Linguistic Research 41(1): 1-26
DOI: 10.17250/khisli.41.1.202403.001

Some acoustic and articulatory characteristics of the three diphthongal vowels with /w/ in speech of young Seoul Korean speakers*

Dae-yong Lee[a] · Jiyeon Song[a] · Sahyang Kim[b] · Taehong Cho[a]*****
(Hanyang University[a] · Hongik University[b])

Lee, Dae-yong, Jiyeon Song, Sahyang Kim, and Taehong Cho. 2024. Some acoustic and articulatory characteristics of the three diphthongal vowels with /w/ in speech of young Seoul Korean speakers. *Linguistic Research* 41(1): 1-26. This study examines acoustic-articulatory properties of three Korean diphthongal vowels of /we/* '외', /wɛ/ '왜', and /we/ '웨' in monosyllabic contexts with and without the onset /h/, produced by young Seoul female and male speakers in their twenties. (/we/* refers to the diphthongized vowel of the previous monophthong /ø/.) Results indicate that the three diphthongs do not significantly differ in various phonetic measures, suggesting that they are largely merged. However, there is some evidence that young speakers still maintain a subtle phonetic difference between the 'new' diphthong (/we/*) and one of the 'old' diphthongs (/wɛ, we/). The distinctions are further constrained by the onset context, complicating the results. For example, a difference between /we/* and /wɛ/ is observed in articulatory Movement duration only with /h/. Additionally, there is an asymmetry between acoustic and articulatory measures such that a more pronounced distinction between /we/* and one of the 'old' diphthongs is observed in the articulatory dimensions, such as Movement duration, Displacement, and Tongue position. Finally, there is no evidence of a gender-related difference that suggests female speakers are leading the merger, while some distinctions between the 'new' and the 'old' diphthongs are made only by the female speakers, possibly due to their propensity to make distinctions reflected in the orthographic representations. These results are further discussed in their implications for the importance of multi-dimensional phonetic studies and the role of gender in understanding the nuanced extent of merger of the diphthongs. **(Hanyang University · Hongik University)**

Keywords Korean diphthongal vowels, diphthong merger, sound change, kinematic measures

---

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5C2A02086884). We thank the two anonymous reviewers for their comments from which our paper has greatly benefited.

** First author
*** Corresponding author

1. Introduction

Korean is commonly believed to have various diphthongs, such as /je, jɛ, ja, jo, ju, jʌ, wi, we, wɛ, wa, wʌ, and ɰi/ (e.g., Lee 1993).[1] However, research focusing on Korean diphthongs with a /w/ suggests a plausible ongoing merger in the phonetic realization of diphthongs /wɛ/ '왜' and /we/ '웨' in Seoul Korean. For example, Yang (1993) examined acoustic characteristics (e.g., fundamental frequency, formant frequencies, duration) of Korean diphthongs including /wɛ/ '왜' and /we/ '웨' produced by speakers of Seoul Korean. The results of the study presented distinct formant frequencies and duration disparities between the diphthongs /wɛ/ '왜' and /we/ '웨', thereby indicating a distinction of these diphthongs in Seoul Korean during the early 1990s. More recent studies, however, showed that young speakers of Seoul Korean do not exhibit acoustic differentiation between these particular diphthongs (e.g., Shon 2022; Zhao et al. 2023), in line with the recent merger of the non-high front vowels /ɛ/ 'ㅐ' and /e/ 'ㅔ' (e.g., Lee and Cho 2021). For example, Shon (2022) analyzed the formant frequencies of the diphthongs /wɛ/ '왜' and /we/ '웨' produced by 20 speakers of Seoul Korean in their twenties and showed that these speakers do not distinguish between these diphthongs in terms of formant frequencies. Additionally, the Korean vowel /ø/ '외', which was traditionally categorized as a monophthong (e.g., Lee 1993), has undergone diphthongization, suggesting a merger with the diphthongs /wɛ/ '왜' and /we/ '웨' in Seoul Korean (e.g., Cho 2003; Chang 2017; Shon 2022). (In the present study, to maintain distinctions as reflected by the Korean orthographic system, we will use three distinct notations: /wɛ/ '왜', /we/ '웨', and /we/* '외'.)[2]

There is therefore a body of literature, based on acoustic studies (e.g., Cho 2003; Chang 2017; Shon 2022; Zhao et al. 2023), suggesting a merger of the diphthongs /wɛ/, /we/, and /we/* in Seoul Korean. However, the precise extent of this merger among the diphthongs remains unclear. Notably, this merger of the three-way contrast involves two distinct processes: the convergence of /ɛ/ 'ㅐ' and /e/ 'ㅔ', and the diphthongization of

1 The debate remains whether each of these sounds can be phonologically classified as a genuine diphthong or a combination of a glide and a vowel (e.g., Cheon 2002; Kang 2003). While this issue is outside the scope of the current study, we refer to these sounds as diphthongs, indicating that they may have two phonetic components, regardless of whether one of them is a glide phonetically or phonologically.

2 Considering the recent sound changes related to /wɛ/ '왜', /we/ '웨', and /ø/ '외', using distinct IPA phonetic symbols for the underlying representations of these sounds could be misleading. This would become especially true if the sound change had fully taken effect in Seoul Korean, wherein a single IPA phonetic symbol, like /we/, might otherwise be used for all three sounds.

/ø/ '외'. The continued use of distinct orthographic representations ('왜', '웨', and '외') in the Korean writing system further complicates matters. Given these complexities, a key question arises as to whether this merging process has completed among younger speakers, or it is still evolving within the Seoul Korean dialect. In other words, considering the distinct orthographic representations learned and the potential phonetic remnants of the three-way contrast that may still exist among older generations in the Seoul speech community, as can be inferred from Lee (1993), one might hypothesize that young Seoul Korean speakers might produce these diphthongs with nuanced phonetic details that reflect these aspects. As introduced above, however, previous studies on the merger in question have already indicated that these diphthongs are produced with no clear distinction. But as these studies have been limited to acoustic domains, it remains undetermined whether any phonetic detail that may underlie the otherwise neutralized three-way contrast among diphthongs is evident in the fine-grained articulatory characteristics.

Some previous studies indeed suggest that articulatory characteristics of speech sounds do not necessarily have a direct mapping to their acoustic characteristics (e.g., De Decker and Nycz 2012; Wieling et al. 2016). For example, Wieling et al. (2016) demonstrated that structural dissimilarities between two dialects did not manifest in formant-based acoustic measurements. Instead, these differences were evident through articulatory measurements. This finding is in line with the non-linear relationship between acoustic and articulatory dimensions as proposed in, for example, the Quantal Theory (Stevens 1989) and various other theoretical approaches. (see Iskarous (2011) and references therein for a review on the non-linearity between the acoustic and articulatory dimensions.) In the context of Korean diphthongs (/we/*, /wɛ/, /we/), this suggests that relying solely on acoustic formant measurements might overlook the phonetic subtleties speakers use to differentiate these sounds. Articulatory efforts to distinguish these diphthongs might not always be evident acoustically. If this is the case, an exclusive focus on acoustics could misrepresent the phonetic characteristics of the merger of these diphthongs. Thus, to address this gap and to assess the exact phonetic nature of the merger, the present study investigates both the acoustic and articulatory attributes of the Korean diphthongs /we/*, /wɛ/, and /we/. In particular, it examines kinematic parameters such as movement velocity and displacement associated with the lip and tongue movements corresponding to the glide and vowel counterparts, which illuminate the dynamic kinematic nature of the diphthongs which have not been previously examined.

In the present study using an Electromagnetic Articulography system, we explore this possibility by examining the phonetic realization of these diphthongs in both acoustic and articulatory dimensions, produced by young Seoul Korean speakers. The study concentrates on the diphthongs when pronounced in isolation, to ensure that they are produced in a clear speech context as the potential differences among the three diphthongs, if present, are likely to be more pronounced in a context of hyperarticulation.

As we investigate the acoustic-articulatory characteristics of the diphthongs produced by young Seoul Korean speakers, the data will also allow us to consider the potential influence of speaker gender on the suggested merger. Some Korean sound changes have been known to be influenced by speaker gender (e.g., Kang 2014; Choi et al. 2020). For instance, differences in Voice Onset Time (VOT) between lenis and aspirated stops in Korean are less pronounced in female speakers compared to their male counterparts (Kang 2014), indicating that the ongoing VOT merger in Korean stop production is primarily driven by female speakers. Given this tendency of sound changes being led by female speakers (cf. Labov 1973), the merger of the diphthongs under investigation may also be led by female speakers. Alternatively, however, given that female speakers generally demonstrate higher intelligibility than male speakers, particularly when adopting a distinct mode of speech known as clear speech (i.e., a form of hyperarticulation; e.g., Ferguson and Morgan 2018), female speakers of the present study might demonstrate a clearer differentiation between these diphthongs compared to male speakers in line with the orthographic representations.

Finally, it is worth noting that the present study also included the examination of /hwe/* '회', /hwɛ/ '홰', and /hwe/ '훼' in our analyses. The glottal fricative /h/ is not specified with an oral place of articulation and does not have a known coarticulatory effect on the vowel. Therefore, these additional syllables provide further data for examining the phonetic nature of the diphthongs under investigation, thus increasing the generalizability of any findings in the present study. However, because /h/, as a voiceless fricative, influences the voicing characteristics at the beginning of the target syllables, it may also affect the formant structures of the diphthongs. Therefore, the presence or absence of /h/ is another factor examined in the present study.

Overall, the present study examines the acoustic and articulatory characteristics of Korean diphthongs /we*, wɛ, we/ as well as /hwe*, hwɛ, hwe/, and explores a possible effect of speaker gender on the production of these diphthongs. The aim is to provide a more detailed phonetic insight into the merger of these diphthongs, assessing whether this sound change has indeed completed without any phonetic remnants of orthographic and historical traces among young speakers in Seoul, as suggested by previous acoustic studies; or whether such phonetic remnants still exist, albeit subtly, when analyzed in both acoustic and articulatory dimensions within a clear speech (hyperarticulation) context.

2. Methods

2.1 Materials

Materials for the present study came from a Korean acoustic and articulatory database, currently under construction at Hanyang Institute for Phonetics and Cognitive Sciences of Language (HIPCS 2022). The database included acoustic recordings of Korean paragraphs and syllables read by native Korean speakers from a variety of dialects and articulatory information that was collected together with the acoustic recordings. Specifically, the database recorded acoustic and articulatory information of seven short paragraphs, 25 sets of monosyllables, and four sets of disyllables for each participant. From the sets consisting of monophthongs and diphthongs, two repetitions of Korean diphthongs /we/* '외', /wɛ/ '왜', /we/ '웨', /hwe/* '회', /hwɛ/ '홰', and /hwe/ '훼' read by 16 Seoul and Gyeonggi Korean speakers (8 female, 8 male) between 19 and 29 years old were analyzed for the current study (i.e., 3 diphthongs × 2 onsets × 2 repetitions × 16 speakers = 192 tokens in total). The diphthongs examined in the present study were all monosyllabic words produced in isolation (i.e., the diphthongs were not produced in carrier sentences).

Acoustic recordings of the materials were conducted in a sound-attenuated room in Hanyang Institute for Phonetics and Cognitive Sciences of Language at Hanyang University using a Tascam HC-P2 digital recorder and a SHURE KSN44 condenser microphone. Simultaneously, articulatory data were collected with an Electromagnetic Articulography system (EMA, Carstens Articulograph AG501). Eight sensor coils were used to track participants' articulatory movements. Particularly, six sensor coils were attached to the tongue dorsum, tongue body, tongue tip, lower incisor, and lower and upper lips around the vermilion border to track the movements of articulators; and two sensor coils that served as reference points were attached to the nasion and upper incisor.

The locations of the sensor coils are shown in Figure 1.

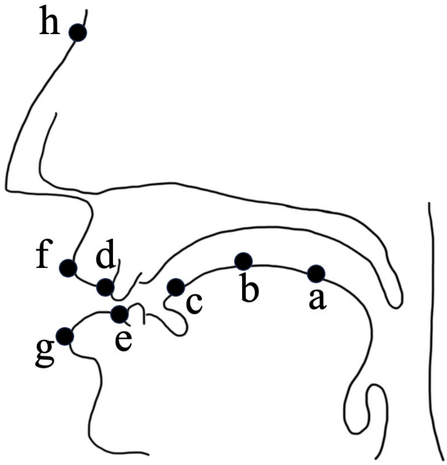

Figure 1. Locations of sensor coils: (a) tongue dorsum, (b) tongue body, (c) tongue tip, (d) upper incisor, (e) lower incisor, (f) upper lip, (g) lower lip, and (h) nasion

2.2 Measurements and data analysis

Formant frequencies (F1, F2, and F3) at 25%, 50%, and 75% points of the acoustic vowel duration and the Duration of the vowels were measured using a Praat script (Barreda 2021) to examine whether Korean diphthongs /we/*, /wɛ/, /we/, /hwe/*, /hwɛ/, and /hwe/ demonstrated acoustically different patterns. The determination of diphthong onsets differed based on the presence or absence of the /h/ onset. In diphthongs without an /h/ onset, we established onset points by visually inspecting the amplitude and formant frequencies on the spectrogram and amplitude on the waveform, identifying a juncture where the precise measurement of formant frequencies became reliable. For diphthongs with an /h/ onset, we identified the onset as the commencement of the glottal fricative /h/. We made this assumption based on the understanding that /h/ is realized as a glottal fricative independently, superimposed on the oral vowel production, and that formants remain reliably observable during the presence of /h/ (out of the 96 diphthongs with an /h/ onset, 11 tokens included measurements of which the 25% point of the vowel duration was within the /h/ portion). The determination of the offset was the same as diphthongs without an /h/ onset. That is, by visually inspecting the spectrogram and waveform, we identified a point where reliable measurements of formant frequencies became difficult. Formant frequencies were converted to z-scores for statistical analyses in order to account for the difference between female and male speakers. (i.e., z-scores were calculated across F1, F2, and F3 per speaker and were calculated by subtracting the mean formant frequency from each target formant frequency and then dividing the value by the standard deviation of the formant frequencies.)

Along with the acoustic analyses, the current study included articulatory analyses of the diphthongs. The articulatory analyses included tongue dorsum movements (movement in terms of tongue dorsum frontness and height), lip aperture (distance between the upper and lower lips), and the lag between tongue dorsum and lip articulations. The present study included analyses of tongue dorsum movements and lip aperture as the tongue dorsum movements and lip aperture are closely related to the articulation of the target diphthongs (/we/*, /wɛ/, /we/, /hwe/*, /hwɛ/, /hwe/) (Gick 2003). Moreover, given that the production of the diphthongs entails a transition in the articulation involving the lips and tongue dorsum, the current study analyzed the temporal discrepancy (i.e., lag) between the articulatory gestures of the tongue dorsum and those of the lips. (Note that we recognize the importance of examining the horizontal movement of the lips for understanding lip production. However, due to challenges in consistently measuring lip protrusion within our dataset, we did not include lip protrusion analysis in the present study. Including this measure in future studies will enhance the completeness of our kinematic analyses.)

Figure 2 shows a schematized tongue dorsum vertical movement trajectory of the diphthongs examined in the present study. Specifically, the upper panel in Figure 2 shows the vertical movement direction of the tongue dorsum during the lip opening movement. The solid black line in the upper panel corresponds to the height of the tongue dorsum throughout the diphthong production. For example, the ascending solid black line indicates elevation of the tongue dorsum within the oral cavity. The solid black line in the lower panel shows the velocity of the tongue dorsum movement. That is, the lower panel shows that velocity increases as it reaches peak velocity (i.e., highest point in terms of velocity) and decreases after reaching peak velocity. Below, we describe the kinematic measurements of the tongue dorsum movement analyzed in the present study:

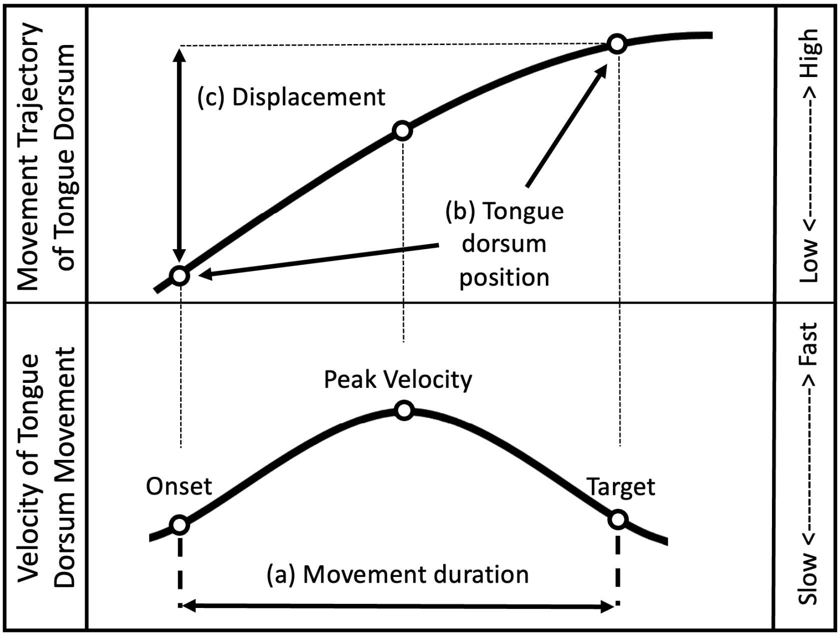

Figure 2. Schematized tongue dorsum movement trajectory measures (upper panel) and tongue dorsum movement velocity (lower panel)

a. Movement duration: the duration between the onset and target of the tongue dorsum movement
b. Tongue dorsum position (onset and target): position of the tongue dorsum (i.e., sensor coil attached to the tongue dorsum) at the onset and target of the tongue dorsum movement trajectory
c. Displacement: the spatial distance in height between the tongue dorsum at the onset and target of the tongue dorsum movement

Kinematic landmarks including onset, peak velocity, and target were determined by assessing the velocity of the tongue dorsum movement by using MVIEW (Tiede 2005). Specifically, peak velocity corresponds to the point in time when the tongue dorsum movement achieves its maximum speed, as illustrated in Figure 2. Further, we established the onset and target of the tongue dorsum movement with reference to this peak velocity. That is, the onset and target were identified as the time points at which the velocity of the tongue dorsum movement reached 20% of the peak velocity. The kinematic measures for assessing the horizontal movement of the tongue dorsum were computed in the same manner. As partially illustrated in Figure 2, the general movement of the tongue dorsum during lip opening showed an upward and forward movement. That is, during vowel production, speakers raised and advanced their tongue dorsum within the oral cavity as they opened it.

Figure 3 shows a schematized lip aperture movement trajectory of the target diphthongs. The upper panel shows the movement trajectory of the lip aperture. In the upper panel, the upward movement of the solid black line corresponds to the reduction in lip aperture (i.e., closing movement of the upper and lower lips) while the downward movement of the line indicates the widening of the lip aperture (i.e., opening movement of the upper and lower lips). The lower panel shows the velocity of the lip aperture movement. Specifically, the ascending trajectory of the solid black line in the lower panel corresponds to an escalation in the velocity of lip aperture movement, while the descending trajectory of the line indicates a deceleration in velocity. We describe the kinematic measurements of the lip aperture included in the present study below:

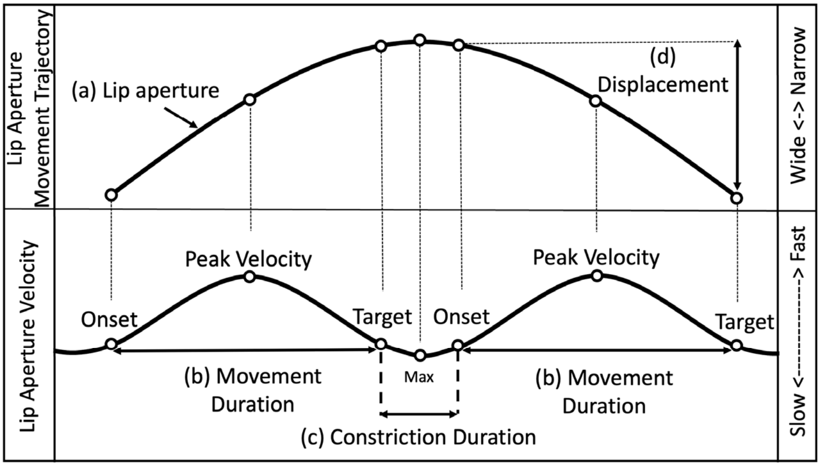

Figure 3. Schematized lip aperture movement trajectory measures (upper panel) and lip aperture movement velocity (lower panel)

a. Lip aperture: the distance between the upper and lower lips represented by the black solid line in the upper panel of Figure 3.

b. Movement duration: the duration between the onset and target of lip movements
c. Constriction duration: the duration of the lip closure (i.e., maximum constriction between the upper and lower lips)
d. Displacement: the difference in lip aperture at the onset and target of lip movements

Lastly, the Lag between tongue dorsum and lip articulations is the duration between the onset of the lip aperture closing movement (i.e., where the lip aperture becomes narrower) and the onset of the tongue dorsum movement. The spatial measures (Position of tongue dorsum, Displacement of tongue dorsum, Lip aperture, and Displacement of lip aperture) were converted to z-scores to mitigate potential variations in sensor locations among speakers. On the other hand, as the possible variation in sensor locations does not have direct effect on the temporal measures (Movement durations of tongue dorsum, Movement and Constriction durations of lip aperture), the temporal measures were reported as raw data.

Results were analyzed with a series of linear mixed-effects regression models using the lme4 package (Bates et al. 2015) within the R computing program (R Core Team 2018). Each of the acoustic and kinematic measures was the dependent variable and fixed effects included in the model were Vowel (/we/*, /wɛ/, or /we/), Onset (absence vs presence of /h/), and Gender (female vs male). The fixed effects were contrast coded with simple coding and the reference levels for Vowel, Onset, and Gender were /we/*, absence of /h/, and female, respectively. Further, each model included two-way and three-way interactions between each of the fixed effects. Random effect structures included the maximal structure to the extent that the model allowed convergence. Additionally, for all linear mixed-effects regression models, post-hoc Tukey tests were conducted using the *emmeans* package (Lenth 2021) within the R computing program when there were significant interactions between the fixed effects. As the main models made a comparison between /we/*-/wɛ/ and /we/*-/we/, but not a comparison between /wɛ/ and /we/, we created separated models to compare /wɛ/ and /we/. Specifically, we made a subset of the data excluding /we/* and created a separate linear mixed-effects regression model in the same way as the main model. All linear mixed-effects regression models and the summary tables of the models are available on the Open Science Foundation (OSF) storage (https://osf.io/urm3c/?view_only=38474eec5c874180b2e871689ec35274).

3. Results

3.1 Acoustic analyses

A series of linear mixed-effects regression models showed that there was a main effect of Vowel on F2 at the 50% and 75% points of the vowel duration and F3 at the 75% point of the vowel duration. Regarding the main effect of Vowel on F2 at the 50% point of the vowel duration, a linear mixed-effects regression model showed that F2 at the 50% point of the vowel duration was lower for /we/* (Mean = 0.25, SD = 0.33) than /wɛ/ (Mean = 0.31, SD = 0.29) (β = 0.06, SE = 0.03, p = 0.02) and /we/ (Mean = 0.32, SD = 0.29) (β = 0.07, SE = 0.03, p = 0.01), respectively, as shown in the left panel of Figure 4. Moreover, there was no significant interaction between Vowel and Gender, indicating that both female and male speakers demonstrated similar patterns, as shown in the right panel of Figure 4. Regarding the difference between /wɛ/ and /we/, a separate linear mixed-effects regression model comparing the two diphthongs showed that there was no significant difference (β = 0.007, SE = 0.03, p = 0.79).

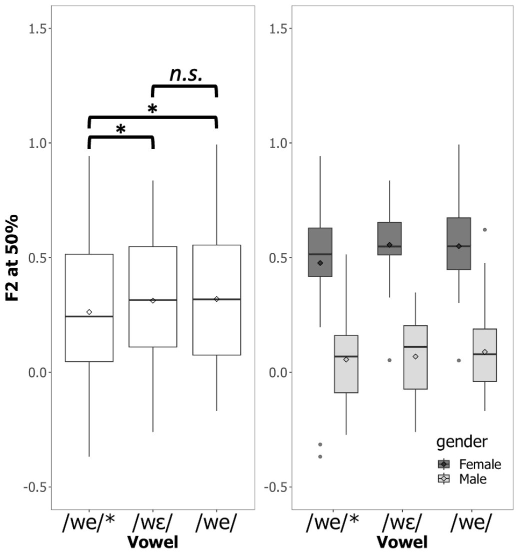

Figure 4. Box plot showing normalized F2 at the 50% point of vowel duration as a function of Vowel (left panel) and as a function of Vowel and Gender (right panel). The diamonds on the boxes represent the mean F2 at the 50% point of the vowel duration for each condition.

In terms of the main effect of Vowel on F2 at the 75% point of the vowel duration, a linear-mixed effects regression model showed that F2 at the 75% point of the vowel duration was lower for /we/* (Mean = 0.36, SD = 0.34) than /wɛ/ (Mean = 0.43, SD = 0.30) (β = 0.06, SE = 0.03, p = 0.04), as shown in the left panel of Figure 5. Moreover, there was a significant three-way interaction between Vowel, Onset, and Gender on F2 at the 75% point of the vowel duration (β = -0.27, SE = 0.12, p = 0.03). Post-hoc pair-wise comparisons revealed that the interaction resulted from the difference in F2 at the 75% point of vowel duration between /hwe/* (Mean = 0.45, SD = 0.42) and /hwɛ/ (Mean = 0.68, SD = 0.11) for female speakers (β = -0.23, SE = 0.06, p = 0.01), as illustrated in the right panel of Figure 5. That is, female speakers' F2 at the 75% point of the vowel duration was higher for /hwɛ/ than /hwe/* suggesting a possibility of a more anterior tongue position for /hwɛ/ than /hwe/* for the female speakers in the present study. Such distinction in the /h/ onset context was not made by male speakers. Further, A separate linear mixed-effects regression model comparing /wɛ/ and /we/ showed no significant difference between the two diphthongs (β = -0.01, SE = 0.03, p = 0.66).

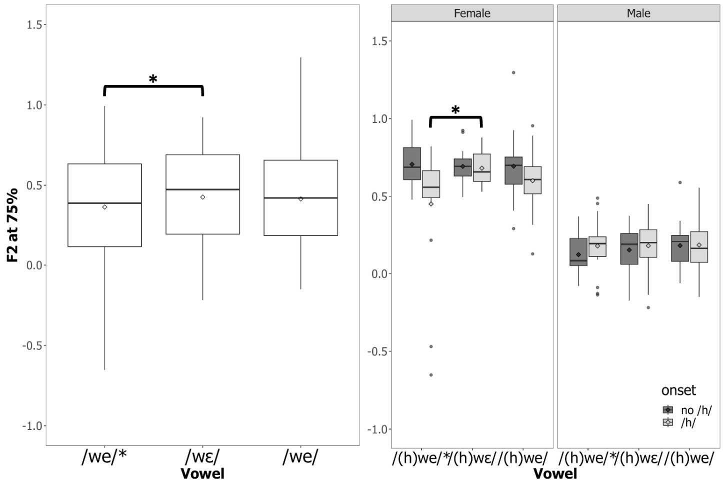

Figure 5. Box plot showing normalized F2 at the 75% point of vowel durations as a function of Vowel (left panel) and as a function of Vowel, Onset, and Gender (right panel). The diamonds on the boxes represent the mean F2 at the 75% point of the vowel duration for each condition.

Regarding the main effect of Vowel on F3 at the 75% point of the vowel duration, a linear mixed-effects regression model showed that F3 at the 75% point of the vowel duration was lower for /we/* (Mean = 1.06, SD = 0.34) than /wɛ/ (Mean = 1.11, SD = 0.33) (β = 0.05, SE = 0.02, p = 0.04), as shown in the left panel of Figure 6. As shown in the right panel of Figure 6, this pattern was consistent for both female and male speakers. All other main effects and their interactions did not show significant differences of formant frequencies between the diphthongs /we/* and /wɛ/ and between /we/* and /we/, respectively. Moreover, a separate linear mixed-effects regression model comparing /wɛ/ and /we/ did not show any significant differences between the two diphthongs (β = -0.03, SE = 0.02, p = 0.19).

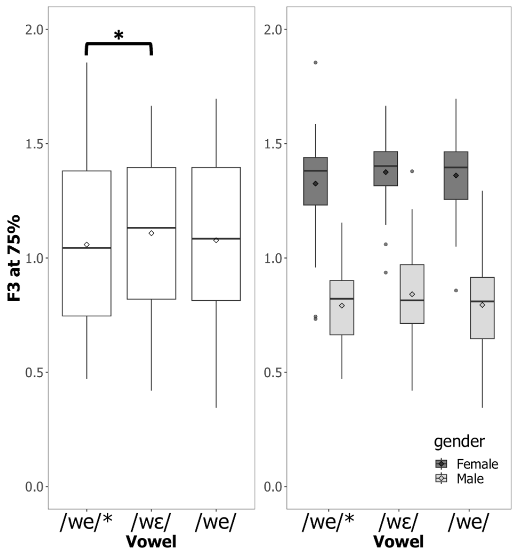

Figure 6. Box plot showing normalized F3 at the 75% point of vowel durations as a function of Vowel (left panel) and as a function of Vowel and Gender (right panel). The diamonds on the boxes represent the mean F3 at the 75% point of the vowel duration for each condition.

Regarding acoustic vowel duration, a linear mixed-effects regression model showed no main effect of Vowel, Onset, and Gender on the acoustic duration of the vowels and there was no significant interaction between the main effects that show any distinction between the three diphthongs. Additionally, a separate linear mixed-effects regression model showed that there was no significant difference between /wɛ/ and /we/. These results suggest that speakers in the present study did not distinguish /we/*, /wɛ/, and /we/ in terms of the acoustic duration of the diphthongs.

In summary, young Seoul Korean speakers in the present study did not distinguish among /we/*, /wɛ/, and /we/ in temporal acoustic dimensions. However, the speakers showed some differences among the diphthongs in the spectral acoustic dimensions. Specifically, the speakers demonstrated significant differences between /we/*-/wɛ/ and /we/*-/we/ on F2 at the 50% point of the vowel duration, with F2 being higher for /wɛ/ and /we/ than /we/*. Speakers also showed a significant difference between /we/* and /wɛ/ on F2 at the 75% point of the vowel duration, with F2 being higher for /wɛ/ than /we/*. No gender-related differences were observed in this context with no /h/. However, in the presence of /h/, female speakers showed a distinction between /hwe/* and /hwɛ/, with F2 being higher for the latter. Moreover, with regards to F3 at the 75% point of the vowel duration, speakers demonstrated higher F3 for /wɛ/ than /we/*.

3.2 Kinematic measures

Tongue dorsum movements were analyzed in terms of both horizontal (i.e., tongue frontness) and vertical (i.e., tongue height) movements.

Regarding horizontal tongue dorsum movements, there were no significant effects observed on Displacement, Peak velocity, and Tongue position in the spatial dimension. However, regarding Movement duration - i.e., the duration between the onset and target of the horizontal tongue dorsum movement, there were significant two-way interactions between Vowel (/we/*-/wɛ/) and Onset (β = 34.54, SE = 13.11, $p < 0.01$) and between Vowel (/we/*-/we/) and Onset (β = 31.65, SE = 13.11, $p = 0.02$). Post-hoc pair-wise comparisons showed that the interaction stemmed from the fact that /hwe/* and /hwɛ/ were produced with different horizontal tongue dorsum movements (β = -31.02, SE = 9.66, $p = 0.03$). That is, speakers showed shorter horizontal tongue dorsum Movement duration when producing /hwe/* (Mean = 173.41, SD = 48.81) than /hwɛ/ (Mean = 204.43, SD = 51.78) as can be seen in the left panel of Figure 7. It is interesting to note that these two vowels showed difference in the acoustic dimension of F2, being lower for /hwe/* than for /hwɛ/. However, while this difference in the acoustic dimension of F2 was demonstrated for only the female speakers, the difference in the articulatory dimension of horizontal tongue dorsum movement was consistent for both female and male speakers, as shown in the right panel of Figure 7.

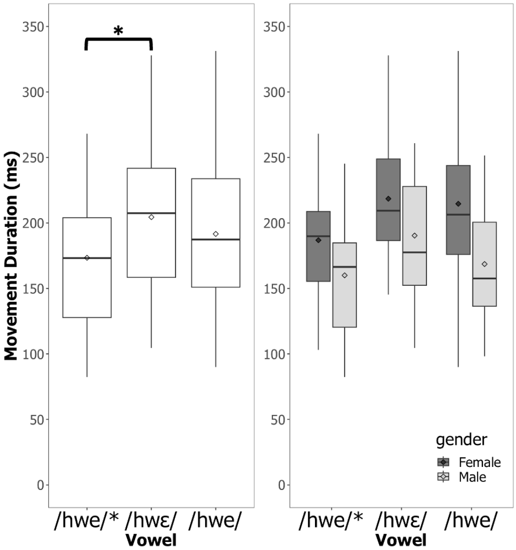

Figure 7. Box plot showing the horizontal tongue dorsum Movement duration as a function of Vowel (left panel) and as a function of Vowel and Gender (right panel). The diamonds on the boxes represent the mean Movement duration for each condition.

In terms of the vertical movement of the tongue dorsum, a series of linear mixed-effect regression models showed no effects related to Movement duration and Peak velocity. However, there was a significant main effect of Vowel on Displacement and Tongue position at the onset of the tongue dorsum movement. Specifically, a linear mixed-effects regression model showed that Displacement (i.e., the difference in tongue position at the onset and target of the vertical tongue movement) was larger for /we/* (Mean = 1.02, SD = 1.52) than /we/ (Mean = 0.64, SD = 1.60) (β = -0.38, SE = 0.17, p = 0.02), as shown in the left panel of Figure 8. Moreover, this difference between /we/* and /we/ was similar for both female and male speakers, as shown in the right panel of Figure 8.

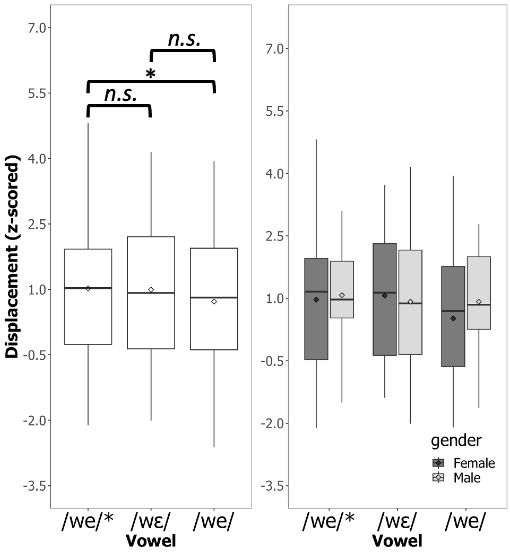

Figure 8. Box plot showing the Displacement of vertical tongue dorsum movement as a function of Vowel (left panel) and as a function of Vowel and Gender (right panel). The diamonds on the boxes represent the mean Displacement for each condition.

The significant main effect of Vowel on the Tongue position at the onset of the tongue dorsum movement ($\beta = 0.35$, $SE = 0.15$, $p = 0.02$) showed that when producing the diphthong /we/*, speakers started producing the diphthong with a lower tongue dorsum position (Mean = -0.98, SD = 1.09) than /we/ (Mean = -0.63, SD = 1.27), as shown in the left panel of Figure 9. Additionally, these patterns were consistent for both female and male speakers, as shown in the right panel of Figure 9.

While /we/* was produced with a lower vertical tongue dorsum position at the onset compared to /we/, this difference, as suggested by a reviewer, might be due to a contextual effect: The height of the item presented to the speakers before the target vowel could influence the starting position of the target vowel. As described in Section 2.1, the data analyzed in the present study were taken from a corpus which included lists of monophthongs and diphthongs. Thus, even though the items were monosyllabic words produced in isolation, each target item was presented after another vowel during elicitation. For example, if a high vowel was presented to the speakers before the target vowel, the target vowel may have begun with a higher tongue dorsum position.

Therefore, a follow-up analysis that controls for the height of the preceding vowel was conducted. Specifically, a linear mixed-effects regression model with a subset of the data including target diphthongs that followed high vowels was conducted. The follow-up analysis demonstrated that the subset of the target diphthongs that followed high vowels continued to show a significant difference between /we/* and /we/ (i.e., /we/* had a lower tongue dorsum position than /we/) ($\beta = 0.52$, $SE = 0.20$, $p = 0.01$), suggesting that the effect was not solely driven by the height of the preceding item. Further, as there was no significant difference between the diphthongs /we/* and /we/ on the Tongue position at the target of the tongue dorsum movement, it is likely that the difference in Displacement between /we/* and /we/ (the left panel of Figure 8) have stemmed primarily from the difference in the Tongue position at the onset of the tongue dorsum movement (the left panel of Figure 9).

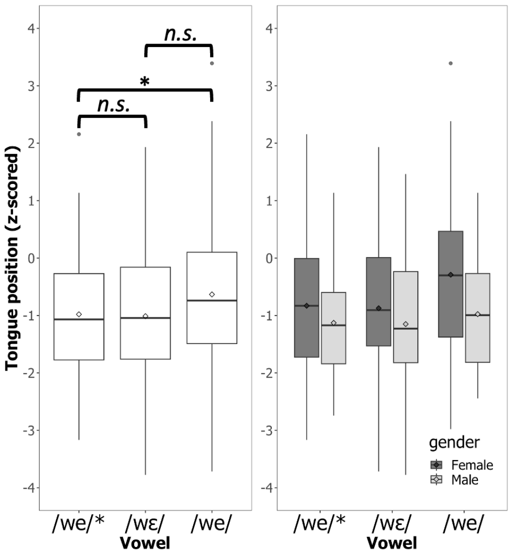

Figure 9. Box plot showing the Tongue position at the onset of the vertical tongue dorsum movement as a function of Vowel (left panel) and as a function of Vowel and Gender (right panel). The diamonds on the boxes represent the mean Tongue position for each condition.

Regarding lip aperture, no factors pertaining to lip aperture provided evidence for the distinction between the three diphthongs. That is, no significant main effects were found for Lip aperture, Movement duration, Displacement, Peak velocity, nor significant interactions between the main effects. Further, no significant main effects were found for the Lag between horizontal tongue movement and lip movement and between vertical tongue movement and lip movement suggesting that speakers of Seoul Korean rely on tongue dorsum movements (e.g., different movement durations and different positions) when making any subtle distinctions between the diphthongs /we/*, /wɛ/, and /we/.

4. Summary and discussion

The present study examined the acoustic and articulatory characteristics of Korean diphthongs /we/* '외', /wɛ/ '왜', and /we/ '웨' in isolation, considering gender (female/male) and onset (presence/absence of /h/) as additional factors. Its goal was to gain a better understanding of the precise extent of the merger of Korean diphthongs /we/*, /wɛ/, and /we/ as produced by young Seoul Korean speakers in their twenties. The acoustic analysis involved the measurement of Formant frequencies (F1, F2, F3) and acoustic Vowel duration; and articulatory kinematic analysis involved Tongue dorsum position, Movement duration, Peak velocity, Displacement in both horizontal and vertical dimensions, as well as lip Movement duration, Constriction duration, Peak velocity, and Displacement.

Many of these measures showed no difference in the production of /we/*, /wɛ/, and /we/. In fact, there was no single measure that differentiated all three diphthongs with a three-way contrast. In particular, no measure, whether acoustic or kinematic, returned any significant difference between the 'old' diphthongs /wɛ/ and /we/ when compared directly regardless of the presence of /h/. This indicates that these two diphthongs are indeed merged among young Seoul Korean, possibly in line with the merger of the vowels /ɛ/ 'ㅐ' and /e/ 'ㅔ' whose merger has been reported to be completed at least among young Seoul Korean (e.g., Lee and Cho 2021).

Crucially, however, the results showed some evidence that the 'new' diphthong /we/* ('외'), which has undergone diphthongization relatively recently (e.g., Yang 1993; Shon 2022), manifested itself somewhat differently in some fine phonetic details from the other two 'old' diphthongs which have been merged relatively earlier along with the merger of the vowels /ɛ/ ‘ㅐ’ and /e/ ‘ㅔ’. In the acoustic dimension, there were spectral measures (i.e., F2 at 50% and 75% points of the vowel duration and F3 at the 75% point of the vowel duration) that showed some spectral differences between the ‘new’ (/we/*) and ‘old’ (/wɛ/ and /we/) diphthongs. Given that the precursor of the newly introduced diphthong was a rounded vowel (/ø/), the F2/F3 lowering may be a phonetic remnant of the roundedness in the preceding vowel, as formants are typically lower when vowels are rounded compared to when they are unrounded. Moreover, a significant difference in F2 at the 75% point of vowel duration was observed in the /h/ context, specifically between /hwe/* ‘회’ and /hwɛ/ ‘홰’, but this was found only in the speech of female speakers. (We will discuss this further later.) In contrast, no other acoustic measures, be they spectral or temporal, showed any differences between /wɛ/ ‘왜’ and /we/ ‘웨’, regardless of the presence or absence of /h/ in the onset. In the articulatory analyses, on the other hand, more evidence emerged regarding the distinction between the ‘new’ diphthong /we/* ‘외’ and the ‘old’ diphthongs /wɛ/ ‘왜’ and /we/ ‘웨’. Specifically, in the horizontal dimension of the tongue dorsum Movement duration, /we/* ‘외’ exhibited a shorter duration than /wɛ/ ‘왜’ but only in the presence of /h/ (i.e., /hwe/* ‘회’ versus /hwɛ/ ‘홰’). In the vertical movement dimension, speakers showed a distinction between /we/* ‘외’ and /we/ ‘웨’ in Displacement. That is, Displacement in the vertical dimension was larger for /we/* ‘외’ than /we/ ‘웨’. Further, the speakers showed lower Tongue position at the onset of the vertical tongue dorsum movement for /we/* ‘외’ than for /we/ ‘웨’, while such distinction was not observed with /h/. These differences in Movement duration, Displacement, and Tongue position suggest that the recently introduced diphthong /we/* ‘외’ may not be entirely merged at least in phonetic terms especially in the articulatory temporal dimension with the ‘old’ diphthongs even among young Seoul Korean speakers. But the distinction was evident differentially as a function of the phonetic context of the onset (presence/absence of /h/).

At the moment, we do not have a clear explanation for why these differences are specific to certain contexts, and we acknowledge that interpreting such an asymmetrical pattern should be approached with caution. Nevertheless, we touch on a rather speculative possibility that these context-specific patterns might be due to traditional categorizations, as reflected in orthographic representations and potentially in interaction with lexical frequency effects. For one thing, one could not entirely dismiss the possibility that the difference between /we/* ‘외’ and /we/ ‘웨’ in terms of Displacement and Tongue position at the movement onset indeed reflects the speaker’s effort to distinguish the two diphthongs. Specifically, given that Seoul Korean speakers likely encounter diverse dialects and various age groups in the Seoul speech community, and given that such distinctions are reflected in the orthographic representations, the young speakers might have been exposed to different diphthong variants (including the one that closely resembles the monophthong variant of /we/*), which in turn could prompt speakers to exert greater effort in distinguishing between the 'new' and 'old' diphthongs. However, it still remains unclear why young Seoul Korean speakers began pronouncing /we/* '외' with a lower tongue position compared to /we/ '웨'. Our analysis of contextual effects suggests that this difference is not attributable to varying prior contexts (i.e., item presented before the target item) between the two. Instead, it might stem from the influence of orthography: 'ㅗ' in '외' (/we/*) corresponds to the mid vowel /o/, and 'ㅜ' in '웨' (/we/) corresponds to the high vowel /u/ in Korean. It is plausible to hypothesize that this orthographic distinction could have influenced young speakers in their attempts to differentiate these sounds. This hypothesis requires further validation.

For another, the difference in the Movement duration between /hwe/* '회' and /hwɛ/ '홰' may be further influenced by the lexical frequency of the diphthongs. High-frequency words are more likely to undergo temporal reduction than low-frequency words (e.g., Aylett and Turk 2004; Pluymaekers, Ernestus, and Baayen 2005; Arnon and Priva 2013). In our study, the observation that speakers produced /hwe/* '회' with shorter tongue dorsum movement durations compared to /hwɛ/ '홰' could be interpreted as being related to the differing lexical frequencies of these two diphthongs with /hwe/* '회' being more frequent. This frequency-related temporal distinction, coupled with the influence of orthographic representations, may account for the context-specific effects observed in our study. However, further validation is needed, especially regarding the relationship between lexical frequency and orthographic effects.

These results taken together reveal an intriguing pattern regarding the phonetic nature of the merger of these diphthongs, indicating that the traditional categorizations of these diphthongs that are clearly reflected in orthographic representations influence their production. While the young Seoul Korean speakers in this study do not show discernible differences in the acoustic and articulatory features between the 'old' diphthongs /wɛ/ '왜' and /we/ '웨', they do distinguish between /we/* '외' and /wɛ/ '왜', as well as /we/* '외' and /we/ '웨', though in a very subtle but significant way in particular contexts. This implies that even young Seoul Korean speakers may still make efforts to maintain distinctions between the 'new' diphthong that has been newly introduced to the sound system (/we/* ‘외’) and the ‘old’ diphthongs (/wɛ/ ‘왜’ and /we/ ‘웨’). The details of these differences have further implications in terms of the merits of studying sound change in both acoustic and articulatory studies and the role of gender in sound change.

First, the results of the present study demonstrate the benefits of incorporating articulatory data when investigating sound change. The present study showed potential incongruities between acoustic and articulatory features. For instance, although there was a difference in the duration of tongue dorsum movement between the diphthongs /hwe/* ‘회’ and /hwɛ/ ‘홰’ (with the former being shorter), this difference did not clearly manifest in the acoustic duration of the diphthongs. Additionally, in the articulatory dimension, there was a distinction observed between /we/* ‘외’ and /we/ ‘웨’ in terms of Displacement and Tongue position at the onset, yet no corresponding difference was found in acoustic formant measures. In contrast, in the acoustic dimension, differences emerged between the ‘new’ (/we/*) and ‘old’ (/wɛ/ and /we/) diphthongs in F2 and F3 at a later point of the vowel (75%), while no corresponding articulatory Tongue position measure exhibited a similar pattern.

As discussed at the outset of the present study, this incongruity between the articulatory and acoustic features of the diphthongs is indeed consistent with previous observations (e.g., De Decker and Nycz 2012; Wieling et al. 2016), and is consistent with the view which posits a non-linear relationship between changes in articulation and the resulting acoustic characteristics (Stevens 1989; Stevens and Keyser 2010; see Iskarous (2011), and references therein for a review on the non-linearity between the acoustic and articulatory dimensions). While the young Seoul Korean speakers in our study seemed to exert articulatory efforts to distinguish between the diphthongs, particularly the ‘new’ and ‘old’ ones, these distinctions in articulatory characteristics may not necessarily align with acoustic features, and vice versa. As the results of the present study suggest that the articulatory and acoustic features of the diphthongs do not consistently align, a nuanced understanding of the phonetic nature of the potential ongoing sound change such as the merger of /we/*, /wɛ/, and /we/ may require an examination of both the acoustic and articulatory features of these diphthongs.

Secondly, while the results of the present study generally indicated comparable diphthong productions across genders, there was some evidence suggesting a potential gender effect. Specifically, the results revealed that female speakers made a small but significant distinction between /hwe/* (‘회’) and /hwɛ/ (‘홰’) in the second formant (F2) at the 75% point of the vowel duration, whereas male speakers did not exhibit this differentiation. That is, F2 was significantly lower in /hwe/* ('회') than /hwɛ/ ('홰') in female speech. It is not clear whether this directionality is just an epiphenomenal way of making the distinction, or if it may have some basis. Here again we lack a definitive explanation for this, but we do not rule out the possibility that the observed directionality is phonetically motivated. As we discussed above, the F2 lowering of /hwe/* ('회') may reflect the roundness of the precursor /ø/ of the newly introduced diphthong.

While this interpretation requires further corroboration, it is also important to exercise caution when drawing generalizations from this specific result. As pointed out by a reviewer, this caution is necessary due to the limited sample size, which consisted of 8 female and 8 male speakers, each producing two repetitions of every vowel. Furthermore, the context was limited to F2 at the 75% point of the vowel duration. Thus, we examined the distinction pattern between /hwe/* '회' and /hwɛ/ '홰' for each speaker at this measurement point to assess the consistency of the observed distinction across individual speakers. All eight female speakers exhibited a lower mean F2 at the 75% point of the vowel duration for /hwe/* '회' compared to /hwɛ/ '홰'. In contrast, five out of the eight male speakers did not demonstrate this pattern (i.e., F2 at the 75% point of the vowel duration was the same between the two diphthongs or lower for /hwɛ/ '홰' than /hwe/* '회'). To the extent that this analysis holds, the difference may be attributed to the tendency of female speakers to enunciate more clearly than male speakers, especially in communicative situations that demand clarity (e.g., Ferguson and Morgan 2018). Specifically, the need to distinguish between the two types of diphthongs seems to be reinforced by the distinctiveness maintained in the orthographic representations of these diphthongs. As was also noted by a reviewer, it is possible that the orthographic difference between the new ('ㅚ') and old diphthongs ('ㅙ/ㅞ') is more noticeable than the difference between the old diphthongs themselves. This could motivate the continued differentiation of the two types of diphthongs, even if it is subtle, in the speech of female speakers who typically prioritize clarity and standard variety over male speakers, all else being equal (cf. Labov 2001). However, the reason for the observed gender effect in the /h/ context (/hwe/* '회' versus /hwɛ/ '홰'), but not in the onsetless contexts (/we/* '외' versus /wɛ/ '왜'), remains unclear. As previously discussed in conjunction with the articulatory durational difference, this may be related to lexical frequency effects, since the difference in lexical frequency is greater for the former pair than the latter, which is evident in the speech of the female speakers. This could also imply that young speakers struggle to distinguish these sounds in a systematic way, further suggesting that the merger of these sounds may have been completed at the level of phonology, while other factors such as orthography and lexical frequency may fine-tune the phonetic realization of the phonologically merged diphthongs.

In conclusion, the present study has demonstrated that the three diphthongs /we/*, /wɛ/, and /we/ are nearly fully merged across various acoustic and articulatory dimensions. It was particularly evident between the 'old' diphthongs /wɛ/ '왜' and /we/ '웨'. But the results showed some evidence that the newly introduced diphthong /we/* '외' is still produced differently, albeit in some subtle ways, compared to the other two diphthongs /wɛ/ '왜' and /we/ '웨'. This distinction might suggest that the merger between the 'new' and 'old' diphthongs is not yet complete. However, it is also plausible that the observed difference is due to orthographic effects, indicating that the merger may be complete at the phonological level. The importance of examining both the acoustic and articulatory dimensions to uncover nuanced phonetic details of otherwise seemingly neutralized contrasts was emphasized. This approach helps illuminate the traditional categorizations of the diphthongs, the orthographic system, and the role of gender in understanding the extent of the three-diphthong merger. However, as we pointed out above, the discussion based on the current findings should be considered preliminary, especially given the limitation of the sample size and its scope. Further research is required to better understand how the observations made in the present study may be generalizable to the broader contexts and across different dialects, and how they may be influenced by lexical factors (e.g., lexical status and frequency) and prosodic structural factors. This includes investigating whether the distinction of possible merger cases may be further conditioned by prosodic boundaries and prominence factors in connected speech, given that sound change is often conditioned by prosodic structure in which it occurs (e.g., Choi et al. 2020). Finally, more detailed analyses of these diphthongal vowels, especially focusing on dynamics of the articulatory movement in multiple dimensions (including lip protrusion characteristics), may further reveal the phonetic versus phonological nature of diphthongs as a combination of a consonantal glide plus a vowel as separate phonological representations or two vocalic components constituting a single phonological representation.

References

Arnon, Inbal and Uriel Cohen Priva. 2013. More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech* 56(3): 349-371.

Aylett, Matthew and Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1): 31-56.

Barreda, Santiago. 2021. Fast Track: Fast (nearly) automatic formant-tracking using Praat. *Linguistics Vanguard* 7(1): 1-10.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1-48.

Chang, Seung-Eun. 2017. Enhancement effects of clear speech and word-initial position in Korean glides. *The Journal of the Acoustical Society of America* 141(6): 4188-4199.

Cheon, Sang Yee. 2002. Glides as consonants in Korean. *Language Research* 38(2): 619-646.

Cho, Sung-Moon. 2003. An acoustic study of Korean vowel system. *Korean Language and Culture* 24: 427-441.

Choi, Jiyoun, Kim Sahyang, and Cho Taehong. 2020. An apparent-time study of an ongoing sound change in Seoul Korean: A prosodic account. *PloS one* 15(10): e0240682.

De Decker, Paul M. and Jennifer R. Nycz. 2012. Are tense [æ] s really tense? The mapping between articulation and acoustics. *Lingua* 122(7): 810-821.

Ferguson, Sarah Hargus and Shae D. Morgan. 2018. Talker differences in clear and conversational speech: Perceived sentence clarity for young adults with normal hearing and older adults with hearing loss. *Journal of Speech, Language, and Hearing Research* 61(1): 159-173.

Gick, Bryan. 2003. Articulatory correlates of ambisyllabicity in English glides and liquids. In John Local, Richard Ogden, and Rosalind Temple (eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI,* 222-236. Cambridge: Cambridge University Press.

Iskarous, Khalil. 2011. Articulatory to acoustic modeling. In Abigail Cohn, Cécile Fougeron, and Marie K. Huffman (eds.), *The Oxford handbook of laboratory phonology,* 472-483. Oxford: Oxford University Press.

Kang, Hyeon-Seok. 2003. A phonetic account of the current inventory and ongoing changes of Korean diphthongs. *Language Research* 39(4): 809-838.

Kang, Yoonjung. 2014. Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics* 45: 76-90.

Labov, William. 1973. *Sociolinguistic patterns.* Philadelphia, PA: University of Pennsylvania Press.

Labov, William. 2001. *Principles of linguistic change: Social factors*, *volume 2*. Oxford: Blackwell.

Lee, Hyun-Bok. 1993. Korean. *Journal of the International Phonetic Association* 23(1): 28-31.

Lee, Yong-cheol and Cho Sunghye. 2021. The merger of /e/ and /ɛ/ in Seoul Korean tracing speakers born from the 1930s to the 1960s. *Studies in Linguistics* 61: 1-16.

Lenth, Russel V. 2021. Estimated marginal means, aka least-squares means [R Package Emmeans Version 1.6.0]. *Comprehensive R Archive Network (CRAN)*.

Pluymaekers, Mark, Mirjam Ernestus, and R. Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America* 118(4): 2561-2569.

R Development Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Shon, Sangmi. 2022. *A study on the formant trajectory of Korean diphthongs*. PhD Dissertation. Hanyang University.

Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17(1): 3-45.

Stevens, Kenneth N. and Samuel Jay Keyser. 2010. Quantal theory, enhancement and overlap. *Journal of Phonetics* 38(1): 10-19.

The Hanyang Institute for Phonetics and Cognitive Science of Language (HIPCS). 2022. *Dynamics of speech production through articulatory DB construction*.

Tiede, Mark. 2005. *MVIEW: Software for visualization and analysis of concurrently recorded movement data.* New Haven, CT: Haskins Laboratories.

Wieling, Martijn, Fabian Tomaschek, Denis Arnold, Mark Tiede, Franziska Bröker, Samuel Thiele, Simon N. Wood, and R. Harald Baayen. 2016. Investigating dialectal differences using articulography. *Journal of Phonetics* 59: 122-143.

Yang, Byeong-Gon. 1993. An acoustical study of Korean diphthongs. *Malsori* 25(1): 3-26.

Zhao, Dan, Jiyeon Park, and Cheoljae Seong. 2023. Acoustic characteristics of glides and nuclear vowels of Korean diphthongs in coarticulation condition. *Hanguel* 84(1): 5-43.

Dae-yong Lee
Postdoctoral Researcher
Hanyang Institute for Phonetics and Cognitive Sciences of Language
Hanyang University
222, Wangsimni-ro, Seongdong-gu
Seoul 04763, Korea
E-mail: daeyonglee@hanyang.ac.kr

Jiyeon Song
Postdoctoral Researcher
Hanyang Institute for Phonetics and Cognitive Sciences of Language
Hanyang University
222, Wangsimni-ro, Seongdong-gu
Seoul 04763, Korea
E-mail: jiyeoni00@hanyang.ac.kr

Sahyang Kim
Professor
Department of English Education
Hongik University
94, Wausan-ro, Mapo-gu
Seoul 04066, Korea
E-mail: sahyang@hongik.ac.kr

Taehong Cho
Professor
Hanyang Institute for Phonetics and Cognitive Sciences of Language
Department of English Language & Literature
Hanyang University
222, Wangsimni-ro, Seongdong-gu
Seoul 04763, Korea
E-mail: tcho@hanyang.ac.kr

Received: 2023. 11. 01.
Revised: 2024. 02. 06.
Accepted: 2024. 02. 07.