learners* (Master's thesis). Hankuk University of Foreign Studies, Seoul, Korea.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK: Blackwell.

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics* (6th ed.). Boston, MA: Cengage Learning.

Park, S. (2004). An analysis of the causes of English mispronunciation by Korean learners and solution for their improvement. *Journal of the Linguistic Society of Korea*, *40*, 113-143.

Park, S., & Jang, T. Y. (2016). Acoustic characteristics of English liquids produced by Korean learners of English. *Studies in Phonetics, Phonology and Morphology*, *22*(2), 289-315.

R Core Team. (2021). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Regier, K. T. (2014). Formant trajectory analysis using dynamic time warping: Preliminary results. *The Journal of the Acoustical Society of America*, *136*(4), 2082.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *26*(1), 43-49.

Sohn, H. S., & Lim, S. (2020). Phonetic implementation of the darkness of English word-final /l/ across prosodic positions: Comparison of native English speakers and EFL Korean speakers. *Korean Journal of English Language and Linguistics*, *20*, 450-474.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*(3), 291-311.

Stevens, K. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *The Journal of the Acoustical Society of America*, *123*(6), 4466-4481.

• **Hyesun Cho,** Corresponding author
Associate Professor, Dept. of Education
Graduate School of Education, Dankook University
152, Jukjeon-ro, Suji-gu, Yongin 16890, Korea
Tel: +82-31-8005-3968
Email: hscho@dankook.ac.kr
Fields of interest: Phonetics, Phonology