An Integrated Dialect Analysis Tool Using Phonetics and Acoustics

Gregory Jones[1], Naghmeh Nadjibzadeh[2], László Károly[3], Mahdi Mohammadpour[1]

[1]Wolfson School, Loughborough University, UK
[2] Johannes Gutenberg University of Mainz, Mainz, Germany
[3] Department of Linguistics and Philology, Uppsala University, Sweden

Abstract

This study aimed to verify a computational phonetic and acoustic analysis tool created in the MATLAB environment. A dataset was obtained containing 3 broad American dialects (Northern, Western and New England) from the TIMIT database using words that also appeared in the Swadesh list. Each dialect consisted of 20 speakers uttering 10 sentences. Verification using phonetic comparisons between dialects were made by calculating the Levenshtein distance in Gabmap and the proposed software tool. Agreement between the linguistic distances using each analysis method was found. Each tool showed increasing linguistic distance as a function of increasing geographic distance, in a similar shape to Seguy's curve. The proposed tool was then further developed to include acoustic characterisation capability of inter dialect dynamics. Significant variation between dialects was found for the pitch, trajectory length and spectral rate of change for 7 of the phonetic vowels investigated. Analysis of the vowel area using the 4 corner vowels indicated that for male speakers, geographically closer dialects have smaller variations in vowel space area than those further apart. The female utterances did not show a similar pattern of linguistic distance likely due to the lack of one corner vowel /u/, making the vowel space a triangle.

Keywords: dialectometry, geographical linguistics, acoustic analysis, phonetics

1. Introduction

Dialectology involves the study of dialects and their variation over geography, demographics and time. Methods have been developed to quantify dialectal characteristics to determine how they differ. The fundamental postulate states that geographically further apart dialects should be less similar than those closer together (Chambers & Trudgill, 1998). Nerbonne proposed that variation in dialect with respect to geography could be described by Seguy's Curve whereby there is a log relationship between geographic and linguistic distance (Nerbonne, 2010).

There are exceptions to the abovementioned geographical-linguistic distance correlation. This behaviour may change for different dialects and other case studies as described by Bakker (2007). He showed that there are many examples of spread of features from one language to another despite the presence of geographical distance, natural or social barriers. This means that smaller linguistic distance might be observed despite bigger geographical distance. Most of these exceptions are found in the phonology.

Phonetic and acoustic analysis are the main methods of analysing dialectal change over a region. An evaluation of phonetic analysis methods conducted by Heeringa (2004) showed the effectiveness of the Levenshtein algorithm as opposed to other methods. First introduced by Kessler when studying Irish Gaelic dialects, the distance aims to find the minimum number of insertions, deletions, and substitutions to turn one phonetic string into another (Levenshtein, 1966; Kessler, 1995).

Gabmap is an online phonetic analysis tool widely used to characterise the difference between dialects. The Levenshtein distance can be calculated between phonetic strings and averaged to achieve an overall linguistic distance between regions of interest. Features are available to conduct cluster analysis, Multidimensional Scaling (MDS) as well as creating reference point maps of the aggregated dataset (Nerbonne, et al., 2011). However, limitations in all phonetic analysis methods stem from the accuracy of the transcription. It has been shown that a transcribers dialect impacts the accuracy of the tabulated phonetic transcription

(Heeringa, 2005). The process of transcription is also highly laborious when working on a large dataset. The large dataset is central to the study of dialects variation and change in the time domain, i.e. the dynamics of dialects. It is also critical in investigating extra-linguistic factors. Extra-linguistic factors in include age, gender, education and social class. Taking in to account these factors requires mass collection and processing of linguistic data. The investigation of extra-linguistic factors can lead to clearer explanation of linguistic phenomenon and language variation and change. Labov (2001) believes that leaders of language variation and changes are women which highlights the importance of taking into account the effect of gender. Keller (1990, 1994) highlights the effect of social class on the language change by providing the "invisible hands" model which emphasizes the effect of extra-linguistic factors on the linguistic phenomenon.

Acoustic analysis methods have been developed that attempted to eliminate the need for phonetic transcriptions. One such method of analysis uses formant frequencies to characterise dialects without the use of phonetics. Formants are resonances created when the shape of sound waves are altered in the vocal tract by articulation of the lips, jaw, tongue and other speech organs (Maddieson & Ladefoged, 1996).

The second formant frequency (F2) has shown to be the most influential factor for conveying accents (Yan & Saeed Vaseghi, 2003). Adank, Van Hout and Smits (2004) found that regional impact on F2 may be much more prominent than the first formant (F1).

Between different dialects, any of the first three formant frequencies may show significant classification results. For example, Birmingham and Liverpool accents, certain vowels represented classification characteristics for all three formants (Zheng, Dyke, Berryman & Morgan, 2012).

Another method of acoustic analysis developed around using the fundamental frequency (F0) to determine dialect variation. The changes in F0 during speech correlates with the rise and fall of someone's voice when speaking. Each language contains its own set of patterns for intonation, stress, and rhythm. It has been shown that for some British accents pitch slope (variation in pitch over the duration of a vowel) plays a role in accent identification although not as large a role as some other factors (Zheng, et al., 2012).

Analysis of English, French and German languages indicate that speakers significantly differ in their intonation slope (Grover, Jamieson & Dobrovolsky, 1987). This correlation has also been shown for certain Indian dialects (Agrawal, Jain & Sinha, 2016).

Many previous studies use the powerful Praat software created by David Weenik and Paul Boersma that can calculate formant tracks, pitch tracks, visualisations of the signal and corresponding spectrograms (Boersma & Weenik, 2001). However, since it is not a devout dialect analysis tool, it does not provide a similar geographical plotting feature that Gabmap provides, neither an aggregated analysis of a larger acoustic dataset to plot formant trajectories of the aggregated data.

Both Pratt and Gabmap were created in part to help researchers investigate the variation of dialects over a geographical area. Although these tools are powerful, they lack an easy to use and integrated approach to dialectology which can be picked up by almost anyone and used to create meaningful and visible results. There is also a plethora of different analysis methods available to researchers of dialect, however, much of the technical content may be considered too hard for these researchers to computationally implement. The developed software provides the researcher with some of the key cornerstones in phonetic and acoustic analysis from which they can investigate any dialect of interest over any given geographical map. There is scope in the future to translate some of the more modern technical methods of analysis such as dialect likelihood recognition methods using Hidden Markov Models (Chen, et al., 2014) or the use of Support Vector Machines (SVMs) to increase dialect recognition (Biadsy, et al., 2010). Fortunately, due to the modular structure of the software the method of adding analysis functionality does not require much more work than developing the computational model for the proposed analysis method. By creating the software on Matlab many of the intrinsic analysis features can also be applied to dialect analysis such as in-built Hidden Markov Model creation methods and SVM support.

This paper aims to present a new software tool created specifically for dialectology using the MATLAB environment, bringing together the phonetic and acoustic analysis techniques into one easy to use tool. Verification of phonetic analysis against the conventional software tool Gabmap was undertaken by analysing a test data set obtained from the TIMIT database. The acoustic element of the software uses vowel formant frequency, pitch, and duration analysis to investigate the dialectal differences between the representative data. Agreement between dialect characteristics has been seen using both forms of analysis that fit the fundamental postulate described by Chambers & Trudgill (1998).

2. Methodology

2.1 Phonetic Analysis

The Levenshtein distance was calculated between each phonetic string for a given word. Insertion and deletion of a character received a score of one and substitution as two since it consists of deletion followed by insertion. An example calculation for the word *morning* shows the minimum number of operations to convert one string into another.

Table 1: An example Levenshtein distance calculation using two phonetic transcriptions of the word "morning"

Word	M	O	R	N	I	N	G
Dialect x	ɱ	ɔ	ɾ	ŋ	ɪ	n	g
Dialect y	m	o	ɾ	ŋ	i	n	-
Change Cost	2	2	0	0	2	0	1
Linguistic Distance	2	4	4	4	6	6	7

The ratio of the linguistic distance to the maximum linguistic distance between word pairs was used to normalise the data. This gave results that were independent of the original string length.

Since each location usually contained more than one phonetic transcription per word, pairwise linguistic distances were calculated and then averaged to obtain the percentage linguistic distance between each location for a given word using the following equation,

$$L_{ap} = \frac{\sum_{i=1}^{N2} \sum_{j=1}^{N1} \frac{L(i,j)}{L_m\ (i,j)}}{N1 * N2} \qquad (1)$$

Where L_{ap} is the average percentage Levenshtein distance, $N1$ is the number of phonetic stri (Chen, et al., 2014)ngs for a word at location A, $N2$ is the number of phonetic strings for a word at location B, $L(i,j)$ is the Levenshtein distance between phonetic transcription i at location B and j at location A and $L_m(i,j)$ is the maximum Levenshtein distance.

To investigate the log linear relationship described in Szmrecsanyi (2012), the log Levenshtein distance was calculated. Since the logarithm of zero is undefined, one was added to the value of the linguistic distance.

$$L_{lap} = \frac{\sum_{i=1}^{N2} \sum_{j=1}^{N1} \frac{ln(L(i,j)+1)}{ln(L_m(i,j)+1)}}{N1 * N2} \qquad (2)$$

Where L_{lap} is the average log-percentage Levenshtein distance between two locations.

The total linguistic distance between each location was then found by calculating the unweighted average of each word's linguistic distance. A triangular matrix of location by location linguistic distances was formed where the diagonal contained zeroes since the linguistic distance between a location and itself is zero. Comparison of percentage linguistic distance values showed the agreement in results between Gabmap and the proposed software tool. Using reference points, the linguistic distances were plotted over geographic distance to inspect the pattern between Gabmap and the proposed solution for the percentage linguistic distance and the log percentage linguistic distance.

2.2 Acoustic Analysis

2.2.1 Acoustic Data Inputs

The user is required to input a speech file from which the formants are to be calculated. A reference to the name, age and gender of the speaker as well as the geographic origin of the speech file must be specified. Where either the age or gender are not provided or kept private an undefined option may be selected.

2.2.2 Formant Algorithm

Formant frequencies are calculated using a common method of Linear Predictive Coding (LPC). LPC analysis calculates the properties of the vocal tract filter that created a speech signal. It works on the principle that if shape of the vocal tract and the output waveform are known, the filter properties that turned one into the other can be calculated. The formant frequencies are calculated by finding the roots of a polynomial generated through LPC analysis (Snell & Milinazzo, 1993). The implementation of this algorithm in Matlab was provided by the Mathworks documentation as well as the aforementioned work on the mathematics of the problem.

The LPC filter which is provided as part of the Matlab Signal Processing Toolbox, can be seen as a function with a set of coefficients. The LPC filter determines these coefficients of a forward linear predictor by minimizing the prediction error in the least squares sense [46].

Initially the input speech signal is processed by applying a Hamming window over the signal to reduce the effects of spectral leakage. A pre-emphasis high pass all-pole AR filter is then applied. The inbuilt “lpc” function determines the set of coefficients of an nth-order finite impulse response (FIR) filter that “predicts the current value of the real-valued time series based on previous data” [44]. The roots of this equation are complex conjugates therefore only the positive imaginary roots are kept, eliminating duplicated results. The angle of the root from the axis (θ) is calculated using simple Pythagoras. This angle can then be converted into a frequency value using the following formula (Snell & Milinazzo, 1993):

$$F(i) = \frac{f_s}{2\pi}\theta_i \; Hz \qquad (3)$$

Where f_s is the sampling frequency and i represents the number of the formant i.e. first, second or third.

2.2.3 Software Formant Frequency Data Output

The individual formant frequencies can be observed on an absolute basis on a map and compared between different regions for specific vowel utterances or groups of vowel utterances. It is also possible for the user to filter all results by age and gender and carry out all the calculations stated below.

The vowel section trajectory length (VSL) describes the variation of the formant over the utterance. The first and second formants at five equidistant points corresponding to 20%-35%-50%-65%-80% are used to find the section specific trajectory length. The first and final 20% are not used to reduce the effect of flanking consonants on the formant frequency. This procedure has been used in previous acoustic studies to investigate spectral change within vowels (Fox & Jacewicz, 2009; Adank, et al., 2004). The VSL for each vowel utterance can be calculated between each point resulting in four vowel section trajectories using the following formula:

$$VSLn = \sqrt{(F1_n - F1_{n+1})^2 + (F2_n - F2_{n+1})^2} \quad (4)$$

The overall trajectory length (TL) can then be found to be the sum of the individual vowel sections:

$$TL = \sum_{n=1}^{4} VSLn \quad (5)$$

The following formula is used to measure the TL rate of change (TLroc) over the vowel sections investigated:

$$TLroc = \frac{VSLn}{0.15 \times duration} \quad (6)$$

Since the F1 relates to the position of the jaw when speaking and F2 relates to tongue position, the vowel space area (VSA) can be used to indicate the position of these articulators (Lee & Shaiman, 2003). The vowels /ɪ, u, ɑ, æ/ are normally used as the corners of the vowel quadrilateral however vowel area may be calculated from any set of three or more vowels using the software. A quantitative analysis using the MATLAB “polyarea” function is used to quantitatively measure the variation between dialects.

2.2.4 Duration

The vowel duration is calculated by simply storing the length of the input speech file. This value may be compared between locations by plotting the data on a map or by creating a histogram. One of the key uses of the vowel duration is in the calculation of the TLroc, shown in equation 5.

2.2.5 Pitch Algorithm

The proposed software uses an algorithm developed by Zahorian & Hu (2008) capable of accurately plotting the pitch track over the duration of an utterance using a combination of time and cepstral based analysis. Yet Another Algorithm for Pitch Tracking (YAAPT) calculates the cross correlation of the speech signal against itself when one frame is shifted in time from a position of no lag to maximum lag defined by parameters. The correlation is then compared to the cepstral-based analysis resulting in two pitch tracks. The optimum track is found using a method of dynamic coding. Full description of the algorithm and its operation can be found in Zahorian & Hu (2008).

2.2.6 Software Pitch Data Output

The central 60% of the pitch track is used to attempt to reduce the effect of flanking consonants as in the formant analysis. The mean pitch over 60% of the vowel is used to quantitatively investigate inter dialectal differences. Both the mean pitch and the pitch track can be plotted on a histogram to see how each variable varies relevant to location. Filtering of data enables the user to produce a range average pitch and pitch track for various data subsets.

2.3 Integration of Algorithms

Each algorithm implemented is a modular part of the software such that implementing a different algorithm in the future or making small changes is as simple as integrating a new module or making changes to an existing module.

When conducting the phonetic analysis, the results of each pairwise comparison are stored in a structure along with the specific information relevant to each utterance such as the word compared, the speakers who uttered the two phonetic phrases and the location of each speaker. Therefore, when two locations are compared for a specific word, the structure can be filtered such that the average Levenshtein distance can be calculated for the given locations and given word.

For the acoustic analysis a data structure is created where each element contains all data for a particular speaker, this includes their location, age and gender as well as a link to the speech file and all results from preliminary analysis such as formant frequencies, vowel duration, pitch and pitch tracks. This way when plotting the aggregated data, the user can simply specify their filter parameters through an intuitive GUI and the data structure is filtered and required values calculated as such.

The presented methodology can be applied to any speech data that follows the file structure that is required to be inputted to the program. The program can be made available via the project's website "*https://dialectech.org/*".

3. TIMIT Test Data

The TIMIT speech corpus consists of eight dialects of America that were used to increase acoustic phonetic knowledge and speech recognition systems. The corpus consists of 630 speakers of different gender and dialect each speaking 10 sentences. The corresponding phonetic transcriptions for each utterance have been verified. Three areas of interest were chosen for the investigation, Northern, New England and Western. The dialect centres were specified using a ".kml" file generated on Google Earth. The data was then uploaded to both Gabmap and the proposed software tool where geographic distances were calculated.

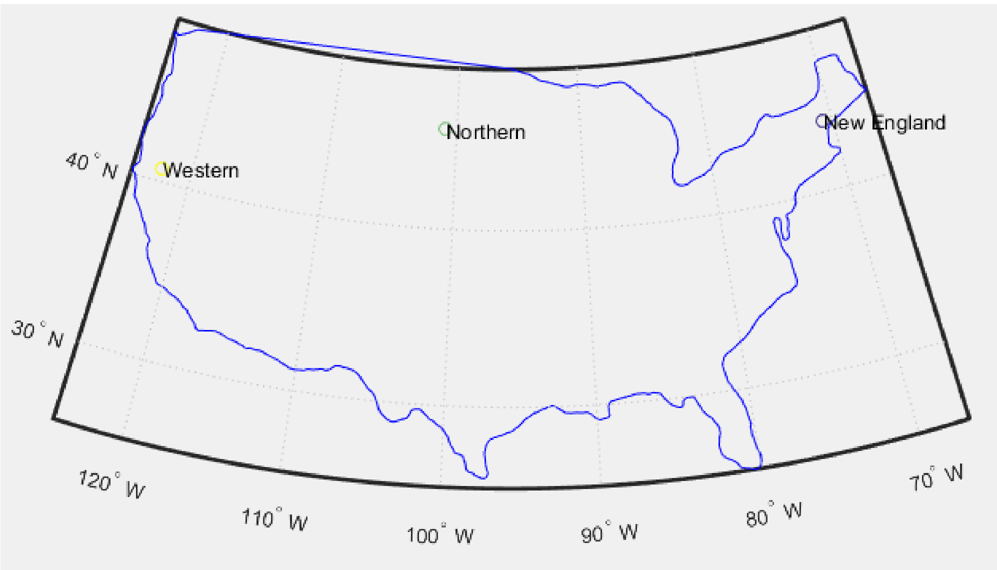

Figure 1: Geographical map of data points. The three points shown indicate the centres of the dialects that have been chosen for the investigation.

From each location, 20 speakers were selected uttering all 10 sentences resulting in 200 sentences. Information on their ages, background, and thickness of dialect was unknown as well as their geographic distribution within these dialect areas. Table 2 shows the male- female ratio for different area. Within the code itself, a random number generator selects the index of data corresponding to a speaker. With regards to the imbalance between male and female, the ratio of male-female speakers reflects the ratio of the sentences uttered by each gender from each location in the complete dataset. The result is a sample set which contains a number of utterances proportional to the total number of utterances produced by each gender.

Table 2: Distribution of the speakers analysed from the TIMIT database across gender and location

Dialect Region	Male	Female	Total
Northern	14 (70%)	6 (30%)	20
New England	13 (65%)	7 (35%)	20
Western	12 (60%)	8 (40%)	20

The words investigated were based on the Swadesh list which consists of 100 words that have been used to analyse the interrelations between languages (Swadesh, 1955). The 100 words were cross referenced against the sentences uttered by the speakers and collated into a dataset.

The 39 words that appeared in both the Swadesh list and TIMIT database as well as the phonetic vowel in the word is listed below as well as the number of utterances of each word for each dialect.

Table 3: Cross referenced words that appear in both the Swadesh list and TIMIT database

I	Long	Moon
You	Small	Water
We	Woman	Night
This	Man	Hot
That	Dog	Cold
What	Mouth	Full
Not	Hand	New
All	See	Good
Many	Know	Dry
One	Die	Name
Two	Walk	
Big	Give	

For cases where a word appeared more than once in each sentence, both utterances were used in the analysis. In total, 443 acoustic measurements were made on the words.

The region of 50Hz - 5500Hz was used to investigate the locations of the first three formants. The formant ceilings for women were specified at 5000Hz and men at 5500Hz to reflect men having a longer vocal tract than women (Escudero, Boersma, Rauber & Bion, 2009). The window length was set at 20ms and the step length was set to 10ms.

The recordings were carried out at 16kHz and down sampled to 11kHz by the proposed software to reflect two times the maximum frequency of interest.

The TIMIT database used for collecting the speech samples contained the start and end time within the speech sample that the vowel was uttered. It was assumed that the transcriptions were accurate and that the times recorded were accurate.

Altogether, 17 vowels were investigated although data for some were too sparse to provide any comparative results between all three dialects. The number of utterances per vowel is shown below.

Table 4: Number of vowel utterances produced by the 60 speakers for the words given above

Vowel	Utterance Count	Vowel	Utterance Count
aɪ	29	ɛ	12
ʉ	35	æ	66
ɨ	13	ʌ	19
i	33	ɑ	44
u	10	ɔ	120
ʊ	8	eɪ	3
ɪ	20	oʊ	10
ə	5	ɝ	16

4. Results

4.1 Phonetics

Initial verification of results between Gabmap and Dialectech was performed as the baseline test for the new tool. This verification shows that for all pairwise utterances compared between locations, the maximum difference is 15.82% when comparing Northern and New England dialects. The variation of results between the analysis tools was found to be lowest for the geographically furthest apart dialects (Western to New England), verifying the phonetic analysis using the proposed software tool.

Table 5: Diagonal matrix of phonetic data results with the percentage difference between Gabmap and Dialectech shown in brackets

	New England		Northern	
	Gabmap	**Dialectech (%Difference)**	**Gabmap**	**Dialectech (%Difference)**
New England	0	0		
Northern	0.1485	0.1764 (15.82)	0	0
Western	0.1693	0.1821 (7.04)	0.15995	0.1881 (14.97)

Figure2 shows how the linguistic distance varies over geographic distance. For Gabmap and the proposed software, there is a difference between New England and the other two dialects. From New England, the linguistic distance increases at a higher rate initially and then at a progressively lower rate as the distance continues to increase representing a logarithmic relationship.

The Northern reference point map indicates increasing linguistic distance up to the first point but then decreasing linguistic distance as the geographic distance increases. This behaviour can be explained by considering the extra-linguistic factors as explained by Bakker (Bakker 2007). The pattern between Gabmap and the proposed analysis tool is the same indicating similar results for both. It should be noted that as mentioned above, the observed correlation between the linguistic and geographical distances are case dependent. The current results confirm this correlation, but it might not be applicable for other dialects. Provided results show the consistency between results of the current analysis and Gabmap as a verification.

The only minor difference in the pattern of results between Gabmap and the proposed software occurs when calculating the linguistic distance from the Western point. Gabmap shows an increase in the linguistic distance from Western to Northern and Western to New England of 1%. However, the proposed tool shows a decrease in linguistic distance by 0.6%. Although the difference in linguistic distance is small, it is exaggerated when the logarithmic linguistic distance is calculated.

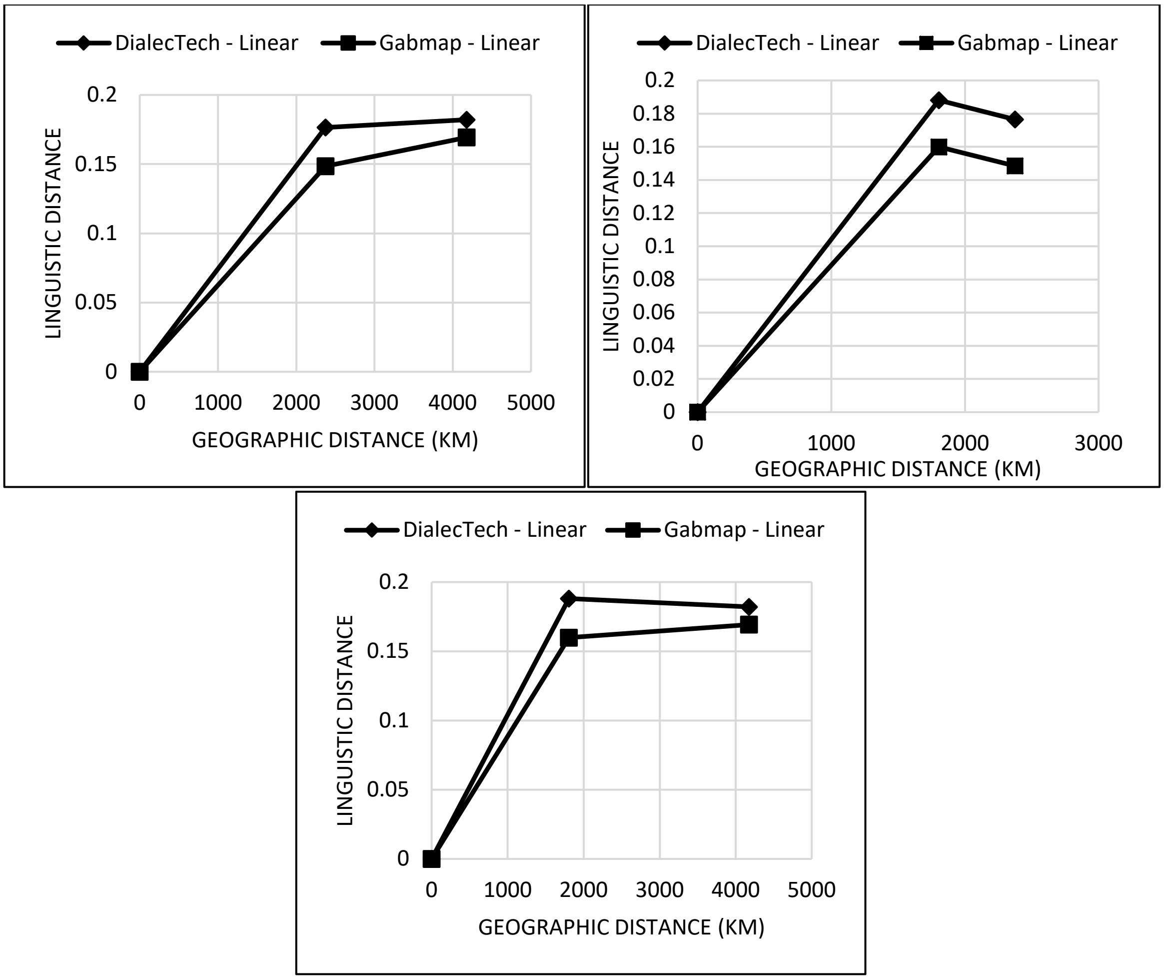

Figure 2: Reference point map from New England (top left), Northern (top right) and Western (bottom) dialect

4.2 Acoustics

4.2.1 Formants

Using the centre of the vowel as the vowel nucleus, absolute formant values were obtained. A repeated measure one-way analysis of variance (ANOVA) for each vowel with dialect and gender as between subject factors. It should be noted that a linear mixed effects model can be used as further enhancement to the presented method. No significant variation in F1 for 16 of 17 of vowels were found between dialect and gender. The mean value of the F1 formant for the vowel /ɔ/ indicated significant difference between the dialects ($p < 0.06$) showing that absolute F1 values were not able to fully characterise dialect dynamics. Similar results were found when analysing F2. One-way ANOVA of F3 indicated a significant difference in dialectal characteristics for /ə/ ($p < 0.0015$).

Of more interest is the variation in the formant frequencies over the duration of the vowel. To investigate this the total TL was calculated using Equation 4 for each utterance. One-way ANOVA of each dialect group was carried out for each vowel to investigate whether there is a significant difference between each dialect. Significant inter dialectal differences were found for /I, u, æ, eɪ, ɝ/ ($p < 0.07$).

The TL could capture certain dynamics over the duration of the formant where the formant nucleus could not. One-way ANOVA analysis of the TLroc values reinforced the importance of /æ, ɝ/ in the characterisation of dialect ($p < 0.01$). No significant difference between dialects for other vowels were found using the TLroc calculation.

To visualise this difference, mean formant trajectory plots for each of these vowels were generated for each dialect using the conventional formant plot used in dialectology. Each vowel formant track represents specific dynamics that are not repeated in other dialects. As expected the spectral roc and absolute formant values vary as the vowel category varies. Due to lack of data for /eɪ/ for the Northern dialect, the formant track could not be plotted. Clear characterisation of dialect is possible when looking at these plots due to the distinct differences in the formant trajectories. The dialectal variance in trajectory for each vowel varies significantly. The /u/ vowel exhibits a very clear variation in the articulation over the duration of the utterance for the male speakers. The /æ/ shows that for all dialects the vowel sound becomes more open and back as the utterance progresses. However, for the Northern dialect there is a slight closure as the utterance nears and end that is not seen in other dialects. The vowel sound /ɝ/ produced by male speakers is more back and open compared to the female speakers which shows the vowel becoming more back and mid as it progresses. For male speakers the overall change in formant frequency for the vowel /ɝ/ is similar for each dialect. The New England dialect shows the highest variation in formant frequency throughout the duration of the vowel sound /ɝ/ for the female speakers. There is significant variation in the front closed vowel sound /i/ for all dialects of both genders. There is a high amount of inter-dialectal and inter-gender formant trajectory variation as indicated in figures 3 and 4 by the change in direction of the arrows. This may indicate that there are significant differences within the dataset chosen which could be as a result of the large area over which the samples were taken.

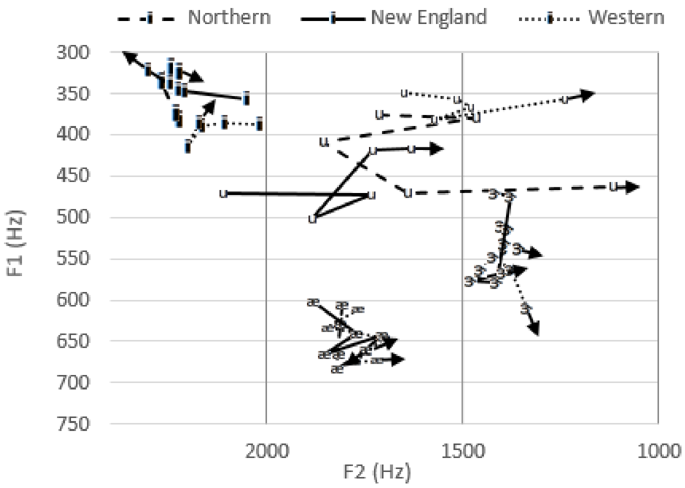

Figure 3: Male vowel formant trajectories. Samples are taken at five equidistant points along position of vowel. Arrows indicate the movement of the formant trajectory throughout the vowel.

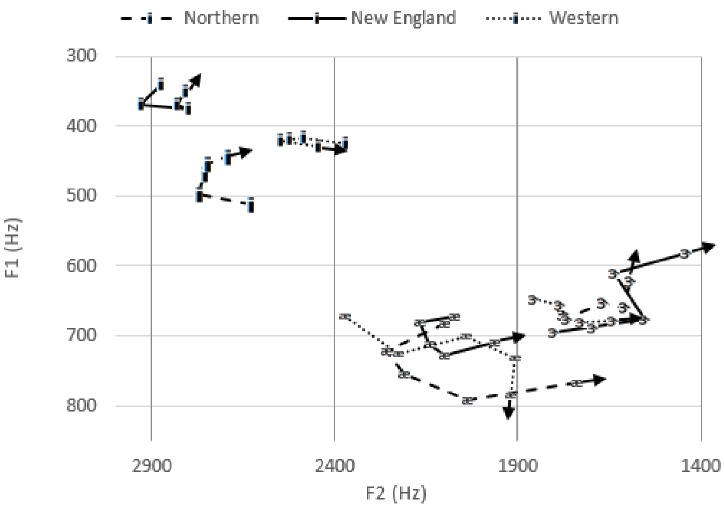

Figure 4: Female vowel formant trajectories. Samples are taken at five equidistant points along position of vowel. Arrows indicate the movement of the formant trajectory throughout the vowel.

By plotting F1 against F2 in the conventional way, a vowel quadrilateral can be drawn. Significant differences between dialects exist for the whole and male data sets. For the male dialects from Western to New England, the back vowel /u/ becomes more fronted. The /ɪ, ʌ, æ/ vowels spoken in the Northern dialect indicate a significant amount of fronting when compared to the other dialects. There is no significant variation in the location of /æ/ between dialects. In the case of the open back unrounded vowel /ʌ/ the Western and New England formants seem to be more similar than that of Western and Northern.

The Female data represents a vowel triangle since there was no data for /u/ analysed. The location of the corners of the triangle are closer than for the male data set. There is also no significant inter-dialectal variation for the open back unrounded vowel /æ/ showing that the TL and TLroc could indicate variation where the absolute formant values were unable.

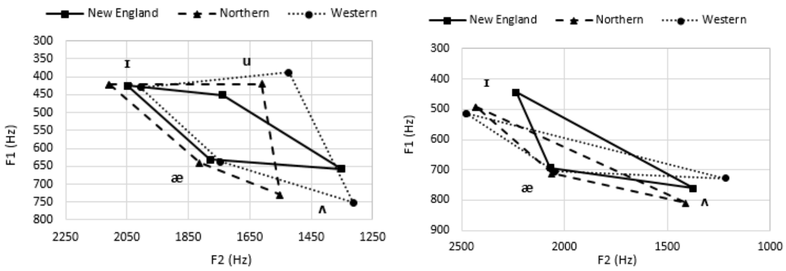

Figure 5: Male (left) and Female (right) VSA plot of each dialect with inverted axes

The variation in VSA relative to the VSA of New England indicated increasing linguistic distance as a function of geographic distance for the mixed and male utterances. The female utterances showed that New England and Western dialects were more similar than that of New England and Northern. This could be because of not taking the VS of /u/ into account which for male data was one of the main sources of VS variation. This could also be a result of extra-linguistic factors and gender effect as pointed out by Labov (Labov 2001) which requires further analysis and investigations.

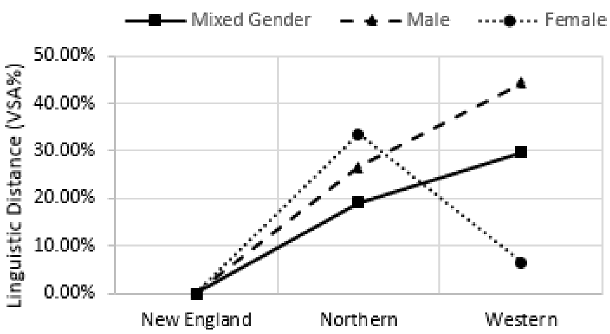

Figure 6: Linguistic distance using VSA using New England as the reference point

4.2.2 Pitch

By analysing the central 60% of the vowel utterance for each dialect and averaging the pitch track obtained, direct comparisons between dialects could be made. In general, there was no distinct difference between the average pitch between each dialect. Significant results were seen for /aɪ, ə, ɔ/ where the pitch of each dialect for these three vowels were not considered similar, indicating that there is a clear difference between the dialects ($p < 0.03$).

5. Discussion

The phonetic analysis indicated that the absolute values between the proposed analysis method and Gabmap showed a good agreement. As generally expected, the pattern of dialect over geography indicated that geographically further apart dialects exhibit a higher linguistic distance. The comparison between Gabmap and the proposed tool use the same transcriptions, therefore potential erroneous transcriptions do not affect the comparative investigation. The dialect areas investigated were wide-ranging with many sub-dialect regions within them. Since specific geographical locations of each speaker was not known, the dialects were measured from an approximate singular point as shown in Figure 1. This may have been the source of some unexpected results in the phonetic analysis whereby geographically further dialects (Northern, Western relationship compared to Northern, New England relationship) appeared more similar.

The acoustic analysis carried out by the proposed software tool showed significant inter dialectal variation when using the TL as a measure of formant dynamics across the duration of a vowel. The differences were characterised by different absolute values of TL for vowels /i, u, æ, eɪ, ɝ/. The TL plot for the vowel sound /i/ indicated that there was a significant variation between dialects of how that vowel was being produced. This could be due to vowel reduction when sounded at the end of the word *man**y*** whereas no reduction when uttered in other words.

The results for TLroc indicated that two vowels, /æ, ɝ/ were significantly different, reinforcing some of the TL findings. Clear differences in the vowel dynamics were found when plotting the formant frequencies at the five equidistant points in the vowel.

Significant inter dialectal differences for the /aɪ, ə, ɔ/ vowel sounds were found when analysing the mean pitch over the middle 60% of the vowels duration.

Quantification of results like those achieved by phonetic analysis can be obtained when evaluating the variation in VSA between dialects. The likely reason for the variation in the

VSA relates to the closeness of the formant quadrilateral to the physical articulation (jaw and tongue position) of each vowel. A plot of the percentage VSA difference using New England (*Figure 4*b) as the starting point showed similar patterns to what was seen in the phonetic analysis, suggesting that geographically further apart dialects loosely follow Seguy's curve when analysing the VSA. Visible differences can also be seen by analysing the VSA plots in *Figure 4* for male speakers. The back vowels contribute the most to the variation in VSA between dialects. For the female utterances the vowel triangle approach where 3 of the corner vowels are used did not exhibit the same variation between dialects that were seen for the males. However, the missing /u/ vowel sound played a large role in the VSA variation for male data and therefore could be the reason for female VSA similarities.

6. Conclusion

The present study sought to provide and verify a new software tool (Dialectech) against already existing software. It also aimed to verify that the software tool uses appropriate acoustic analysis techniques to reveal inter dialectal characteristics.

Using data from the well-known TIMIT database, three dialect regions were specified each consisting of 34 words. Each word contained multiple phonetic transcriptions within the dialect regions as well as numerous vowel utterances.

Analysis showed that Dialectech gives similar phonetic results (maximum difference 15%) to Gabmap with both analysis tools following the trends expected as geographic separation between dialects increase.

Acoustic analysis of the TL and TLroc, as well as pitch, indicated that the dialects analysed were significantly different for 7 out of the 17 vowels sampled. Significant results were not obtained when using the absolute formant values as a measure of dialect dynamics. The results showed the capability of systematically using acoustic analysis for dialectometric purposes. This can add significant ability and flexibility in analyzing large scale datasets quickly whilst being able to capture time domain variation and extra-linguistics effects through the use of age and gender filters. This also eliminated the potential errors in the transcription process.

The results verify the capability of Dialectech as a phonetic and acoustic analysis tool, showing the capability of acoustic analysis to be used instead of or as well as phonetic analysis.

References

Adank, P., Van Hout, R. & Smits, R., 2004. An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical society of America,* 116(3), pp. 1729-17388.

Agrawal, S. S., Jain, A. & Sinha, S., 2016. Analysis and modeling of acoustic information for automatic dialect classification. *International Journal of Speech Technology,* 19(3), pp. 593-609.

Biadsy, F. et al., 2010. *Discriminative Phonotactics for Dialect Recognition Using Context-Dependant Phone Classifiers.* s.l., Odyssey.

Boersma, P. & Weenik, D., 2001. Praat, a system for doing phonetics by computer. *Glot International,* 5(9/10), pp. 341-345.

Chambers, J. K. & Trudgill, P., 1998. *Dialectology.* s.l.:Cambridge University Press.

Chen, N., Tam, S., Shen, W. & Campbell, J., 2014. Characterizing phonetic transformations and acoustic differences across English dialects. *IEEE- TASLP.*

Escudero, P., Boersma, P., Schurt Rauber, A. & Bion, R. A., 2009. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America,* 126(3), pp. 1379-1393.

Fox, R. A. & Jacewicz, E., 2009. Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America,* 126(5), pp. 2603-2618.

Grover, C., Jamieson, D. G. & Dobrovolsky, M. B., 1987. Intonation in English, French and German: perception and production. *Language and Speech,* 30(3), pp. 277-295.

Heeringa, W., 2005. Dialect variation in and around Frisia; classification and relationships. *Us Wurk,* 54(3-4), pp. 125-167.

Heeringa, W. J., 2004. *Measuring dialect pronunciation differences using Levenshtein distance,* s.l.: PhD diss., University Library Groningen.

Kessler, B., 1995. *Computational dialectology in Irish Gaelic.* Dublin, Morgan Kaufmann Publishers Inc..

Lee, J. & Shaiman, S., 2003. Relationship between articulatory acoustic vowel space and articulatory kinematic vowel space. *Acoustical Society of America,* 132(3).

Levenshtein, V., 1966. Binary codes capable of correcting deletions, insertions and reversals. *Soviet physics doklady,* 10(8), pp. 707-710.

Maddieson, I. & Ladefoged, P., 1996. In: *The sounds of the world's languages.* Malden: Blackwell Publishing.

Nerbonne, J., 2010. Measuring the diffusion of linguistic change. *Philisophical Transactions of the Royal Society B: Biological Sciences,* 365(1559), pp. 3821-3828.

Nerbonne, J. et al., 2011. Gabmap. *Dialectologia,* Volume II, pp. 65-89.

Snell, R. C. & Milinazzo, F., 1993. Formant location from LPC analysis data. *IEEE transactions on Speech and Audio Processing ,* 1(2), pp. 129-134.

Swadesh, M., 1955. Towards greater accuracy in lexicostatistic dating. *International journal of American linguistics,* 21(2), pp. 121-137.

Szmrecsanyi, B., 2012. *Grammatical variation in British English dialects: A study in corpus-based dialectometry.* s.l.:Cambridge University Press.

Yan, Q. & Saeed Vaseghi, 2003. *Analysis, modelling and synthesis of formants of British, American and Australian accents.* Hong Kong, China, s.n.

Zahorian, S. A. & Hu, H., 2008. A spectral/temporal method for robust fundamental frequency tracking. *The Journal of the Acoustical Society of America,* 123(6), pp. 4559-4571.

Zheng, D. C., Dyke, D., Berryman, F. & Morgan, C., 2012. A new approach to acoustic analysis of two British regional accents—Birmingham and Liverpool accents. *International Journal of Speech Technology,* 15(2), pp. 77-85.