http://dx.doi.org/10.17959/sppm.2021.27.3.463

GAMM-based modeling of dynamic phonetic trajectories*

Tae-Jin Yoon
(Sungshin Women's University)

Yoon, Tae-Jin. 2021. GAMM-based modeling of dynamic phonetic trajectories. *Studies in Phonetics, Phonology and Morphology* 27.3. 463-481. Dynamic trajectories of F0s and formants influence the perception of speech sounds. For this paper, the dynamic trajectories of two calibration sentences read by all participants (more than 600) in the TIMIT database are preprocessed by the general-purpose Python programming language and the third-party packages Parselmouth, TextGridTools, and Pandas. Parselmouth and TextGridTools function as interfaces to the underlying Praat software, which is a de facto standard tool for phonetic research but lacks the functionality of general-purpose computer programs. The data preprocessed through the interfacing packages are formed into a data frame and then fed into the R environment, which still has superior capability to deal with statistical modeling in comparison to Python, to model the dynamic phonetic trajectories using the Generalized Additive Mixed Model (GAMM). The F0 trajectories over all calibration sentences were also inspected across dialects and different education levels. The approach taken in this paper will further our understanding of the phonetic trajectories in the shaping of phonetic and phonological patterns. **(Sungshin Women's University, Associate Professor)**

Keywords: GAMM, Parselmouth, TextGridTools, Pandas, Python interfaces to Praat, Dynamic Phonetic Trajectories, Formants, F0, The TIMIT database

1. Introduction

The purpose of this paper is to examine and model the dynamic properties of speech sounds using a corpus-based approach. The speech sounds that humans use for communication have long been the subject of research under separate disciplines of phonetics and phonology. Speech sounds are an intriguing subject of research that has a duality in that they have both physical and continuous properties on one hand and abstract and categorical properties on the other hand. The 20th century

* This work was supported by the Sungshin Women's University Research Grant of 2021.

phonological study has developed by explicitly or implicitly accepting the distinction between phonetics and phonology (Ladd, 2014). In classical phonology, a postulated phonological presentation consists of phonological features and can be changed by phonological rules and constraints. However, examples of phonological alternations typically viewed as categorical and phonological phenomena have also been challenged in several previous studies (e.g., Nolan, 1992).

Over the years, theories to explain both categorical and gradient phonological phenomena have been proposed. Such alternative theories include the Exemplar theory and Corpus Phonology. Exemplar theory was originally proposed as a methodology that could classify multidimensional stimuli in psychology (Hintzman 1986; Nosofsky, 1986), and the theory has been applied to phonology and has been active since the mid-1990s. In brief, exemplar theory concerns the representation and processing of categories in which stimuli are processed by comparing them to a set of previous exemplars in memory. The learning of statistical distributions in categories is influenced by the frequency and recency of exposure to exemplars and factors such as gender, age, and social categories. Thus, along with the advent of exemplar theory, the need for corpus phonology as a modern phonological research method has increased since corpus phonology can meet the call for being embedded within an extensive framework of social, cognitive, and biological science (Ernestus & Baayen, 2011; Durand, Gut, & Kristoffersen, 2014; Cho et al. 2021).

In this paper, continuing the tradition of linking the phonetic details into phonological patterning using speech corpus, I attempt to construct statistical models with phonetic trajectories of formants and F0s to better understand how much information in dynamic trajectories explain variation in phonological representations. To be specific, I present GAMM-based approaches to statistical modeling of dynamic phonetic trajectories using the TIMIT database (Zue & Seneff, 1996). The TIMIT database has long been used in the realm of speech recognition and phonetics research (cf. Yoon 2019). The corpus is accompanied by detailed meta information. The pieces of information, together with dynamic trajectories of speech samples, can provide a valuable test case for evaluating the validity of exemplar theory. In this paper, I first demonstrate how the acoustic and textual information can be processed, mutated, and merged so that contextual information can be used as input features to statistical modeling. The manipulation will be made using third-party packages such as Parselmouth (Jadoul, Thompson, & de Boer, 2018), TextGridTools (Buschmeier & Włodarczak, 2013), and Pandas (McKinney, 2011). Among the Python packages,

Parselmouth and TextGridTools function as interfaces to the Praat software, which is considered a *de facto* standard tool for linguistic phonetic research. In this paper, I will choose a couple of meta information to build statistical models. As for the statistical modeling, Generalized Additive Mixed Model (or GAMM for short; Wood, 2015) has been applied to the so-called calibration sentences in the TIMIT database, as the two sentences are uttered by all participants (n > 600).

2. Methods

2.1 Data

The current study employs corpus-based studies using the phone-balanced connected speech corpus of TIMIT. The TIMIT database is accompanied by detailed meta information, including broadly defined dialectal areas of participating speakers, gender, age, height, year of birth, and educational background, which can provide a valuable test case for evaluating the validity of exemplar theory. Because of the quantity of speech and the fine-grained segmentation and labeling qualities, the TIMIT database provides an unusual corpus for phonetic research. The TIMIT database includes 630 talkers and 2342 different sentences, comprising over five hours of speech. Among the sentences, two sentences, called calibration sentences, were produced by all participants. The two calibration sentences are presented in (1).

(1) a. She had your dark suit in greasy wash water all year. (SA1)
b. Don’t ask me to carry an oily rag like that. (SA2)

According to Zue & Seneff (1996) and Byrd (1994), the calibration sentences were designed to incorporate phonemes in contexts where significant dialectal differences were anticipated. Byrd (1994) tried to tease apart the dialectal differences by comparing the duration of the two utterances. In this study, I will examine whether the intonational patterns of the two sentences in (1) show differences among dialects. Note that calibration sentence one has 13 syllables, and calibration sentence two has 12 syllables. Thus, two separate models will be constructed for each sentence.

2.2 Acoustic feature extraction

The extraction of features in the TIMIT database was aided by using the Python programming language and its third-party libraries such as Parselmouth (Jadoul et al. 2018) and TextGridTools (Buschmeier & Włodarczak, 2013). The rationale for choosing Python and its third-party libraries lies in their powerful and efficient handling of large-scaled phonetic data.

In the past decades, data analysis in phonetics and phonology research more than often relied on the functionality of Praat (Boersma & Weenink 2018). Praat is an extensive phonetic analysis software package, and it has been widely used by phoneticians and phonologists for analyzing speech sounds. Praat is very useful due to its scripting function, which can significantly facilitate acoustic feature extraction by automatizing repetitive, tedious, and time-consuming tasks of extracting acoustic features (Buschmeier & Włodarczak, 2013; Jadoul et al. 2018).

As neatly explained in Jadoul et al. (2018), although the Praat scripting language is suitable for automating repeated works, it is limited compared to a general-purpose computer programming language such as Python, which can integrate various computational utilities. On the other hand, it is difficult or time-consuming to conduct the analysis of linguistic phonetic studies with the general-purpose programming language. The necessary functionality is "often unavailable or dispersed over multiple unrelated and sometimes incompatible libraries (Jadoul et al. 2018: 2)." In this study, I used Praselmouth (Jadoul et al. 2018), as it provides access to Praat's functionality as well as the full-range computational power of Python, with one shortcoming.

The shortcoming is that Parselmouth has a very limited and primitive capability of dealing with TextGrid files. In linguistic phonetic research, the time-stamped information of speech samples annotated with textual information is routinely used to understand the segmental and suprasegmental nature of speech sounds in a variety of contexts. To cope with the limitation, another Python package TextGridTools (Buschmeier & Włodarczak, 2013) has been utilized together with Parselmouth. TextGridTools is designed to "offer functions to parse, manipulate and query Praat annotations (Buschmeier & Włodarczak, 2013)."

In this study, I extracted from the training portion of the TIMIT database acoustic and textual information using Parselmouth and TextGridTools in the Jupyter Notebook environment, a web-based open-source interactive environment that can be

used to create documents containing Python codes, the output, and visualization, among others. In the Jupyter Notebook, I wrote Python codes to go through each folder and subfolder and select the two calibration files (a pair of wave file and TextGrid file named SA1 and SA2)[1]. When two instances of the SA1 and SA2 files are encountered, TextGrid objects and Sound Objects are instantiated using the functionality of Parselmouth and TextGridTools. F0 and the first two formants are extracted at ten equal points for the duration of each vocalic phone. Text information was also used to get phone and word information, among other things. Part of the processed data in the format of Pandas DataFrame is shown in Figure 1. Note that due to the space limit in the Jupyter Notebook, some columns in the middle were shown to be truncated in the screenshot.

	Dialect	Speaker	Filename	Idx	Sidx	Duration	WDur	StartTime	EndTime	PhoneIndex	...	Pitch	Intensity	F1	F2
0	DR6	MSDS0	SA2	0	0	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
1	DR6	MSDS0	SA2	0	1	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
2	DR6	MSDS0	SA2	0	2	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
3	DR6	MSDS0	SA2	0	3	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
4	DR6	MSDS0	SA2	0	4	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
5	DR6	MSDS0	SA2	0	5	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
6	DR6	MSDS0	SA2	0	6	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
7	DR6	MSDS0	SA2	0	7	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
8	DR6	MSDS0	SA2	0	8	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
9	DR6	MSDS0	SA2	0	9	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
10	DR6	MSDS0	SA2	1	10	121.25	167.81	0.166	0.288	MSDS0ow_1	...	11.91	66.53	6.12	9.93
11	DR6	MSDS0	SA2	1	11	121.25	167.81	0.166	0.288	MSDS0ow 1	...	11.91	66.53	6.12	9.93

Figure 1. Pandas DataFrame of the extract features

The TIMIT database is accompanied by text files containing meta information such as gender, recorded data, birth date, height, race, and education. The files containing these pieces of meta information are processed and converted into another Pandas DataFrame as in Figure 2.

[1] In the TIMIT database, annotated labels are saved as PHN (for time aligned phone information) and WRD (for time aligned word information) files in the format of plain text. These PHN and WRD text files were converted to TextGrids with corresponding filenames using a custom-made Praat script.

	Dialect	Speaker	Filename	Idx	Sidx	Duration	WDur	StartTime	EndTime	PhoneIndex	...	Pitch	Intensity	F1	F2
0	DR6	MSDS0	SA2	0	0	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
1	DR6	MSDS0	SA2	0	1	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
2	DR6	MSDS0	SA2	0	2	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
3	DR6	MSDS0	SA2	0	3	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
4	DR6	MSDS0	SA2	0	4	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
5	DR6	MSDS0	SA2	0	5	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
6	DR6	MSDS0	SA2	0	6	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
7	DR6	MSDS0	SA2	0	7	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
8	DR6	MSDS0	SA2	0	8	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
9	DR6	MSDS0	SA2	0	9	21.56	167.81	0.145	0.166	MSDS0d_0	...	13.54	50.01	6.42	11.65
10	DR6	MSDS0	SA2	1	10	121.25	167.81	0.166	0.288	MSDS0ow_1	...	11.91	66.53	6.12	9.93
11	DR6	MSDS0	SA2	1	11	121.25	167.81	0.166	0.288	MSDS0ow 1	...	11.91	66.53	6.12	9.93

Figure 2. Processed DataFrame of TIMIT metadata

These two DataFrame objects in Figure 1 and Figure 2 are merged to be used for further data analysis. I chose a subset of columns in the merged DataFrame to test whether phonetic trajectories such as F0, F1, and F2 over the duration of different phones show legitimate and expected patterns. Then I tested whether we could find any dialectal differences among the F0 patterns over each calibration sentence.

2.3 Statistical modeling

Dynamic trajectories of F0 and formants are statistically modeled using the Generalized Additive Mixed Model (GAMM; Wood, 2015). GAMM is used due to its outstanding functionality of modeling nonlinearity (Sóskuthy, 2017, 2021; Wieling, 2018). Measurements on human speech often show nonlinear patterns. Formant trajectories and pitch contours are well-known examples of nonlinear patterns. For example, pitch contour typically does not develop linearly over time. F0 contour over a stretch of sentence can be quite fluctuating or wiggly. Unlike a common practice of considering a pre-defined subset of measurements, such as maximum or minimum pitch values, GAMM extends the generalized linear mixed model with a large array of tools for modeling nonlinear dependencies between a response variable and one or more numeric predictors (Wood, 2015). However, there is a shortcoming in nonlinear regression models. In contrast with linear regression models, one cannot interpret the shape of the regression line from the summary in nonlinear regression models. Therefore, visualization is an essential tool for

interpreting nonlinear regression models (van Rij, Wieling, Baayen & van Rijn, 2015).

The Generalized Additive Mixed Model (GAMM) will be used with the help of R and its packages, especially tidyverse (Wickham, 2017), mgcv (Wood, 2015), itsadug (van Rij et al., 2015), because the statistical modeling implemented as R packages can capture the underlying dynamic patterns as well as the effects of random factors[2]. The two packages, mgcv and itsadug, are specifically designed for the GAMM modeling and its visualization. The package mgcv is for estimating penalized Generalized Linear Models and includes the implementation of 'gam' (generalized additive model) that contains automatic smooth estimation using penalized regression splines (Wood, 2015). An additional package itsadug provides a set of useful functions that help facilitate the evaluation, interpretation, and visualization of GAMM models constructed via the mgcv package (van Rij et al., 2015).

The tidyverse library is a whole suite of packages that include widely used ggplot2, dplyr, and tidyr packages for data manipulation and visualization in the R environment (R Core Team, 2019). Instead of loading each package individually, I ran the command library(tidyverse) in the Jupyter Notebook environment.

3. Results

3.1 Phone duration

In this section, I present the results of analysis and visualization of vowel duration, vowel formant, and F0 trajectories over the range of vowel duration, and then the analysis of F0 contours over the whole calibration sentences (SA1 and SA2).

The data points for each vowel are not limited to only one or two time points but extended to 10 time-normalized points. Table 1 shows the number of tokens for each phone observed in the two calibration sentences SA1 and SA2. Note that since the transcription in the TIMIT database is made for each token of the utterance, the number of vowels is not evenly balanced.

[2] My experience with the two languages of Python and R is that though there are packages that connects between R and Python (e.g., rpy2), Python is limited in fully entertaining the statistical modeling power of R. I had to change to the R environment for the GAMM-based statistical modeling.

Table 1. Number of vowel tokens (in parentheses) of the calibration sentences in the training portion of the TIMIT database

[ɑ] "aa" (776)	[æ] "ae" (1515)	[ʌ] "ah" (37)	[ɔ] "ao" (980)	[aʊ] "aw" (1)
[ə] "ax" (67)	[ɚ] "axr" (866)	[ɛ] "eh" (521)	[ɜɹ] "er" (320)	[e] "ey" (11)
[ɪ] "ih" (725)	[ɨ] "ix" (1147)	[i] "iy" (2076)	[o] "ow" (427)	[ɔɪ] "oy" (339)
[ʊ] "uh" (33)	[u] "uw" (25)	[u] "ux" (437)	[aɪ] "ay" (405)	

Figure 3 is a bar plot that illustrates the mean duration of each phone. Keating et al. (1994) reported that TIMIT distinguished between full and reduced vowel qualities, with reduced [ə] ("ax") and [ɨ] ("ix") used only for very short vowels of unstressed quality. Figure 3 confirms that the [ə] and [ɨ] are the reduced vowels as observed in the calibration sentences. The figure also illustrates that the diphthongs [aɪ] ("ay"), [ɔɪ] ("oy") and [aʊ] ("aw") together with the low front vowel [æ] ("ae") exhibit the longer phone duration than the other vowels. A longer duration of [æ] is expected, given that the low front vowel has inherently long vowel duration associated with the physiological factor of more extended jaw opening and closing.

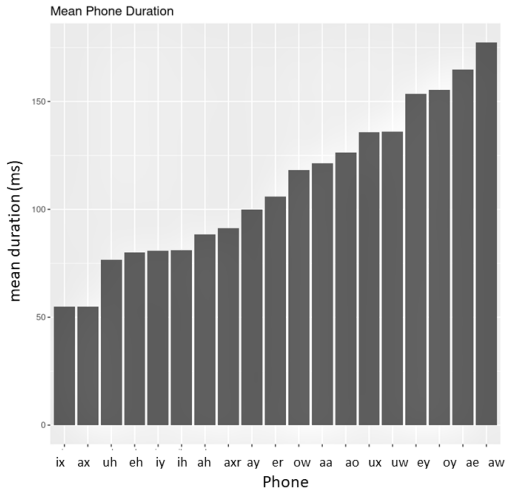

Figure 3. Mean phone duration

3.2 Formants

Figure 4 illustrates raw F1 & F2 contour plots for each of the vowel types observed in the calibration sentences. The measurement unit is converted from Hertz to Bark.

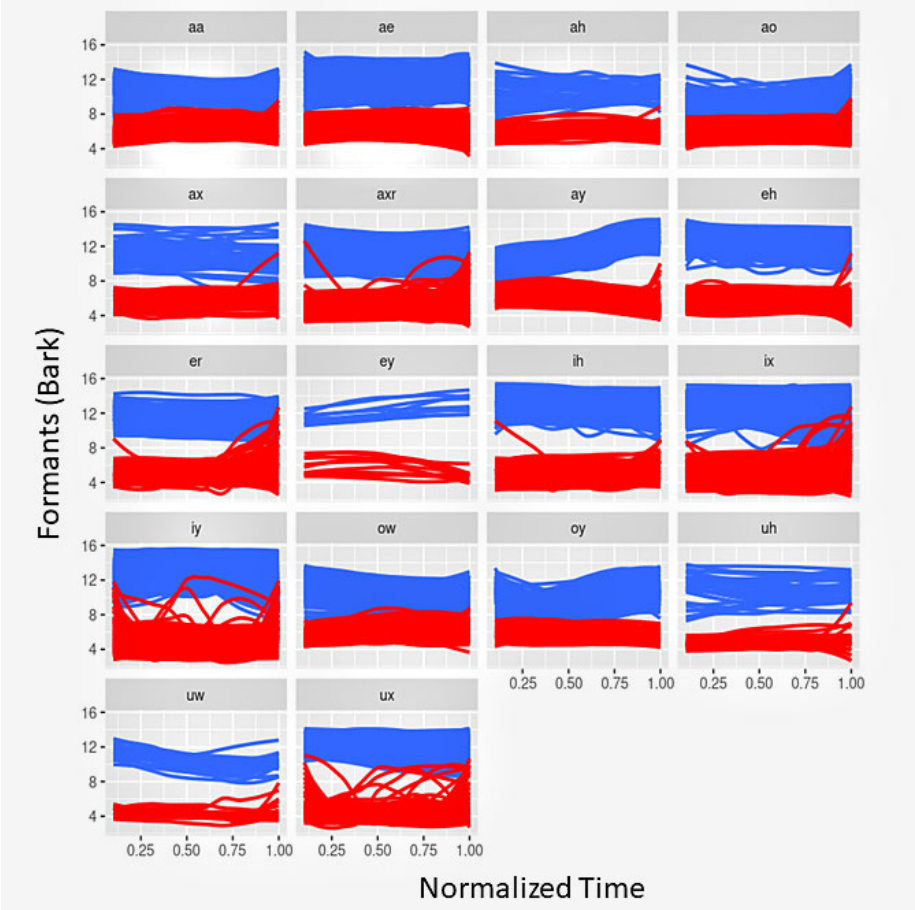

Figure 4. Raw F1 and F2 trajectories

The raw formant values were used as an input to the GAMM model. The predicted variables are F1 and F2 formant trajectories, respectively. The explanatory variables

are phones and gender, together with formant values extracted from ten evenly spaced time points. The ten data points were subject to a cubic spline basis function.

Overlaid formant contours over the normalized time intervals are shown in Figure5. Figure 5 illustrates the F1 and F2 formant contours predicted by the model. Due to the sheer number of phones, instead of a pairwise comparison of differences between pairs of possible phones, only r-squared values and deviance explained by the model are presented in this paper. As for the F1 trajectories, the adjusted r-squared value is 0.614, meaning that the model can explain 61.4% of the deviance[3]. As for F2, the adjusted r-squared value is 0.734, meaning that the model can explain 73.4% of the deviance[4].

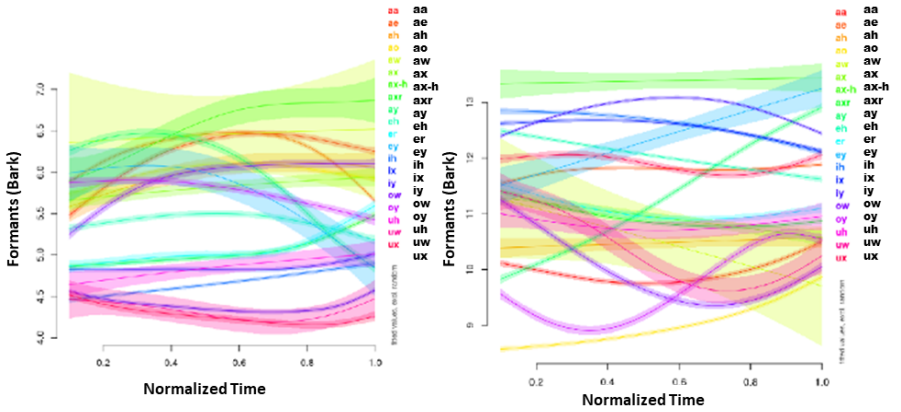

Figure 5. F1 and F2 formant contours predicted by the model

3.2.1 Visualization of formant trajectories with F1 and F2 combined

The functionalities of GAMM seem to lack a way of modeling multivariate outcomes (such as F1 and F2 values as a simultaneous outcome variable). Alternatively, I made

[3] The GAMM model for predicting F1 trajectories:

```
timit_phone_F1.gam <- bam(F1s ~ Phone  + Gender + s(FIndex2,
  by=Phone, bs="cr"), data=timit_v, method="ML")
```

[4] The GAMM model for predicting F2 trajectories:

```
timit_phone_F2.gam <- bam(F2s ~ Phone  + Gender + s(FIndex2,
  by=Phone, bs="cr"), data=timit_v, method="ML")
```

a combined contour plot for each phone with the predicted formant trajectories. In Figure 6, the first panel is for front vowels and the second panel for back vowels.

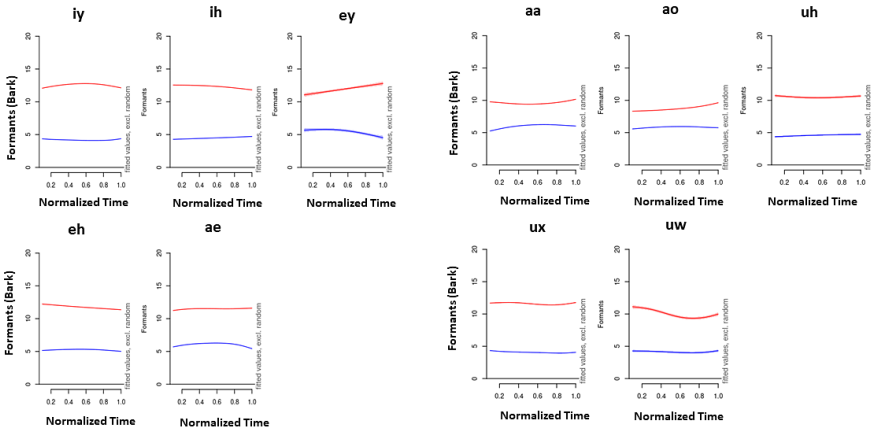

Figure 6. GAMM-Modeled F1 and F2 contours (front vowels in the left panel and back vowels in the right panel)

Figure 7 shows the formant trajectories for diphthongs. As for the diphthongs, [aʊ] "aw" and [oʊ] "ow" show a trend of f2 lowering over time, and [ɔɪ] "oy" shows the F2 rising patterns.

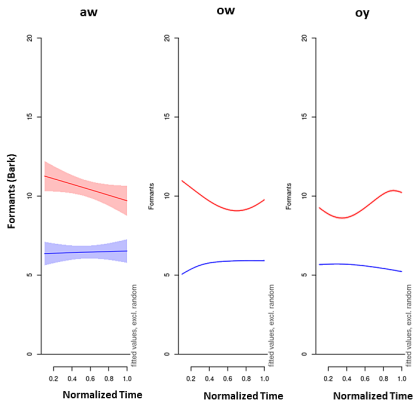

Figure 7. GAMM-Modeled F1 and F2 contours for diphthongs

3.3 F0 contours (in semitone)

Figure 8 illustrates the F0 contours over each phone. Since prosodic contexts in which the phones were located were not controlled, segmental effects on F0 values will not be discussed. Nevertheless, it is worth noting that the reduced vowels ("ax" and "ix") exhibit F0 contours in the lowest F0 ranges, which is expected given the lack of stress on the reduced vowels.

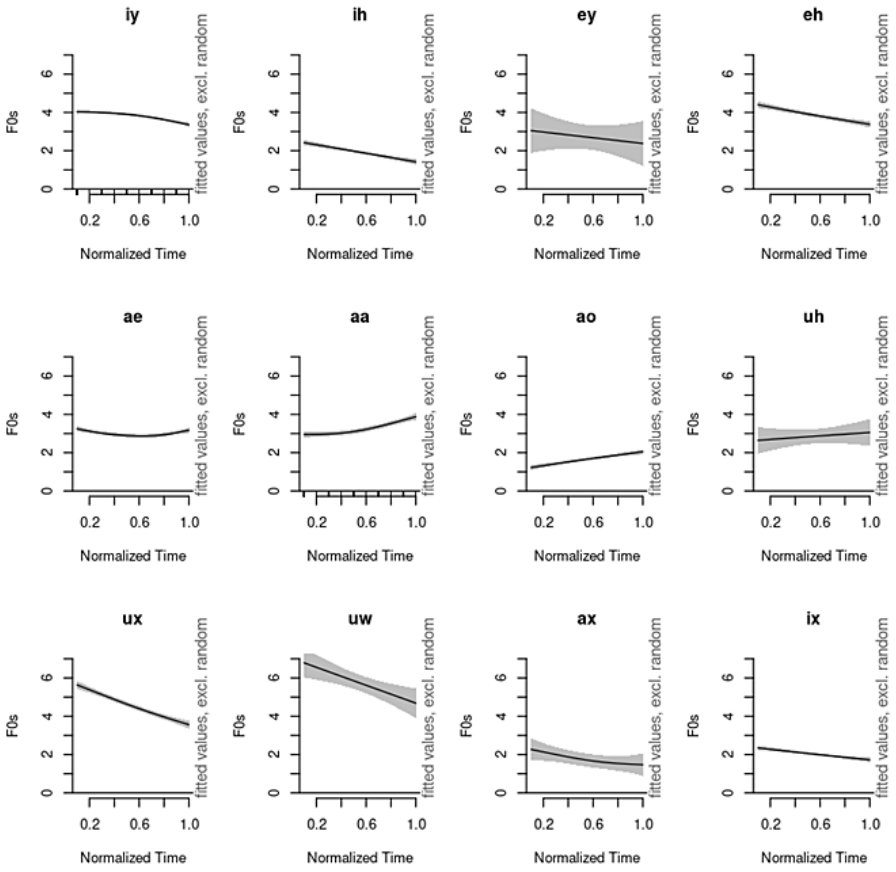

Figure 8. GAMM-modeled F0 contours for each phone

3.4 Sentences

The two calibration sentences were said to be designed to "incorporate phonemes in contexts where significant dialectal differences are anticipated" (Byrd, 1994). According to Zue & Seneff (1996), these two sentences were "designed by Jared Bernstein of SRI in order to compare dialectal and phonological variations across speakers. (p. 515)" Broadly speaking, seven geographic dialect region are regions of North, North Midland, North East, NY City, South, South Midland, and West (Byrd, 1994)." An additional dialect division called "Army Brat" was in the TIMIT database, but it is excluded from further analysis due to the small number of speakers. Byrd (1994) found that the duration of both calibration sentences (or speaking rate) significantly affected the dialect region. She found that "dialects range from slowest to fastest in the order of South, South Midland, NY City, North, West, North Midland, and North East."

In this paper, I'd like to see whether dialectal differences can be observed for the intonational patterns over utterances. To answer this question, I formulated the GAMM model as in (2) for each of the two calibration sentences. In (2), the formula is for SA1. In the formula, F0 contours for SA1 are predicted with explanatory variables of dialect and gender. The normalized time (as indicated by Sidx2) is subject to a cubic spline basis function alone and depending on dialects.

(2)
```
timit_SA1.gam.diff <- bam(F0s ~ DialectNo.ord +
       Gender + s(Sidx2, bs="cr") +
       s(Sidx2,by=DialectNo.ord, bs="cr"),
       data=timit_v_SA1, method="ML")
```

Table 2 shows the parametric coefficients for the GAMM model for SA1. The first sentence SA1 results in that the baseline (New England) dialect shows that its F0 contour is significantly different from the F0 patterns except for dialect 3 (the Northern Midwest) and dialect 7 (the Western United States).

Table 2. Parametric coefficients for the GAMM-model for the SA1 sentence

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	11.17126	0.05010	222.989	< 2e-16	***
DialectNo.ord2	0.20847	0.05594	3.727	0.000194	***
DialectNo.ord3	-0.06251	0.05619	-1.113	0.265898	
DialectNo.ord4	0.44494	0.05730	7.766	8.25e-15	***
DialectNo.ord5	0.25715	0.05669	4.536	5.75e-06	***
DialectNo.ord6	0.11478	0.06793	1.690	0.091076	*
DialectNo.ord7	0.04600	0.05603	0.821	0.411716	
GenderM	-8.41326	0.02945	-285.673	< 2e-16	***

R-sq.(adj) = 0.633 Deviance explained = 63.4%

Table 3 is obtained using the same formula in (2), except that the data is from SA2. The first sentence SA2 results in that the baseline (New England) dialect shows that its F0 contour is significantly different from the F0 patterns in other dialects except for dialects 2(the Atlantic seabord), 3(the northern MidWest), and 5(the Southeastern). The final line in Table 2 and Table 3 below shows the goodness-of-fit statistics. 63.4% of the deviance for SA1 and 61.3% of the deviance for SA2 in the calibration sentences can be explained by the model in (2).

Table 3. Parametric coefficients for the GAMM-model for the SA2 sentence

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	10.8191207	0.0546855	197.842	< 2e-16	***
DialectNo.ord2	0.0767778	0.0606130	1.267	0.2053	
DialectNo.ord3	0.0008724	0.0610957	0.014	0.9886	
DialectNo.ord4	0.2801049	0.0622437	4.500	6.81e-06	***
DialectNo.ord5	0.0882217	0.0617566	1.429	0.1531	
DialectNo.ord6	0.1644641	0.0727612	2.260	0.0238	*
DialectNo.ord7	-0.1310973	0.0610374	-2.148	0.0317	*
GenderM	-8.3293040	0.0320138	-260.179	< 2e-16	***

R-sq.(adj) = 0.613 Deviance explained = 61.3%

Figure 9 shows GAMM-modelled overlaid F0 contours of two calibration sentences in the TIMIT database. Visual inspection of the modeled F0 patterns for SA1 and SA2, respectively, reveals that one dialect (New York City) in SA1 stands out in that the ending portion of the F0 rises up significantly, which is typical of question type of utterances or utterances with uncertainty, or uptalks. In the case at hand, since the type of sentence is declarative, the rising F0 pattern may be due to uncertainty. Systematic inspection of the data in this dialectal region is needed. It also reveals that the differences between dialects may come from the variances at the end of the F0 patterns. Visualization of the F0 patterns among different education levels seems to show similar patterning regarding overall F0 patterns.

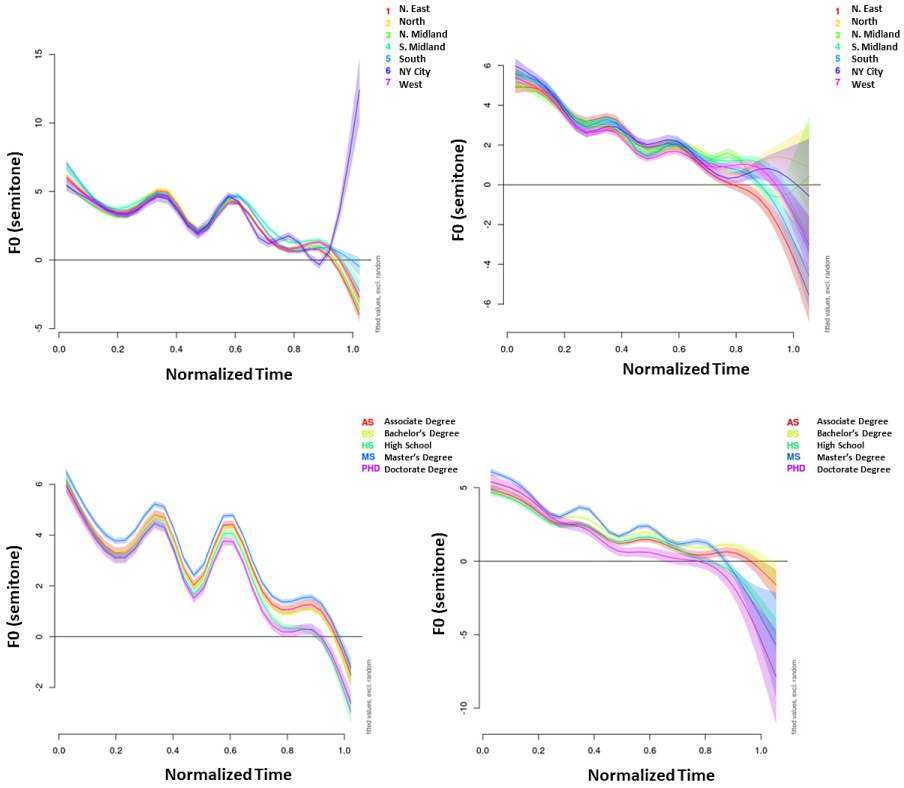

Figure 9. Model F0 contours per dialects (above) and education (down), for SA1 (left) and SA2 (right)

The following two panels in Figure 10 show individual F0 patterns per dialect. Once again, New York City shows sharp rising patterns of F0 at the end of the sentence SA1. However, given the patterns of the F0 in SA2, which belongs to the declarative type of sentence, and given the absence of rising patterns when the patterns are modeled after the education level, the unique rising pattern as we observed in SA1 for the New York dialect requires future explanation.

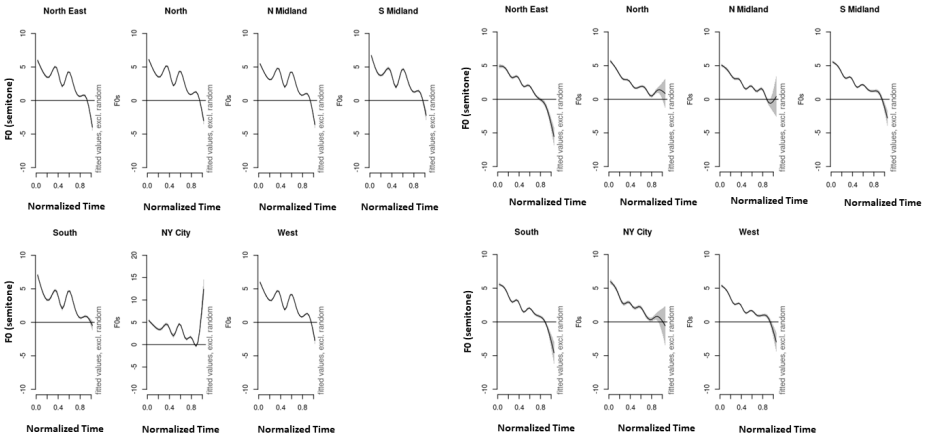

Figure 10. Individual F0 contours over SA1 and SA2, respectively, per dialect

4. Results and conclusion

In this paper, GAMM-based approaches to dynamic trajectories of phonetic features are presented using the TIMIT database. Using third-party packages such as Parselmouth, TextGridTools, and Pandas, I demonstrated how the acoustic and textual information could be processed, mutated, and merged so that contextual information could be used as input features to statistical modeling. As for the statistical modeling, Generalized Additive Mixed Model (GAMM) has been applied to the calibration sentences in the TIMIT database, as the two sentences (SA1 and SA2) are read by all participants. Due to the sheer number of comparisons to make for each pair of phones, I have not tried any systematic pairwise comparison. Instead, visualization of the F0, F1, and F2 models was made for each phone based on GAMM. The visualization revealed legitimate and expected patterns. To observe

whether the calibration sentences exhibit dialectal differences on the intonation level, I made separate GAMM-based models of F0 contours for each calibration sentence. Even though it requires further systematic investigation, dialectal differences, if any, seem to be most vivid at the ending part of the utterance. Apparently, only a very small subset of the meta information available in the TIMIT database, both sound files and meta information, has been used in this study. Nevertheless, the approach taken in this paper looks promising in furthering our understanding of the phonetic trajectories in the shaping of phonetic and phonological patterns. More systematic and comprehensive analysis using the available meta information will be left for future research.

REFERENCES

BOERSMA, PAUL and DAVID WEENINK. 2018. Praat: Doing phonetics by computer (Version 6.0.40) [Computer program]. http://www.praat.org.

BUSCHMEIER, HENDRIK and MARCIN WŁODARCZAK. 2013. TextGridTools: A TextGrid Parsing and Analysis Toolkit. *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung* 2013, 152-157.

BYRD, DANI. 1994. Relations of sex and dialect to reduction. *Speech Communication* 15, 39-54.

CHO, SUNGHEY, NAOMI NEVLER, SANJANA SHELLIKERI, NATALIA PARJANE, DAVID J IRWIN, NEVILLE RYANT, SHARON ASH, CHRISTOPHER CIERI, MARK LIBERMAN, and MURRAY GROSSMAN. 2021. Lexical and acoustic characteristics of young and older healthy adults. *Journal of Speech, Language, and Hearing Research* 64.2, 302-314.

DURAND, JACQUES, ULRIKE GUT, and GJERT KRISTOFFERSEN. 2014. *The Oxford Handbook of Corpus Phonology*. OUP Oxford.

ERNESTUS, MIRJAM and R. HARALD BAAYEN. 2011. Corpora and exemplars in phonology. *The Handbook of Phonological Theory* (2nd ed.), 374-400. Wiley-Blackwell.

HINTZMAN, DOUGLAS. L. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93.4, 411.

JADOUL, YANNICK, BILL THOMPSON, and BART DE BOER. 2018. Introducing parselmouth: A python interface to praat. *Journal of Phonetics* 71, 1-15.

KEATING, PATRICIA. A., DANI BYRD, EDWARD FLEMMING, and YUICHI TODAKA. 1994. Phonetic analyses of word and segment variation using the TIMIT corpus of American English. *Speech Communication* 14.2, 131-142.

LADD, D. ROBERT. 2014. *Simultaneous Structure in Phonology*. Oxford: Oxford University Press.

MCKINNEY, WES. 2011. Pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing* 14.9, 1-9.

NOLAN, FRANCIS. 1992. The descriptive role of segments: Evidence from assimilation. *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, 261-280. Cambridge: Cambridge University Press.

NOSOFSKY, ROBERT. M. 1986. Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General* 115.1, 39.

R CORE TEAM. 2019. R: A Language and Environment for Statistical Computing (Version 3.6.3) [Computer program]. https://www.R- project.org.

SÓSKUTHY, MÁRTON. 2017. Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. *arXiv.* 1703.05339 [stat:AP].

_____. 2021. Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics* 84, 1-22.

VAN RIJ, JACOLIEN, MARTIJN WIELING, R HARALD BAAYEN, and DIRK VAN RIJN 2015. Itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version 1.0.1. https://cran.r-project.org/web/packages/itsadug.

WICKHAM, HADLEY. 2017. Tidyverse: Easily install and load the 'tidyverse'. R package version 1.2.1. https://cran.r-project.org/web/packages/tidyverse.

WIELING, MARTIJN. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70, 86-116.

WOOD, SIMON. 2015. mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation. R package version 1.8. https://cran.r-project.org/web/packages/mgcv.

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

ZUE, VICTOR W. and STEPHANIE. SENEFF. 1996. Transcription and alignment of the TIMIT database. In HIROYA FUJISAKI (ed.). *Recent Research Towards Advanced Man-Machine Interface Through Spoken Language*, 515-525. Elsevier.

Tae-Jin Yoon (Associate Professor)
Department of English Language and Literature
Sungshin Women's University
34 Da gil 2 Bomun-ro, Sungbuk-gu
Seoul 02844, Republic of Korea
e-mail: tyoon@sungshin.ac.kr

received: October 30, 2021
revised: December 04, 2021
accepted: December 16, 2021