


5. Methods of acoustic analysis

 The *Atlas of North American English* [ANAE] presents two principal kinds of data on the vowels of North American English: the presence or absence of phonemic distinctions between vowels, and the precise place of articulation of vowels in phonological space. The data on mergers and splits come mainly from participants' productions and perceptions of word pairs, which are coded as "different", "close", or "the same", by means of auditory impressionistic analysis (Chapter 4). The data on place of articulation, and on the operation of chain shifts affecting the articulation of whole sets of vowels, come from acoustic analysis. Acoustic analysis also serves to clarify cases of merger or split where auditory impressionistic analysis is not decisive. This chapter describes the methods of acoustic analysis used for ANAE.

5.1. The philosophy of measurement involved

The Telsur project and its product, ANAE, were driven by a philosophy of measurement that requires greater accuracy and also greater efficiency than is normally demanded in laboratory research. For much experimental work on categorization, discrimination, habituation, etc., margins of error of ± 50 Hz are often satisfactory, and are usually obtained by measuring vowels at the mid-point of the resonant portion, or averaging over the whole nucleus. The close study of variation across dialects and age groups within dialects needs finer resolution, both in the location of the central tendency of formants and in locating the point in time of measurement. Methods for obtaining this increased accuracy are discussed below.

The goal of the Telsur project was to represent the ongoing sound changes in the urbanized areas of North America, and the project interviewed 805 persons, of whom 762 were ultimately selected as satisfying Telsur criteria for local speakers. The goal of the acoustic measurement program was to measure the vowel systems of as many of these subjects as possible and, at the same time, obtain a complete and accurate inventory of the phonemes and allophones involved in sound change. This meant raising the number of vowels measured from the 150 characteristic of the early studies of Labov, Yaeger & Steiner [LYS] to a typical level of 300, or in some cases much more. In the final analysis, 439 speakers and 134,000 vowels were measured. This entire data bank of measured formants is available to users of the Atlas on the accompanying CD.


This increase in the volume and accuracy of the data was in part the result of the efficiency and accuracy of the CSL system used for LPC measurement. It was also the result of decisions made early in the project to collect for the great majority of vowels a single F1/F2 measurement as the best indication of the central tendency of each nucleus. In the interests of describing the widest possible range of communities and sound changes, measurements of F3, F0, duration, intensity, and bandwidth were not collected. As noted below, a great deal of supplementary information is contained in the Plotnik vowel files that are available to ANAE users. The field registering lexical information contains special codes indicating the presence, absence and direction of glides; stylistic context; observations

of the analyst on marked auditory qualities of the signal, number of poles used in measurement, and other information bearing on the reliability of the signal.


Much of the time spent on measurement consists of locating the words of interest and storing these segments. More than one member of our research staff has projected a program for automatic location, segmentation, and measurement of vowel nuclei, but so far, all such attempts have led to an increase in gross error rates of several orders of magnitude. At present, we find there is no effective substitute for the careful examination and measurement of the formant trajectories of each individual vowel token by an analyst relying on both auditory and visual information, double-checking the computer's analysis against auditory impressions. More recent software programs like Praat reduce the time required for segmentation, but the same combination of auditory and visual inspection is necessary to reduce gross errors.

The discussion that follows assumes a basic knowledge of acoustic phonetics. This chapter will be principally concerned with issues surrounding the selection of a single point of measurement that best represents the central tendency of a vowel. Readers who need to review basic principles of sound spectrography and vowel formant identification are referred to an introductory phonetics textbook such as Ladefoged (1993).

5.2. Equipment

 All of the Telsur interviews were conducted over the telephone and recorded by means of a telephone signal splitting device, first one sold by Radio Shack and later a Hybrid Coupler made by Gentner Communications Corporation. The early interviews were recorded on analog reel-to-reel tape using a Nagra IV, a Nagra E, or a Tandberg Model 9021. The later interviews were recorded on digital cassette tapes (DAT) using SONY WMD6C DAT recorders.

All acoustic analysis was carried out with the Computerized Speech Lab (CSL) program developed by Kay Elemetrics. The version used was 4300B, running in DOS. The interview tapes were digitized at a sampling rate of 11,000 Hz using the CSL digitization hardware and software.

The use of the telephone was an essential element of the Telsur method, permitting the collection of speech samples from across North America over a period of a few years without incurring the long delays and high costs of sending field workers to every city in North America. This benefit did not come without a cost. The telephone line limits the frequency range of the transmitted signal to about 300 to 3,000 Hz, and it also significantly reduces the dynamic range. Still, the signal is satisfactory for conversation all over the world.  Although the quality of sound obtained by recording from the telephone line is clearly not comparable to that obtained in face-to-face interviews recorded with a high-quality microphone. However, in the vast majority of cases, the sound quality of the digitized speech signals was found to be high enough to permit acoustic analysis with a satisfactory degree of confidence and reliability. The signals were often accompanied by varying levels of background or mechanical noise, yet spectrograms made from these signals usually produced clearly interpretable formant structures.

5.3. Acoustic analysis of telephone interviews compared to face-to-face interviews

The study of language change and variation in Philadelphia utilized a series of 60 telephone interviews to obtain a geographically random sample of the city (Hindle 1980). Comparisons of these recordings with recordings of face-to-face interviews are reported in Labov 1994: Ch. 5. Telephone recordings were shorter and more formal than the face-to-face neighborhood recordings and obtained results that were less advanced in the direction of the sound changes being studied. To the extent that this finding applies to the data of the Telsur survey, the findings on the extent of sound changes in progress may be understated.

The Philadelphia study found the most significant differences in measurements of the high vowels, which were lower in the telephone recordings by 30–50 Hz. For the Telsur survey, a face-to-face interview was conducted with one speaker who had been interviewed by telephone, a 32-year-old man in Cedar Rapids, Iowa. This study confirmed the previous finding that telephone recordings registered lower values for F1. The mean difference between telephone recording and face-to-face recording values for F1 was 41 Hz. Insofar as this tendency is general, it will not affect the results of the analysis, since all comparisons are made across telephone interviews, and the normalization routine discussed in Section 5.6 will compensate for any skewing from one telephone handset to another. There are two exceptional cases to be noted. For /e/ before nasals, telephone recordings showed higher F1 values. Thus raising of /e/ in this environment is apt to be understated by the effect of telephone recording. However, the major sound change affecting this allophone is the merger of /i/ and /e/ before nasals, which is traced through minimal pair judgments rather than acoustic measurement (Chapter 9).

The largest differences in the comparisons made between telephone and direct recording are found in /iy/, where F1 was lower and F2 higher in telephone recordings. Lowering and backing of the nucleus of /iy/ is a defining feature of the third stage of the Southern Shift, so the bias of telephone recording will understate the extent of that sound change. The bias will be most important when face-to-face recordings are compared directly to telephone recordings without normalization, where we can expect face-to-face recording to show stronger movement in this third stage.

5.4. Selection of tokens for analysis

Segmentation. The words containing the target vowels were extracted from 20-second sections of digitized speech and stored in CSL's .NSP format in a directory established for each speaker. They were of four types, each representing a different style of speech (see Chapter 4 for the structure of the Telsur interview):

1. elicited minimal pairs (e.g. *cot* and *caught*; *pin* and *pen*);
2. elicited semantic differential items (e.g. *unhappy* and *sad*; *pond* and *pool*);
3. elicited word lists (e.g. counting from 1 to 10; days of the week; breakfast foods);
4. spontaneous speech (e.g. responses to demographic questions and discussion of issues of local interest, such as the state of downtown).

In the spontaneous speech category, only fully stressed tokens, bearing the primary stress of a phrase as well as primary syllable-stress within the word, were selected for analysis. This was to ensure that automatic processes of vowel reduction and centralization in non-primary stress environments would not interfere

with the analysis of regional patterns and that each token studied would provide an opportunity to observe the maximum extent of the sound changes under study.

Given that much of the data came from spontaneous speech, it was not possible to obtain an identical set of data from each speaker. The selection of tokens for analysis was constrained by the set of words that occurred in 20–30 minutes of conversation. Within this constraint, the analyst aimed at segmenting a similar balance of vowels and allophones for each speaker. As a general principle, each vowel phoneme or allophone was represented by no fewer than three tokens. In most cases, five to ten tokens of each vowel and allophone were collected. Collection of the most frequently occurring allophones was limited to approximately ten tokens, in order to prevent skewing of the representation of the speaker's vowel space by an over-representation of one or two vowels. By these methods, approximately 300 tokens were selected for each speaker. Some speakers had as few as 200 tokens, where conversation was limited, or low signal quality prevented the analysis of parts of the interview. Others had 400 to 500 tokens, where sound quality was good and conversation lengthy. The total number of measurements for 439 speakers was 134,000, an average of 305 tokens per speaker.

5.5. Selection of points of measurement

Once tokens from a speaker's interview had been digitized and saved as .NSP files, each token was called up in turn for spectrographic and linear predictive coding (LPC) analysis. The bandwidth of the spectrograms was 500 Hz, and the LPC analysis was computed at either 8, 10, 12, or 16 poles, depending on the strength of the signal. Where formants appeared to be missing, the number of poles was increased; where there were too many formants, the number was decreased.

While it is possible to measure many different aspects of vowel articulation using a spectrogram, Telsur accepted the findings of DeLattre et al. (1952), Cooper et al. (1952), and Peterson and Barney (1952), that the quality of most English vowels can be adequately represented by the frequency of their first and second formants, reflecting their height and advancement, respectively. Duration, rounding, nasality, pitch, tone, and laryngeal tension can also play an important role in vowel quality, but LYS demonstrated that a plot of F1 against F2 illustrates the most salient regional and social differences in the pronunciation of the vowels of North American English, including both vowel shifts and differences in phonemic inventory.

The general principle followed by Telsur is that no means of instrumental analysis can be considered reliable without some degree of auditory confirmation. LPC analysis is more precise than auditory impressions in some respects, but it is also subject to errors much greater than those found with auditory analysis, particularly when an incorrect number of formants is identified. Analysts continually use their knowledge of acoustic–auditory relations in deciding whether an appropriate number of formants has been located and in choosing the correct point in the time series for measurement (see below). Nevertheless, it is not possible for the analyst to recognize some gross errors until the analysis is completed and the entire vowel system is projected. For each of the 439 speakers analyzed acoustically, the F1/F2 plots produced by Plotnik (see below) were closely compared with auditory impressions. Two types of measured values were examined most closely. Outliers from the main distribution were re-played and compared to samples from the main distribution. They were accepted as valid tokens only if the auditory impressions differed in ways comparable to the measured differences. Secondly, special attention was given to cases where vowels from different

word classes showed the same F1/F2 values. (In such cases, word class assignment is typically disambiguated by an offglide.) Though in most cases these were valid indications of merger, there are configurations where differences are heard that do not correspond to F1/F2 differences, indicating the limitations of the two-formant axes in defining vowel timbre.¹

There are many possible approaches to the measurement of F1 and F2. A series of paired measurements taken at every pitch period would provide a wealth of detail on every movement of the tongue over the course of the vowel, including the nature of opening and closing transitions, and of on-glides and off-glides. While it is easy to plot an array of sequential measurements of a single vowel, plotting 300 such trajectories for a single speaker would obscure any pattern and preclude the goal of describing the vowel systems of North America. Moreover, inter-speaker comparisons, the central concern of dialectological or sociolinguistic research, are not feasible with trajectories, since precise points of comparison would be difficult to establish and quantitative analysis is problematic. For these reasons, the Telsur project followed the practice of LYS in representing the central tendency of each vowel with a single pair of F1/F2 values. The best choice of a single point of measurement therefore became the central methodological issue in the acoustic analysis that underlies the Atlas.

One approach to the representation of a vowel with a single measurement of F1 and F2 would be to take an average of the frequency of these formants over the whole course of the vowel's nucleus. While this technique has the advantage of reducing the likelihood of erroneous measurements, it runs the risk of missing important information about details of vowel articulation that can distinguish one region or speaker from another. Where a vowel's nucleus is characterized by a steady state in both formants, a nuclear average would seem adequate, as long as it did not include pre- or post-nuclear transitional values. However, many vowels involve a clear point of inflection in one or both formants at a specific point in the nucleus. A point of inflection indicates the moment when the tongue stops its movement away from an initial transition into the vocalic nucleus and begins moving away from the nucleus, either into a glide (in the case of a diphthong) or toward the position required for the next segment. As such, it is also the best representation of the vowel's overall quality, and gives a more accurate portrayal of the extent to which a speaker participates in a sound change than a nuclear average. Listeners appear to be sensitive to such points of inflection, perhaps because they are the best indication of the vowel's target.

The identification of points of inflection depends on an analysis of the central tendency of each vowel – the main trajectory of the tongue during its articulation. The central tendency of most short vowels and many long upgliding vowels is a downward movement of the tongue into the nucleus, followed by a rise out of the nucleus into the glide or following segment. The acoustic reflection of this fall and rise is a rise and fall in F1, with a maximal value of F1 representing the lowest point reached by the tongue. Vowels displaying this tendency were therefore measured at the point where F1 reached its maximal value. F2 was then measured at the same point, since measuring it at any other point would suggest a vowel quality that did not in fact occur.

The major exception to the principle of using the F1 maximum as a point of measurement occurs with those vowels whose central tendency is not so much a lowering and raising of the tongue as a movement of the tongue towards and then away from the front or rear periphery of the vowel space; these are ingliding vowels. In these cases, a point of inflection in F2, indicating maximum displacement toward the front or back periphery, was used as the point of measurement, with F1 measured at the corresponding point. Vowels whose tendency was movement toward and away from the front periphery were measured at their F2 maxima; those moving toward and away from the rear periphery were measured at their F2 minima.

In North American English, ingliding vowels typically arise in two situations. The first type comprises both historically long and ingliding vowels, like /æh/ and /oh/ in the Mid-Atlantic region, and originally short vowels that have been tensed and raised along the peripheral track, like /æ/ in the Northern Cities Shift, and /e/ and /i/ in the Southern Shift. The second case is that of high upgliding vowels followed by liquids (*fear*, *pool*). The liquids are articulated in mid-central position and therefore have some of the same characteristics as central inglides. Depending on the height of the nucleus and inglide of ingliding vowels, the maximum value of F1 may in fact occur in the glide rather than in the nucleus. A point of inflection in F2 rather than the F1 maximum is therefore the best measure of their nuclear quality. The trajectory of F2 was also used in some cases to identify a more precise point of measurement within a steady-state in F1, especially when a point of inflection in F2 appeared to indicate the maximal distance from consonantal transitions on either side of the vowel.

The most obvious inadequacy of single-point nuclear measurement is its failure to indicate the presence and quality of offglides. While some offglides are purely phonetic, having no contrastive function, others have phonemic status and play an essential role in distinguishing one vowel from another, as in the contrast between /ay/ and /aw/ in many English dialects. Moreover, while many of the most striking differences between English dialects involve variation in the position of the nucleus, others – including some of the best known – involve variation in the presence and quality of glides. The monophthongization of /ay/ in the Southern United States is the most obvious example, but subsequent chapters will reveal several other cases in which glides are as important as nuclei – in a few cases more important – in the differentiation of North American English dialects. Despite the importance of glides, in most cases it was found that the presence or absence and quality of glides could be effectively indicated with a code included in the comments attached to the measurements of nuclear quality, and that an actual measurement of the glide target was not necessary. These codes were used where the nature of the glide deviated from the norm for the vowel class or dialect in question, as when an upgliding vowel was monophthongal or a short vowel had developed an inglide. They were also used where the presence of a glide was one of the local features under study, as with the monophthongization of /ay/ in the South, or of /aw/ in Pittsburgh, or the development of a back upglide in Southern pronunciations of /oh/.

Though the normal practice was not to measure the endpoint of glides, the vowel files do include several thousand such measurements.² Glide measurements were made particularly for back glides that are shifted frontwards,³ the midpoints and endpoints of “Southern breaking”,⁴ and the “Northern breaking” of short-*a* into two morae of equal length.⁵

¹ In such residual cases, the normal course is to consider additional measurements of duration, F0, F3, or bandwidths, but the use of these well-known parameters has not in general proved useful in accounting for anomalies in F1/F2 measurements.

² In the vowel files provided with the accompanying ANAE CD, this coding appears in curly brackets following the word identification. The codes {f,b,i,m} represent front upgliding, back upgliding, ingliding and monophthongal vowels respectively. {s} represents shortened monophthongs, {br} the second half of a broken /æ/. The notation {g} is used whenever the measurement represents the endpoint of a glide.

³ Chapter 12 notes that “The 7036 Telsur records of /uw/ include 42 tokens where such a fronted upglide was noted by the analyst”.

⁴ Often referred to as the Southern drawl; see Chapter 18.

⁵ Chapter 13 presents a detailed analysis of this phenomenon, which includes the 1,025 measurements of the second half of such tokens.

5.6. Format and content of vowel files

Log files of the vowel analysis conducted using CSL were produced, facilitated by macros written for the Telsur implementation. These log files were transformed into the six-field format used by the Plotnik program, which produces the displays of vowel systems in this volume. These 439 files are found in the **Telsur/pln** folder on the ANAE CD,

The input format for PLOTNIK is a comma-delimited text file which may be read with a text editor like Word, or a spreadsheet like Excel, saving the data as text with comma delimiters between items (Excel's CSV format). The file begins with the following format:

```

line
1 Thelma M., 31, Birmingham, AL TS 341
2 560,6.992571
3 480,1808,,1.1118,1,slip {i}
4 539,1531,,1.1121,1,fi2 {g} -5-
5 364,2188,,1.1121,1,fi2 {i} -5-
6 378,2246,,1.14264,1,kidney2 {i}
7 451,2173,,1.16123,1,mixed2 -- 8p
8 524,2173,,1.16123,1,mixed -- 8p; hi pitch; F1 from spectrogram"
.. ....
    
```

The first line is a **header** with information the speaker's name, age, place of origin and an identifying number.⁶ The second line gives the number of vowels measured (number of tokens) and the group log mean for normalization. The third and all following lines contain the tokens themselves. Each token consists of six items separated by a comma.

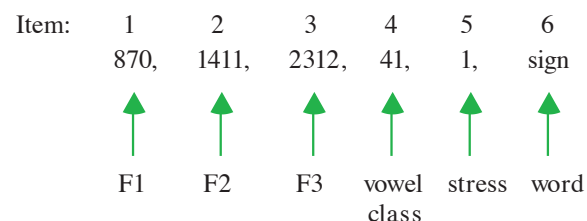


Figure 5.1. Format of data token

The first three items are integral formant measurements in Hertz; F3 is blank in the Telsur files.

The fourth item is the vowel class, a number or letter code for the structural category of which the particular token is an instance. These numerical codes, based on the subsystems of the initial position in Chapter 2, are explained in detail in the documentation for the ANAE CD.

The fifth item in the string is reserved for impressionistic ratings of **Stress**, with values of 1 (primary), 2 (secondary), or 3 (unstressed). This is always 1 in Telsur files.

The sixth item is a descriptive comment. It begins with the standard orthographic representation of the word in which the vowel token occurs, and also contains information on the presence or absence of a glide along with its direction, contextual style, and observations of the analyst on any unusual auditory or acoustic aspects of the signal. The sixth item might read:

bad {i} -4- 10p; interp. F2.

This indicates that the word being analyzed is *bad*; that there was an inglide after the nucleus; that the contextual style was 4, the semantic differential (see Chapter 4); that the analysis was done at 10 poles, rather than the 12 poles that was the default filter order for this speaker; and that the measurement of F2 was interpolated between two neighboring LPC points, because the F2 point corresponding in time to the desired F1 point was missing. The majority of entries are not this complex, and contain no more than the numerical and orthographic identification of the token.

5.7. Normalization

An essential feature of all ANAE analyses and comparisons of vowel systems is normalization, the adjustment of all vowel systems to a common framework that eliminates differences in acoustic realization that are due to differences in vocal tract length. Studies such as Peterson and Barney 1952 illustrate the fact that men, women, and children have very different physical realizations of vowels that sound "the same" to a listener. The task of normalization is to find a mathematical function that does the same work as the normalizing ear of the listener, compensating for the physical differences in articulatory systems. At the same time, we must preserve those differences in phonetic realization that are actually present in the speech community; the sound changes that ANAE is designed to study may be realized as actual differences between the speech of men, women, and children.

Although several studies have shown that the relationship between men's, women's, and children's vowel systems is not exactly linear, several linear functions give a good approximation. One of these is the log-mean normalization explored by Nearey (1977). Labov (1994) reports the studies of four normalization methods by the Philadelphia project on language change and variation. Of the various methods tested, the log-mean normalization was most effective in eliminating male-female differences due to vocal tract length and preserving the social stratification of stigmatized variables that had been established by auditory impressions.

The log-mean normalization is a uniform scaling factor based on the geometric mean of all formants for all speakers.

$$G = \frac{\sum_{k=1}^p \left(\sum_{j=1}^m \left(\sum_{i=1}^n \ln(F_{i,j,k}) \right) \right)}{m * \sum_{i=1}^p n_i}$$

Here *p* is the number of speakers measured; *m* is the number of formants for the Telsur data is 2; and *n* is the number of tokens measured for a given speaker. To normalize any given speaker, this group log mean *G* is subtracted from the individual log mean *S* for that speaker:

$$S = \frac{\sum_{j=1}^m \left(\sum_{i=1}^n \ln(F_{i,j}) \right)}{m * n}$$

⁶ More complete identifying information is found on the Telsur/Master.wks spreadsheet under the TS number.

The anti-log of this difference is the **uniform scaling** factor F for that individual.

$$F = \exp(G - S)$$

For a man, the scaling factor F will be a number greater than 1, and his system will be expanded; for a woman, F will be less than 1, and the system will be contracted. The end result is a series of vowel systems that can be superimposed on a single grid, where differences in the means of different vowels display the course of the sound change in progress.

In the course of the Telsur project, the parameter G was successively updated as the number of subjects increased. Beyond $n = 345$, no significant change in G was found, and the group log mean was kept at the figure calculated for these 345 subjects, $G = 6.896874$.

Unnormalized Telsur files have the extension .plt; normalized files are identified with the extension .pln.

For a recent view and comparison of methods of normalization, see Adank 2003.

5.8. Analyzing and displaying vowel systems with the Plotnik program

Plotnik is a program developed at the University of Pennsylvania Linguistics Laboratory by W. Labov for the display and analysis of complex vowel systems in English and other languages. The vowel charts found in Chapters 12–20 of

ANAE are outputs of the Plotnik program, which is included on the ANAE CD along with a tutorial, and internal and external documentation. At present it is compatible only with Macintosh operating systems, and is supplied in both OS 9 and OS X versions.

Plotnik normally takes as input a Telsur file with the extension .plt or .pln. The program then displays all vowel tokens, tokens for a single vowel or any subset, with or without means or median values displayed. The program automatically codes each token for environmental features, reading from the orthographic representation. A single keystroke will display any of the subsystems of the initial position of Chapter 2. Function keys highlight vowels before nasals, liquids, before voiceless consonants, or in final position.

Plotnik calculates and displays means and standard deviations for all or some vowels, and for any two vowel means, it calculates a t-test of the statistical significance of the difference between any two vowel means. The program operates upon any subset defined by environment, style or stress. Endpoints of glides may be plotted and connected with their nuclei.

For the rapid analysis of a given subset of vowels across the entire population being studied, Plotnik will open new files and plot only the last set of vowels examined. This permits a rapid survey of a given phonological feature for many individual speakers.

Specific configurations that are labeled and equipped with a legend may be saved and retrieved.

