9 More on vowels[1]

Plotting and normalization

Dominic Watt, Anne Fabricius, and Tyler Kendall

Introduction

In this chapter, we present theoretical and practical issues behind the topics of plotting and normalization, and describe some of the strategic choices that must be made in the course of a sociophonetic investigation. We discuss plotting methods and design desiderata, and a typology of normalization methods (inspired by Adank 2003 and Thomas and Kendall 2007) providing brief descriptions of some of the better-known techniques.

Plotting

The transformation of numerical counts or measurements into graphical formats "provides a front line of attack, revealing intricate structure in data that cannot be absorbed in any other way. We discover unimagined effects, and we challenge imagined ones" (Cleveland 1993: 1). Or, less poetically, "Nothing beats a picture" (K. Johnson 2008: 6). So indispensable are graphical techniques for representing quantitative and qualitative data in linguistics that it is hard to imagine how the field could have progressed without them.

However, many current sociolinguistic methodology texts (for example, Milroy and Gordon 2003; Tagliamonte 2006, Litosseliti 2010) give little information on this important topic. The value of carefully chosen graphical techniques is rarely spelled out, even though a well-designed graphic and the integration of the graphic with the prose can be crucial to the impact of a study's findings. Within sociophonetics there are established graphical conventions for data presentation, but best practice is also still in the process of formation. It is easier than ever to experiment and be creative with graphical techniques, owing to the flexibility of current software packages (see the companion website). Of course, well-designed visualizations are only meaningful if used in support of a sound piece of linguistic analysis; the former is never a substitute for the latter.

Joos (1948) is credited with the first published example of a plot in which the axes representing the first and second formants are oriented such that the plot mimics the traditional vowel quadrilateral (as in the figure in this text). Much more widely reproduced than Joos' figure is Peterson and Barney's (1952: 182) plot for American English monophthongs. Despite their age, these data are still used as reference values for American English (e.g., Cebrian 2006). For reasons unknown, Peterson and Barney chose not to follow Joos' recommendation and displayed their formant data with F1 on the *x*-axis and F2 on the *y*-axis without reversing the scales, though earlier in their paper they use a plot (their Figure 3) which is configured in the now conventional manner.

Formant plots were first used in sociophonetic research in the LYS survey of American

English, and the conventions they established remain largely unchanged, at least in North America. *ANAE* and Thomas (2001) *Acoustic Analysis of Vowel Variation in New World English* contain numerous examples of plots allowing direct visual comparison of samples for speakers of a wide range of accents. It is straightforward to represent within-speaker variation (resulting from a change in speech style, for instance) using judiciously chosen markers and symbols (see "Interpreting vowel plots," below). Similarly, change-in-progress can be detected by comparing patterns of vowel distribution between speakers, styles, and generations. Scatterplots have been used to good effective to establish the relative chronology and directions of stages of the Northern Cities Vowel Shift (NCVS) (Labov 1994: 185–200), for example, and also in the proliferation of work on other chain shifts, mergers, and splits in North American English vowel systems (e.g., Labov 1994, 2001). Plotting allows the inference of directionality and the conceptualization of phenomena-like chain shift as coordinated processes.

For all its obvious utility, and although virtually *de rigueur* in North American sociophonetics, formant plotting has been rather slower to catch on elsewhere. Cases where the method has been used in analyses of British English include Wells (1962), Harrington, Palethorpe, and Watson (2000), Watt and Tillotson (2001), Kamata (2006), Fabricius (2007), and Kerswill, Torgersen, and Fox (2008). The technique has also been used recently in Australia and New Zealand (e.g., Watson, Harrington, and Evans 1998; Cox 2006; Maclagan and Hay 2007).

There are favored and more or less conventionalized ways of representing the results of vowel analyses. For your own analysis it is worth taking time to evaluate the pros and cons of different practices. In the next section we examine some of the practical considerations to take into account when designing your own figures.

Practicalities of plotting

Before generating your own plots, there are a number of practicalities to consider beyond specifics related to the nature of vowels, acoustic space, and linguistic analysis. Straightforward issues—such as whether you can use color or whether you can (or need to) include IPA fonts—have implications for the presentation of your data.

- Are your plots intended for exploratory data analysis? If so, something "rough and dirty" may suffice and the main criteria should be that you can generate the plots easily and make changes without much manual work. You don't want to finalize a plot until it is clear no further changes will be necessary.
- Are your plots for on-screen presentation through means such as PowerPoint? If so, you have flexibility in terms of format and the use of color, but you should ensure that the points of interest are well indicated so that the audience can quickly focus on key aspects of the plot. Make sure that your colors, fonts, and sizes are intelligible when viewed on a projection screen. Animation—e.g., with different vowel series appearing sequentially—might be used to good effect.
- Are your plots for inclusion on a presentation handout? If so, the possibility that the final figure is small and produced on a low quality photocopier should be considered. It is frustrating to have created "perfect" plots only to find they are illegible on the photocopies you distribute.
- Are your plots for a publication? If so, the publisher's resolution requirements must

be taken seriously. Can you—or are you required to—submit vector graphics, which resize seamlessly? Which labeling conventions can or must you use—IPA (e.g., /æ/), lexical frame/keyword (e.g., BAT; Wells 1982a, b, and c; Chapter 8), SAMPA/darpabet (e.g., @, used by most speech engineers), or something else entirely? Publishers' constraints may ultimately dictate your decisions more than anything else.

Plotting decisions should ultimately be driven by the needs of your audience and the medium as well as the specifics of your message. There is also a balance between following conventions (such as using F1–F2 vowel plots, as discussed above and exemplified in Figure 9.1 and weighing the purpose of your work.

Questions about vowel measurements were discussed in the previous chapter, but many such questions carry over into decisions about plotting. In conventional F1–F2 vowel plots, such as in Figure 9.1, monophthongs are often plotted as single points, while diphthongs and triphthongs (vowels considered as continuous moving trajectories) raise more complicated questions.

These are questions of both convention and theory. The actual F1 and F2 values for a monophthong will change during its articulation, but it is conventional to portray monophthongs as single F1 and F2 values indicating the steady state (or, failing this, the point of maximum formant displacement, or simply the temporal midpoint, as discussed in Chapter 8) of the vowel's articulation.[2] For diphthongs, it is important to indicate the direction of movement of the vowel over the token's time course. This is usually accomplished by plotting diphthongs as vectors with arrowheads (as in Figures 9.1 and 9.2) or by symbolizing the glide differently from the nucleus. We may plot monophthongal productions of vowels typically considered diphthongal in the same way as we plot (phonetic) diphthongs in order to fully illustrate a speaker's or group's level of monophthongization. For example,

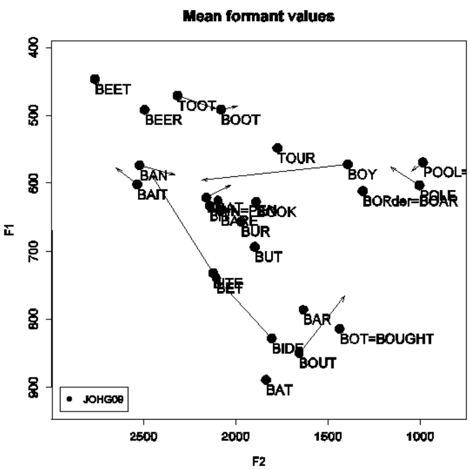

Figure 9.1 F1–F2 vowel plot showing mean values for a US White female (JOHG09) from central Ohio. (Generated by NORM from NORM's sample data). Diphthong trajectories are indicated by arrows. The positions of some point labels have been adjusted for clarity.

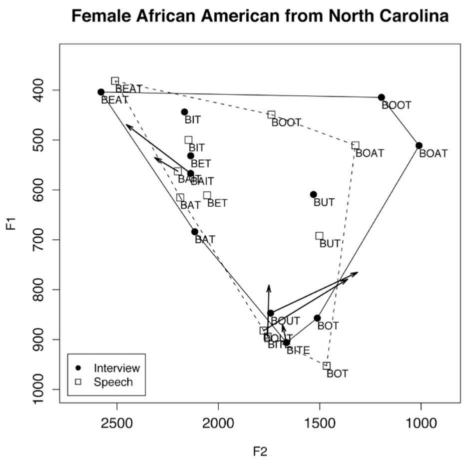

Figure 9.2 F1–F2 plot of vowel means for an African American female in the southern US, in two speech settings. (Generated by NORM; adapted from Kendall and Wolfram 2009.)

(ay) (in, e.g., *my time*) in Southern American English is often monophthongal, but is typically marked using a diphthong symbol even when produced by speakers perceived to use monophthongs exclusively.

In terms of the conventional F1–F2 plot, whether to plot individual vowel measurements or central tendencies (typically mean values) is an open question. Some scholars (e.g., *PLC*) prefer to plot "clouds" of individual vowel tokens, while others (e.g., Thomas 2001) plot vowel means, as in Figures 9.1–9.3. The solution depends on the goals of the analysis and presentation. Plotting individual vowels represents more accurately the fluidity of F1–F2 vowel space and overlapping vowel production, and makes outliers and

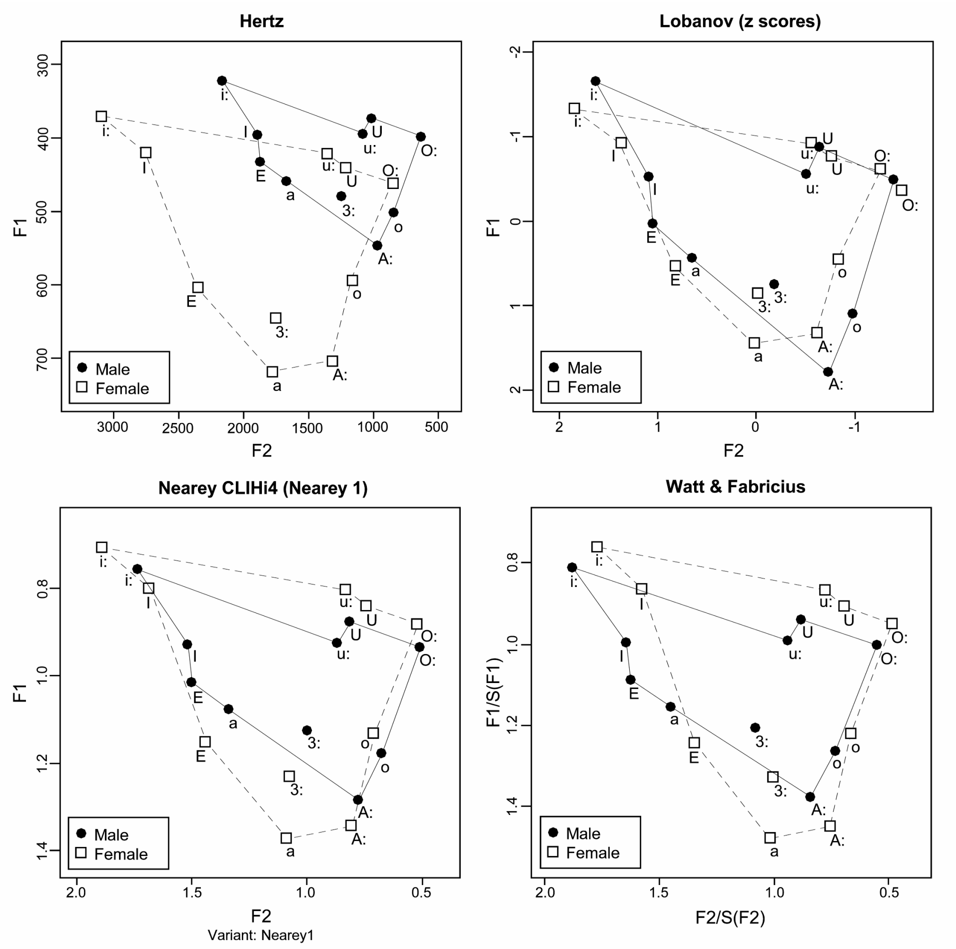

Figure 9.3 NORM-generated plots of F1–F2 means for monophthong vowels spoken by two speakers of Southern Standard British English (male = filled circles; female = open squares; data from Hawkins and Midgley 2005 and Moreiras 2006, reproduced with permission). Lines linking the peripheral vowel means have been superimposed to assist comparisons within and between plots. *Top left* = means in Hertz; *top right* = Lobanov normalized (z scores); *bottom left* = Nearey CLIH$_{i4}$ (Nearey1 in NORM) normalized; *bottom right* = Watt and Fabricius normalized. The positions of some point labels have been adjusted for clarity.

measurement errors easier to spot. Furthermore, a scatter of points for a particular vowel helps identify the direction of any ongoing changes. If we assume a particular vowel's pronunciation is stable, we expect its field of dispersion of points to be approximately circular in shape, reflecting the inevitable variation in the vowel's articulation that arises from, for example, the influence of neighboring consonants. Elongation of the cloud of points in one direction or another, however, might suggest that the vowel's target is moving in articulatory space (back vowel fronting, for example). Plotting mean scores gives a simplified representation of overall tendencies within the vowel space, on the other hand. Naturally, these two options are not in opposition; there are ways to incorporate the benefits of each. For example, error bars (e.g., Maclagan and Hay 2007) or standard deviation ellipses (e.g., Harrington 2006: 447) can enhance plots of mean scores by indicating the range of variation in the data.

Interpreting vowel plots

It is tempting to interpret differences between vowel productions of speakers or groups based solely on differences visible in vowel plots, but such diagrams should be interpreted cautiously. For example, Figure 9.2 presents the mean vowel productions for a female politician from a town in North Carolina across two speech events: a sociolinguistic interview with two White fieldworkers and a public speech to her local, African American constituency (Kendall and Wolfram 2009). There are both similarities and differences in the two vowel spaces in Figure 9.2. Most noticeably, the speaker's BOAT and BOOT vowels are more front in her public speech than in her interview, as indicated by differences in F2, which a t-test confirms to be statistically significant. Statistical tests also indicate that the height of BIT is different between the speech events, but that the differences in the heights of BOT—which appear to be about the same distance apart as those for BIT—are not significant. That is, visible differences do not always imply statistical differences. Further, to argue that these differences are socially meaningful, we should go beyond quantitative differences and question whether they are perceptible.

It is also just as possible to under-interpret differences among varieties based on similarities between vowel plots. F1 and F2 are just two of many acoustic features of vowels, as discussed in the previous chapter, and a simple plotting of the vowel midpoint may not fully specify the information that is of perceptual importance to hearers. Generally, you should be wary of visible differences that are inaudible, and if an audible difference is not visible, it is likely that the plot has not represented the right information (as discussed in other chapters, you should always double-check your measurements. However, Chapter 8 also warns us that there are audible differences that may be important linguistically, but may be below our usual level of conscious awareness). We reiterate that best practices are still being developed and that there is ample scope for new approaches beyond the conventional F1–F2 plot.

Normalization

Normalization refers here to the factoring out of physical (i.e., acoustic) differences in vowel production resulting from anatomical differences between speakers. It has long been assumed that listeners process incoming speech signals in this way, by unconsciously compensating for the higher formant frequencies for a given vowel phoneme—say, [i]—of a female speaker compared to a male one (those for a child being higher still). The listener will nonetheless report hearing an [i] vowel in each case. The process by which we can recover intended vowel identities, at times based on just a single vowel, is still not fully understood, however (see Pisoni 1997; Adank 2003; Nearey and Assmann 2007).

The major motivation for normalizing data is that through normalized data one can directly and *quantitatively* compare speakers' and speaker groups' vowel productions with one another; one can, for example, make statements like "in community X, for every decade we add to a speaker's year of birth, vowel Y is Z units fronter," a practice originating in Matthew Lennig's analysis of vowels in Parisian French (Lennig 1978). We should not attempt to make this kind of generalization using unnormalized measurements if we are dealing with data from both men and women, as will most often be the case in community studies.

Normalization of vowel formant frequency data has thus become the standard practice in sociophonetics. If researchers have chosen to leave their data as raw Hertz values, some explanation of this fact is now expected. Should it be found that vowel data for a same-gender group varies markedly in terms of the areas of individuals' vowel spaces, within-gender comparisons of adult speech may also necessitate normalization, or at least conversion of Hz values to a psychoperceptual scale such as the equivalent rectangular bandwidth (ERB) scale (Moore and Glasberg 1983), as in Hawkins and Midgley's (2005) study of 20 male RP speakers.

Mathematical transformations that are used on acoustic data are of two main types. The two types reflect the stages of processing carried out on incoming acoustic signal, first by the peripheral auditory system, then by the auditory processing centers of the brain.

The first type approximates the non-linear frequency response of the sound-detecting hair cells of the inner ear, which are much more sensitive to changes in frequency at the lower end of the frequency spectrum than to equivalent changes in frequency higher in the spectrum. These algorithms convert raw (linear) measurements in Hertz units into their psychoperceptual equivalents. Bark units, based upon the critical bandwidth sensitivity of the inner ear (Traunmüller 1990, 1997), are especially frequently used. One critical bandwidth is 100 Hz in the range between 150 and 250 Hz, but 350 Hz between 2,150 and 2,500 Hz,[3] similar to the Western musical scale of semitones and octaves. The correspondence between Hertz and Bark is approximately linear between 0 Hz and 1 kHz, but the two scales diverge markedly in the frequency bands above this. Traunmüller's formula for converting Hz values into Bark units is:

(1) $z = [26.81/(1 + 1960/f)] - 0.53$

where f is frequency in Hertz.

Conversion of Hertz values into Bark results in a picture of the vowel space in which the higher frequencies seen in F2 and higher formants are "compressed" relative to F1, such that the perceptual primacy of changes in F1 over changes of similar size in higher formants is emphasized.[4] Conversion into Bark gives a stronger sense of how differences in vowel quality are perceived by listeners: e.g., large differences in frequency in the higher formants of the speech of women and children may actually be imperceptible. We consider Bark-transformed data again below. Note that other non-linear psychoperceptual warpings of the Hertz scale (e.g., mels and Koenig units; see Table 9.1 on the book website) have been proposed in the past, and several continue to be used. The ERB scale mentioned above, which, like Bark, approximates the frequency sensitivity of the inner ear, is used increasingly often among phoneticians, as it is considered to model frequency response more accurately than Bark.

Normalization proper

We devote the remainder of this chapter to discussion of *normalization proper*, by which we mean the application of a mathematical procedure to sets of data from different individual speakers. This (at least loosely) replicates auditory processing in the brain. We can follow Thomas and Kendall (2007), quoting Disner (1980) and Thomas (2002), and describe the effects of normalization through four general goals:

(a) to eliminate variation caused by physiological differences among speakers,
(b) to preserve sociolinguistic/dialectal/cross-linguistic differences in vowel quality,
(c) to preserve phonological distinctions among vowels,
(d) to model the cognitive processes that allow human listeners to normalize vowels uttered by different speakers.

The balance between goals (a) and (b) is central. While it is well known that adult male speakers have, on average, larger and longer vocal tracts than adult female speakers (see, for example, Goldstein 1980), there can also be cases where sociolinguistic variation differentiates male and female speakers in a community. An analysis might also seek to examine perceptible differences in pronunciation between age groups, thus identifying changes in progress. Intra-speaker style differences or regional differences could also be brought into focus (e.g., Fox and Jacewicz 2008). These differences are all interesting, and need to be preserved in normalized data. Point (c) has often been used as a benchmark for comparisons of vowel normalization methods, such as in Adank (2003) and Deterding (1990). It is not often employed in descriptive sociophonetic studies where the phonological distinctions between vowels or conditioned allophonic differences (such as tokens of tensed /a/ in Northern American English dialects, or of British English /əʊ/ in pre-lateral contexts, e.g., *goal* versus *goat*) can be presupposed (although this is of course not so in the case of putative mergers in progress).

Point (d) can also be relevant. The processing mechanisms that derive equivalent percepts from acoustically distinct inputs are of clear sociolinguistic interest and relevance. A speaker grows up within and functions as a member of a speech community not only through his or her sociolinguistic productions, but also through sociolinguistic perception, learning, for instance, how to categorize vowel tokens into their appropriate phonological categories. Sociolinguistic aspects of perception are made most noticeable through cross-dialectal misunderstandings (*PLC* and Labov, forthcoming).

The abundance of normalization algorithms in the literature has led to a series of studies carrying out different evaluation procedures on the various normalization methods; see, for example, Nearey (1977/8); Hindle (1978); Disner (1980); J. Miller (1989); Deterding (1990); Rosner and Pickering (1994); Adank (2003); Adank, Smits, and van Hout (2004); Clopper (2009). These works can be highly technical and challenging for those lacking a mathematical background, but the last three in particular can be recommended.

Given the array of algorithms available, how does an investigator know when a normalization routine has done a satisfactory job? An inductive approach to this question would mean running through a range of routines and comparing their effects on data before settling on a single procedure. We can also ask: when is the normalization task complete? Is there a gold standard choice of algorithm that would provide perfectly normalized data, and what would those data look like? The answers to these questions depend on what we seek within the transformed data. Sociophoneticians want to eliminate the effects of physiological differences (especially those of divergent vocal tract lengths) from their data, but typically nothing else; for their purposes socially-important variability should not be "tidied up" much by normalization. Laboratory phoneticians, on the other hand, usually seek to remove vocal tract length differences *and* minimize the scatter of points around mean values for individual vowels resulting from features other than phonological context, speech rate, and the like, depending on what the experiment is seeking to show.

A typology of normalization methods

One way to understand the implications of a normalization method is to determine what type of input information the algorithm itself incorporates. Normalization methods operate on vowel formants, most typically F1 and F2; some also utilize F3, others the fundamental frequency, F0. Formants differentiate vowel tokens, and normalization methods either operate mathematically on formant information from tokens of one vowel at a time, or from tokens of multiple vowels each time the algorithm is performed. Finally, data come from individual speakers, and a range of normalization methods derive their input information from one speaker at a time, while others use input from a population of speakers. Each of these parameters—*formant*, *vowel*, and *speaker*—provides input information to a normalization algorithm; the typological difference hinges on whether one or multiple inputs are involved at each level. The former is defined as an "intrinsic" method, the latter an "extrinsic" method. The parameters are also cross-cutting, in that, for example, a particular normalization algorithm can be vowel-extrinsic but formant-intrinsic. Table 9.1 (on the book website) provides a classification of the transforms and normalization algorithms discussed in this chapter, according to their intrinsic/extrinsic dimensions, and clustered into groups based on typological similarity.

Bark, mel, Koenig, and ERB

This group contains the psychoperceptual transform scales discussed earlier. Adank (2003: 13) notes that "the majority of the intrinsic procedures [were] developed in the field of perceptual phonetics with the purpose of modelling human vowel perception." These transforms are most often employed as normalization methods in acoustic phonetics-oriented studies. Harrington (2006) and Harrington, Palethorpe, and Watson (2000) used Bark in their studies of real-time vowel change in Queen Elizabeth II's speech. ERB is used in Hawkins and Midgley's (2005) study of vowel qualities in RP across time. Use of the mel scale, which divides the 0–1 kHz band into 1,000 mel units on the basis of listeners' judgments of equivalent steps in pitch, is rare in the sociophonetic literature; it has, however, been used in studies such as Potter and Steinberg (1950) and Hillenbrand et al. (1995). Mathematical algorithms for all the above transforms can be found in the references in Table 9.1 and online at http://www.ling.su.se/STAFF/hartmut/bark.htm#refs. The Koenig scale has not been used in any sociophonetic research we are aware of; examples of studies using this algorithm can be found in Putnam and O'Hern (1955) and Hillenbrand and Gayvert (1993).

Syrdal and Gopal

Syrdal and Gopal (1986) describe a vowel-intrinsic, formant-extrinsic procedure which appears in modified form on the NORM webpage as the "Bark Difference Metric." Hertz values are converted into Bark and then used to obtain differences which model degrees of advancement (conventionally the horizontal dimension in vowel plots) and vowel height (the vertical dimension). NORM modifies the Syrdal and Gopal method to avoid using F0 because of the susceptibility of F0 to aging and other factors making it unreliable to include in a normalization algorithm (Thomas and Kendall 2007). Ladefoged (2001: 200) also employs a simple "difference metric," plotting F2 minus F1 on the horizontal axis, but in the same book and in later publications (e.g., Ladefoged 2005b) he makes use of plots in which F1 is scaled linearly and F2 is warped using an unspecified non-linear transform.

Gerstman, Lobanov, Nearey CLIH$_{i4}$, and Watt and Fabricius

All vowel-extrinsic and formant-intrinsic procedures operate with a conception of the entirety of the vowel space as contributing to the normalization of a single vowel token, one formant at a time. One underlying acoustic assumption made in this type of procedure is that intervals between different speakers will be constant for all individual vowel comparisons in the vowel system, although modulated between different formants. This assumption would not be accepted by all acousticians (see, for example, Fant 1973), and these methods are less commonly employed in experimental acoustic studies.

The various algorithms differ as to how the formants of other vowels contribute mathematically to the procedure: by means of ranges (Gerstman), z-scores (Lobanov), individual log-means (Nearey) or centroids (Watt and Fabricius). The group includes the two most successful methods examined in Adank's (2003) comparison of procedures for sociolinguistic purposes, Lobanov and Nearey CLIH$_{i4}$, a reduced version of which is available as Nearey 1 on the NORM webpage).[5] Interest in Lobanov's procedure has increased since the publication of Adank, Smits, and van Hout (2004), which recommends it.

The Watt and Fabricius (2002) procedure is inspired by earlier studies utilizing a vowel space centroid such as Koopmans-van Beinum (1980). It has been used in studies of modern RP short vowels (Fabricius 2007), London English vowel systems (Kamata 2006), vowel fronting in South African English (Mesthrie 2010), and, in a modified form, Southern Illinois English (Bigham 2008). The procedure was specifically developed for use with sociophonetic data, and can be used inductively and modified to suit the variety under study. The method is also potentially less sensitive to wide variations in sample size than other algorithms. Testing of the implications of the Watt and Fabricius method is in its early stages, but Fabricius, Watt, and Johnson (2009) found it to outperform Nearey1 and 2, although not Lobanov, on measures of improvement of interspeaker vowel space area agreement and overlap. Nearey1, Lobanov, and Watt and Fabricius procedures seemed to perform similarly on vowel juxtapositions (relative planar locations) within the vowel space.

Nordström and Lindblom, Nearey CLIH$_{s2}$

Nordström and Lindblom's (1975) vocal tract scaling transformation and Nearey's shared log-mean model make up the fourth speaker-intrinsic group in Table 9.1. These vowel-extrinsic and formant-extrinsic methods use input into the algorithm from all vowels and all (measured) formants to normalize a single token. These algorithms thus provide "uniform scaling factors" to map speakers' data sets onto each other.

The formant-extrinsic version of Nearey's CLIH procedure is the one most commonly employed in North American sociolinguistic studies. NORM's version is actually CLIH_{s2}, as it uses only F1 and F2, not the full range from F0 to F3 as in Nearey's original (1977/8) study. As far as we can ascertain, CLIH_{s2} is also the basis for the Plotnik software's normalization procedure.

Labov ANAE

The ANAE methods are variations on Nearey CLIH_{i4}, and CLIH_{s2}, with a speaker-extrinsic overlay in the form of a population-derived value known as the G value. This is the only group of methods classified as speaker-extrinsic, employing information from more than a single speaker at a time. Use of this type of procedure presupposes a large population of speakers, and indeed the Telsur G value (G = 6.896974) was determined after sampling of 345 speakers of North American English (ANAE: 39). Acoustic phonetic studies (as distinct from sociophonetic studies) almost never operate with such large numbers of speakers, which is partly why speaker-extrinsic methods are generally absent from the (non-sociophonetic) acoustics literature.

Plotting normalized formant values

Figure 9.3 illustrates the ways in which a selection of different normalization algorithms reconfigure the relative positions of points on the F1–F2 plane. The mean formant values in Hertz in the top left panel show very little overlap in the speakers' systems, and comparability between the two speakers' relative vowel positions is obscured. The normalized plots in the other three panels, by contrast, have the double effect of equalizing vowel space areas for the two individuals *and* superimposing the two vowel systems on each other, so that direct comparisons between a male and a female speaker can be made. However, all three normalization methods represented in Figure 9.3 (Lobanov, Nearey CLIH_{i4}, and Watt and Fabricius) result in different outputs, reflecting their individual mathematics. All three bring the relative areas of the male and female vowel polygons into closer agreement with one another, with Lobanov's method giving the best result. The corner or anchor points (shown here using the NORM-compatible symbols [i:], [O:], and [A:], representing the vowels of the BEET, BOUGHT, and BART lexical sets) are considerably closer together in the plots for all three methods than in the Hertz plot; only in the Lobanov plot are the male and female means for the long central monophthong [3:] (BURT) in very close proximity. The Nearey and the Watt and Fabricius methods appear to compress the F1 range for the male speaker relative to the extent of the female's polygon, and the latter method "stretches" F2 in the front vowel series. On the other hand, Nearey and Watt and Fabricius apparently preserve the original shape of the male's polygon better than Lobanov, and also approximate the [o] (BOT) means for the two speakers more closely. All three methods markedly improve the mapping of the two speakers' systems, however, so despite the discrepancies in Figure 9.3, it is clearly worth normalizing the raw data, whichever particular method is chosen. It must be remembered that there is no "right" way of doing this—after all, the vowels are physically different to start with, and their auditory qualities may in any case not be exactly alike (remembering that sociophoneticians also ultimately have an interest in perceptual differences). It remains obvious, for instance, that the male speaker's [a] (BAT) vowel is much closer to his [E] (BET) vowel than is the case for the female, whose BAT vowel is by any measure backer and considerably more open, and closer to her BART mean.

As an overarching perspective on the process of vowel normalization, we concur with Wolfram (1993: 203), who writes: "[I]t is important for language variationists to be good linguists and good sociolinguists, not simply good collectors of data or good number crunchers." That is, the final product of an analysis will always come about through an interplay of listening and observing on the one hand, and normalization and plotting on the other. It should also be stressed that it is important that sociophonetic researchers tackle the task of normalizing data as appropriate for the material at hand; for the majority of comparisons of adult speech, normalization will be necessary. The selection of any particular normalization algorithm is best made by establishing the implications of the procedure for the data and weighing the advantages and disadvantages of each methodological choice. While Thomas (2002: 174) is correct to comment that choosing a normalization method is "a matter of deciding which drawbacks are tolerable for the study at hand," a well-considered and well-argued choice of normalization method improves the confidence with which you can present convincing results.

Exercises

Exercise 1: making an F1–F2 vowel plot

As demonstrated in Figures 9.1–9.3, the basic F1–F2 vowel plot is generated by plotting F2 values on the *x*-axis and F1 values on the *y*-axis. Importantly, the directions of the axes must be flipped, so that F2 values increase from right to left and F1 values increase from top to bottom, to maximize the plot's similarity to the traditional vowel quadrilateral.

Figure 9.1 was generated using the CentralOhioFemaleNORM.txt file available on the NORM website (http://ncslaap.lib.ncsu.edu/tools/norm/samples.php). This file contains the formant measurements for a female speaker, *JOHG09*.

1. Review the section on plotting programs on the companion website.
2. Then download CentralOhioMaleNORM.txt, which contains data from a male speaker, *OHDMTV_M*. This file is in tab-delimited text format—that is, it is a simple ASCII text file, which uses tabs to separate fields. It should open as a spreadsheet within Microsoft *Excel* or other spreadsheet software.
3. Generate mean values for F1 and F2 for each vowel type in the data set. The easiest way to do this will be to use a spreadsheet application, like Microsoft Excel. (Use spreadsheet functions to calculate the means to avoid calculation and entry errors.)
4. Plot the mean values using a spreadsheet application, or even by hand, to get a feel for the process. Use the software of your choice, such as Excel. Try to make your plot look as much like Figure 9.1 as you can (of course, the vowel data come from different speakers, so will not be identical).

Exercise 2: comparing F1–F2 vowel plots

Compare the plot you made in Exercise 1 for the male speaker from Ohio, *OHDMTV_M*, with the plot for the female speaker, *JOHG09*, shown in Figure 9.1 on the book website.

1. How do these two vowel spaces compare? Are you able to make any judgments regarding vocalic differences between these two speakers based on a visual inspection of their *separate* vowel plots? You may want to review Thomas (2001: Chapter 3), which provides a good survey of vowel variation in Ohio.
2. The data from both speakers, *OHDMTV_M* and *JOHG09*, are available in a combined datasheet in the file CentralOhioNORM.txt on the companion website. Plot mean values for the vowels for each of these two speakers on the same plot, either by hand or through the software used for Exercise 1.
3. Are there any problems with plotting the data from these two speakers together as you were asked to do in question 2?
4. Finally, note the absolute differences in terms of the formant values (e.g., comparing BEET, BAT, BOUGHT, . . .) for the two speakers. Can you explain what accounts for these differences?

Exercise 3: normalizing formant data

Use the NORM website to test different normalization techniques on the vowel data from Exercises 1 and 2. To use NORM, point your browser to http://ncslaap.lib.ncsu.edu/tools/norm/, and follow the links to learn more about NORM, or to normalize and plot vowel data.

1. Experiment with various normalization techniques on the data in the CentralOhio NORM.txt file. How do these techniques compare in their visual effect on these data?
2. More specifically, as in the discussion of Figure 9.3 above, compare an unnormalized plot to plots of Lobanov, Nearey1, and Watt and Fabricius normalized vowels. Can you come up with a principled argument for why one of these methods might be more (or less) appropriate for assessing sociolinguistically-relevant differences between these two speakers than the other methods?

Notes

1. We wish to thank Doug Bigham, Marianna Di Paolo, Bernhard Fabricius, Robert Fox, Frans Gregersen, Jonathan Harrington, Eva Jacewicz, Christian Jensen, Daniel Ezra Johnson, Inger Mees, Geoffrey Morrison, Francis Nolan, Jennifer Nycz, Nicolai Pharao, Tom Purnell, Erik Thomas, Roeland van Hout, and Malcah Yaeger-Dror for comments and discussion of the contents of this chapter.
2. However, Chapter 8 recommends at least three measurement points even for both presumed monophthongs as well as diphthongs.
3. See the conversion table provided at http://HztoBark.notlong.com.
4. Labov (2001: 167–68) contends that English speakers tend to utilize F1 differences principally for the purposes of phonological contrast, while F2 differences are more liable to acquire socio-indexical functions. Irrespective of their putative perceptibility, and regardless of the language under analysis, it is vital that F2 (and F3) differences of potential or known sociophonetic importance not be obscured when any transformation of the original Hz data is carried out.
5. CLIH$_{i4}$ (constant log interval hypothesis) indicates that the individual log-means for four dimensions—F0, F1, F2, and F3—are calculated. See Nearey (1977/8) and Adank (2003: 22).