Systematic centralized reduction at the offset of Korean monophthong signals in spontaneous speech*

Soonhyun Hong
(Inha University)

Hong, Soonhyun. 2020. Systematic centralized reduction at the offset of Korean monophthong signals in spontaneous speech. *Studies in Phonetics, Phonology, and Morphology* 26.3. 545-565. The present paper reports the observation that the signals of all Korean monophthong types in spontaneous speech are characterized by a unique spectral pattern over the temporal domain. Over the course of production, Korean monophthong signals showed quite stable F1/F2 frequencies from 20% to 50% of vowel duration, but drastically moved toward the center of the acoustic vowel space from 50% to 80% of vowel duration. This unique spectral pattern was shared by signals of all monophthong types produced by both males and females. It was also found that centralized vowel reduction at the offset was more frequent in shorter signals than longer ones. The unique spectral centralization locally at the offset in Korean signals did not pattern together with reports in the literature that across-the-board articulatory undershoot occurs in the entire temporal domain of vowel signals in English and French. **(Inha University, Professor)**

Keywords: vowel undershoot, reduction, centralization, Korean spontaneous speech

1. Introduction

Research on articulatory and acoustic characteristics of vowels has been conducted in two directions. First, coarticulation with neighboring phones and phonetic reduction resulting from the limited temporal domain is the result of failing to reach the vowel target under the influence of neighboring phones over the course of vowel production (Moon and Lindblom 1994, Hillenbrand et al. 2001). Second, shorter vowel duration

---

* This work was supported by INHA UNIVERSITY Research Grant. Special thanks should go to Jongpyo Hong (CDS) who first observed systematic centralized reduction at the offset of signals of all monophthong types while writing a Praat script to extract spectral frequencies at cross sections over the temporal domain of monophthong signals from segmented and labelled spontaneous speech in Seoul Corpus in a project with the author (Hong and Hong 2019).

hinders articulators from reaching the vowel target, resulting in vowel undershoot (Lindblom 1963, 1983). The resulting undershoot vowel signals get reduced and centralized closer to the center of the acoustic vowel space, which in turn gets compressed (Fourakis 1991, Moon and Lindblom 1994, Guenther 1995). Vowel reduction has been frequently reported in spontaneous and conversational speech (Meunier and Espesser 2011) and speech at faster speech (Van Son and Pols 1992).

A large corpus study on vowel reduction is found in Gendrot and Adda-Decker (2005, 2007), who investigated the relationship between French vowel duration and spectral change in French corpus of recorded French radio broadcast news. They reported that shortened vowels have more reduced and centralized spectral measurements on the acoustic vowel space, confirming that the frequency of vowel reduction is correlated with shortened duration.

The present paper is also a corpus study on vowel reduction in Korean spontaneous speech. The focus is to be placed on vowel reduction in relation to the shortened vowel duration. Most studies on vowel reduction assumed that there is an ideal vowel target in vowel production (DiCanio et al. 2015). They thought that static spectral properties sampled at a steady-state central cross section along the temporal domain of signals could characterize vowels. As a result, they were focused on whether vowel signals hit the vowel target over the temporal domain under the influence of coarticulation and shortened vowel duration.

However, the present study begins with analyzing the patterns of vowel inherent spectral change (VISC) over the course of production of Korean monophthong signals. The VISC approach said that variations in formant frequencies over the temporal domain help to characterize vowel categories. It has been reported that dynamic spectral patterns of vowels in both English monophthongs (and diphthongs) affect vowel identification (Nearey and Assmann 1986, Hillenbrand et al. 1995, and Hillenbrand 2013).

Hong and Hong (2019) first observed through a corpus study of Korean monophthongs in spontaneous speech that monophthong signals are characterized by one unique dynamic spectral pattern: Stable over the first half of vowel duration then dynamic over the second half. The present paper tries to verify this observation with more evidence and further propose that Korean monophthong signals are produced consistently with one systematic spectral pattern: The approximate target is hit on at the onset, the target hitting is stably sustained until the central cross section, and finally deviation from the approximate target occurs at the offset, resulting in vowel reduction. And such vowel reduction is realized as vowel centralization on the acoustic vowel space. It will be also shown that the production sequence of target hitting followed by centralized vowel reduction within the single temporal domain is quite systematical for all monophthong types.

In the experiment, F1-3 frequencies are to be extracted from all monophthong signals in the spontaneous speech in Seoul Corpus (Yun et al. 2015) at 20%, 50%, and 80% of vowel duration along with vowel duration and talkers' gender. Then the vectors of the averaged F1/F2 measurements of signals of all vowel types along the three cross sections over the temporal domain, were to be plotted on the acoustic vowel space to see the spectral patterns of signals.

2. Experiment

2.1 Subjects

The Korean Corpus of Spontaneous Speech (Seoul Corpus) (Yun et al. 2015) are forty-hours-long speech signals which were recorded at 44kHz with 16-bit quantization. It contains speech signals spoken by four groups of ten Seoul Korean talkers in their teens, twenties, thirties, and forties, which were segmented and labeled using Praat (Boersma and Weenink 2018). Each talker group has five females and five males who produced spontaneous speech for one hour for each talker in conversations with interviewers.

2.2 Materials

From Seoul Corpus, 498,204 monophthong signals (Table 1) were extracted, using a Praat script (Yun et al. 2015, Hong 2019), along with the talkers' gender, monophthong duration, F1, F2 and F3 frequencies sampled at the cross sections of 20%, 50%, and 80% of vowel duration. The spectral sampling at 20% and 80% was done to avoid the coarticulation effects with preceding or following phones.

Table 1. Extracted signals of seven monophthong vowel types

IPA	Symbols in Seoul Corpus	Korean alphabet	No. of signals
i	ii	ㅣ	77,594
e	ee	ㅔ	73,217
a	aa	ㅏ	115,082
ɨ	xx	ㅡ	80,571
ə	vv	ㅓ	72,132
u	uu	ㅜ	36,009
o	oo	ㅗ	43,599
		total	498,204

3. Results

Vowels may be characterized by F1 and F2 frequencies on the acoustic vowel space sampled at cross sections along the temporal domain. Table 2 shows the average F1 and F2 frequencies sampled at 20%, 50%, and 80% of vowel duration for males' and females' monophthong signals.

Table 2. The average F1 and F2 frequencies sampled at 20%, 50%, and 80% of vowel duration for signals of each vowel type produced by both males and females[1]

Gender	Monophthongs	F1_20	F2_20	F1_50	F2_50	F1_80	F2_80
Males	i (/ii/)	380	1960	368	1972	355	1893
	e (/ee/)	371	1804	378	1804	368	1746
	a (/aa/)	508	1337	512	1343	458	1363
	ɨ (/xx/)	391	1467	389	1464	375	1471
	ə (/vv/)	387	1159	387	1132	376	1188
	u (/uu/)	401	1267	384	1247	369	1311
	o (/oo/)	346	1005	345	994	347	1108
Females	i (/ii/)	395	2291	394	2308	382	2199
	e (/ee/)	437	2133	442	2137	414	2060
	a (/aa/)	626	1521	637	1531	547	1556

[1] "F1_20" refers to average F1 frequency sampled at 20% of vowel duration, and "F2_20" to average F2 frequency sampled at 20%.

	ɨ (/xx/)	431	1692	434	1696	416	1699
	ə (/vv/)	441	1265	444	1234	414	1290
	u (/uu/)	404	1346	398	1348	384	1426
	o (/oo/)	379	1112	379	1092	364	1182

Figure 1 below illustrates males' and females' monophthongs on the acoustic vowel space, based on the average F1 & F2 frequency measurements sampled at 50% of vowel duration under the assumption that vowel signals are static and stable over the course of vowel production.

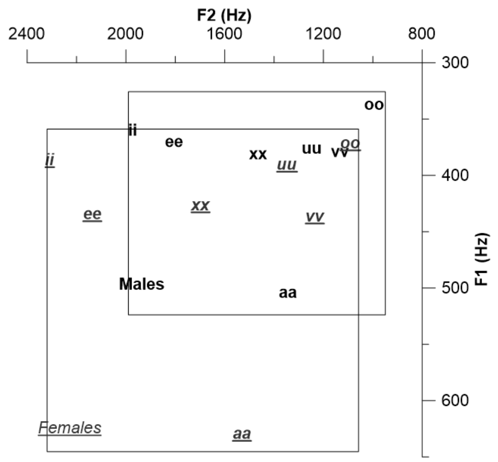


Figure 1. Males' and females' monophthongs plotted with average F1/F2 frequencies sampled at 50% of vowel duration (females' vowels italicized and underlined in the lower rectangle)

However, we will show that spectral patterns of Korean monophthongs in spontaneous speech are not static but dynamic across the temporal domain and hence Korean monophthongs cannot be properly characterized by the static spectral measurements at 50% of vowel duration.

This time, on the other hand, males' and females' seven monophthongal vowel types are represented on the acoustic vowel space with spectral vector plots beginning with average F1 and F2 measurements at 20% and ending with those at 50% of vowel duration on the one hand, and with spectral vector plots beginning with average F1 and F2 measurements at 50% and ending with those at 80%. Figure

2 and 3 illustrate males' spectral vowel vectors from 20% to 50% of vowel duration (Figure 2) and males' spectral vectors from 50% to 80% (Figure 3), on the acoustic vowel space. Figure 4 and 5 illustrate females' spectral vectors from 20% to 50% (Figure 4) and females' spectral vectors from 50% to 80% (Figure 5). Note in the figures that each spectral vector begins with a vowel symbol and ends with an arrow.

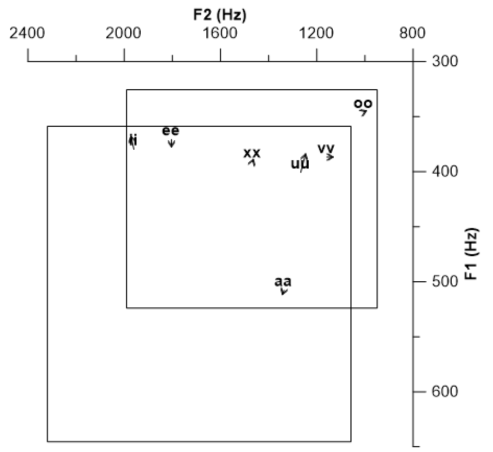


Figure 2. Males' monophthongs on the acoustic vowel space plotted with average F1/F2 vectors from 20% to 50% of vowel duration (each spectral vector begins with a vowel symbol and ends with an arrow; the lower rectangle refers to females' vowel space for comparison purpose.)

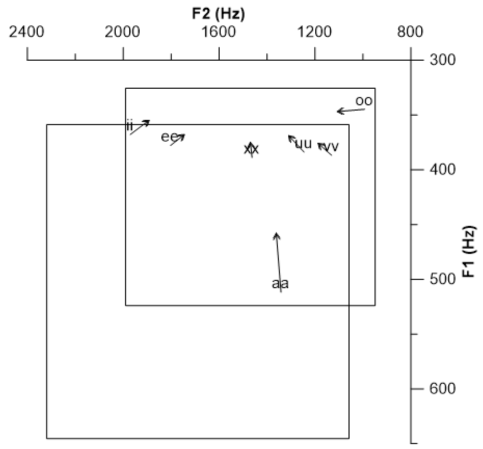


Figure 3. Males' monophthongs on the acoustic vowel space plotted with average F1/F2 vectors from 50% to 80% of vowel duration

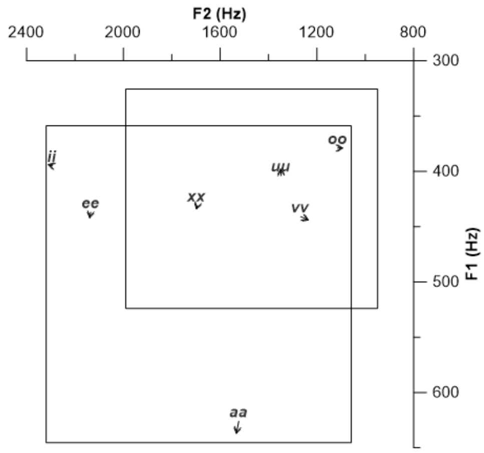


Figure 4. Females' monophthongs on the acoustic vowel space plotted with average F1/F2 vectors from 20% to 50% of vowel duration

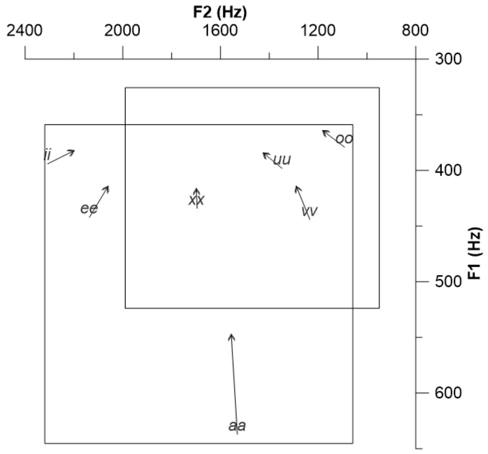


Figure 5. Females' monophthongs on the acoustic vowel space plotted with average F1/F2 vectors from 50% to 80% of vowel duration

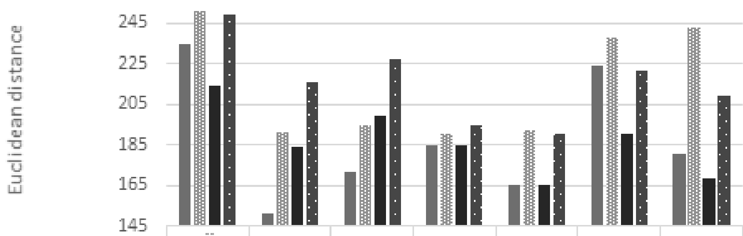
From Figures 2-5, it is observed that both males' and females' spectral vectors from 20% to 50% of vowel duration were far shorter for all vowel types than those from 50% to 80%. The spectral change of both males' and females' monophthongs was minimal and static from 20% to 50% of vowel duration (hereafter, 20to50) whereas the spectral change of both males' and females' monophthongs from 50% to 80% (hereafter, 50to80) was drastic and dynamic, moving to the center of the acoustic vowel space. It seems that static spectral properties of signals at 20% (the onset) and 50% (the central cross section) approximate spectral vowel targets. However, drastic spectral change of both males' and females' monophthongs deviate from the vowel targets while moving toward 80% (the offset). Such drastic deviation may be treated as target undershoot at the offset, resulting in systematically centralized vowel reduction.

The occurrence of vowel reduction at the offset of signals of all monophthong types may be verified by comparing the 20to50 spectral vectors to the corresponding spectral 50to80 vectors in length. Given signals of one monophthong type, if the spectral 50to80 vectors turn out to be significantly longer than the corresponding 20to50 spectral vectors, it can be said that vowel reduction occurred at the offset for that monophthong type.

The average length of spectral vectors from 20% to 50% and from 50% to 80% for males' and females' monophthongs was calculated using Euclidean distance as follows:

D1 = square root ((F1_50 – F1_20)^2 + (F2_50 – F2_20)^2)
D2 = square root ((F1_80 – F1_50)^2 + (F2_80 – F2_50)^2)
where F1_20, F1_50. F1_80, F2_20, F2_50, and F2_80 are the F1 and F2 frequencies in Hz at 20%, 50% and 80% of vowel duration, respectively, and D1 and D2 are the length of dynamic spectral changes of a monophthong type over 20to50 and 50to80, respectively.

Figure 6 and Table 3 below show that spectral 50to80 vectors were significantly longer for all males' and females' monophthongs than corresponding 20to50 spectral vectors.



	ii	ee	aa	xx	vv	uu	oo
Males 20to50	234.8	151.7	172.1	184.6	165.6	224.1	180.3
Males 50to80	250.4	191.4	194.8	190.6	191.7	237.4	242.6
Females 20to50	213.9	183.6	199.8	184.9	165.6	190.8	168.5
Females 50to80	249.5	215.5	227.1	194.2	190.4	221.1	209.6

Males 20to50 Males 50to80 Females 20to50 Females 50to80

Figure 6. Length of spectral vectors from 20% to 50% and from 50% to 80% of the temporal domain of all monophthongs produced by males and females

Table 3. Results of paired t-test on length difference between 50to80 and 20to50 vectors across the temporal domain of males' and females' monophthongs

Males	20to50	50to80	Paired t-test	Females	20to50	50to80	Paired t-test
ii	234.8 (343.2)	250.4 (341.3)	t(43335)=-8.0, p<0.01	ii	213.9 (335.2)	249.5 (345.8)	t(34248)=-16.1, p<0.01
ee	151.7 (207.1)	191.4 (246.5)	t(38585)=-29.0, p<0.01	ee	183.6 (247.2)	215.5 (259.4)	t(34622)=-19.8, p<0.01
aa	172.1 (214.4)	194.8 (226.9)	t(62611)=-22.3, p<0.01	aa	199.8 (212.2)	227.1 (218.5)	t(52460)=-24.8, p<0.01
xx	184.6 (276.3)	190.6 (278.9)	t(44243)=-4.1, p<0.01	xx	184.9 (269.6)	194.2 (283.3)	t(36312)=-5.8, p<0.01
vv	165.6 (241.7)	191.7 (280.6)	t(37915)=-16.2, p<0.01	vv	165.6 (200.3)	190.4 (230.0)	t(34183)=-17.9, p<0.01
uu	224.1 (312.8)	237.4 (321.5)	t(19081)=-4.8, p<0.01	uu	190.8 (260.4)	221.1 (280.7)	t(16920)=-12.4, p<0.01
oo	180.3 (276.0)	242.6 (355.2)	t(25530)=-25.0, p<0.01	oo	168.5 (225.7)	209.6 (272.1)	t(18064)=-18.2, p<0.01

The length difference between 50to80 and 20to50 vectors was significant for all males' and females' monophthong types: Paired t-test: 20to50 < 50to80, p<0.001 for all monophthongs. This means that the spectral properties of signals of all monophthong types keep steady state over the first half of vowel duration whereas the spectral properties abruptly change from 50% to 80%, as was observed first in Hong and Hong (2019).

These results indicate that the spectral properties of all monophthongs produced by males and females deviated from the spectral targets at the offset, resulting in vowel reduction after the approximate spectral target had been hit over the first half of vowel duration. These reduced vowels are realized as centralized on the acoustic vowel space. Vowel centralization becomes more evident when F1/F2 vectors of all monophthongs from 20% directly to 80% are plotted on the acoustic vowel space. In Figures 7 and 8, F1/F2 vectors of all monophthongs of males' (Figure 7) and females' (Figure 8) from 20% to 80% of vowel duration are plotted on the acoustic vowel space, where it is clearly observed that all monophthongs get centralized at the offset.

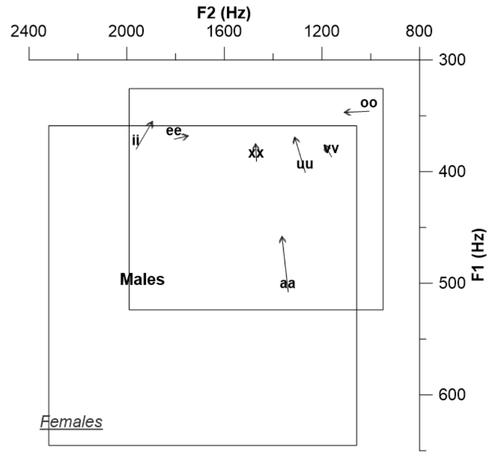


Figure 7. Spectral vectors from 20% directly to 80% of vowel duration for all monophthongs produced by males

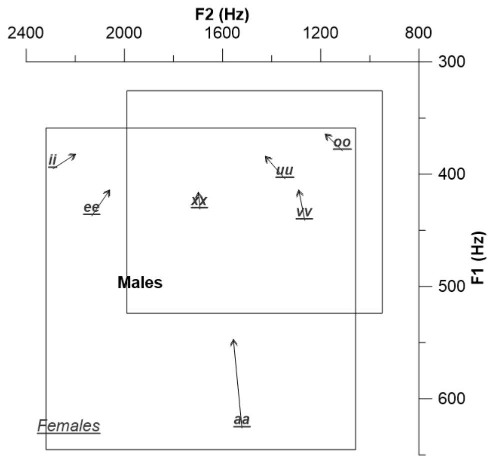


Figure 8. Spectral vectors from 20% directly to 80% of vowel duration for all monophthongs produced by females

To summarize, vowel reduction or articulatory undershoot at the offset was realized as centralization, as F1 and F2 frequencies of all males' and females' monophthongs deviated, at the offset, from the appropriate spectral target and moved toward the center of the acoustic vowel space.

4. Discussion

4.1 Vowel centralization at the offset verified by neural network classifier modeling

Recent studies have tried to build up spectral pattern recognition models of vowel signals for better identification accuracy with an emphasis on temporal spectral movement patterns of signals. Proposed were pattern recognition models which were fitted to spectral measurements sampled singly at one cross section along vowel signals or spectral measurements sampled multiply at two or more cross sections (Assmann et al. 1982, Nearey and Assmann 1986, Zahorian and Jagharghi 1993, Hillenbrand et al. 1995, Hong 2019, 2020). Hillenbrand et al. (1995) demonstrated that discriminant pattern recognition models fitted to spectral measurements sampled doubly at 20% and 80% performed better than at 50%, for classification of the American English (AE) vowel signals in the hVd syllable. They suggested that the temporal spectral movement patterns play an important role in correct identification of AE vowel signals and it can be captured best by spectral samples picked up doubly at 20% and 80% of vowel duration.

On the other hand, most studies on vowel undershoot or reduction tacitly assumed that there is an ideal vowel target over the course of vowel production (DiCanio et al. 2015). Hillenbrand et al. (2001) and Moon and Lindblom (1994) reported that vowel target undershoot results from coarticulation with neighboring phones. Lindblom (1963, 1983) and Meunier and Espesser (2011) argued that vowel target undershoot results from shorter vowel duration.

The present study tacitly assumes that there may be an approximate vowel target at least at one cross section over the temporal domain of vowel signals. And it is further assumed, following the VISC approach (Nearey and Assmann 1986, Hillenbrand et al 1995, Hillenbrand 2013), that the spectral patterns across the temporal domain are additionally important to characterize vowels. Under these two hypotheses, we hope to verify, through modelling study of Korean monophthongs, that the approximate vowel target is formed over the first half of the temporal domain and vowel undershoot or reduction occurs at the offset.

If vowel undershoot or vowel reduction really occurs at the offset, it may be naturally hypothesized that the approximate vowel target had been reached over the first half. Then it can be conjectured that listeners perceived monophthong signals as intended monophthongs by referring, though not entirely, more to the spectral properties over the first half. Then it would naturally follow that a pattern recognition model fitted to the spectral properties at the offset (80% of vowel duration) would turn out a poorer vowel identification accuracy than models fitted to the spectral properties sampled at one cross section (20% or 50%) over the first half. If this is the case, it can be said that vowel reduction occurred at the offset.

For this purpose, neural network (NN)-based static vowel classifiers (Yoon (2019), Hong (2020) for English vowels) were fitted to the F1-3 measurements sampled singly at 20%, 50% or 80% of vowel signals for correct identification rates. Considered were three static spectral NN pattern recognition models[2] 1) fitted to F1-3 frequencies sampled at 20% (F123_20), 2) fitted to F1-3 sampled at 50% (F123_50), and 3) fitted to F1-3 sampled at 80% (F123_80). The NN vowel classifier was trained with two repetitions on 66% of the data and tested with two repetitions on the rest of the data, in a supervised mode through stratified random sampling.

The purpose was to find a static model turning out the worst identification performance when fitted to spectral properties sampled singly at different temporal cross sections. If F123_80 performed worse than F123_20 and F123_50, it indicates that vowel reduction occurred at the offset and the approximate vowel target may have been reached elsewhere.

Table 4. Static neural network spectral classifiers fitted to F1-F3 measurements sampled at different temporal cross sections of vowel duration

Static Models	F123_20	F123_50	F123_80
Description	F1-F3 sampled at 20% of vowel duration	F1-F3 sampled at 50% of vowel duration	F1-F3 sampled at 80% of vowel duration

[2] In the present study, used was the neural network widget in Orange 3.21 (Demsar et al. 2013) which uses sklearn Multi-layer Perception algorithm that can learn non-linear models as well as linear: 100 Neurons per hidden layer, the rectified linear unit function for activation, stochastic gradient-based optimizer Adam as a solver for weight optimization, 0.00010 for L2 penalty (regularization term) parameter Alpha, and 200 maximum number of iterations (Hong 2019, 2020). For more information, please refer to https://docs.biolab.si//3/visual-programming/widgets/model/neuralnetwork.html.

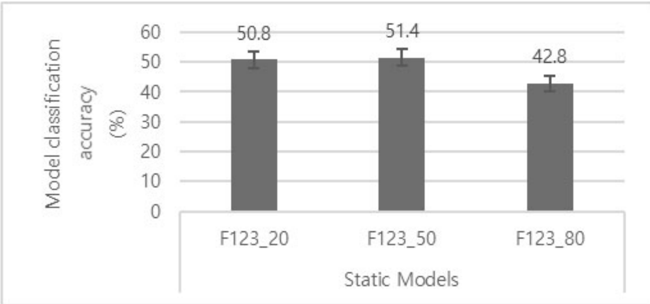


Figure 9. Identification performance (% correct) of static models

Figure 9 shows that the static spectral classifier with spectral measurements sampled singly at 80% of vowel duration (F123_80: 42.8% correct) identified vowel signals as intended vowels far worse than those with spectral measurement at 20% (F123_20: 50.8%) and at 50% (F123_50: 51.4%). However, F123_20 and F123_50 showed relatively little difference in their identification performance. This result suggests that the spectral vowel target is formed over the first half of vowel duration and vowel reduction occurred at the offset.

It seems that monophthong signals stably shoot the approximate spectral target at 20% of vowel duration and the target shooting was stably sustained at 50%. Afterwards they moved away from the spectral target, getting centralized at 80%.

Either across-the-board articulatory undershoot or perceptual overshoot occurring in the entire temporal domain of vowel signals in spontaneous speech has been frequently reported in the literature. However, Korean monophthong signals in spontaneous speech systematically shared the peculiar spectral pattern within a single temporal domain: The approximate vowel target is stably reached over the first half of vowel duration and then articulatory undershoot or vowel reduction occurs at the offset. Such a peculiar spectral pattern of Korean monophthong signals has never been reported in the literature.

4.2 Role of vowel reduction in vowel identification

So far it has been shown that a static spectral model fitted to F1-3 frequencies sampled singly at 80% of vowel duration showed worse identification accuracy than models fitted to F1-3 sampled singly at 20% or at 50%. In this section, it will be further shown that spectral properties of vowel centralization at the offset (80%) are secondary cues for vowel identification, suggesting that vowel centralization at the offset constitutes an additional spectral characteristic to identify Korean monophthong signals as intended monophthongs.

A dynamic classifier was fitted to F1-3 frequencies sampled doubly or triply at 20%, 50% and 80% of vowel duration, as described in Table 5.

Table 5. Dynamic neural network spectral classifiers fitted to F1-F3 measurements sampled doubly or triply at temporal cross sections of vowel duration

Dynamic Models	F123_2050	F123_5080	F123_2080	F123_205080
Description	F1-F3 sampled at 20% and 50% of vowel duration	F1-F3 sampled at 50% and 80% of vowel duration	F1-F3 sampled at 20% and 80% of vowel duration	F1-F3 sampled at 20%, 50%, and 80% of vowel duration

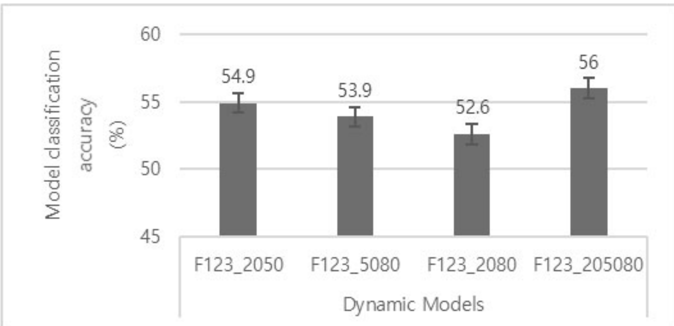


Figure 10. Identification performance (% correct) of dynamic models

In Figure 10, the spectral pattern recognition model fitted to F1-3 frequencies sampled triply at 20%, 50% and 80% of vowel duration (F123_205080: 56% correct) performed better than models fitted to F1-3 frequencies doubly at 20% and 50% (F123_2050), at 50% and 80% (F123_5080), or at 20% and 80% (F123_2080). Notice that F123_2080 performed worst whereas F123_205080 performed best. And F123_2050 performed worse than F123_205080. This means that all the spectral information at 20%, 50% and 80% of vowel duration should be referred to for the best identification performance unlike the best performing F123_2080 model for the identification of AE vowel signals in the hVd syllable in Hillenbrand et al. (1995). [3] Hillenbrand et al. 1995 further reported that a pattern recognition classifier fitted to F1-3 frequencies sampled triply at 20%, 50%, and 80% did not have the edge in identification performance over the model fitted to F1-3 sampled doubly at 20% and 80%.

However, the spectral pattern of Korean monophthongs did not pattern together with the spectral pattern of AE vowel signals in the hVd syllable. In identification of Korean monophthong signals, F123_205080 (56%) had the clear edge over F123_2080 (52.6%), F123_2050 (54.9%), or F123_5080 (53.9%). The difference between Korean and English monophthong signals in model identification may result from the unique and systematic spectral pattern of Korean monophthong signals over the single temporal domain: Static at the onset and still static at the central cross section, and then abruptly reduced and centralized at the offset. And this unique spectral pattern of Korean monophthong signals seems to be captured better with F123_205080 than with F123_2080, F123_5080 or F123_2050. It seems that F123_2080 and F123_5080 failed to capture the static spectral properties over the first half of vowel duration. F123_2050 failed to capture the dynamic vowel reduction over the second half. However, the best-performing F123_205080 seems to capture the static properties over the first half and the dynamic vowel reduction over the second half.

These results suggest that each of Korean monophthongs has its own centralizing pattern over the second half of vowel duration which constitutes a secondary cue to monophthong identification in addition to static spectral cues at 20% and 50%. It is concluded that Korean monophthong signals can be characterized best by a shared spectral pattern over the course of production: Stable onset, stable central cross section, and centralized offset.

---

[3] AE monophthong signals in spontaneous speech cannot be directly compared to Korean signals in spontaneous speech since they are realized completely reduced in the temporal domain.

4.3 Inverse relationship between length of vowel duration and frequency of vowel reduction at the offset

Talkers tend to minimize their articulatory effort and hence speak less clearly in less formal situations than in more formal situations. Koopmans-van Beinum (1980) reported that vowel reduction is more frequently found in conversational speech. Meunier and Espesser (2011) reported that shorter vowel duration of vowel signals leads to more frequent spectral vowel reduction on the acoustic vowel space, based on the corpus study on vowel signals in conversational French. The shared idea in the literature is that speech production is governed by a principle of economy of effort (Lindblom 1990). Namely, in vowel production, low-cost behaviors are preferred to solutions with higher energy expenditure.

Korean monophthong signals also showed that shorter vowel duration leads to more frequent vowel centralization at the offset, though the approximate vowel target is still reached over the first half of vowel duration. In the graphs in Figure 11 below, the x-axis refers to vowel duration whereas the y-axis to length of spectral vectors from 50% to 80% of vowel duration. Note that longer 50to80 distance indicates the more deviation of vowel signals from the approximate target at the offset.

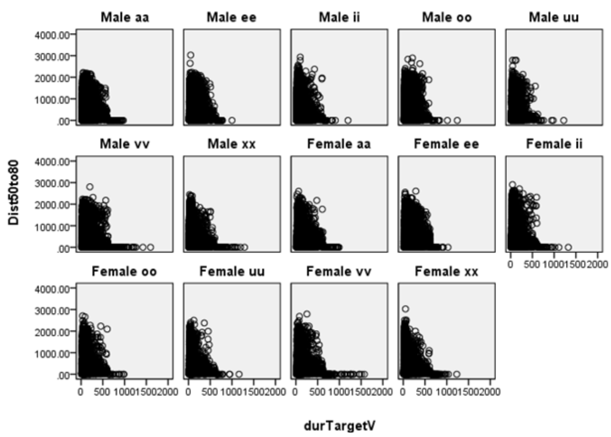


Figure 11. The relationship between length of vowel duration (durTargetV) and length of F1/F2 vectors from 50% to 80% of vowel duration (Dist50to80) for all monophthong types

Figure 11 shows that as monophthong signals became shorter in duration (durTargetV), the more frequent cases of vowel reduction at the offset were observed for all males' and females' monophthongs. Shorter duration of monophthong signals leads to more signals deviating from the approximate target at the offset: Pearson correlation $p<0.01$ for all monophthong types. These results confirmed the shared observation in the literature that the frequency of vowel reduction is correlated with shortened duration.

5. Conclusion

So far, we tried to show that Korean signals of all monophthong types in spontaneous speech shared the unique and systematic spectral pattern over the single temporal domain: Static at the onset and still static at the central cross section, and then abruptly reduced and centralized at the offset. And this spectral pattern occurs over the single temporal domain unlike across-the-board vowel reduction or undershoot over the entire temporal domain observed in conversational speech in English or French. The spectral characteristics of Korean monophthong signals observed in the present study are summarized as follows:

1. Signals of all monophthong types in spontaneous speech were characterized by a unique spectral pattern: Target reached both at the onset and at the central cross section, and then target undershoot abruptly at the offset.

2. The abrupt vowel reduction as the result of vowel target undershoot at the offset, was realized as spectral centralization on the acoustic vowel space.

3. Vowel reduction at the offset occurred more frequently over the shorter temporal domain.

References

ASSMANN, PETER F., TERRANCE M. NEAREY and JOHN T. HOGAN. 1982. Vowel identification: Orthographic, perceptual, and acoustic aspects. *The Journal of the Acoustical Society of America* 71, 975-989.

BOERSMA, PAUL and DAVID WEENINK. 2018. Praat: Doing phonetics by computer (Version 6.0.37) [Computer program]. Retrieved 14 March 2018 from http://www.praat.org.

DEMSAR, JANEZ, TOMAZ CURK, ALES ERJAVEC, CRT GORUP, TOMAZ HOCEVAR, MITAR MILUTINOVIC, MARTIN MOZINA, MATIJA POLAJNAR, MARKO TOPLAK, ANZE STARIC, MIHA STAJDOHAR, LAN UMEK, LAN ZAGAR, JURE ZBONTAR, MARINKA ZITNIK and BLAZ ZUPAN. 2013. Orange: Data mining toolbox in python. *Journal of Machine Learning Research* 14, 2349-2353.

DICANIO, CHRISTIAN, HOSUNG NAM, JONATHAN D. AMITH, REY C. GARCÍA and DOUGLAS H. WHALEN. 2015. Vowel variability in elicited versus spontaneous speech: evidence from Mixtec. *Journal of Phonetics* 48, 45-59.

FOURAKIS, MARIOS. 1991. Tempo, stress, and vowel reduction in American English. *The Journal of the Acoustical Society of America* 90, 1816.

GENDROT, CEDRIC and MARTINE ADDA-DECKER. 2005. Impact of duration on F1/F2 formant values of oral vowels: An automatic analysis of large broadcast news corpora in French and German. *INTERSPEECH* 2005. Lisbon, Portugal. September 4-8, 2005.

____________________ 2007. Impact of duration and vowel inventory size on formant values of oral vowels: An automated formant analysis from eight languages. *ICPhS XVI*. Saarbrücken, Germany. August 6-10, 2007.

GUENTHER, FRANK H. 1995. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review* 102, 594-621.

HILLENBRAND, JAMES M. 2013. Static and dynamic approaches to vowel perception. In Geoffrey S. Morrison and Peter F. Assmann (eds.). *Vowel Inherent Spectral Change*, 9-30. Berlin-Heidelberg: Springer Verlag.

HILLENBRAND, JAMES M., MICHAEL J. CLARK and TERRANCE M. NEAREY. 2001. Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America* 109, 748-763.

HILLENBRAND, JAMES M., LAURA A. GETTY, MICHAEL J. CLARK and KIMBERLEE WHEELER. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America* 97, 3099-3111.

HILLENBRAND, JAMES M. and TERRANCE M. NEAREY. 1999. Identification of resynthesized /hVd/ utterances: Effects of formant contour. *The Journal of the Acoustical Society of America* 105, 3509-3523.

HONG, SOONHYUN. 2019. Perceptual influence of non-spectral cues on spectral properties of Korean vowel signals in Seoul Corpus: A neural network modeling study. *Studies in Modern Grammar* 102, 135-164. The Society of Modern Grammar.

_______________. 2020. The relative contributions of non-spectral cues to static and dynamic spectral model identification of Korean monophthong signals in Seoul Corpus. *Studies in Phonetics, Phonology, and Morphology* 26.1, 159-184. The Phonology-Morphology Circle of Korea.

HONG, SOONHYUN and JONGPYO HONG. 2019. Half static and half dynamic spectral characteristics of Korean vowels in Seoul Corpus. *Proceedings of Spring Conference 2019 at Daegu Catholic University*. The Society of Modern Grammar.

KOOPMANS-VAN BEINUM, FLORINA J. 1980. *Vowel Contrast Reduction, an Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions*. PhD. Dissertation. University of Amsterdam.

LINDBLOM, BJÖRN. 1963. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35, 1773-1781.

_________________ 1983. Economy of speech gestures. In Peter F. MacNeilage (ed.). *The Production of Speech*, 217-245. New York: Springer-Verlag.

_________________ 1990. Explaining phonetic variation: A sketch of the H&H theory. In William J. Hardcastle and Alain Marchal (eds.). *Speech Production and Speech Modeling*, 413-415. Dordrecht: Kluwer.

LINDBLOM, BJÖRN and MICHAEL STUDDERT-KENNEDY. 1967. On the role of formant transitions in vowel recognition. *The Journal of the Acoustical Society of America* 42, 830-843.

MEUNIER, CHRISTINE and ROBERT ESPESSER. 2011. Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics* 39, 271-278.

MOON, SEUNG-JAE and BJÖRN LINDBLOM. 1994. Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical Society of America* 96, 40-55.

NEAREY, TERRANCE and PETER F. ASSMANN. 1986. Modeling the role of vowel inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America* 80, 1297-1308.

VAN SON, ROB and LOUIS C. POLS. 1992. Formant movements of Dutch vowels in a text, read at normal and fast rate. *The Journal of the Acoustical Society of America* 92, 121-127.

YOON, TAE-JIN. 2019. Two-layer neural network-based vowel classification experiments using formant trajectory. *Studies in Phonetics, Phonology and Morphology* 25.1, 95-112. The Phonology-Morphology Circle of Korea.

YUN, WEONHEE, KYUCHUL YOON, SUNWOO PARK, JUHEE LEE, SUNGMOON CHO, DUCKSOO KANG, KOONHYUK BYUN, HYESEUNG HAHN and JUNGSUN KIM. 2015. The Korean corpus of spontaneous speech. Daegu: Industry-Academic Cooperation Foundation, Keimyung University (Distributor).

ZAHORIAN, STEPHEN A. and AMIR J. JAGHARGHI. 1993. Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America* 94, 1966-1982.

Soonhyun Hong (Professor)
Department of English Language and Literature
Inha University
100 Inha-ro, Michuhol-gu,
Incheon 22212 Republic of Korea
e-mail: shong@inha.ac.kr

Received: September 29, 2020.
Revised: November 28, 2020.
Accepted: December 4, 2020.