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 the 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.