



Robust measurement of English vowel formants in stressed monophthongs: A comparison of single- and multi-point methods

Hyunsong Chung*

Department of English Education, Korea National University of Education, Chungbuk, Korea

Abstract

This study investigates the most robust method for measuring vowel formant frequencies in stressed monophthongs. Five approaches were compared using the Aix-MARSEC database: measurement at the midpoint (Mid), at the one-third point (OneThird), at three equally spaced points (Three), across all points between one-third and two-thirds of the interval (MidAll), and across all points throughout the vowel interval (Full). Measurements of F1Bark, F2Bark, and Euclidean distances of individual vowels from the centroid of each vowel were analyzed using standard deviations, correlation tests, and linear mixed-effects regression models. Results indicate that the Full method—averaging formants at 10-ms time steps across the entire vowel interval—consistently yields the most stable and reliable measurements. This robustness is evident once vowels following /j/, /w/, or /ɹ/, and those preceding /l/, are excluded. In this study, outliers are included or removed, and only vowels exceeding 40 ms in duration are retained. The findings challenge the traditional reliance on single-point measurements and demonstrate that multi-point analysis, particularly the Full method, offers a more reliable representation of vowel formant frequencies. These results highlight the importance of comprehensive temporal sampling in acoustic phonetics.

Keywords: Euclidean distance, Bark, F1, F2, formant frequency

1. Introduction

This study seeks to identify the method that provides the most robust measurement of vowel formant frequencies in monophthongs produced in stressed syllables. Prior research has typically assessed vowel formants either at the midpoint (Ferragne & Pellegrino, 2010; Harrington, 2010; Harrington et al., 2000; Renwick & Ladd, 2016) or at the one-third point (Yang, 1996) of the vowel interval, positions traditionally regarded as the most stable portions of the vowel. A single midpoint is often chosen because it minimizes the

influence of surrounding sounds. The vocal tract is often maximally open (F1) for non-high vowels, and F2 often reaches a maximum in /i/ or a minimum in /u/ near the temporal midpoint (Harrington, 2010). Measurements at the onset or offset of the vowel interval have generally been avoided due to potential coarticulatory influences from adjacent segments. Ideally, the steady-state point of a vowel should be visually identified. However, this approach is impractical for large corpora, as it is difficult to maintain consistency in identifying steady states within dynamic formant movements, and the procedure does not readily lend itself to

* hchung@knue.ac.kr, Corresponding author

Received 16 August 2025; Revised 5 September 2025; Accepted 8 September 2025

© Copyright 2025 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

automation.

Renwick & Ladd (2016) reduced formant tracking errors by setting different numbers of formants and different maximum formant frequencies for male and female speakers: 4,500 Hz with five formants for males, and 5,000 Hz with four formants for females. Yang (2022) notes that when extracting lower formant frequencies, setting the number of formants to five increases the likelihood of misidentifying two formants as one, particularly for rounded back vowels.

Bauer (1985), by contrast, measured multiple points at 10-ms intervals throughout the steady-state portion of vowels to ensure a more reliable picture of formant frequencies.

When comparing monophthongs with diphthongs, multiple measurements are often taken at 20% and 80%, and sometimes also at 50% of vowel duration.

The objective of the present study is to evaluate the relative reliability and validity of single-point approaches by comparing them with alternative methods that incorporate multiple points, or all points, across the vowel interval in formant frequency measurement.

2. Database

In this study, the Aix-MARSEC database (Auran et al., 2021) was used for analysis. This database is an expanded version of the Spoken English Corpus (SEC), which contains approximately five and a half hours of recordings from BBC broadcasts. The SEC initially collected recordings from 17 male speakers and 36 female speakers in the 1980s. The Aix-MARSEC database currently provides sound files for 29 male speakers and 10 female speakers, along with corresponding TextGrid files containing eight annotation tiers: phonemes, syllables, feet, rhythm units, words, intonation units, intonation patterns, and target pitches (Figure 1).

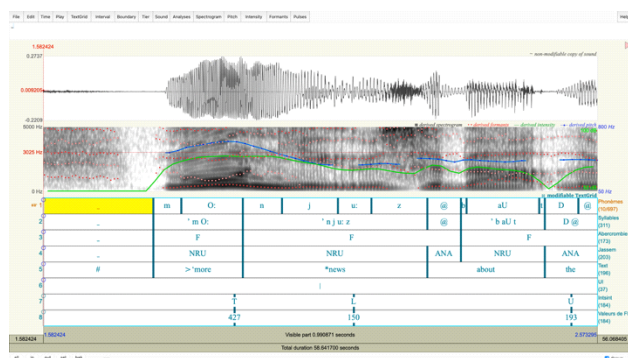


Figure 1. Soundwave and TextGrid of the Aix-MARSEC database.

Because multi-level annotations were manually produced in the Aix-MARSEC project, no additional processing was required for this study. We used the tiers containing information on phonemes and syllables. Phonemes are transcribed in Speech Assessment Methods Phonetic Alphabet (SAMPA; Wells, 1997). A Praat script (Appendix 1) was written to obtain Bark values of the first formant frequency (F1Bark) and the second formant frequency (F2Bark) at several time points: the midpoint (Mid), the one-third point (OneThird), three equally spaced points between one-third and

two-thirds (Three), all points between one-third and two-thirds (MidAll), and all points from the beginning to the end of the vowel interval (Full). If multiple points were used, the average was extracted. Diphthongs were excluded from the analysis, as their formant frequencies fluctuate over time. Table 1 summarizes the measurement methods used in this study.

Table 1. Methods for measuring F1Bark and F2Bark

Method	Description
Mid	Midpoint of the vowel
OneThird	One-Third point of the vowel
Three	Average of one-third, midpoint, and two-thirds points of the vowel
MidAll	Average of all points at 10-ms time steps between one-third and two-thirds
Full	Average of all points at 10-ms time steps from the start to the end of the vowel

In the script, a list was generated to read all sound and TextGrid files in a specified folder using the “Create Strings as file list” function in Praat. Then, in a loop, the sound files and their corresponding TextGrids were imported one by one. Each sound file was converted into a formant object using the “To Formant (burg)” function. Following Renwick & Ladd (2016), different formant settings were implemented in the script. For male speakers, five formant frequencies were extracted with a frequency ceiling of 4,500 Hz, whereas for female speakers, four formant frequencies were extracted with a frequency ceiling of 5,000 Hz, as shown in Table 2.

Table 2. Formant settings based on speaker sex

```
selectObject: mySound

# Formant settings based on the speaker's sex
if inFolder$ = "M"
  myFormant = To Formant (burg): 0.01, 5, 4500, 0.025, 50
elseif inFolder$ = "F"
  myFormant = To Formant (burg): 0.01, 4, 5000, 0.025, 50
endif
```

Within an embedded loop, only monophthongs were selected, and the time interval for each vowel was extracted. From these intervals, vowel and syllable labels, formant frequencies (in Bark), and vowel durations were obtained. No undefined values were encountered in the measurements. The syllable tier in the Aix-MARSEC database also encodes stress information. For example, the word “good” bears primary stress on [ˈgʊd], whereas “unification” carries secondary stress on [ˌjʊnɪfɪkəˈʃən]. This information was used to identify vowels in stressed syllables.

All output files were then merged, and only stressed vowels with a duration greater than 40 ms were retained. Because at least three time points were required to calculate formants for multi-point methods, vowels shorter than 40 ms were excluded, given the 10-ms time step of the formant analysis. To minimize coarticulatory effects, vowels following /j/, /w/, or /ɹ/ and those preceding /l/ were also excluded (Deterding, 1997; Harrington et al., 2000).

Table 3. Number of vowel tokens in the two datasets

Vowel	With outliers		Without outliers	
	M	F	M	F
i	1,320	477	1,186	432
ɪ	1,605	663	1,463	616
ɛ	2,057	850	1,896	807
æ	1,368	558	1,229	508
ɑ	822	382	766	350
ɒ	1,046	417	972	386
ɔ	967	296	897	277
ʊ	188	82	182	75
u	403	183	384	164
ʌ	1,216	521	1,117	472
ɜ	501	193	470	178
Total	11,493	4,622	10,562	4,265

Two datasets were prepared to calculate the means of F1Bark, F2Bark, and the Euclidean distances of individual vowels from the centroid of each vowel by sex. The first dataset included outliers, while the second excluded them. This allowed us to assess whether extreme formant values could influence the results. Harrington (2010) manually identified and removed outliers in their study; in contrast, we removed outliers using z-score filtering (± 2.5 SD) applied to each vowel for each speaker. This threshold covers about 98.76% of data points, leaving about 1.24% as potential outliers (Pollet & van der Meij, 2017), and thus provides a good balance. Outliers in any measurement were removed across all data, ensuring that the same tokens were analyzed regardless of method. Finally, the SAMPA transcriptions were converted into IPA symbols. Table 3 presents the number of vowel tokens in the two datasets.

3. Results

3.1. F1Bark and F2Bark

Tables 4–7 present the means and standard deviations of F1Bark and F2Bark for each vowel in datasets without outliers.

Table 4. F1Bark of each vowel produced by male speakers for each method in the dataset without outliers

Vowel	Mid		OneThird		Three		MidAll		Full	
	F1	SD	F1	SD	F1	SD	F1	SD	F1	SD
i	3.22	.47	3.39	.64	3.33	.54	3.25	.51	3.56	.75
ɪ	4.21	.89	4.24	1.02	4.23	.92	4.21	.88	4.23	.93
ɛ	5.33	.94	5.27	1.01	5.29	.93	5.31	.92	5.19	.95
æ	6.64	1.24	6.61	1.16	6.62	1.12	6.62	1.15	6.36	1.07
ɑ	6.56	1.17	6.54	1.17	6.55	1.08	6.54	1.09	6.31	1.07
ɒ	5.89	1.15	5.84	1.19	5.85	1.13	5.86	1.12	5.71	1.12
ɔ	4.74	.80	4.72	.83	4.73	.78	4.75	.78	4.77	.84
ʊ	4.48	.98	4.43	.94	4.45	.92	4.46	.94	4.40	.86
u	3.65	.96	3.75	.98	3.72	.83	3.66	.81	3.83	.88
ʌ	6.38	1.15	6.23	1.15	6.28	1.08	6.35	1.08	6.06	1.05
ɜ	5.14	.71	5.10	.82	5.11	.71	5.14	.73	5.03	.79

Table 5. F1Bark of each vowel produced by female speakers for each method in the dataset without outliers

Vowel	Mid		OneThird		Three		MidAll		Full	
	F1	SD	F1	SD	F1	SD	F1	SD	F1	SD
i	3.30	.35	3.37	.39	3.35	.36	3.32	.36	3.70	.72
ɪ	4.11	.59	4.11	.62	4.11	.59	4.11	.57	4.18	.74
ɛ	5.61	.96	5.48	.95	5.52	.92	5.59	.93	5.36	.88
æ	7.40	1.25	7.39	1.17	7.40	1.14	7.38	1.20	7.05	1.12
ɑ	7.10	1.12	7.02	1.02	7.04	.97	7.11	1.01	6.84	.89
ɒ	6.68	1.37	6.62	1.40	6.64	1.29	6.65	1.27	6.56	1.14
ɔ	5.13	1.23	5.05	1.18	5.08	1.10	5.13	1.08	5.26	.91
ʊ	4.58	1.35	4.39	1.21	4.45	1.19	4.53	1.20	4.55	1.05
u	3.55	.55	3.61	.63	3.59	.49	3.56	.45	3.82	.61
ʌ	6.88	1.51	6.74	1.47	6.79	1.43	6.84	1.47	6.50	1.36
ɜ	5.70	.92	5.62	1.02	5.65	.90	5.67	.86	5.48	.85

Table 6. F2Bark of each vowel produced by male speakers for each method in the dataset without outliers

Vowel	Mid		OneThird		Three		MidAll		Full	
	F2	SD	F2	SD	F2	SD	F2	SD	F2	SD
i	13.14	1.37	13.05	1.26	13.08	1.18	13.10	1.17	12.82	.97
ɪ	11.94	1.11	11.86	1.16	11.89	1.10	11.92	1.09	11.90	.98
ɛ	11.60	1.08	11.56	1.08	11.58	1.01	11.59	1.02	11.57	.92
æ	11.20	.97	11.22	.90	11.21	.86	11.20	.88	11.20	.77
ɑ	9.71	1.38	9.82	1.36	9.78	1.29	9.74	1.31	9.95	1.22
ɒ	9.15	1.59	9.32	1.55	9.27	1.51	9.17	1.51	9.43	1.42
ɔ	8.16	2.50	8.33	2.44	8.27	2.37	8.20	2.34	8.66	2.02
ʊ	9.85	1.48	9.66	1.44	9.72	1.36	9.84	1.39	9.93	1.21
u	10.30	1.29	10.62	1.21	10.51	1.19	10.32	1.23	10.56	1.09
ʌ	10.14	.92	10.25	.96	10.21	.88	10.16	.88	10.24	.91
ɜ	10.76	.64	10.76	.71	10.76	.63	10.76	.63	10.80	.64

Table 7. F2Bark of each vowel produced by female speakers for each method in the dataset without outliers

Vowel	Mid		OneThird		Three		MidAll		Full	
	F2	SD	F2	SD	F2	SD	F2	SD	F2	SD
i	14.82	.55	14.74	.56	14.77	.52	14.79	.53	14.51	.46
ɪ	13.34	.70	13.29	.68	13.31	.68	13.33	.69	13.34	.64
ɛ	12.89	.71	12.88	.75	12.89	.72	12.89	.69	12.89	.65
æ	11.89	.63	11.92	.65	11.91	.60	11.91	.60	12.02	.59
ɑ	10.32	1.07	10.37	1.00	10.36	.95	10.34	.96	10.75	.83
ɒ	10.42	1.96	10.47	1.83	10.45	1.71	10.44	1.76	11.00	1.41
ɔ	9.75	3.04	9.47	2.77	9.57	2.62	9.77	2.53	10.35	1.87
ʊ	11.07	1.61	10.84	1.48	10.91	1.43	11.08	1.55	11.29	.96
u	11.45	1.35	11.80	1.14	11.68	1.12	11.49	1.23	11.89	.99
ʌ	10.84	.94	10.98	.92	10.93	.88	10.86	.92	11.12	.88
ɜ	11.75	.46	11.83	.65	11.80	.52	11.79	.44	11.92	.54

Figure 2 illustrates the ellipses of each vowel within 2.5 standard deviations in the Full method. The letter representing each vowel marks the centroid of that vowel. Euclidean distances were calculated from this centroid to each token of the corresponding

Table 12. Lmer results of contrasts between methods on F1Bark in the dataset without outliers

Contrast	M			F		
	estimate	z-ratio	p-value	estimate	z-ratio	p-value
Mid-OneThird	.013	1.206	.748	.057	2.849	.036
Mid-Three	.009	.805	.929	.038	1.897	.319
Mid-MidAll	.008	.715	.953	.013	.649	.967
Mid-Full	.082	7.377	.000	.104	5.218	.000
OneThird-Three	-.004	-.401	.995	-.019	-.952	.876
OneThird-MidAll	-.005	-.491	.988	-.044	-2.200	.179
OneThird-Full	.069	6.171	.000	.047	2.369	.124
Three-MidAll	-.001	-.090	1.000	-.025	-1.249	.723
Three-Full	.073	6.572	.000	.066	3.321	.008
MidAll-Full	.074	6.662	.000	.091	4.569	.000

In the male dataset without outliers in F1Bark (Table 12), all significant differences were observed only when contrasts were made against Full ($p=.000$). In the female dataset, significant contrasts were identified for Mid–OneThird ($p=.036$), Mid–Full ($p=.000$), Three–Full ($p=.008$), and MidAll–Full ($p=.000$) in F1Bark.

Table 13. Lmer results of contrasts between methods on F2Bark in the dataset without outliers

Contrast	M			F		
	estimate	z-ratio	p-value	estimate	z-ratio	p-value
Mid-OneThird	-.034	-2.260	.158	-.002	-.085	1.000
Mid-Three	-.023	-1.504	.560	-.001	-.054	1.000
Mid-MidAll	-.003	-.187	1.000	-.007	-.295	.998
Mid-Full	-.063	-4.205	.000	-.166	-7.016	.000
OneThird-Three	.011	.757	.943	.001	.030	1.000
OneThird-MidAll	.031	2.073	.232	-.005	-.210	1.000
OneThird-Full	-.029	-1.945	.294	-.164	-6.931	.000
Three-MidAll	.020	1.317	.681	-.006	-.241	.999
Three-Full	-.041	-2.701	.054	-.164	-6.961	.000
MidAll-Full	-.060	-4.018	.001	-.159	-6.721	.000

In the male dataset without outliers, significant differences were observed for Mid–Full ($p=.000$) and MidAll–Full ($p=.001$) in F2Bark (Table 13). In the female dataset, significant contrasts were identified for Mid–Full ($p=.000$), OneThird–Full ($p=.000$), Three–Full ($p=.000$), and MidAll–Full ($p=.000$) in F2Bark.

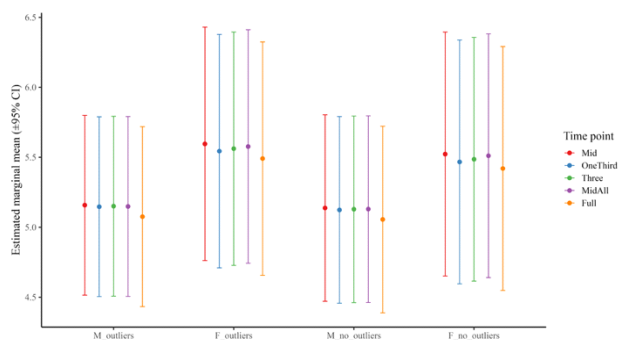


Figure 3. Boxplot of global means of F1Bark by time point.

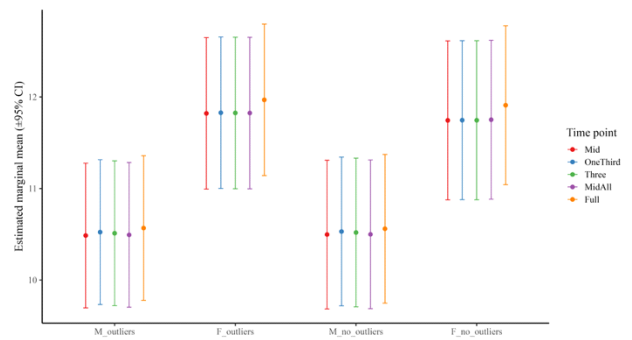


Figure 4. Boxplot of global means of F2Bark by time point.

Figures 3 and 4 illustrate the global means of F1Bark and F2Bark by time point. The results indicate that the majority of significant contrasts involved comparisons with the Full method, and these contrasts were more pronounced in the female dataset in F2 Bark.

3.2. Euclidean distances

Using the outputs of F1Bark and F2Bark measurements, we also calculated the Euclidean distances of each vowel within the same speaker (Chung, 2024). Euclidean distance has been widely used to measure vowel dispersion from the centroid of the vowel system (Chung, 2024; Deterding, 1997; Harrington, 2010).

To accomplish this, the centroid of each vowel was first calculated for every speaker. The Euclidean distance from each individual vowel token to the centroid (the mean of F1Bark and F2Bark for that vowel) was then computed. Because these distances represent deviations from the group means, shorter distances indicate greater consistency in vowel production. Table 14 presents the global means of Euclidean distances for each method by sex.

Table 14. Global means and standard deviations of Euclidean distances for each method

Method	With outliers				Without outliers			
	M		F		M		F	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mid	1.064	1.08	1.256	1.19	.914	.89	1.068	.93
OneThird	1.066	1.05	1.213	1.14	.934	.88	1.032	.88
Three	.981	.94	1.134	1.02	.855	.80	.970	.81
MidAll	.972	.93	1.147	1.03	.850	.79	.982	.79
Full	.864	.73	1.031	.76	.786	.64	.916	.63

Across all datasets, the mean Euclidean distances were shortest in the Full method. Consistent with the results for F1Bark and F2Bark, the smallest standard deviations were consistently observed in the Full method. These findings support the view that the Full method is more reliable and robust than other methods, as it exhibits the least variability in measuring the same vowel. Furthermore, the boxplots in Figure 5 illustrate that the Full method generally yields shorter Euclidean distances than other methods.

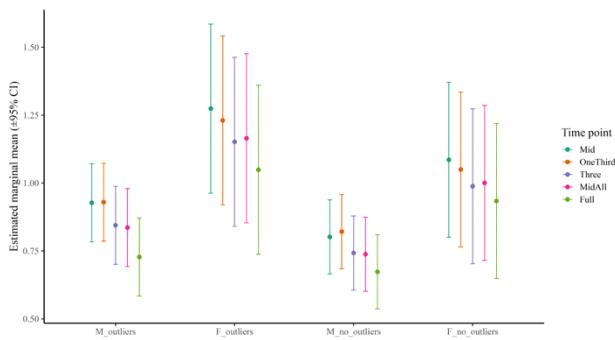


Figure 5. Boxplot of global means of Euclidean distances by time point.

We also examined the global correlations of Euclidean distances between methods. All methods were significantly ($p < .001$) and strongly correlated with one another (Tables 15 and 16).

Table 15. Global correlations of Euclidean distances between methods in the male dataset without outliers

	Mid	OneThird	Three	MidAll
OneThird	.73			
Three	.81	.94		
MidAll	.89	.76	.85	
Full	.70	.69	.76	.79

Table 16. Global correlations of Euclidean distances between methods in the female dataset without outliers

	Mid	OneThird	Three	MidAll
OneThird	.76			
Three	.83	.93		
MidAll	.88	.76	.85	
Full	.63	.64	.69	.71

In the dataset with outliers, the highest correlations were observed between OneThird and Three (M: $r = .94$; F: $r = .94$). The lowest correlations were found between OneThird and Full ($r = .69$) in the male dataset and between Mid and Full ($r = .65$) in the female dataset. In the dataset without outliers, the highest correlations were again observed between OneThird and Three (M: $r = .94$; F: $r = .93$), whereas the lowest were between Mid and Full (M: $r = .70$; F: $r = .63$).

Across both datasets, the Full method was distinct from the other methods in consistently showing lower correlations.

A further linear mixed-effects regression (lmer) analysis was conducted to assess differences in global mean Euclidean distances across methods. In this analysis, Euclidean distance served as the dependent variable, measurement methods as fixed effects, and speakers and vowels as random effects.

In the dataset with outliers, most contrasts were statistically significant, except for Mid–OneThird (M: $p = .999$; F: $p = .170$) and Three–MidAll (M: $p = .944$; F: $p = .965$). In the dataset without outliers (Table 17), a few contrasts were not significant, including Mid–OneThird ($p = .233$) and Three–MidAll ($p = .986$) in the male dataset, as well as Mid–OneThird ($p = .119$) and Three–MidAll ($p = .921$) in the female dataset. All other contrasts were highly

significant.

Table 17. lmer results of Euclidean distances in the dataset without outliers

Contrast	M			F		
	estimate	z-ratio	p-value	estimate	z-ratio	p-value
Mid–OneThird	–.019	–2.070	.233	.036	2.387	.119
Mid–Three	.059	6.276	.000	.098	6.510	.000
Mid–MidAll	.064	6.792	.000	.085	5.677	.000
Mid–Full	.129	13.653	.000	.152	1.129	.000
OneThird–Three	.079	8.346	.000	.062	4.123	.000
OneThird–MidAll	.083	8.863	.000	.049	3.290	.009
OneThird–Full	.148	15.723	.000	.116	7.741	.000
Three–MidAll	.005	.517	.986	–.012	–.833	.921
Three–Full	.069	7.377	.000	.054	3.619	.003
MidAll–Full	.065	6.860	.000	.067	4.452	.000

Across both datasets, contrasts involving the Full method were most frequently significant (14 contrasts in total). The results further indicated that Euclidean distances were shorter in the dataset without outliers than in the dataset with outliers. Within the dataset without outliers, the Full method yielded the shortest Euclidean distances, whereas the longest distances were associated with either the OneThird or Mid method (male and female datasets without outliers; see Table 18). These findings suggest that shorter Euclidean distances reflect reduced variability in formant frequency measurements, while longer distances indicate greater variability.

Table 18. Summary of the order of Euclidean distances of the dataset without outliers, from shortest to longest

Sex	Order
M	Full < MidAll < Three < Mid < OneThird
F	Full < Three < MidAll < OneThird < Mid

4. Discussion and Conclusion

Because the same vowels from the same speakers were analyzed in this study, the variations found are attributable solely to the different measurement methods, aside from inherent differences between vowels.

The analysis of F1Bark and F2Bark across measurement methods showed that all methods are highly correlated overall; however, mixed-effects models revealed systematic differences among them, most frequently in contrasts with the Full method. This suggests that, in principle, formant frequencies can be measured at any point within the vowel interval. Notably, the strongest correlations were observed between the OneThird and Three methods, indicating that whether formants are measured at a single one-third time point or averaged across three equally spaced points between one-third and two-thirds of the interval makes little difference.

However, when linear mixed-effects regression models were applied to assess differences across vowels within sex, the Full method consistently emerged as the most distinct. Contrasts involving the Full method showed stronger and more consistent differences than those involving other methods. These contrasts

were more frequently significant in the female dataset for F2Bark. Because female speakers have shorter vocal tracts and higher fundamental frequencies, the harmonic spacing in their speech signal is wider. This reduces spectral resolution around formant peaks, especially in higher formants, and makes automatic tracking more sensitive to the precise temporal window used. Consequently, differences between single-point and multi-point methods are magnified in the female dataset, since averaging across multiple points helps stabilize estimates in contexts where sparse harmonic sampling can obscure or shift the apparent formant peak. This explains why contrasts between single- and multi-point methods were more prominent in the female dataset for F2Bark. These results indicate that although correlations between methods are generally strong, different approaches may still produce systematically different measurements.

The Euclidean distance analysis further confirmed the robustness of the Full method. Across all datasets, with and without outliers, measuring formants using all points throughout the vowel interval produced the shortest Euclidean distances, suggesting greater consistency. In contrast, the longest distances were observed when using only a single point, either at the midpoint or the one-third point.

Taken together, these results demonstrate that measuring vowel formant frequencies at multiple points, and especially using the Full method, is more effective and reliable than relying on a single measurement point. Earlier assumptions that formant values fluctuate excessively at the boundaries of the vowel interval due to coarticulatory effects were not supported here. Once vowels adjacent to /j/, /w/, /ɪ/, and /l/ were excluded and outliers removed, fluctuations were minimal. Under these conditions, averaging formant values across the entire vowel duration in 10-ms steps provided compact and precise measurements, superior to single-point approaches. Furthermore, advances in computational tools such as Praat scripts now make large-scale multi-point analysis both practical and efficient, a task that would have been far more challenging only a decade ago.

This study is not without limitations. First, the larger number of measurement points in the Full method may have contributed to shorter average Euclidean distances. Although the number of vowel tokens was the same across all methods, the number of measurement points differed substantially. For example, in a vowel lasting 100 ms, the Full method involves eleven measurement points, while the Mid and OneThird methods involve only one, and the Three method involves three. Second, the findings have not yet been tested on external corpora. Future research should replicate this procedure with American English or L2 English corpora to determine whether the robustness of multi-point formant measurement generalizes across languages and speaker populations. Third, we did not compare manual measurements of steady-state formant frequencies with the semi-automatic method used here, primarily due to the size of the corpus. Such a comparison would be valuable for assessing how closely automated methods align with manual measurements. Fourth, it may be inappropriate to generalize the results of this study to all vowels, as preprocessing excluded vowels following /j/, /w/, /ɪ/, as well as those preceding /l/, and vowels shorter than 40 ms. For very short vowels, measurement necessarily relies on a single point.

In conclusion, despite certain limitations, this study provides clear evidence that vowel formant measurements are more consistent and reliable when calculated from multiple points rather

than from a single point, with the Full method proving the most robust. These findings not only refine methodological practices in acoustic phonetics but also highlight the advantages of modern computational tools in handling large-scale, fine-grained phonetic data.

Acknowledgment

We are grateful to the three anonymous reviewers for their insightful and constructive comments on an earlier version of this paper.

References

- Auran, C., Bouzon, C., de Looze, C., & Hirst, D. (2021). Aix-MARSEC database [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage). Retrieved from <https://www.ortolang.fr/market/corpora/sldr000033?lang=en>
- Bauer, L. (1985). Tracing phonetic change in the received pronunciation of British English. *Journal of Phonetics*, 13(1), 61-81.
- Chung, H. (2024). Analysis of monophthongal vowel formants in the Queen's and King's Christmas broadcasts. *Phonetics and Speech Sciences*, 16(4), 43-52.
- Deterding, D. (1997). The formants of monophthong vowels in standard southern British English pronunciation. *Journal of the International Phonetic Association*, 27(1-2), 47-55.
- Ferragne, E., & Pellegrino, F. (2010). Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association*, 40(1), 1-34.
- Harrington, J. (2010). *The phonetic analysis of speech corpora*. Hoboken, NJ: Wiley-Blackwell.
- Harrington, J., Palethorpe, S., & Watson, C. (2000). Monophthongal vowel changes in received pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association*, 30(1-2), 63-78.
- Pollet, T. V., & van der Meij, L. (2017). To remove or not to remove: The impact of outlier handling on significance testing in testosterone data. *Adaptive Human Behavior and Physiology*, 3, 43-60.
- R Core Team. (2025). R: A language and environment for statistical computing (version 4.5.1) [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Renwick, M. E. L., & Ladd, D. R. (2016). Phonetic distinctiveness vs. lexical contrastiveness in non-robust phonemic contrast. *Laboratory Phonology*, 7(1), 1-29.
- Wells, J. C. (1997). SAMPA computer readable phonetic alphabet. In D. Gibbon, R. Moore, & R. Winski (Eds.), *Handbook of standards and resources for spoken language systems: Volume 4: Spoken language reference materials* (pp. 60-107). Berlin, Germany: Mouton de Gruyter.
- Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, 24(2), 245-261.
- Yang, B. (2022). Measuring vowel. In R. A. Knight, & J. Setter (Eds.), *The Cambridge handbook of phonetics* (pp. 261-284). Cambridge, UK: Cambridge University Press.

- **Hyunsong Chung**, Corresponding author
Professor, Dept. of English Education
Korea National University of Education
250 Taeseongtabyeon-ro, Gangnae-myeon, Heungdeok-gu,
Cheongju 28173, Korea
Tel: +82-43-230-3554
Email: hchung@knue.ac.kr

Appendix 1. Praat script for extracting F1Bark and F2 Bark

```

# Extract F1Bark, F2Bark, and the duration of each vowel from all WAV and TextGrid files in a specified folder
# Different formant settings based on the speaker's sex
# Obtain formant values at:
# - Midpoint
# - One-third point
# - Average of one-third, midpoint, and two-thirds points
# - Average of all points from one-third to two-thirds of the interval
# - Average of all points from the start to the end of the interval

form Get F1Mid, F2Mid, F1oneThird, F2oneThird, F1Three, F2Three, F1midAll, F2midAll, F1Full, F2Full
# Specify the paths for infolder, outFolder
sentence inFolder M
sentence outFolder type\your\outfolder\here\
sentence txtName results.txt
integer tier 1
integer syltier 2
endform

# Check whether the output file already exists
if fileReadable(outFolder$ + txtName$) = 1
pause The file already exists. If you click continue it will be overwritten.
endif
echo 'outFolder$'

# Index files for looping
myList = Create Strings as file list: "myList", "your\path\" + inFolder$ + "\" + "*.wav"
nFiles = Get number of strings

# Loop through all files in a folder
for file to nFiles
selectObject: myList
nameFile$ = Get string: file
base$ = nameFile$ - ".wav"
myTextGrid = Read from file: "your\path\" + inFolder$ + "\" + base$ + ".TextGrid"
# base name
myTextGrid$ = selected("TextGrid")
mySound = Read from file: "your\path\" + inFolder$ + "\" + myTextGrid$ + ".wav"
selectObject: myTextGrid
nOfIntervals = Get number of intervals: tier
Convert to Unicode

selectObject: mySound

# Formant settings based on the speaker's sex
if inFolder$ = "M"
myFormant = To Formant (burg): 0.01, 5, 4500, 0.025, 50
elseif inFolder$ = "F"
myFormant = To Formant (burg): 0.01, 4, 5000, 0.025, 50
endif

# Specify the name of the output file
outfile$ = base$ + ".Formant.txt"

# If the output file already exists, overwrite it.
deleteFile: outFolder$ + outfile$

# Write column headings to the output file
writeFileLine: outFolder$ + outfile$, "fileName" + tab$ + "nInterval" + tab$ + "Vowel" + tab$
... + "Syllable" + tab$ + "PreLabel" + tab$ + "PostLabel" + tab$
... + "F1Mid" + tab$ + "F2Mid" + tab$
... + "F1oneThird" + tab$ + "F2oneThird" + tab$
... + "F1Three" + tab$ + "F2Three" + tab$
... + "F1midAll" + tab$ + "F2midAll" + tab$
... + "F1Full" + tab$ + "F2Full" + tab$ + "Duration"

# Loop through all intervals
nInterval=1
for nInterval to nOfIntervals
selectObject: myTextGrid
labelOfInterval$ = Get label of interval: tier, nInterval

# Perform actions only for vowels
if index(labelOfInterval$, "i:") < 0 or
... index(labelOfInterval$, "I") < 0 or
... index(labelOfInterval$, "e") < 0 or
... index(labelOfInterval$, "E") < 0 or
... index(labelOfInterval$, "a:") < 0 or
... index(labelOfInterval$, "A") < 0 or
... index(labelOfInterval$, "o:") < 0 or
... index(labelOfInterval$, "O") < 0 or
... index(labelOfInterval$, "@U") < 0 or
... index(labelOfInterval$, "U") < 0 or
... index(labelOfInterval$, "u:") < 0 or
... index(labelOfInterval$, "@") < 0 or
... index(labelOfInterval$, "v:") < 0 or
... index(labelOfInterval$, "3:") < 0 or
... index(labelOfInterval$, "al") < 0 or
... index(labelOfInterval$, "aU") < 0 or
... index(labelOfInterval$, "OI") < 0 or
... index(labelOfInterval$, "@@") < 0 or
... index(labelOfInterval$, "e@") < 0 or
... index(labelOfInterval$, "U@") < 0

# Retrieve labels of the intervals preceding and following the target vowel
preInterval = nInterval - 1

```

```

postInterval = nInterval + 1

# Assign '_' if preInterval is out of range
if preInterval < 1
preLab$ = "_"
else
preLab$ = Get label of interval: tier, preInterval

# Handle the case where the label is empty
if preLab$ = ""
preLab$ = "_"
endif
endif

# Assign '_' if postInterval is out of range
if postInterval > nOfIntervals
postLab$ = "_"
else
postLab$ = Get label of interval: tier, postInterval

# Handle the case where the label is empty
if postLab$ = ""
postLab$ = "_"
endif
endif

# Extract the timing of the interval
endPoint = Get end point: tier, nInterval
startPoint = Get starting point: tier, nInterval
durInterval = endPoint - startPoint
oneThird = startPoint + durInterval/3
midInterval = startPoint + durInterval/2
twoThird = startPoint + 2*durInterval/3
durIntervals = durInterval*1000
durIntervals$ = fixed$(durIntervals, 0)

# Extract the label of the word
syllInterval = Get interval at time: syltier, midInterval
syllLabel$ = Get label of interval: syltier, syllInterval

# Extract formants
selectObject: myFormant
nPoints = Get number of frames

f1allsum = 0
f2allsum = 0
numUsed = 0
f1midallsum = 0
f2midallsum = 0
nUsed = 0

# Extract formants at the midpoint and one-third of the interval
f1Mid = Get value at time: 1, midInterval, "Bark", "Linear"
f2Mid = Get value at time: 2, midInterval, "Bark", "Linear"
f1oneThird = Get value at time: 1, oneThird, "Bark", "Linear"
f2oneThird = Get value at time: 2, oneThird, "Bark", "Linear"

f1Three = (f1oneThird + f1Mid + f1oneThird)/3
f2Three = (f2oneThird + f2Mid + f2oneThird)/3

f1Mid$ = fixed$(f1Mid, 2)
f2Mid$ = fixed$(f2Mid, 2)
f1oneThird$ = fixed$(f1oneThird, 2)
f2oneThird$ = fixed$(f2oneThird, 2)

f1Three$ = fixed$(f1Three, 2)
f2Three$ = fixed$(f2Three, 2)

# Extract formants across the range from one-third to two-thirds of the interval
for point to nPoints
time = Get time from frame number: point

if time >= startPoint and time <= endPoint
f1all = Get value at time: 1, time, "Bark", "Linear"
f2all = Get value at time: 2, time, "Bark", "Linear"

if f1all < undefined and f2all < undefined
f1allsum = f1allsum + f1all
f2allsum = f2allsum + f2all
numUsed = numUsed + 1
endif
endif

if time >= oneThird and time <= twoThird
f1midall = Get value at time: 1, time, "Bark", "Linear"
f2midall = Get value at time: 2, time, "Bark", "Linear"

if f1midall < undefined and f2midall < undefined
f1midallsum = f1midallsum + f1midall
f2midallsum = f2midallsum + f2midall
nUsed = nUsed + 1
endif
endif
endif
endfor

```

```

if numUsed > 0
    f1allavg = f1allsum / numUsed
    f2allavg = f2allsum / numUsed
else
    f1allavg = 0
    f2allavg = 0
endif

if nUsed > 0
    f1midallavg = f1midallsum / nUsed
    f2midallavg = f2midallsum / nUsed
else
    f1midallavg = 0
    f2midallavg = 0
endif

f1Full$ = fixed$(f1allavg, 2)
f2Full$ = fixed$(f2allavg, 2)
f1midAll$ = fixed$(f1midallavg, 2)
f2midAll$ = fixed$(f2midallavg, 2)

# Write results to a text file
appendFileLine: outFile$, base$ + tab$ + string$(nInterval) + tab$ + labelOfInterval$ +
tab$
... + sylLabel$ + tab$ + preLab$ + tab$ + postLab$ + tab$
... + f1Mid$ + tab$ + f2Mid$ + tab$
... + f1oneThird$ + tab$ + f2oneThird$ + tab$
... + f1Three$ + tab$ + f2Three$ + tab$
... + f1midAll$ + tab$ + f2midAll$ + tab$
... + f1Full$ + tab$ + f2Full$ + tab$ + durIntervals$
endif
# End loop for intervals
endfor
removeObject: myTextGrid, mySound, myFormant

# End loop for files
endfor
removeObject: myList
echo Done.

```