Korean Journal of English Language and Linguistics, Vol 22, December 2022, pp. 1443-1464
DOI: 10.15738/kjell.22..202212.1443

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

Production of Coda Obstruent Clusters by Korean and English Speakers: Acoustical and Dynamic Time Warping Analyses

Hyesun Cho (Dankook University)

Received: November 07, 2022
Revised: December 18, 2022
Accepted: December 30, 2022

Hyesun Cho
Associate Professor,
Department of Education,
Graduate School of Education,
Dankook University
E-mail: hscho@dankook.ac.kr

ABSTRACT

Cho, Hyesun. 2022. Production of coda obstruent clusters by Korean and English speakers: Acoustical and dynamic time warping analyses. *Korean Journal of English Language and Linguistics* 22, 1443-1464.

Coda obstruent clusters in English are known to be difficult for Korean learners of English to produce due to phonotactic differences between Korean and English syllables. In English, the coda obstruent clusters undergo reduction, so the medial consonant in CCC clusters is deleted by native speakers of English. In this study, coda obstruent clusters produced by Korean and English speakers are compared for their acoustic properties (center of gravity and intensity) and the similarity distance obtained by the dynamic time warping (DTW) algorithm. The CoG and intensity was overall lower in Korean speakers than in English speakers. The DTW similarity distance between the clusters produced by English speakers and those produced by Korean speakers was greater than the distance between the clusters produced by English speakers only. In addition, the DTW similarity distance between the clusters produced by English speakers and the error tokens by Korean speakers was greater than the distance between the clusters produced by English speakers and the non-error tokens by Korean speakers. The current study further employed the K-Nearest Neighbors (KNN) classifier for L1 and error detection using the DTW distance measures. The results showed that the DTW similarity distance was an adequate measure to capture the differences due to speakers' L1 and error production.

KEYWORDS

coda obstruent clusters, dynamic time warping, intensity, center of gravity

1. Introduction

1.1 English Coda Obstruent Clusters

One source of difficulties that Korean EFL learners have lies in the phonotactic differences between Korean and English (Lee 2000, Cho and Lee 2005). Korean does not allow consonant clusters to surface in syllable onset and coda, whereas English allows up to three consonants in the onset (e.g., *string*) and four consonants in the coda position (e.g. *contexts*) (Gimson 1989, Yavaş 2020). Studies show that Korean learners' difficulties are related to syllabic position, manner of articulation, and cluster length (Cho 2005, Lim 2021). In terms of syllabic position, Korean learners of English have more difficulties in producing consonant clusters in coda position than in onset position (Lee, Joh and Cho 2002, Cho 2005, Cho and Lee 2005). In terms of manner, clusters with only obstruents are more difficult to produce than clusters containing sonorants (Prator and Robinett 2009). Lim (2021: 217) showed that Korean speakers made more errors in fricative-fricative (*sixth*, *leaves*) and stop-fricative (*maps*, *kicks*) clusters than in nasal-fricative (*seventh*, *month*), nasal-stop (*seemed*, *pink*) clusters. In terms of cluster length, biconsonantal clusters are more difficult for Korean EFL learners to produce than triconsonant clusters (Cho 2005).

Based on these, we can infer that English consonant clusters with more than two obstruents in coda position will be the most difficult for Korean learners to pronounce (e.g., *asked* /skt/). The difficulties will be added if the cluster contains phonemes that do not exist in the Korean phonemic inventory, such as interdental fricative /θ/ (e.g., *sixths* /ksθs/). The present paper focuses on coda consonant clusters with at least two consecutive obstruents produced by Korean speakers.

1.2 Repair Strategies for Coda Obstruent Clusters

Coda obstruent clusters violate the Sonority Sequencing Principle in English, which states that sonority must decrease toward the end of a syllable (Clements 1990), so codas such as /kp/ are unattested. Nevertheless, alveolar obstruents are allowed to occur multiple times at the right edge of a syllable, e.g., *contexts* /kʌtɛksts/, licensed by the Prosodic-word (PrWd) node instead of the syllable node (Roca and Johnson 1999). Kreidler (2004: 236) also noted that consonant clusters in *asks, risked,* and *sixths* undergo cluster simplification. In the surface form, these coda obstruent clusters undergo reduction, e.g. *acts* /ækts/ [æks], *lifts* /lɪfts/ [lɪfs], *asked* /æskt/ [æst], *depths* /dɛpθs/ [dɛps] (Prator and Robinett 2009: 182, Celce-Murcia et al. 2010: 107). According to Celce-Murcia et al. (2010: 107), English has the following reduction rules in coda position.

(1) a.	skt→ st	asked	/æskt/	[æst]
	sks → ss	asks	/æsks/	[æss][1]
	sts → ss	lists	/lɪsts/	[lɪss]
	kts → ks	facts	/fækts/	[fæks]
	pts → ss	scripts	/skrɪpts/	[skrɪps]
	fθs→ fs	fifths	/fifθs/	[fifs]
b.	ksθs → ks	sixths	/sɪksθs/	[sɪks]
	ntθs → n(t)s	tenths	/tɛntθs/	[tɛn(t)s]

In (1a), the triconsonantal clusters in coda position (CCC#) are reduced to CC by deleting the medial consonant ($C_1C_2C_3 \rightarrow C_1C_3$). In (1b), when there are four consonants (CCCC#), it is reduced to C(C)C ($C_1C_2C_3C_4 \rightarrow C_1(C_2)C_4$). Celce-Murcia et al. (2010) recommend teaching the way native speakers simplify coda consonant clusters like these, though omissions should be restricted because too many omissions may give listeners the impressions that the speaker is uneducated (Prator and Robinett 2009).

There are two repair strategies in producing coda consonant clusters: deletion or epenthesis. As can be seen in (1), native speakers of English reduce coda obstruent clusters by deleting consonants in the middle. Note that adult native English speakers do not insert epenthetic vowels in their production of the clusters. In L1 acquisition by English children, simplification of a consonant cluster (*blue* [bu] in onset) or deletion of the entire consonant cluster (*milk* [mɪ] in coda) are commonly observed (Lee 2003:341, Weinberger 1994), due to their inability of articulating consonant clusters. In contrast, L2 adult learners of English rely more on vowel epenthesis to resolve consonant clusters (*trick* [tɨr]) (Lee, Joh and Cho 2002). This is to preserve all the consonants in the cluster, referred to as the recoverability principle (Weinberger 1994). The present study will examine how the different repair strategies emerge focusing on the production of coda obstruent clusters.

Coda obstruent clusters are more difficult to produce when they contain sounds that are not phonemes in their L1. For example, Lee, Joh and Cho (2002) noted that /s/-initial clusters are easier than /f/ and /ʃ/-initial clusters for Korean speakers whose L1 lacks the latter two phonemes. Likewise, we can expect that coda obstruent clusters that contain interdental fricative /θ/ will also be difficult for Korean speakers due to the lack of the phoneme in the Korean phonemic inventory. According to Hong et al. (2014: 65), Korean speakers replaced 20.7% of the English /θ/ sounds with other phonemes: /s/ (10.4%), /t/ (5.4%), /d/ (3.8%), and other minor variations (1.1%). Given this, we can expect that Korean speakers will also replace /θ/ with /s/ in a cluster most often, but without the

[1] An anonymous reviewer noted that he or she cannot agree with the transcriptions for *asks* [æss] and *lists* [lɪss]. As indicated above, the transcriptions are from Celce-Murcia et al. (2010: 107) and will be assumed as is in this paper. The transcriptions here seem to be broad transcription. It is possible that the articulatory gesture for the cluster-medial consonants (*asks, lists*) is hidden by gestural overlap with the preceding and following consonants, rather than being completely deleted. For example, in *perfect memory*, the medial consonant [t] is not audible ([..fəkmɛm..]), but the alveolar closure gesture for [t] remains (Browman and Goldstein 1989: 216). Likewise, we can imagine that the velar gesture for [k] in *asks* exists but it is not audible.

knowledge of the reduction rule that deletes /θ/ as in (1), they will be uncertain how to distinctively pronounce /θ/ and /s/ when they are immediately adjacent (e.g., *mon<u>ths</u>*).

Against this backdrop, the present study examines Korean speakers' coda obstruent clusters, comparing those with English speakers, focusing on the clusters with three or four consonants ((C)VCCC, (C)VCCCC) as presented in (1). Previous studies on Korean speakers' production of consonant clusters mostly focused on biconsonantal clusters (CVCC) (Kim 2015, Kwon 2008, Lim 2021), though Cho (2005) looked at biconsonantal and triconsonantal clusters together. Previous studies on this topic usually focused on the types and frequencies of errors such as insertion, deletion, and replacement (e.g., Cho 2005), instead of their acoustic characteristics. The present study examines the acoustic characteristics and overall acoustic similarities of the clusters produced by Korean and English speakers, along with the traditional error analysis.

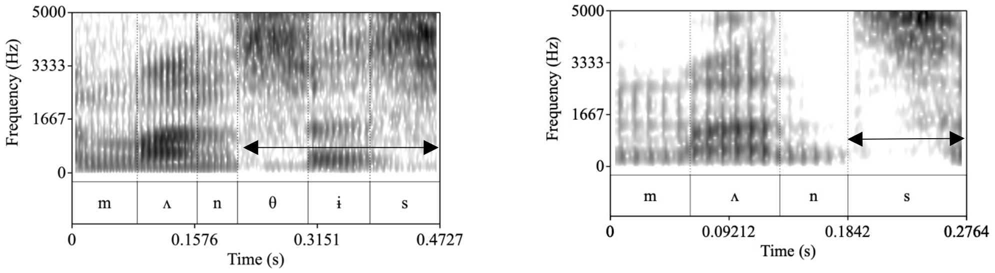
1.3 Acoustic Correlates: Center of Gravity and Intensity

To examine the acoustic properties of the coda obstruent clusters, we look at the intensity and center of gravity in the frication noise. The intensity of frication noise tends to be higher in voiceless fricatives than in voiced fricatives (Balise and Diehl 1994; Silbert and de Jong 2008). /s/ is a sibilant ([+strident]), so it has a higher frication intensity than a non-sibilant fricative /θ/ ([-strident]) (Gussenhoven and Jacobs 2011). Center of gravity (CoG) is a measure of how high the frequencies are in the given spectrum, and it is a useful measure for fricatives (Gordon et al. 2002). In American English, CoG of sibilant fricatives (/s, ʃ/) is higher than that of non-sibilant fricatives (/f, θ/) (Park 2021). In Korean, CoG of the frication portion of voiceless plain /s/ and tense /s'/ is lower than English voiceless fricatives /s/ and /ʃ/ (Sung and Cho 2010: 40). Thus, by transfer effects, it can be expected that the Korean speakers' production of English /s/ will have a lower CoG than the English speakers.

1.4 Similarity Distance by Dynamic Time Warping

The overall similarity between Korean and English speakers' coda obstruent clusters was measured by the dynamic time warping (DTW) algorithm (Giorgino 2009, Sakoe and Chiba 1978). Dynamic time warping is an algorithm that compares two time series data of different lengths, such as two sets of sound waveforms. The algorithm finds the optimal path to match two set of time-series data. Cho et al. (2021) used the dynamic time warping algorithm in order to measure phonetic similarity between two words starting with the 'f' letter. They extracted the 13 mel-frequency cepstral coefficients (MFCCs). The MFCC is a method to extract feature vectors from speech signals, used in speech recognition (Muda et al. 2010). In the present study, the MFCC vectors will be compared between the pairs of obstruents using the DTW method.

While acoustic properties such as CoG and intensity can only capture the local characteristics of a certain segment at a time, the DTW algorithm can be applied over a span longer than a segment, such as words and sentences (Sun et al. 2014). The algorithm can be applied regardless of the types of errors, such as substitution, deletion, or vowel epenthesis because it can be used to compare the similarity of a series of segments, not just one segment. One of the most frequent errors by Korean learners of English is vowel epenthesis (Cho 2005, Cho and Lee 2015). Even when the error production contains epenthetic vowels, Korean speakers' errors can be quantitatively compared with English speakers' production. For example, Figure 1 shows the spectrograms of *months* produced by Korean and English speakers in the present study. The arrows indicate the range where the DTW similarity distance is compared. The two ranges, corresponding to the obstruent codas of *mon<u>ths</u>*, are compared to each other by the DTW algorithm. Figure 2 shows an example of how the two ranges from Figure 1 are compared. The blue line indicates the optimal path that minimizes cost in warping.



(a) (b)

Figure 1. Spectrogram of *months* Produced by (a) Korean Speaker (K3) and (b) English Speaker (E5) (The DTW distance is computed by comparing the MFCC features in the ranges indicated by the arrows.)

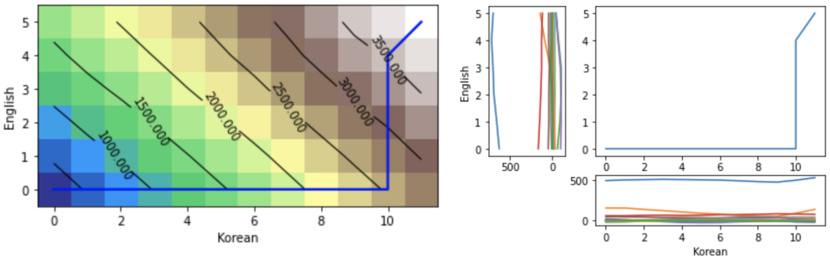


Figure 2. Example of DTW Plots (Left: Density Plot, Right: Threeway Plot) Showing the Optimal Path (Blue Line) to Compute Similarity Distance between the Two Ranges (Arrows) of the Sound Waveforms in Figure (1a) and (1b). (The DTW distance value is 216.7 in this example.)

The dynamic time warping algorithm has been used in the field of speech signal processing such as automatic speech recognition (Sun et al. 2014) and pronunciation error detection (Kanters et al. (2009) for Dutch, Zhao et al. (2012) for Chinese, and Bugdol et al. (2014) for Polish). Bugdol et al. (2014) used the DTW algorithm to detect pronunciation errors, but they used artificial errors by native speakers of the Polish language. Bugdol et al. (2014) considered pronunciation error detection as a classification problem. In the present study, the DTW distance values were fed to the K-Nearest

Neighbors (KNN) algorithm, one of the most commonly used classification algorithms (Cover and Hart 1967, Cunningham and Delany, 2007, Fix and Hodges 1951), and the KNN algorithm classifies errors and non-errors based on the DTW distance values.

2. Research method

2.1 Subjects

The subjects were 6 native speakers of English and 6 native speakers of Korean. The speakers were all male. The Korean speakers were college students in their 20's. None of the Korean speakers have been to American-speaking countries before. Their English proficiency was intermediate to high-intermediate (TOEIC scores 650-875). The English speakers were all from America, one mid-western (Chicago), two western (California), two southern (Texas), and one eastern (Western Massachusetts). The English speakers were in their late 20's to late 30's.

2.2 Speech Materials

The speech material was a dialogue passage from Celce-Murcia et al. (2010: 108) (see Appendix). The dialogue contains words with various coda obstruent clusters, designed as a pronunciation drill for English learners. The dialogue is a conversation between two people (a veterinarian and a pet owner), but in our research, one speaker read both roles. It is expected that dialogue reading will provide a more natural setting than reading target words in a carrier phrase, which would facilitate the application of the reduction rules. The target words contained in the passage are shown in Table 1, arranged according to the medial consonant. The dialogue contained words with other clusters such as (*listless, textbook, asked*), but the scope of this study is limited only to the words with word-final coda clusters ending in /s/ to control for the context.

Table 1. Target Words in the Speech Materials

	Consonant cluster			Medial (deleted) consonant
a.	acts	/kts/	[ks]	t
	lifts	/fts/	[fs]	t
	facts	/kts/	[ks]	t
	bursts	/rsts/	[rss]	t
b.	months	/nθs/	[n(t)s]	θ
	strengths	/ŋθs/	[ŋs]	θ
	fourths	/rθs/	[rs]	θ
	fifths	/fθs/	[fs]	θ

The second column (medial consonant) in Table 1 shows the consonant that is deleted according to the cluster reduction rules in (1). All the words end with /s/, which is supposed to remain after reduction. They all contain at least two consecutive obstruents in word-final position. Four words contain sonorants in the cluster (*bursts, months, strengths, fourths*). In the surface form, all the clusters are supposed to end with frication noise corresponding to [s]. The acoustic analysis in this paper is conducted with this word-final frication noise.

2.3 Recording Procedure

The subjects were recorded reading the dialogue three separate times. Korean speakers were recorded in a sound-attenuated recording studio at a university. English speakers recorded online using Vocaroo (http://vocaroo.com), an online recorder, due to the corona pandemic. The English speakers were given instructions by email. They were asked to record the speech materials in a quiet room and submit a shareable URL link to their recordings. All speakers were compensated for their recordings. The total number of the words to be analyzed is 144 for Korean speakers (8 words × 6 speakers × 3 repetitions) and 144 for English speakers (8 words × 6 speakers × 3 repetitions).

2.4 Analysis Methods

2.4.1 Segmentation and error classification

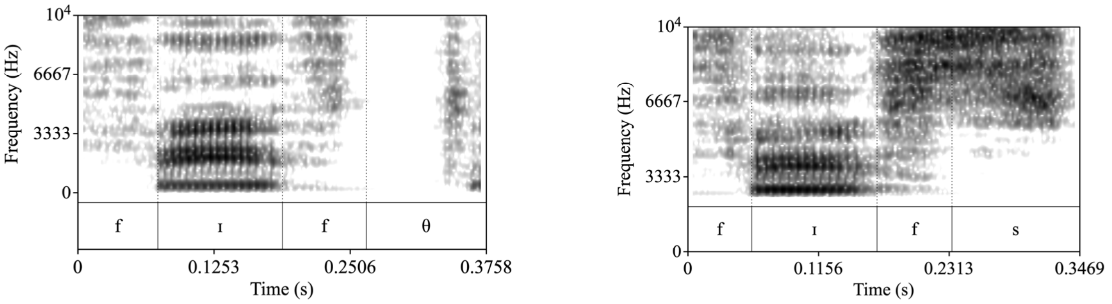
The tokens recorded by Korean speakers were classified into two types: error and non-error. The recorded tokens were all manually examined by the author using Praat (Boersma and Weenink 1992-2020, version 6.1.35)[2]. Tokens that do not follow the reduction rule in (1) were considered errors. Error tokens and types of errors were determined visually and auditorily based on audio sounds and spectrograms[3].

In Figure (3a), C_3 is deleted instead of C_2 (*[fɪfθ], instead of [fɪfs]), which is considered an error. As shown in the spectrogram, the [θ] sound is pronounced with an interval of closure and release. The formant frequencies of the frication noise in this token have the characteristics of [θ] (F1: 1199, F2:

[2] An anonymous reviewer pointed out a problem of determining errors by one rater (the author) alone which may undermine the reliability of error classification. The reviewer suggested, alternatively, re-rating the recordings by the same author with some time interval between the two ratings and comparing the results. Following the suggestion, I classified the same tokens again into errors and non-errors. The first rating was done on 19 August, 2022 and the second rating was done on 21 December, 2022, which is about four-months time interval. The two ratings matched 100% regarding whether a token is an error or non-error. One mismatch was found regarding the type of the error: the error type of K1's first recording of *fourth* was C_2 & C_3 deletion in the first rating, but in the second rating, it was only C_3 deletion with C_2 retained. Based on the sound and spectrogram, the latter was considered correct. The results in this paper have been revised accordingly.

[3] When coda /ts/ sounds (*facts, acts, lifts, bursts*) became an affricate, instead of deleting the medial /t/, they were not considered as errors.

2007, F3: 3079, F4: 4690), and it auditorily sounds like [θ] rather than [s]. Typically, [θ] has F4 above 4000 Hz and has energy in a lower frequency range than [s] (Ladefoged 2005:57-58), as is observed here. On the other hand, in Figure 3(b), C_2 is deleted, while C_3 is retained ([fɪfs]), which is correct. The frication in Figure (1b) is [s], considering its frequency range (F1: 1810, F2: 4337, F3: 4959, F4: 5690). According to Ladefoged (2005), [s] has intense energy in upper-frequency range above 5000 Hz and little energy below 3500 Hz, corresponding to the frequency range of the frication noise shown in Figure 3(b).



(a) (b)

Figure 3. Spectrograms of *fifths* [fɪfs] Produced by a Korean Speaker (K1): (a): C_3 ([s]) Deletion Error, (b): Non-error (C_2 ([θ]) is reduced). (The frequency range in the spectrograms is set up to 10000 Hz to illustrate the characteristics of frication noise for [θ] and [s].)

The presence of an epenthetic vowel was determined by the presence of clear formants (F1, F2 and F3) (Shin and Iverson 2014) and regular pitch periods. By these criteria, there were six epenthetic vowels identified (all produced by speaker K3). A spectrogram for this is shown in Figure (1a) in Section 1.4. The mean vowel duration for epenthetic vowels was 80 ms (*SD* = 9), with the minimum of 72 ms. The mean is very similar to the mean vowel duration (*M* = 84 ms) perceived as epenthesis by Korean male speakers (Kim 2018: 16). Epenthetic vowels are generally shorter than lexical vowels (Davidson 2006), and the mean duration of speaker K3's epenthetic vowels is significantly shorter than that of his lexical vowels (*M* = 145, *SD* = 35) ($t(27.84) = 7.97$, $p < 0.0001$). These epenthetic vowels all have regular pitch periods, so their pitch values were all automatically measurable by Praat (*M* = 132, *SD* = 6.7), ensuring the presence of a vowel.

2.4.2 Acoustical characteristics: Center of Gravity (CoG) and intensity

Center of Gravity and intensity for the word-final frication noise were automatically collected using Praat scripts. Before collecting intensity values, all the recorded files were RMS (Root Mean Square) equalized to have the same peak intensity because speakers were recorded in different places. RMS equalization rescales sound files with different amplitudes to have the same maximum amplitude. A Praat script was used for the RMS equalization (Beckers 2002).

2.4.3 Similarity distance: Dynamic time warping with MFCC features

Similarity distances between coda obstruent clusters produced by English and Korean speakers were measured using the dynamic time warping (DTW) algorithm. Unlike CoG and intensity values, similarity distance values were collected from the entire range of the obstruent clusters, as illustrated in Figure 1. This may include closure and epenthetic vowels altogether, if any, allowing direct comparison between clusters beyond a segment. Python codes[4] were used to extract DTW distance values using the dtw-python package (Giorgino 2009) on the Google Colaboratory platform. The *librosa* library was used to extract MFCC (Mel-Frequency Cepstral Coefficient) values from consonant cluster portions. MFCC is a set of vectors extracted from acoustic features of speech signals, which effectively represents a given speech signal, so is commonly used in speech signal processing (Muda et al. 2010).

Table 2. Description of the Pairs Where DTW Distances Were Measured

Pairs		Number	Predicted DTW distance
E-E	(a) English – English	2160	smallest
E-KN	(b) English – Korean (Non-error)	2286	between (a) and (c)
E-KE	(c) English – Korean (Error)	306	greatest

Table 2 summarizes the pairs where DTW distances were measured. DTW distance values were computed between the pairs of English speakers' tokens and other English speakers' tokens (E-E), between the pairs of English speakers' tokens and Korean speakers' non-error tokens (E-KN), and between the pairs of English speakers' tokens and Korean speakers' error tokens (E-KE). There were 2160 English-English pairs (6 English speakers × 3 repetitions × 5 English speakers[5] × 3 repetitions × 8 words) and 2592 English-Korean error or non-error pairs (6 English speakers × 3 repetitions × 6 Korean speakers × 3 repetitions × 8 words). Among the latter, there were 306 pairs of English speakers' tokens and Korean speakers' error tokens (E-KE), leaving 2286 pairs of English speakers' tokens and Korean speakers' non-error tokens (E-KN).

The hypothesis is that the DTW distance will be greater in the order of E-E < E-KN < E-KE. As mentioned earlier, even non-error tokens produced by Korean speakers may have different characteristics from English speakers' tokens in terms of CoG and intensity. To test this hypothesis, the DTW distance values between English speakers' tokens (E-E) were compared with the DTW distance values between English speakers' tokens and Korean speakers' non-error tokens (E-KN), hypothesizing that the distance of the latter will be greater. Then, the distance between English speakers' tokens and Korean speakers' non-error tokens (E-KN) was compared with the distance between English speakers' tokens and Korean speakers' error tokens (E-KE). It is expected that error tokens will have a greater distance

[4] The codes were adapted from Sunghye Cho's tutorial Python codes used at the Winter Workshop hosted by the Korean Society of Speech Sciences (December 8, 2021).

[5] Words from the same speakers were not paired up. Only the tokens between different speakers were paired.

from English speakers' tokens than non-error tokens.

In addition, K-Nearest Neighbors Classifier (Cunningham and Delany 2020), a machine-learning classification algorithm, was applied to examine how well DTW distances can be used to detect speakers' L1 and pronunciation errors.

2.4.3 Statistical analysis

The CoG, intensity, and DTW distance values collected as described above were analyzed using R statistical software (R Core Team 2022). Linear mixed-effects regression analyses were conducted with the measurement values (CoG and intensity) as dependent variables and language (Korean, English) and medial consonants (/t, θ/) as fixed effects. For CoG and intensity, random intercepts were speakers and words. For DTW distance values, random intercepts were words, because the speakers were paired up.

3. Results

3.1 Error Analysis: Types and Frequency of Errors

In this section, the types and frequency of errors in the Korean speakers' coda obstruent clusters are presented.

Table 3. Error Types and Frequency of Errors

Medial	Word		C_3 deletion $[C_1C_2]$	V insertion $[C_1C_2VC3]$	C_2&C_3 deletion $[C_1]$	No error
[θ]	fourths	/rθs/	1	3	2	12
	months	/nθs/		3		15
	fifths	/fθs/	2			16
	strengths	/ŋθs/				18
	Total		3 (2%)	6 (4%)	2 (1%)	61(42%)
[t]	bursts	/sts/	3			15
	lifts	/fts/	2		1	15
	facts	/kts/				18
	acts	/kts/				18
	Total		5 (3%)	0 (0%)	1 (1%)	66 (46%)
Total			8 (6%)	6 (4%)	3 (2%)	127 (88%)

Table 3 shows the types and frequency of errors, classified by the medial consonant. The total error rate was 12% (17/144). Overall, the most frequent type of error was C_3 deletion with C_2 retained (6%), followed by vowel epenthesis after C_2 (4%) and C2 and C_3 deletion (2%). The error rate was higher in [θ] (46%) than in [t] (42%), as expected. [θ] is not a phoneme in Korean, so Korean speakers have the most difficulty with the clusters containing [θ]. The most frequent error type for clusters with [θ] (*fourths* and *months*) was vowel insertion, consistently found in one speaker (K3) (Figure (1a) in

Section 1.4).

Even the tokens that are classified correct may likely have different acoustic properties compared with those of English speakers. The next section presents the analyses of the acoustic properties in the frication portion of the clusters that are classified as non-error (127 out of 144, those in the last column of Table 3). Section 3.3 presents the analysis of similarity distance using dynamic time warping.

3.2 Acoustic Analysis: CoG and Intensity

3.2.1 Center of Gravity

In this section, the acoustic characteristics of the word-final frication noise in the coda obstruent clusters produced by the Korean speakers are compared with those produced by the English speakers. In particular, the center of gravity ('CoG') and intensity (in dB) of the noise portion of obstruent clusters are compared, which represent major characteristics of the frication noise. Figure 4 shows the center of gravity of Korean and English speakers arranged by the medial consonant. From the figure, it can be seen that the English speakers have overall higher CoG with smaller variation than the Korean speakers.

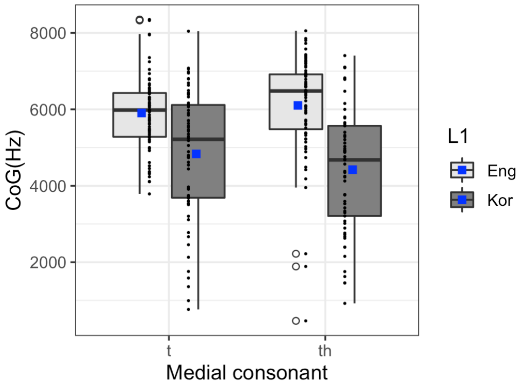


Figure 4. Center of Gravity by Speaker L1 and Medial Consonant

Table 4. Mixed Effects Regression Results for CoG

	B	*S.E*	*df*	*t*	*p*
(Intercept)	5903.88	385.44	14.22	15.32	<0.0001***
Lang:Kor	-1110.45	506.29	11.86	-2.19	<0.05*
Cons: [θ]	199.05	285.98	10.00	0.70	0.50
Lang:Kor × Cons: [θ]	-519.37	290.31	252.56	-1.79	0.07

A linear mixed-effects regression model was fitted with CoG as a dependent variable, L1 and medial consonant and their interactions as fixed effects, with random intercepts for speakers and words. The results in Table 4 suggest that CoG is significantly lower in Korean speakers than in English speakers

(t(11.86) = -2.19, $p < 0.05$) regardless of the underlying medial consonant. There were no significant effects of the medial consonant (t(10.00) = 0.70, p = 0.5). That is, regardless of the medial consonant, CoG was not significantly different. Korean speakers' medial consonant [θ] had a negative coefficient (-519.37), but it was not significantly different from zero (t(252.56) = -1.79, p = 0.07).

3.2.2 Intensity

Figure 5 shows the intensity of the frication portion of coda obstruent clusters produced by Korean and English speakers. The figure shows that English speakers have an overall higher intensity than Korean speakers.

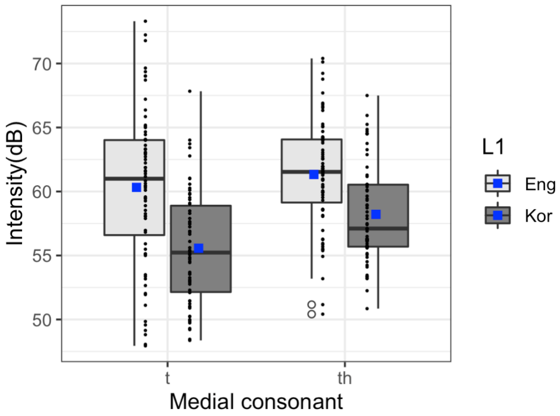


Figure 5. Intensity (dB) by Speaker L1 by the Medial Consonant

Table 5. Mixed Effects Regression Results for Intensity (dB)

	B	*S.E.*	*df*	*t*	*p*
(Intercept)	60.32	1.87	15.62	32.20	< 0.0001***
Lang:Kor	-4.67	2.00	10.60	-2.34	< 0.05*
Cons:[θ]	1.01	1.80	6.39	0.56	0.60
Lang:Kor × Cons:[θ]	1.64	0.68	251.36	2.42	< 0.05*

A linear mixed-effects regression model was fitted to the data with intensity as a dependent variable, L1, medial consonant, and their interaction as fixed effects, with random intercepts for speakers and words. The results in Table 5 show that the coefficient for Korean speakers is –4.67, suggesting that the intensity of Korean speakers' frication was significantly lower than that of English speakers (t(15.62) = -4.67, $p < 0.05$). The interaction term shows that for Korean speakers, frication intensity was slightly higher when the medial consonant is [θ] (t(251.36) = 2.42, $p < 0.05$).

To summarize, both CoG and intensity were significantly lower in Korean speakers than in English speakers, conforming to the previous studies (Park 2021, Sung and Cho 2010). For Korean speakers, intensity is slightly higher when the medial consonant is [θ]. The overall result suggests that even in the tokens where the reduction rule was correctly applied, there were still significant acoustical differences between English and Korean speakers. For English speakers, the frication noise in the surface form was not affected by the underlying medial consonant, but for Korean speakers, the frication noise had different intensities depending on the medial consonant.

3.3 Similarity Distance Obtained by the Dynamic Time Warping Algorithm

3.3.1 Differences between Korean and English speakers

Similarity distances measured by the dynamic time warping algorithm are analyzed in this section, comparing English and Korean speakers. Figure 6 shows that the distance between English speakers' and Korean speakers' non-error tokens (E-KN) was overall greater than the distance between English speakers' tokens (E-E). According to the Figure, for the E-KN pairs, the medial consonant [θ] had greater distance values than [t] whereas, for English-English pairs, the distance was not very different depending on the medial consonant.

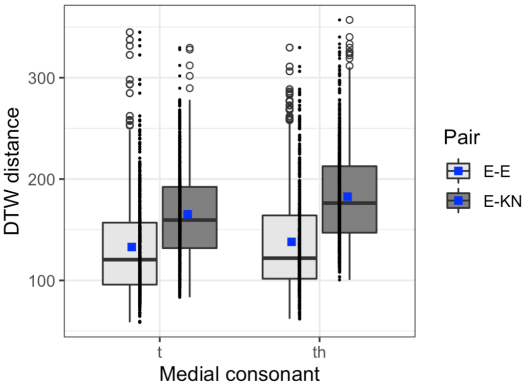


Figure 6. DTW Distance by the Medial Consonant and Speaker L1

Table 6. Mixed Effects Regression Results for DTW Distance

	B	*S.E.*	*df*	*t*	*p*
(Intercept)	132.94	7.64	6.22	17.39	0.00***
Lang:Kor	32.97	1.91	4436.30	17.22	< 0.0001***
Cons: [θ]	5.09	10.81	6.22	0.47	0.65
Lang:Kor × Cons: [θ]	11.05	2.74	4436.55	4.04	< 0.0001***

A linear mixed-effects regression was conducted with the DTW distance values as a dependent variable, L1, medial consonant, and their interactions as fixed effects, with random intercepts for words. Speakers were paired up according to their L1, so random intercepts for speakers were not included in the model. The results in Table 6 suggest that the DTW distance was significantly different by speakers' L1 ($t(4436.30) = 17.22$, $p < 0.0001$). This means that the similarity distance between English speakers' tokens and Korean speakers' non-error tokens was significantly greater than the distance between the tokens by English speakers only. The interaction of the medial consonant and L1 was significant. The distance between English speakers' tokens and Korean speakers' non-error tokens was significantly greater when the medial consonant was [θ] ($t(4436.55) = 4.04$, $p < 0.0001$). To summarize, Korean speakers' word-final obstruent clusters were more different from English speakers' when the cluster contained [θ] which is not a phoneme in Korean.

3.3.2 DTW distance for error and non-error tokens

We can hypothesize that the distance between English speakers' tokens and Korean speakers' error tokens (E-KE) will be greater than the distance between English speakers' tokens and Korean speakers' non-error tokens (E-KN). Figure 7 shows the DTW distance values between English speakers' tokens and Korean speakers' non-error tokens (E-KN) and those between English speakers' tokens and Korean speakers' error tokens (E-KE) for each medial consonant separately. In the figure, we can see that error tokens overall have greater distance values than non-error tokens, verifying the hypothesis.

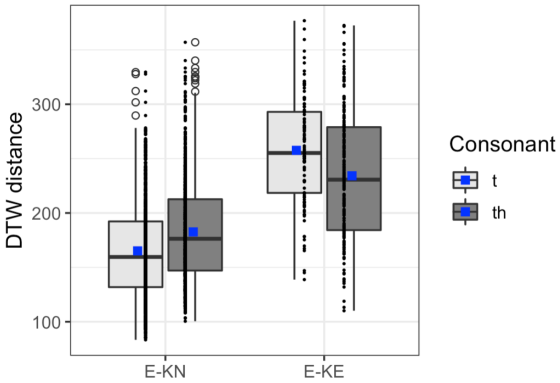


Figure 7. DTW Distance by Error Depending on Medial Consonant
(E-KN: distance between English speakers' tokens and Korean speakers' non-error tokens, E-KE: distance between English speakers' tokens and Korean speakers' error tokens)

To test the significance of the differences between errors and non-errors, a mixed-effects linear regression model was fitted to the data. The dependent variable was distance values in the pairs E-KN and E-KE, and fixed effects were the type of errors, medial consonant, and their interactions. There were random intercepts for words. Error type had four levels (C_3 deletion, C_2 retention & vowel insertion, C_2 & C_3 deletion, and no error). Among these, the reference level was no error. The results are presented in Table 7.

Table 7. Mixed Effects Regression Results for DTW Distance and Error Types

	B	*S.E.*	*df*	*t*	*Pr(>|t|)*
(Intercept)	166.94	10.25	6.01	16.29	<0.0001***
error type: C_3 deletion	77.20	4.77	2583.51	16.17	<0.0001***
error type: C_2 & C_3 deletion	42.39	10.16	2581.71	4.17	<0.0001***
error type: C_2 retention & vowel insertion	22.35	4.45	2584.38	5.02	<0.0001***
medial consonant: [θ]	15.81	14.50	6.01	1.09	0.32
error type × cons: C_2 & C_3 deletion, cons[θ]	54.41	12.03	2583.14	4.52	<0.0001***
error type × cons: C_3 deletion, cons[θ]	-7.57	8.53	2583.26	-0.89	0.38

From the results above, we can see that all three error types have significantly greater distance than non-error tokens ($p < 0.0001$). The positive coefficient values for the three error types indicate that error tokens have a greater distance from English speakers' tokens than non-error tokens. Among the errors, C_3 deletion has the greatest distance from the English tokens (t(2583.51) = 16.71, $p < 0.0001$). C_2 and C_3 deletion (e.g. *lifts* /lɪfts/ *[lɪf] for [lɪfs]) has the second greatest distance (t(2581.71) = 4.17, $p < 0.0001$). C_2 retention accompanied by vowel insertion has the smallest distance (e.g. *months* /mʌnθs/ [mʌnθɨs] for [mʌns]) (t(2584.38) = -5.02, $p < 0.0001$). Distance is not significantly different depending on the medial consonant alone. The coefficient for consonant [θ] is not significantly different from zero (t(6.01) = 1.09, p = 0.32), but it is positive (15.81), so it is in the expected direction following our hypothesis (the more difficult, the greater the distance). The interaction between error type and medial consonant (C_2 and C_3 deletion where C_2 was [θ]) is significant (e.g. *fourths* /fɔːrθs/ *[fɔːr] for [fɔːrs]) (t(2583.14) = 4.52, $p < 0.0001$).

3.3.3 Classification using the K-Nearest Neighbors (KNN) Classifier

In this section, an attempt is made to classify errors from non-errors, and native and nonnative, based on DTW distance values. A machine learning algorithm, K-Nearest Neighbors (KNN) is used (Cover and Hart 1967, Cunningham and Delany, 2007, Fix and Hodges 1951). KNN is one of the simplest and most commonly used classification algorithms that classifies members of a category based on geometric distance. It classifies a new data point based on the classes of the nearest neighboring data points. The class of the new data point is determined as the category to which a majority of the nearest K data points belong, that is, majority voting. KNN algorithm is a nonparametric regression that does not make a strong assumption about the shape of the regression function (Altman 1992). Thus it can be used to classify data without a priori knowledge about the shape of the regression curve. It is a simple classifier but its performance is comparable to other more complex classifiers (Alfeilat et al. 2019, Cunningham and Delany 2007, 2020, Steinbach and Tan 2009).

In the present study, the kNNeighborsClassifer[6] in the *Scikit-learn* library (v.1.1.2) is used for the classification of the data. The hyperparameter was the number of neighbors, which was chosen through random search using the RandomizedSearchCV[7] in the *Scikit-learn* library (v.1.1.2) with 5-fold

[6] https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html

[7] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html

cross-validation. The codes were written in Python 3 on Jupyter Notebook (v.6.0.3).

For error detection, only English-Korean pairs (E-KN, E-KE) were used as the training and testing data. Outliers were removed for the best performance[8]. The predictors were DTW distance (numerical variable) and medial consonant (categorical variable). Categorical variables were one-hot encoded and concatenated with the numerical variable. The dependent variable was binary labels, error or non-error. The combined data were split into train and test data. From a random search, the best parameter value for the number of neighbors was 9. With this, after learning with training data, the training accuracy of classification was 69.4%. The accuracy for the test data was 61%, which means that the algorithm correctly classified 61% of the new data that was not in the training set using only the information about DTW distance and medial consonants. From this result, we can see that DTW distance can reflect characteristics of error tokens from non-error tokens for coda obstruent clusters.

DTW distance can also effectively determine whether given obstruent clusters are produced by English or Korean speakers. The classification was conducted under the same procedure above. The dependent variable was binary labels, English or Korean. The hyperparameter (number of neighbors) was set by default. With this, the training accuracy was 80.3% and the classification accuracy for the test data was 76.4%. To summarize, DTW distance was a useful measure for both error and L1 detection. In particular, the KNN algorithm resulted in higher accuracy in the classification of L1 detection than error detection.

4. Discussion

The present study adopts a novel approach, the DTW algorithm, to the analysis of coda obstruent clusters produced by Korean learners of English and native speakers of English, along with the analysis of traditional acoustic properties for fricatives, CoG, and intensity. CoG of the frication noise was lower in the coda obstruent clusters produced by Korean speakers than those produced by English speakers. This is due to the L1 transfer effect where Korean speakers' alveolar fricatives have lower CoG than English /s/. In addition, Korean speakers have a wider variation in CoG values than English speakers. This may mean that Korean speakers have different proficiency levels and some speakers were more unsure of how to produce English fricatives than other speakers.

The intensity in the frication noise was lower in Korean speakers than in English speakers. The intensity had a significant interaction with the medial consonant. Korean speakers had a slightly but significantly higher intensity in the frication when the medial consonant was [θ] (e.g. *months* [mʌns]), whereas English speakers did not show any such differences depending on the medial consonant. Our results suggest that English speakers do not differentiate the coda [s] in the surface form (e.g. *months* [mʌns̲]) depending on the underlying, deleted medial consonant, but Korean speakers differentiate the

[8] Following the conventional criterion, i.e., data points outside the first and third quantiles + 1.5 × interquantile range

coda [s] frication in the surface form depending the underlying medial consonant. It is possible that Korean speakers hyperarticulated the frication when they tried to produce a cluster with a difficult sound, [θ], resulting in a higher intensity. However, it should be noted that the current study had a relatively small number of words and error tokens, so a further study with larger data may be needed to generalize this finding.

Whereas CoG or intensity values show only partial aspects of acoustic properties in the consonant clusters, DTW similarity distance values make it possible to examine the entire consonant clusters, including epenthetic vowels. The DTW similarity distance analysis showed that first, within-language distance (E-E) is smaller than between-language distance (E-KN and E-KE), even for those classified as non-errors. In addition, the E-KN distance was significantly greater when the medial consonant was [θ]. This confirms that Korean speakers will have more difficulty in pronouncing coda obstruent clusters when the cluster contains [θ], which is not a phoneme in Korean.

Second, Korean speakers' non-error tokens were significantly more similar to English speakers' obstruent clusters than error tokens (E-KN vs. E-KE). The distance was significantly much smaller for non-error tokens than for error tokens. Among the error tokens, similarity distance was the greatest when both C_2 and C_3 were deleted where C_2 is [θ]. Here again, we can see that clusters containing a non-native phoneme are more difficult to produce. The least severe error, based on DTW distance, was vowel insertion after C_2. In future research, it would be interesting to look into whether these results correlate with native speakers' judgments.

As Bugdol et al. (2014) suggest, a pronunciation error detection task is considered a classification problem. The KNN classifier is used for classifying errors and non-errors based on DTW distance values. An advantage of using a machine learning algorithm is its ability to make predictions on the new data that the model has not yet seen during the training phase. The KNN classifier classified new data (i.e., test data) with 61% accuracy. The accuracy was not low but not very high, probably due to the relatively small data size, so the model did not have enough chances to learn what errors are. With larger training data, the accuracy can improve. Nevertheless, for now, we can conclude that DTW similarity distance can be a reasonably effective measure for L1 detection and error detection.

The DTW algorithm has been rarely adopted in the field of general linguistics including second language phonetics. In addition, I took one step further by feeding the DTW distance values to the KNN classifier to examine how accurately the distance can make predictions on new data. The importance of applying techniques used in data science in linguistic study has recently been highlighted. Pater (2019) emphasized the use of findings in neural studies in the study of linguistics. Park (2022) reviewed the recent trend in linguistic research using methods in data science. Against this backdrop, it is valuable and timely to adopt techniques from data science, which have been mathematically proven and whose performances are tested in various fields.

5. Conclusion

In this study, coda obstruent clusters produced by Korean learners of English and native speakers of English were compared in terms of traditional acoustic properties, CoG and intensity in the frication noise and in terms of similarity distance obtained by the Dynamic Time Warping algorithm, which has not been attempted in the previous literature. The acoustic properties in the obstruent clusters produced by Korean speakers showed a transfer effect, as expected. The DTW similarity distance between the clusters produced by English speakers and those produced by Korean speakers was greater than the distance between the clusters produced by English speakers only. In addition, the DTW similarity distance between the clusters produced by English speakers and the error tokens by Korean speakers was greater than the distance between the clusters produced by English speakers and the non-error tokens by Korean speakers. The classification result using the KNN classifier showed that the DTW similarity distance was an adequate measure to capture the differences between L1s and between error and non-error productions. As in this paper, exploring new methods in the study of phonetics and phonology will illuminate the patterns and insights that have not been captured by traditional methods only.

References

Altman, Naomi. S. 1992. An introduction to Kernel and Nearest Neighbor nonparametric regression. *The American Statistician* 46(3), 175-185.

Alfeilat, Haneen Arafat Abu, Ahmad B.A. Hassanat, Omar Lasassmeh, Ahmad S. Tarawneh, Mahmoud Bashir Alhasanat, Hamzeh S. Eyal Salman and V. B. Surya Prasath. 2019. Effects of distance measure choice on K-Nearest Neighbor Classifier performance: A review. *Big Data* 7(4), 221-248. https://doi.org/10.1089/big.2018.0175

Balise, Raymond and Randy Diehl. 1994. Some distributional facts about fricatives and a perceptual explanation. *Phonetica* 51, 99-110.

Beckers, Gabriel, J. L. 2002. RMS equalize Praat scripts. http://wwwbio.leidenuniv.nl/~eew/G6/staff/beckers/beckers.html

Boersma, Paul and David Weenink. 1992-2020. Praat (version 6.1.35): Doing phonetics by computer.

Browman, Catherine P. and Louis Goldstein. 1989. Articulatory gestures as phonological units. *Phonology* 6, 201-251.

Bugdol, Marcin, Zuzanna Miodonska and Micha Kręcichwost. 2014. Pronunciation error detection using dynamic time warping algorithm. In E. Pietka, J. Kawa and W. Wieclawek, eds., *Information Technologies in Biomedicine* 4, 345-354. DOI: 10.1007/978-3-319-06596-0_32

Celce-Muercia, Marianne, Donna M. Brinton and Janet M. Goodwin. 2010. *Teaching Pronunciation: A Course Book and Reference Guide*, 2nd ed. Cambridge: Cambridge University Press.

Cho, Mi-Hui. 2005. Cluster reduction by Korean EFL students: insertion vs. deletion strategies. *The Journal of the Korea Contents Association* 6(1), 80-84.

Cho, Mi-Hui and Shinsook Lee. 2005. Repair strategies of English biconsonantal coda clusters: An optimality-theoretic account in conjunction with P-map. *Studies in Phonetics, Phonology and Morphology* 11(2), 191-214.

Cho, Sunghye, Naomi Nevler, Natalia Parjane, Christopher Cieri, Mark Liberman, Murray Grossman and Katheryn A. Q. Cousins. 2021. Automated analysis of digitized letter fluency data. *Frontiers in Psychology* 12, Article 654214.

Clements, George. N. 1990. The role of the sonority cycle in core syllabification. In John Kingston and Mary E. Beckman, eds., *Papers in Laboratory Phonology I: Between the grammar and the physics of speech*, 283-333. Cambridge: Cambridge University Press.

Cover, Thomas M. and Peter E. Hart. 1967. Nearest Neighbor pattern classification. *IEEE Transactions on Information Theory* 13(1), 21-27.

Cunningham, Pádraig and Sarah Jane Delany. 2007. *K-Nearest Neighbour Classifiers*. Technical Report UD-CSI-2007-4. March 27, 2007.

Cunningham, Pádraig and Sarah Jane Delany. 2020. *K-Nearest Neighbor Classifiers*, 2nd ed. (with Python Examples). arXiv:2004.04523 [cs.LG]

Davidson, Lisa. 2006. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34(1), 104-137.

Fix, Evelyn and J. L. Hodges, Jr. 1951. *Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties*. Technical Report 4, USAF School of Aviation Medicine, Randolph Field.

Gimson, Alfred C. 1989. *An Introduction to the Pronunciation of English*, 4th ed. London: Edward Arnold.

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

Gordon, Matthew, Paul Barthmaier and Kathy Sands. 2002. A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32, 141-174.

Gussenhoven, Carlos amd Jacobs Haike. 2011. *Understanding Phonology: Understanding Language*, 3rd ed. London: Hodder Education.

Hong, Hyejin, Sunhee Kim and Minhwa Chung. 2014. A corpus-based analysis of English segments produced by Korean learners. *Journal of Phonetics* 46, 52-67.

Kanters, Sandra, Catia Cucchiarini, Helmer Strik. 2009. The goodness of pronunciation algorithm: A detailed performance study. In *Proceedings of Speech and Language Technology in Education* (SLaTE 2009), 49-52

Kim, Jungsun. 2015. The effect of word frequency on the reduction of English CVCC syllables in spontaneous speech. *Journal of the Korean Society of Speech Sciences* 7(3), 45-53.

Kim. Jungyeon. 2018. Production of English final stops by Korean speakers. *Phonetics Speech Sciences* 10(4),11-17. https://doi.org/10.13064/KSSS.2018.10.4.011

Kreidler, Charles W. 2004. *The Pronunciation of English: A Course Book*, 2nd ed. Malde:Blackwell Publishing.

Kwon, Bo-Young. 2008. The effects of gestural overlap on the acquisition of English word-final obstruent clusters. *English Language and Linguistics* 25, 149-170.

Ladefoged, Peter. 2005. *Vowels and Consonants*, 2nd ed. Malden:Blackwell Publishing.

Lee, Ho-Young. 2000. The pronunciation of English consonant clusters by Koreans. *Malsori* 40, 79-89.

Lee, Shinsook. 2003. A comparison of cluster realizations in first and second language. The *Journal of Studies in Language* 19(2), 341-357.

Lee, Shinsook, Jeongsoon Joh and Mi-Hui Cho. 2002. Acquisition of English consonant clusters among Korean EFL learners. *Korean Journal of Linguistics* 27(3), 439-472.

Lim, Youngshin. 2021. Repair strategies employed by Korean speakers in producing English final consonant clusters. *Journal of Linguistic Studies* 26(1), 205-221.

Muda, Lindasalwa, Mumtaj Begam, and I. Elamvazuth. 2010. Voice recognition algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) techniques. *Journal of Computing* 2(3), 138-143.

Park, Jin-Sook. 2021. Acoustic and spectral characteristics of English voiceless fricatives. *The Journal of Studies in Language* 37(1), 7-19.

Park, Sunwoo. 2022. Towards grammar research based on data science. *Grammar Education* 44, 1-28.

Pater, Joe. 2019. Generative linguisitcs and neural networks at 60: Foundation, frication, and fusion. *Language* 95(1), e41-e74.

Prator, Jr., Clifford Holmes and Robinett Betty Wallace. 2009. *Manual of American English Pronunciation*, 4th ed. Seoul: Heinle Cengage Learning.

Roca, Iggy and Wyn Johnson. 1999. *A Course in Phonology.* Malden: Blackwell Publishers.

R Core Team. 2022. *R: Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

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

Sung, Eunkyung and Yunjeong Cho. 2010. An acoustic study of Korean and English voiceless sibilant fricatives. *Phonetics and Speech Sciences* 2(3), 37-46.

Silbert, Noah and Kenneth de Jong. 2008. Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustical Society of America* 1.2.4.5, 2769-2779.

Shin, Dong-Jin and Paul Iverson. 2014. Phonetic investigation of epenthetic vowels produced by Korean learners of English. *Journal of the Korean Society of Speech Sciences* 6(4), 17-26.

Steinbach, Michael and Tan, Pang-Ning. 2009. kNN: k-Nearest Neighbors. In Xindong Wu and Vipin Kumar, eds., *The Top Ten ALgorithms in Data Mining*, 151-161. Boca Raton: Chapman and Hall/CRC.

Sun, Xihao, Yoshikazu Miyanaga and Baiko Sai. 2014. Dynamic time warping for speech recognition with training part to reduce the computation. *Journal of Signal Processing* 18(2), 89-96,

Sung, Eunkyung and Yunjeong Cho. 2010. An acoustic study of Korean and English voiceless sibilant fricatives. *Phonetics and Speech Sciences* 2(3), 37-46.

Weinberger, Steven. 1994. Functional and phonetic constraints on second language phonology. In M. Yavas, eds., *First and Second Language Phonology*, 283-302. Singular Publishing Group, Inc.

Yavaş, Mehmet. 2020. *Applied English Phonology*, 4th ed. Malden: Wiley-Blackwell.

Zhao, Tongmu, Akemi Hoshino, Masayuki Suzuki, Nobuaki Minematsu and Keikichi Hirose. 2012. Automatic Chinese pronunciation error detection using SVM trained with structural features. *IEEE Spoken Language Technology Workshop* (SLT), 2012, 473-478. doi: 10.1109/SLT.2012.6424270.

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

Appendix

Speech Materials

A Trip to the Veterinarian (Celece-Murcia et al. 2010, p.108; Target words were underlined)

Vet:What seems to be the problem with Peppy?
Pet owner:Well, he just isn't very lively, Doc. He acts so tired all the time. He just lifts his head up and sighs.
Vet: And this started two months ago? Can you give me some more facts?
Pet owner:Sure. One of Peppy's big strengths as a guard dog is his bursts of energy. I asked him to fetch the newspaper yesterday, and he left three-fourths of it on the doorstep. What does your medical textbook say about that?
Vet:Well, let me look it up under "listless dogs." It says here that "four fifths of all listlessness in dogs is due to poor diet." Why don't I give you some pep pills? Feed him one every day, and we'll see how he acts next week.