Novel Features of Vowel Space for Distinction of English L1 and L2 Speech

Hwamin Kim, Jeong-Sik Park, Tae-Yeoub Jang*

Hankuk University of Foreign Studies (HUFS)
ghkals504@hufs.ac.kr, parkjs@hufs.ac.kr, tae@hufs.ac.kr

ABSTRACT

This study introduces novel metrics (VCI and VSE) that measure the centrality of L2 speakers' vowel space. The features are intended to capture the tendency of L2 speakers' passive articulation and its consequence of shrunk vowel space. A set of English read speech data, produced by native and Korean EFL speakers, was used to test the validity of the features. For a statistical analysis, the Mixed Effects Models were fit with the language groups, including three EFL proficiency levels and a native group, as the fixed effect variable. The result indicated that the features facilitated a significant differentiation between L1 and L2 speech while the distinction among proficiency levels was not obvious. It is suggested that the proposed metrics are practically implemented as features for automated assessment or classification of non-native speech.

Keywords: Formant frequency, vowel space, L1-L2 classification, automated evaluation

1. INTRODUCTION

It is widely admitted that the formant frequency of a vowel, among other features, is the critical information in vowel recognition. Especially, the first formant (F1) and the second formant (F2) are known to cover the most relevant cues to vowel distinction, which are height and backness of the articulator.

Due to various differences in L1 and L2 speech, learners have difficulty producing correct segments of the target language they learn. In L2 speakers' vowel production, there are two distinct characteristics that cause incorrect pronunciations: first, interference of the speaker's L1 vowel system and, second, passive gesture of articulators. The former is hard to measure and predict considering diverse language specific vowel inventories while the latter is a general property of L2 speech. Thus, attempts have been made to demonstrate the characteristics of L2 speech vowel space including the case of Korean EFL (KE) speech. [1] found that English vowels produced by KE speakers are more variable than those by natives implying that the vowel space of L2 has become flexible as compared to L1. [2] demonstrated that the vertical space of KE speech was considerably shrunk as compared to native English speakers (NE). It was also found in [3] that KE pronunciation has lax vowels less centralized than NE speech due probably to a tendency not to reduce vowels in unstressed syllables. Although these studies are in agreement when demonstrating the KE vowel characteristics, a concrete numerical measure that can be applied to automated L1-L2 distinction or L2 assessment has not been developed. Another restriction of these studies is that their conclusion was based on experiments with a relatively small set of controlled data, with less than two hundred tokens.

There was an effort as in [4] to quantify vowel characteristics in Hebrew and Arabic using the Euclidean distance measure. Also, for American English, the metrics for vowel space area (VSA) were used in detecting the degree of vowel centralization of the impaired speech. These metrics were based on the polygon shape using three or four corner vowels such as /iy/, /uw/, /ae/ and /aa/ [5]. However, their intention was to conduct L1 vowel comparison focusing on quantifying the degree of vowel reduction of unstressed syllables in continuous speech.

Filling the gap of the previous research, the current study aims to introduce new metrics (VCI and VSE) that can be used for enabling L1-L2 vowel space differentiation. The proposed metrics of vowel space centrality are intended to be directly implemented to an automated system of L2 pronunciation evaluation or L1-L2 distinction. In order to achieve this goal the metrics will be tested with a relatively large amount of L1-L2 parallel data tokens.

2. PROCEDURE

2.1. Data

For the analysis, we used part of a L1-L2 parallel data set constructed in previous research [6]. The current data comprise a set of 108 sentences read by each of 8 native English speakers and 18 Korean EFL learners which are classified into three proficiency

* Corresponding author

groups: Advanced, High Intermediate and Intermediate, based on such criteria as self-assessed confidence, the length of learning English and the length of experience in English speaking countries. The ratio of males and females are balanced in each group, except Advanced that includes 1 male and 5 female speakers. The format of sentences varies including declarative, interrogative, exclamation, imperative and tagged questions. The data are provided with CMUbet phone labels whose boundaries were demarcated through the GMM-based processing of forced alignment. Despite some obvious auto-segmentation errors, no manual adjustment was attempted so that the validity of proposed features could be tested for automated evaluation.

2.2. L1 and L2 Formants

As the new metrics are defined in terms of the first two formants of each vowel, extraction of F1 and F2 is an essential process. From each phone label, two formant frequencies were extracted using functions available in Praat [7], designating a gender specific limit of maximum frequency such as 5000 Hz for male speech and 5500 for female speech. From each vowel interval, only 40% of the total duration located in the middle was used for calculating average F1 and F2 values. This is an effort to exclude transition periods on both sides of vowel onset and offset. The most and least frequent vowel was /ax/ (3734) and /uh/ (520), respectively.

Table 1 and Figure 1 demonstrate a rough comparison between NE and KE formants. Regarding F1, there is a general tendency that KE has a lower F1 for non-high vowels, reflecting that the jaw does not move as low as NE, which can also be interpreted as a passive articulatory gesture. Meanwhile, F2 seems to vary depending upon interaction between factors of nativity and gender. To sum, it can be inferred that the vowel space of speakers can be differentiated or predicted based on speaker's nativity.

Table 1: Compared distribution formants. '+' means the formant value of NE is greater than KE; '-' means the opposite. The blank means no significant difference between two samples. Significance scale: * <.05, **<.01, ***<.001.

F1	MALE		FEMALE	
vowel	NE-KE	Sig.	NE-KE	Sig.
iy	-	**	-	***
ih	+	***	0	
ey	0		+	*
eh	+	***	+	***
ae	+	***	+	***
ax	0		0	
aa	+	***	+	***
ao	+	***	+	***
ow	+	***	0	
uh	0		0	
uw	0	0	0	

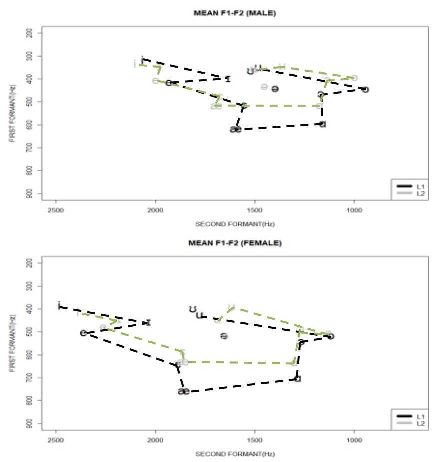
F2	MALE		FEMALE	
vowel	NE-KE	Sig	NE-KE	Sig
iy	-	*	+	***
ih	-	***	+	***
ey	-	**	+	***
eh	-	***	0	
ae	-	***	0	
ax	-	***	0	
aa	0		0	
ao	0		0	
ow	-	***	0	
uh	+	***	+	***
uw	0		+	**

3. VOWEL SPACE MEASURES

Based on the previously described observations on F1 and F2, new features for NE-KE distinction were contrived. First, Vowel Centrality Index (VCI) captures differences between NE and KE in that NE space has more flexibility of lax vowels as they are frequently reduced to a schwa-like centralized position. Second, Vowel Space Eccentricity (VSE) differentiates NE and KE through overall shape of the vowel space.

To compute shape or position based features, reference points should be designated. The three vowels /iy/, /ow/ and /aa/ were found to be positioned at the edge of the vowel space and, consequently, had the maximum and minimum values for F1 and F2, respectively. Thus, these vowels were regarded as reference in the current study.

Figure 1: The mean values of each vowel in inversed F1-F2 plane.



3.1. Vowel Centrality Index

Only F2 values are relevant for this measure. Above all, the two vowels /iy/ and /ow/ of L1 are employed as anchor points. VCI of an individual vowel can be computed as shown in (2). In the case of NE-KE comparison, VCI of the two lax vowels /ih/ and /uh/ are calculated as they are the most radically reduced vowels in NE speech while no such reduction occurs in KE speech. The formula (2) represents the procedure of calculating VCI of a vowel.

$$(1)\ MAX_{f2} = \mu_{f2(iy)}, MIN_{f2} = \mu_{f2(ow)}$$

$$(2)\ VCI(v) = \frac{\mu_{f2(v)} - MIN_{f2}}{MAX_{f2} - MIN_{f2}}$$

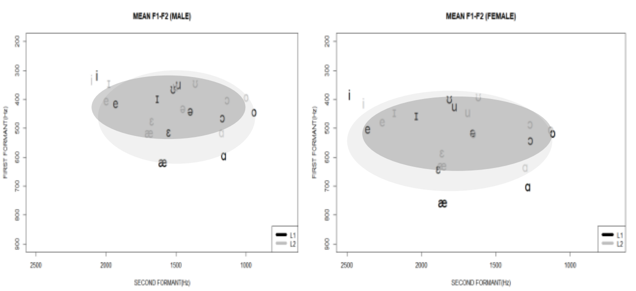
where $\mu_{f2(v)}$ *denotes the* $f2$ *mean values of* v

When a vowel is fully reduced and positioned at the center of the vowel space, its VCI will be 0.5. The VCI approaches closer to 1 when it gets near the vowel /iy/, and 0 when near /ow/. For example, for the /ih/ vowel, its VCI value of KE would be higher than that of natives because of the less centralized tendency of L2. Conversely, for the /uh/ vowel, its values would be lower than that of natives.

3.2. Vowel Space Eccentricity

As for F1, the maximum and minimum values are on /aa/ and /ih/ among the native vowels. On the contrary, as shown in Figure 1, KE vowel space has tilted up in general. It can be described that if NE vowel shape were to be compared in the form of a circle, KE would be more like an ellipse as simulated in Figure 2.

Figure 2: From the /iy/-/ow/ in horizontal distance and /iy/-/aa/ in vertical distance, comprehensive oval shapes are driven in vowel space.



Based on this point, we formed the feature VSE. To compute the eccentricity of an oval, both vertical and horizontal anchor points are necessary. Again the edge vowels /iy/ and /ow/ were used for reference F2 values, while the lowest vowel /aa/ was employed to get a reference point of F1. The formula to obtain VSE based on this procedure can be described as (3).

$$(3)\ E = \sqrt{1 - \frac{b^2}{a^2}},\ VSE = \sqrt{1 - \frac{\left(\mu_{f1(aa)} - \mu_{f1(iy)}\right)^2}{\left(\mu_{f2(iy)} - \mu_{f2(ow)}\right)^2}}$$

where $\mu_{f1(v)}$ *denotes the* $f1$ *mean values of* v *and* $\mu_{f2(v)}$

It can be hypothesized that the VSE values in Korean learners' speech are lower than that of English native speakers.

3.3. Validity Check

To test validity of these two measures, VCI and VSE, a Mixed Effect Model was fit with these metrics as response variables and speaker information as a random effect variable while gender and language groups (L1 and L2) and proficiency groups (NE, AD, IH and IM) were set as fixed effects.

4. RESULTS AND DISCUSSION

The summarized results of statistical significance are given in Table 2.

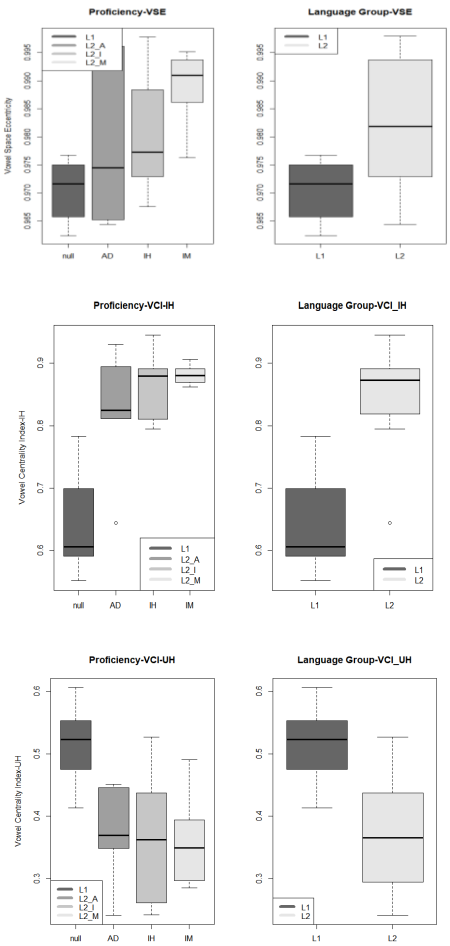
Table 2: The result of LME. Specific features of VCI-EY and VCI-UW are excluded as they are not designated boundary measures.

LME	Gender	Language	Proficiency
VSE	-	*	**
VCI-IH	-	***	***
VCI-UH	-	***	***

Four vowels were tested to check whether VCI successfully discriminates language groups. Above all, there was no gender effect. Language and proficiency effects were confirmed when vowels /ih/ and /uh/ were given, meaning that KE speech did not reduce the lax vowels as much as native speech. VSE was also effective in discrimination of Language and Proficiency groups. Subsequent Post Hoc tests further reveal that groups NE and IM were the only proficiency pair significantly differentiated from each other by VSE. On the other hand, the two VCI features VCI-IH and VCI-UH were able to differentiate NE and each of sub-groups of KE (i.e., AD, IH and IM, standing for advanced, intermediate-high, and intermediate, respectively). Based on these results, it can be stated that the two features VSE and VCI, in concert, successfully demonstrate the fact that native speakers' vowel space is different from vowels space of learners of all proficiency levels.

Results are also visualized through boxplots in Figure 3.

Figure 3: The boxplots for VSE and VCI in discriminating pre-defined proficiency groups and language groups.



5. CONCLUSION

It is confirmed that both metrics of VCI and VSE are useful for distinguishing between speech of native speakers of English (NE) and Korean learners of English (KE). A practical advantage of these features is that as they are designed to be computed automatically, they can be directly implemented into automated systems of language classification and/or autoscoring of L2 proficiency levels.

There are limitations of the current approach. First, all the measures are based on vowels demarcated by auto-segmentation processing. It is expected that increased boundary detection will enhance the validity of the proposed features. Second, as shown in Table 1, formant magnitudes appear to vary depending upon the factor of gender, further investigation seems necessary with more gender balanced data. Third, the current analysis is based upon pre-defined proficiency classifications. Obviously, more systematic level definition, if available, will affect the results of the proposed features.

Presumably, the characteristics of reduced vowel space in non-native speech is a language independent phenomenon. Thus, we assume that the proposed features are widely useful for any L1-L2 situation. Extended experiments with diverse non-native language data will confirm this.

6. ACKNOWLEDGEMENT

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A1A09000903), and the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2019-2016-0-00313) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation).

7. REFERENCES

[1] Hong, H., Kim, S., Chung, M. 2014. A corpus-based analysis of English segments produced by Korean learners. *Journal of Phonetics*. 46, 52-67.

[2] Jeong, S., Cho, M. 2016. An acoustic analysis of Korean EFL students' English vowel production: Focused on 'new' English vowel categories. *Korea Journal of English Language and Linguistics.* 16(2), 405-424.

[3] Lee, S., Nam, H., Kang, J., Shin, D., Kim, Y. 2017. The impact of language-learning environments on Korean learners' English vowel production. *Phonetics and Speech Sciences*. 9(2), 69-76.

[4] Amir, N., Amir, O. 2007. Novel measures for vowel reduction. *Proc. 16th ICPhS* Saarbrucken, 849-852.

[5] Fletcher, A., McAuliffe, M., Lansford, K., Liss, J. 2017. Assessing vowel centralization in dysarthria: A comparison of methods. *Journal of Speech, Language, and Hearing Research*. 60, 341-354.

[6] Chung, H., Jang, T., Yun, W., Yun, I., Sa, J. 2008. A study on automatic measurement of pronunciation accuracy of English speech produced by Korean learners of English. *Language and Linguistics*. 42, 165-196.

[7] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved on 8 September 2018 from http://www.praat.org/.