Part II

Key Life Stage Events Across the Lifespan

5 Longitudinal Sociophonetic Analysis

What to Expect When Working With Child and Adolescent Data

Mary Kohn and Charlie Farrington

1. Introduction: Childhood and Adolescence Are Central to Theories of Language Change

Childhood and adolescence have long captured the attention of linguists due to the dynamic transitions that occur during the first 20 years of life. Hypotheses about everything from sound change (Labov 2001) to the ways in which life stages influence style (Eckert 1997) all hinge on the ways in which individuals adapt and change their speech during this time period. The importance of individual dynamicity during childhood and adolescence motivates longitudinal research for this life stage. Yet, questions about whether child data can be compared to adult data have frequently prevented the inclusion of childhood data in longitudinal and apparent-time studies alike. When working with data spanning different developmental stages the investigator must "normalize out" acoustic correlates of physiological change in order to identify linguistic changes potentially related to social variation. What is the youngest age group that can be included in longitudinal quantitative acoustic analysis? When are comparisons of acoustic data justified and when might they be misleading?

We argue that childhood data can be included in longitudinal studies if researchers are cognizant of developmental patterns for vocal tract morphology and the subsequent influence these patterns have on acoustic correlates. We have designed this chapter to familiarize the longitudinal researcher with trends associated with developing vocal tracts, approaches towards accommodating child vowel data in longitudinal studies, as well as potential pitfalls of working with very young children. We first describe typical paths of development observed in cross-sectional/apparent time data in which different age cohorts are compared as a proxy for change over time. We then utilize our longitudinal corpus to confirm these developmental paths, thus illustrating that changes to the vowel space are predictable across developmental stages. We use this information to discuss when normalization procedures can be applied to child data to reduce physiological variation but to maintain sociolinguistic variation. Finally, we conclude with a brief discussion of statistical analyses for longitudinal acoustic data.

Although child vowel spaces are in essence moving targets, we demonstrate that variation between adult and child time points attributed to physical differences can be anticipated and reduced using the same kinds of methods sociolinguists use to compare adult male and female speakers. Still, these methods are not appropriate for the youngest age groups due to issues including extreme intra-speaker variation and widely spaced harmonics. Because child data is crucial for understanding social development, methodological concerns should not prevent researchers from including child data in longitudinal studies. An understanding of typical developmental trajectories allows researchers to proceed with appropriate caution when working with child data time points.

2. Population Under Analysis

To illustrate typical acoustic paths of development we compare cross-sectional findings with longitudinal data from the Frank Porter Graham (FPG) project, a study that began in 1990 and followed 67 African American children from infancy through early adulthood in Durham and Chapel Hill, North Carolina (see Van Hofwegen and Wolfram, this volume, for background information about the study, and Cieri and Yaeger-Dror, this volume, for more on using data from non-sociolinguistic studies). Here, we focus on a subsample of ten male and ten female speakers. The speakers in this subsample are also analyzed by Van Hofwegen and Wolfram (2010; this volume) and were originally selected because they had the highest-quality recordings of the sample for the time points used (Kohn and Farrington 2012). Four time points were analyzed for each speaker, including ages 10, 14, 16 and 20.[1] For half of the speakers (five males, five females) who produced sufficient data for analysis, we also include an age four time point for analysis. These time points range from prior to puberty to after puberty, providing a picture of development across childhood and adolescence. Formal and informal speech were analyzed at each time point.[2] Each set of recordings includes an informal language elicitation task with the child's mother (ages four and ten) or with a peer (ages 14 and 16), as well a formal language task, such as reading aloud, sentence repetition or the performance of a speech.

Vowels were manually segmented and formant settings were selected on a vowel-by-vowel basis. Vowels were extracted using a Praat script that documented formant settings and measured the first three formants at five equally spaced intervals throughout the vowel using Linear Predictive Coding (LPC) analysis. Only vowels between obstruents over 0.06 seconds were measured unless otherwise noted (e.g. BOAR are pre-rhotic tokens). Over 19,000 tokens of BEET, BIT, BAIT, BET, BAT, BAN, BITE/BIDE, BUT, BOT, BOUGHT, BOAT, BOWL and BOAR[3] were collected, resulting in 80–200 total vowel tokens per speaker per time point (see Table 5.1, below). Further information regarding data analysis can be found in Kohn (2014).

Researchers who are interested in additional information about measuring child vowels in Praat, along with other issues related to child data collection, should consult Khattab and Roberts (2011).

3. Acoustic Correlates of Development

The most common acoustic variables measured in sociophonetic research, including duration, pitch and formants, all display developmental trends across early childhood, and for some variables, into early adolescence due to a number of factors including increasing motor control, changing vocal tract morphology and acquisition of adult-like phonology (Vorperian and Kent 2007: 1511; Hodge 2013). The ways in which these factors interact produce independent maturational trajectories for different acoustic correlates. For example, duration is likely to correlate with developing motor control, while pitch values may show an interaction between changing vocal fold lengths and acquisition of adult language norms.[4] In addition, sex distinctions, as well as possible gender distinctions related to the acquisition of cultural norms, lead to distinct developmental patterns for boys and girls (Vorperian et al. 2011; Fitch and Giedd 1999). Sex differences that emerge after puberty are striking, with voice changes initiating between approximately age 12.5 and 14.5 for the majority of males and lasting for six or more months (Hollien 2014). While these multiple developmental paths may appear daunting to the researcher who would like to document changing sociolinguistic variables across the lifespan, many of these trajectories can be anticipated, and thus controlled for, thanks to the rapidly expanding documentation of child speech and vocal tract morphology. Below, we draw from some of these large-scale cross-sectional studies supplemented with our own longitudinal data to illustrate predictable developmental patterns found across child speech for Vowel Space Area (VSA) (section 3.1), F1 and F2 for corner vowels (section 3.2), as well as vowel duration and dynamics (section 3.3).

3.1 Vowel Space Area and Vowel Dispersion Measures

VSA refers to the acoustic space that an individual uses to produce their full vowel system. Recently, measures of VSA have received increasing attention due to their clinical and theoretical utility. VSA measures can serve as a diagnostic for atypical speech development, grant insight into how children adjust vowel systems as they adapt to developing vocal tract morphology (Flipsen and Lee 2012), and capture variance related to hyperarticulation or undershoot (Pettinato, Tuomainen, Granlund, and Hazan 2016). Childhood and adolescent trajectories of change for VSA values are affected by the growing vocal tract anatomy as well as articulatory factors, so that VSAs shrink over the course of childhood and adolescence. During the process of maturation, the resonant frequencies of the vocal tract decline in a non-uniform fashion as the vocal tract becomes larger. As a result, unnormalized differences between high and low vowels, as well as front and back vowels shift over time. Articulatory issues such as lack of fine motor control are likely to cause overshoot, producing larger VSAs for smaller children (Pettinato et al. 2016). The impact of these changes is evident in the vowel spaces of the five time points included in our data in which VSAs shrink at each subsequent time point. Figure 5.1 displays vowel quadrilaterals based on median corner vowel F1 and F2 measures averaged across males and females in this study.[5] This figure illustrates acoustic correlates of growth spurts as children age. A large skip between age ten and age 14 for males, for example, correlates with the descent of the larynx that occurs during puberty. A similar but more gradual decline in overall vowel space is also evident for females. Our longitudinal exploration of VSA confirms patterns found in cross-sectional data, as will be discussed below.

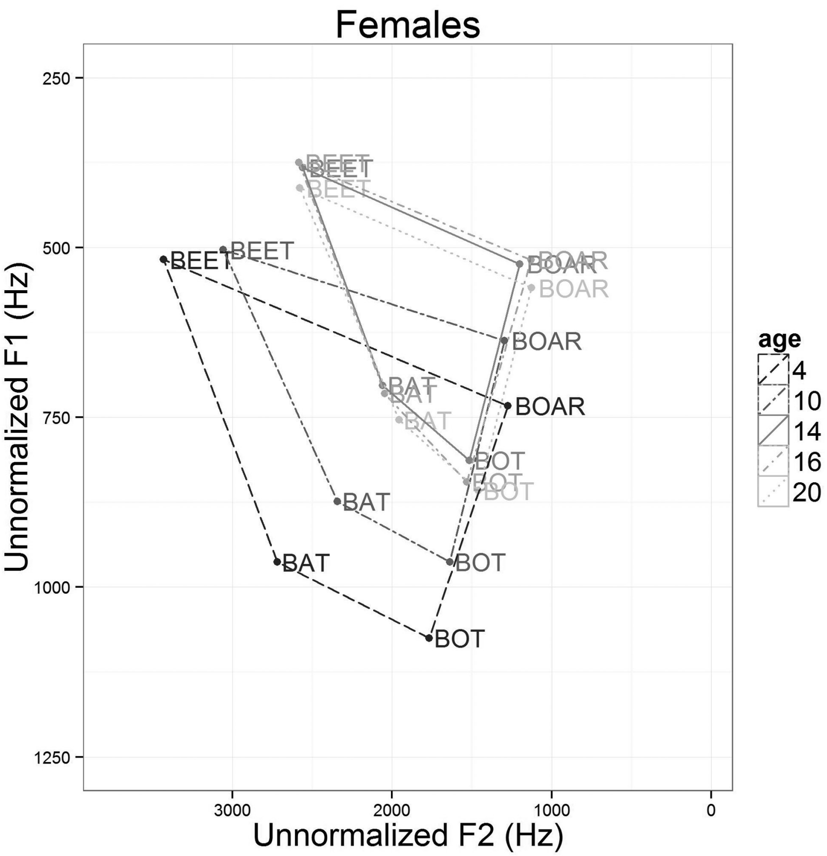

Figure 5.1 Vowel space quadrilaterals for females and males across 5 time points.

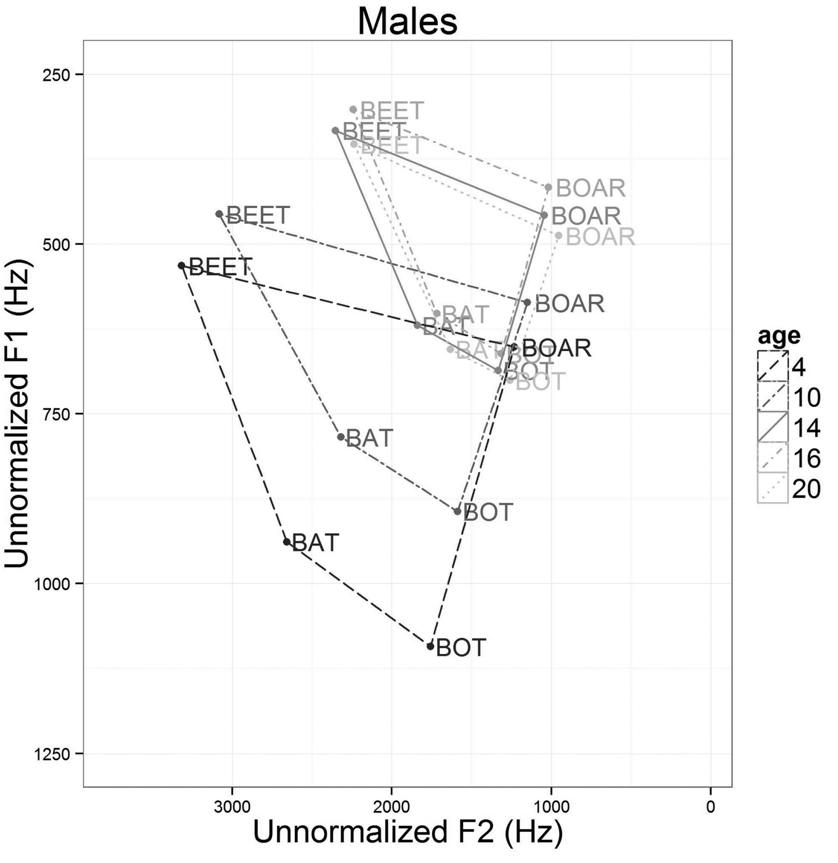

Figure 5.1 (Continued)

An analysis of VSA can provide a holistic measure of developmental change across childhood. Flipsen and Lee (2012) reanalyzed data from the landmark study by Lee, Potamianos and Narayanan (1999) by calculating vowel quadrilateral space to establish typical benchmarks in changes to VSA over time. Within this data set, sex differences in VSA are not consistent until age 16. After that age, male vowel spaces are significantly smaller than female vowel spaces. Surprisingly, Flipsen and Lee (2012) found what appears to be a "dip" between ages 13 and 16 for both sexes in which vowel spaces shrink to proportions that are smaller than adult vowel spaces, with a subsequent expansion between age 16 and adulthood. More information about changes to the vocal tract anatomy during adolescence is necessary to tease apart whether this pattern reflects anatomical differences or stems from another source, such as social influences (e.g., lack of commitment to careful speech in a testing environment) or articulatory issues with adjusting to adult-like anatomy (Hodge 2013). As noted by the authors, "[. . .] the cross-sectional nature of the current data set means we cannot rule out age cohort differences" (Flipsen and Lee 2012: 931). In other words, in a cross-sectional (apparent time) study, differences between age cohorts may be due to social experiences or other extraneous factors that could potentially set apart each cohort, rather than predictable maturation patterns. Our longitudinal study verifies patterns observed in Flipsen and Lee (2012), suggesting that age cohort differences are an unlikely explanation for this dip.

We take a closer look at VSA change using a Vowel Dispersion Measure (VDM) (Bradlow, Torretta, and Pisoni 1996; Pierrehumbert, Bent, Bradlow, Munson, and Bailey 2004). We calculated the VDM across time points for each speaker by first establishing a central tendency for the vowel spaces. This central measure was calculated by taking the grand mean of the median values[6] for BAT, BOT, BOAR[7] and BEET classes, representing extreme low, back, and front/high positions in the vowel space, for both F1 and F2. Using this central tendency measure, the Euclidean distance[8] was then calculated for each individual token for the four corner vowel classes. The median Euclidean distance for the corner vowel classes was then averaged together to produce the VDM. Token counts for each vowel class and time point included in this and subsequent analyses can be found in Table 5.1.

We preferred this method to a calculation of vowel quadrilateral space, such as was performed in Flipsen and Lee (2012), because it is more rigorous across dialects with different vowel configurations. Dialect differences are relevant for our data set as the BAT vowel class tends to be highly variable, with participants from more densely populated African American communities, including the Durham field site, having more raised BAT classes, producing a more triangular vowel space (Kohn 2014). This variation would cause participants with raised BAT classes to have artificially smaller vowel quadrilaterals unrelated to physical development. While the average VDM for Durham participants is smaller than the average VDM for Chapel Hill (608 versus 660), this difference was not significant (t-value = 1.37, p = 0.17) in a linear regression model that included age, sex, and city, indicating

Table 5.1 Token counts included in this analysis.

	Ages 3–4	*Age 10*	*Age 14*	*Age 16*	*Age 20*	*Total*
BEET	50	309	295	242	304	1150
BAT	127	335	298	295	472	1527
BOAR	36	170	191	195	207	799
BOT	52	240	239	236	274	1041
Total	215	1054	1023	968	1257	4517

that the VDM can accommodate some dialectal differences in vowel configurations. It should be noted that, while the VDM is less prone to sociolinguistic noise due to its reliance on dispersion from a central tendency, rather than assuming a quadrilateral vowel space, researchers should proceed with caution when selecting methods and vowel sets for calculating VSA (see Van der Harst 2011 for various considerations). Figure 5.2 illustrates the Euclidean distances for a male (speaker 256) and a female (speaker 1070), both at the age 20 time point, who have disparate BAT pronunciations. The shift in the central tendency from which the Euclidean distances are derived helps control for this dialect difference. Within a longitudinal data set, such control is necessary to anticipate possible changes in dialectal pronunciation over the lifespan. The exterior shape of the overall vowel space may fluctuate, making dispersion measures a better choice than quadrilateral measures for comparisons across time.

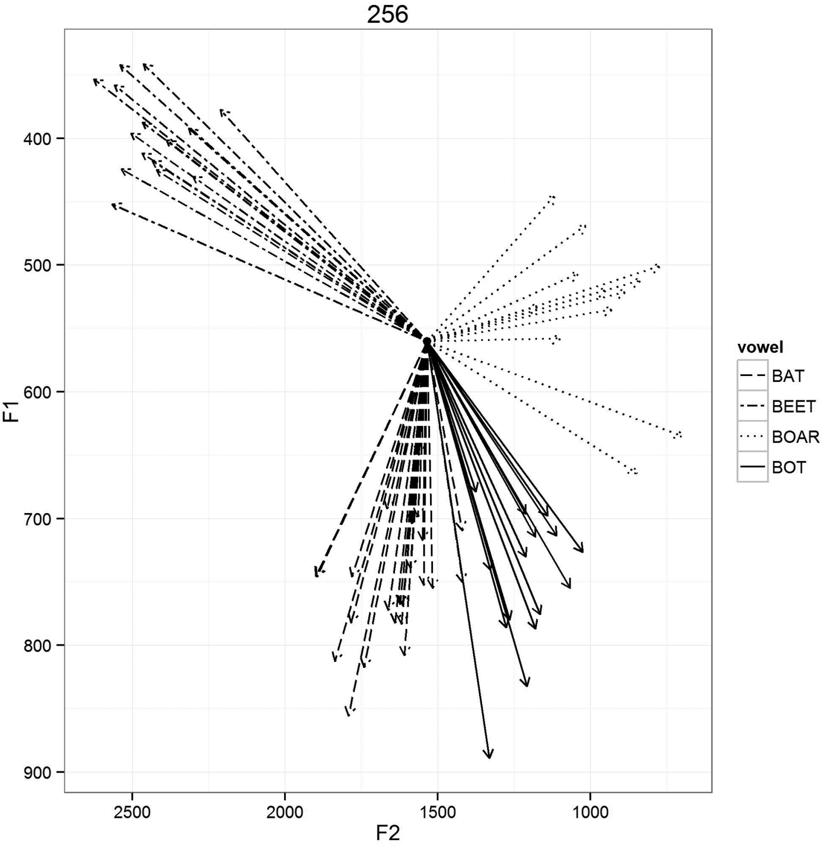

Figure 5.2 Illustration of VDM calculation for speakers 256 and 1070.

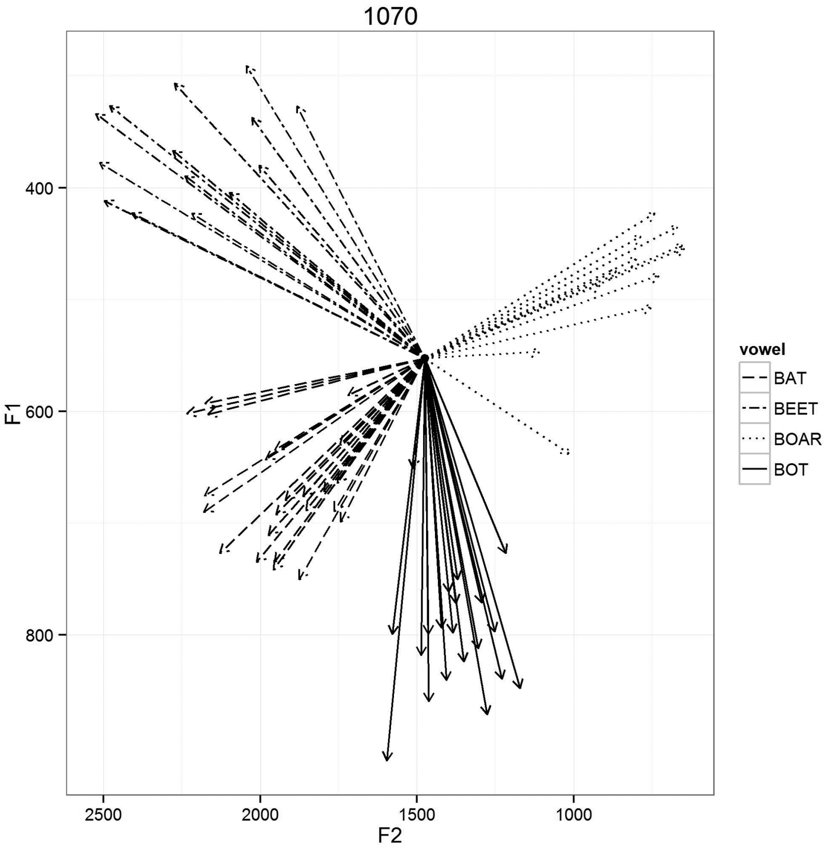

Figure 5.2 (Continued)

VDM values for all participants are plotted by speaker in Figure 5.3.

As will be discussed below, measures taken for the age four time point should be interpreted with caution. However, overall declines across childhood and adolescence show "leaps" around time periods associated with growth spurts. VDMs decline dramatically between ages 10 and 14, a time period associated with puberty. These changes are largest for males, which may be expected given the descent of the larynx between 12.5 and 14.5 years of age (Hollien 2014), while females may show an earlier and more gradual decline across time periods. Changes between age 14 and 20 are much smaller and more idiosyncratic in nature, likely reflecting small adjustments related to non-physiological factors, such as attention to speech or interlocutor accommodation. These observations indicate that data collected from

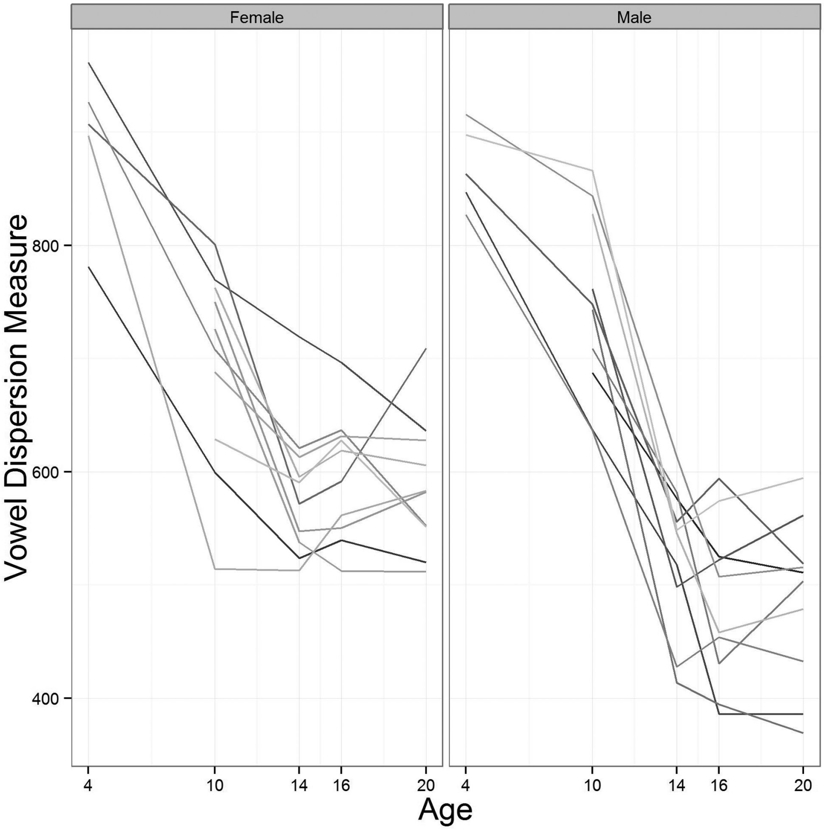

Figure 5.3 VDM trajectories for each speaker (*N* = 20) by sex and age.

individuals ages 16 and under may be especially prone to physiological or articulatory noise associated with growing vocal tracts and motor control. Particularly, between the ages of 10 and 16, there may be some variation resulting from the variable onset of puberty for males. While this concern is also relevant for females, overall shifts across adolescence are not as extreme as for males, mitigating the potential effect of adolescence on the female vowel space. Consistent with observed changes in VSA, the "dip" found by Flipsen and Lee (2012: 930) also appears among our age 14 cohort, but is less pronounced and highly idiosyncratic.

Figure 5.4 displays changes to the VDM for boys and girls as a group. Loess (local regression) curves with standard errors have been applied over the data. These curves are non-parametric and are thus appropriate for

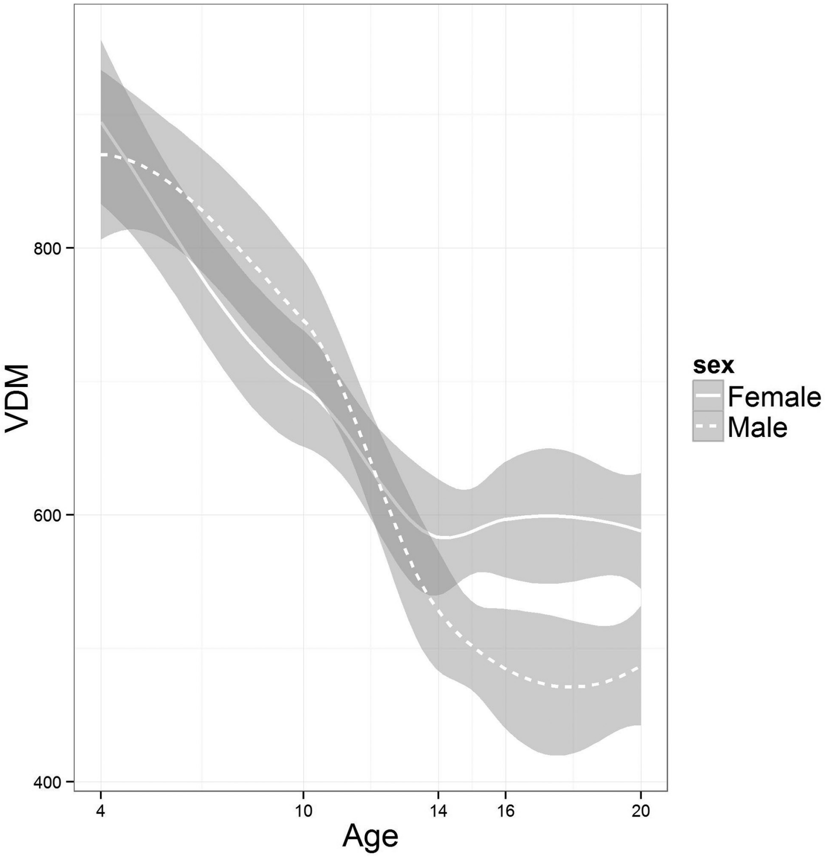

Figure 5.4 Average decline in VDM for all males and females ($N = 20$). Shaded areas represent standard errors for the fitted curve.

graphing non-linear change as may be expected for correlates of physical maturation.[9]

We find no reliable sex differences in VDMs before age 16, corroborating findings in Flipsen and Lee (2012). It is encouraging that our VDM measure confirms the patterns found for VSA in Flipsen and Lee (2012), despite using a different measure. It is also noteworthy that their cross-sectional patterns are confirmed in our longitudinal data. These data offer a general benchmark with respect to when developmental processes are likely to affect the size of the overall vowel space, as well as provide some typical maturation trajectories, with large drops in the size of the vowel space around puberty and sex differences reaching statistically significant levels by age 16.

3.2 Formant Measures

Declines in VSA over childhood hint at the extent to which changing vocal tract anatomy influences production. These morphological developmental patterns also impact trajectories for the first three formants as the speaker grows older. Given that formant values are the most commonly studied acoustic measures in sociolinguistics (Thomas 2011), an understanding of how physical changes affect formant values is crucial to any longitudinal study that includes pre-pubescent children. This section begins with an overview of research on child vocal tract anatomy, followed by a discussion of how patterns of growth impact formant values. We then compare F1 and F2 values across sex and age within our corpus, with particular attention to the pre-pubescent age ten time point. Our findings confirm previous research in illustrating independent paths of change for F1 and F2 values across vowel classes, with largest changes occurring between ages ten and 14, during which time sex differences emerge as significant.

Knowledge of child vocal tract development has massively improved in the last 30 years due to large-scale cross-sectional MRI studies (Vorperian et al. 2011; Fitch and Giedd 1999). These studies confirm that the vocal tract develops in non-uniform patterns, with distinct trajectories for the oral and the pharyngeal portions. Recent MRI studies show that the oral tract approaches maturity prior to the pharynx, with sex differences emerging in the horizontal plane (the vocal tract) prior to the vertical plane (the pharynx) (Vorperian et al. 2011; Hodge 2013). While early studies identified no difference in male and female vocal tract anatomy before age 11 (Fitch and Giedd 1999), Vorperian et al. (2011) found a more complex pattern of sexual dimorphism, with components of the vocal tract anatomy displaying sex differences at disparate points during maturation. The vertical plane of the vocal tract significantly differs across sexes past age 12 with differences lasting into adulthood, but the horizontal plane shows sexual dimorphism at about ages three years to seven years, only to disappear and reemerge after age 12. These studies indicate that morphological differences are not uniformly distinct across time periods. Rather, change occurs in complex ways, even as sexual dimorphism may produce early acoustic differences.

Changes to different components of the vocal tract affect formant values in distinct ways. For example, it has been hypothesized that growth in the pharynx disproportionally affects F2 values compared to F1 and F3 values (Fant 1975). Under this hypothesis, Vorperian et al. (2011: 10) suggest that growth patterns described above may result in non-uniform developmental patterns across the vowel space. For example, changes to the low vowel region may differ from the high vowel region as "low vowels require increased constriction of the pharyngeal region" (*ibid.*).

Sex differences emerge for formant values due to differences in maturation patterns for girls and boys. For example, changes to F2 values are likely to be more extreme for males undergoing puberty compared to their female counterparts due to the more dramatic changes to the pharynx observed for males during this developmental stage. Further, sex differences for F2 values will emerge later than F1 or F3 due to the later development of the pharynx. Below we compare group trajectories of changes to F1 and F2 values to findings from cross-sectional acoustic studies in order to identify common patterns related to maturation.

Figures 5.5 and 5.6 display Loess curves for F1 and F2 values across all five time points included in the study. Shaded areas represent standard errors for the fitted curve. Males and females have been plotted separately to illustrate developmental differences across sexes. We focus on corner vowels to identify trends for different segments of the vowel space, although rhoticism will obviously influence our back vowel measure BOAR. Means and standard deviations for each time point are given in Tables 5.2 and 5.3.

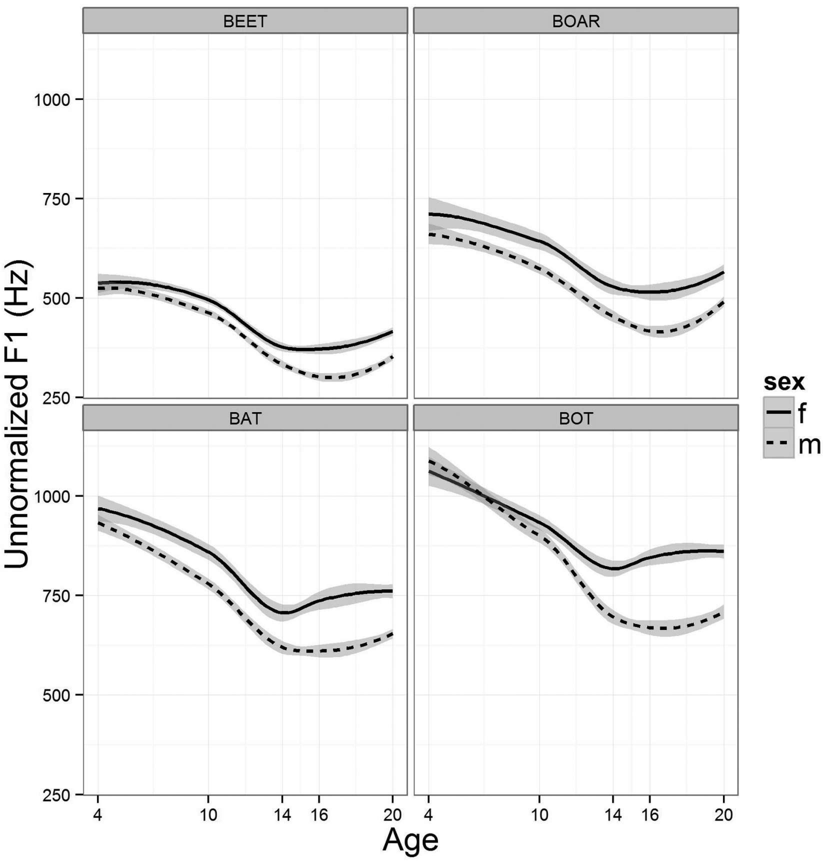

Figure 5.5 LOESS curves for F1 formant trajectories for BEET, BAT, BOAR and BOT facetted by sex. Shaded areas represent standard errors for the fitted curve.

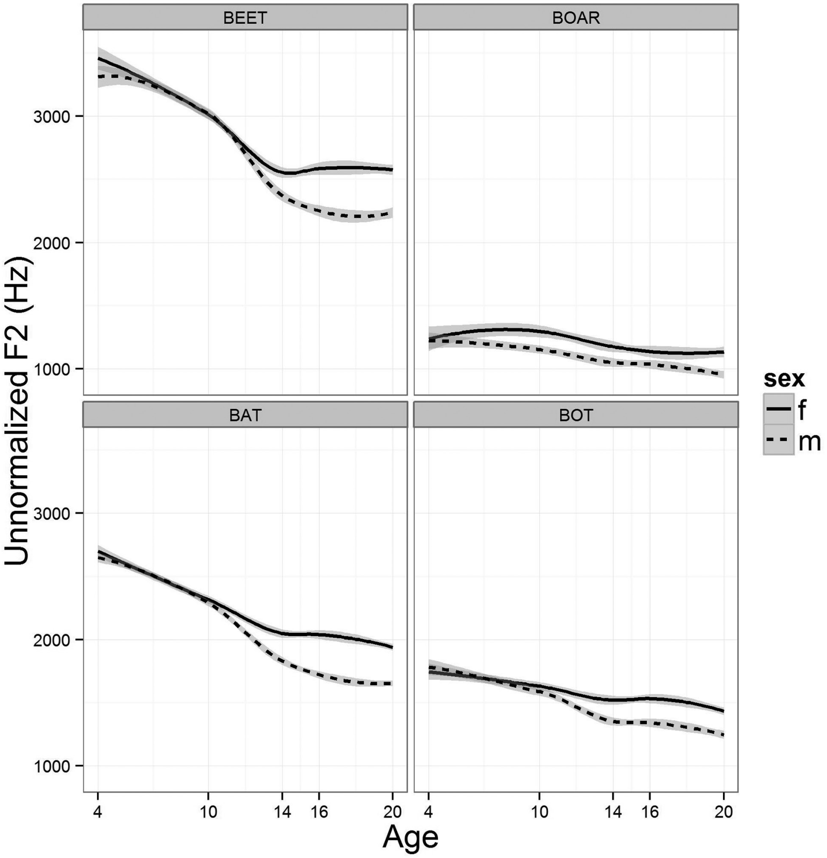

Figure 5.6 LOESS curves for average F2 formant trajectories for BEET, BAT, BOAR and BOT facetted by sex. Shaded areas represent standard errors for the fitted curve.

Several patterns are apparent. First, trajectories differ for each vowel. The decline in F2 values for BEET, and to a lesser extent, BAT, is steeper than those of the back vowels BOAR and BOT (see Table 5.3). Changes to the F1 values of low vowels are also more extreme than changes to the F1 of the high vowel (Table 5.2). These independent trajectories of change appear predictable across studies. Disproportionally large declines in the F2 of the high front vowel and the F1 of low vowels were also observed in cross-sectional studies summarized in Vorperian and Kent (2007).

For the purposes of informing child longitudinal studies, it is worthwhile to take a closer look at our age ten time point. While mean values for males are lower than mean values for females across all vowels by our age ten time point,

Table 5.2 F1 means and standard deviations (in parentheses) by vowel and sex with total change between age 4 and age 20. F1 values are in Hertz. Change = difference between Ages 3–4 value and Age 20 value.

	Sex	*Ages 3–4 (n = 10)*	*Age 10 (n = 20)*	*Age 14 (n = 20)*	*Age 16 (n = 20)*	*Age 20 (n = 20)*	*Change*
BEET mean (St Dev)	F	531 (99)	507 (87)	380 (61)	371 (57)	417 (61)	114
	M	524 (82)	463 (68)	335 (51)	302 (46)	353 (40)	171
BAT mean (St Dev)	F	960 (158)	880 (164)	706 (124)	731 (131)	761 (124)	199
	M	930 (148)	790 (107)	620 (91)	610 (88)	654 (74)	276
BOT mean (St Dev)	F	1057 (141)	949 (131)	817 (112)	846 (104)	860 (96)	197
	M	1088 (133)	900 (111)	696 (103)	669 (90)	709 (71)	379
BOAR mean (St Dev)	F	711 (135)	643 (107)	528 (99)	515 (115)	566 (65)	145
	M	659 (100)	581 (76)	424 (68)	417 (64)	491 (40)	168

Table 5.3 F2 means and standard deviations (in parentheses) by vowel and sex with total change between age 4 and age 20. F1 values are in Hertz. Change = difference between Ages 3–4 value and Age 20 value.

	Sex	*Age 4 (n = 10)*	*Age 10 (n = 20)*	*Age 14 (n = 20)*	*Age 16 (n = 20)*	*Age 20 (n = 20)*	*Change*
BEET mean (St Dev)	F	3434 (271)	3054 (280)	2551 (245)	2583 (284)	2575 (282)	859
	M	3311 (301)	3019 (291)	2370 (213)	2251 (241)	2238 (237)	1073
BAT mean (St Dev)	F	2694 (255)	2335 (224)	2049 (200)	2040 (219)	1940 (178)	754
	M	2645 (280)	2315 (218)	1831 (178)	1722 (128)	1655 (151)	990
BOT mean (St Dev)	F	1742 (236)	1644 (225)	1520 (215)	1533 (190)	1433 (158)	309
	M	1781 (297)	1589 (225)	1352 (176)	1342 (156)	1248 (87)	533
BOAR mean (St Dev)	F	1239 (344)	1297 (245)	1176 (196)	1137 (219)	1135 (205)	104
	M	1221 (211)	1163 (191)	1050 (152)	1036 (134)	954 (153)	267

sex differences in our data set are not statistically significant across all vowels on all dimensions until the age 14 time point. This pattern aligns with patterns observed in Lee et al. (1999), where formant frequency patterns for males and females emerged as distinct around age 11. The age ten time point also shows more overlap between males and females for BEET than BAT on the F1 dimension, consistent with findings from Busby and Plant (1995), even as there is extensive overlap on the F2 dimension for BEET, BAT, and BOT, contra to findings in Busby and Plant (1995). Such variation across studies has been noted by other researchers, but is also expected due to variation in methodologies, sample sizes and statistical analysis (Vorperian and Kent 2007). Still, generalizations are apparent. For example, even during this early time point, sex differences emerge as a relevant factor for consideration, with differences emerging earlier in some dimensions of the vowel space. Given Vorperian et al.'s (2011) findings of early sexual dimorphism, this pattern is unsurprising.

By age 16, formant values for vowels reach adult-like levels, showing only minor change between the age 16 and 20 time points. Sex differences for low vowel F1 values, and the F2 for front vowels, are larger than those for the F1 of high vowels or F2 of back vowels, confirming patterns identified in Vorperian and Kent (2007). Formant values do not always reach their lowest point at the adult time point, but, rather, rebound from what appears to be a developmental "dip" in adolescence, as observed in Lee et al. (1999). In our data set, this is particularly evident for F1 values.

3.3 Additional Developmental Changes

Changes to motor control affect several linguistic correlates across childhood, including vowel duration and within-speaker variance. Cross-sectional studies show no significant vowel duration differences for sex at any age and no significant differences from adult data from around age 11 (Lee et al. 1999), with the duration progressively declining across childhood (Eguchi and Hirsh 1969) until about age 15.[10] While no sex differences have been observed for duration (Assmann, Nearey and Bharadwaj 2013), different phonemes follow different developmental paths, with tense vowels showing greater total change in duration across acquisition (Hubbard, Kiefte, Hossain and Assmann 2013).

Within-speaker variance for vowel production also declines across childhood (Kent 1976; Lee et al. 1999; Eguchi and Hirsh 1969), presumably as children gain motor control skills (Green, Moore and Reilly 2002). The point at which within-speaker variability reaches adult-like levels is debated, with estimates for F1 as early as age three (Nittrouer 1993). Eguchi and Hirsh (1969) identify adult-like variability for F1 and F2 between ages 11 and 13, while Lee et al. (1999) conclude that all formants reach adult-like variability levels by age 14. Even as duration and variability show changes across age groups, vowel formant contours appear largely adult-like for children as young as 5–8 (Assmann et al. 2013). While durations are longer and targets are more variable, the same formant contour is present in young children as in adults.

The cumulative cross-sectional research on child vocal tract anatomy and acoustic development illustrates that there are predictable patterns to maturation. Longitudinal researchers should expect that data spanning childhood and adolescence will be characterized as follows: Vowel spaces decline predictably, but not linearly, with differences across sex. Individual vowel formants do not change in a uniform manner across age or sex, but such patterns are also predictable and therefore can be addressed in normalization. Children under the age of 11, and especially under the age of seven, produce longer vowels with more variable targets. These observations have direct implications for normalization and statistical analysis, as we'll discuss in sections 4 and 5, respectively. Given these patterns researchers must choose normalization procedures with care, as well as determine whether statistical analysis for the youngest age groups is appropriate. The following sections will provide guidance on these issues.

4. Implications for Normalization

Sociolinguistic studies typically employ normalization procedures to control for the influence of physical differences on acoustic correlates, particularly those that surface across sexes in adult populations (see Clopper 2009 for an excellent review of various normalization techniques). To address these concerns, a number of algorithms have been extensively evaluated for their ability to control for variation attributed to physical differences while retaining sociolinguistically meaningful variation associated with region (Adank, Smits, and van Hout 2004; Van der Harst 2011), generation (Fabricius, Watt and Johnson 2009; Hindle 1978) and social class (Labov 2001). Given the influence of physical development on child vowels, normalization is even more important when making comparisons across childhood and adolescence in both longitudinal studies and cross-sectional studies. Yet, research on child vowel normalization has only recently been undertaken (Kohn and Farrington 2012).

Because individual vowels and formants change in distinct ways over childhood and adolescence, "uniform scaling factors are not entirely adequate" (Vorperian and Kent 2007: 1527). A simple log transformation or transformation based on estimates of the vocal tract (c.f. Nordström and Lindblom 1975) is insufficient to control for acoustic correlates of vocal tract growth. However, because child data presents similar challenges for normalization to those for adult data, normalization procedures that perform well on adult data are likely to also be robust for child data, a hypothesis confirmed in Kohn and Farrington (2012). Still, comparisons between child data and adult data are more sensitive in a few respects, particularly when normalization procedures rely on relationships between formants. We discuss below why normalization is appropriate for some (but not all) child data, as well as potential pitfalls for normalizing child data.

Observations of child and adult vocal tract studies reveal the obvious fact that sex differences in the vocal tract are largest during adulthood. As such, sex differences, such as differences in the ratio of the oral tract to the pharynx, affect both child and adult studies. Further, a pre-pubescent male compared to his post-pubescent counterpart presents many of the same issues of comparing an adult male and female as he has not yet experienced the descent of the larynx that will have such a dramatic impact on his F2 values. There are two main differences that normalization procedures must confront to align male and female vowel spaces: Differences in overall VSA and differences in the shape of the vowel space. Specifically, male vowel spaces are smaller than female vowel spaces, female F1 values for low vowels are disproportionally higher than male values and sex differences in F2 values are greatest for the high front vowel. These latter differences are the most challenging aspect of normalization in adult studies. Scaling techniques that rely on an independent transformation for F1 and F2, also known as formant intrinsic methods, are thus better suited than scaling techniques that rely on a single transformation, also known as formant extrinsic methods, to address the distinct adult sex differences on the F1 and F2 planes. Further, techniques that utilize a central tendency and a normalized range are more rigorous against non-uniform differences across the vowel space than those that rely on a normalized range alone. This has been confirmed in a number of studies in which such normalization procedures prove more effective than normalization based on formant ratios (e.g., Bark Difference Metric) or normalized ranges alone (e.g., Gerstman) (Adank et al. 2004; Clopper 2009; Fabricius et al. 2009; Van der Harst 2011; Kohn and Farrington 2012). As such, Lobanov (1971) and Nearey's (1978) vowel extrinsic normalization procedure have become standards in the field.

Because F1 and F2 values show independent trajectories of change, formant intrinsic normalization procedures should be preferred for longitudinal studies that include child time points as well. Kohn and Farrington (2012) found that Lobanov normalization, a z-score normalization procedure (see Table 5.4), effectively aligned vowel spaces for age groups ranging from ages 10 to 20 in their study of child vowel normalization, confirming that normalization procedures that are effective within adult studies also can be used in studies that include childhood time points.

However, some normalization procedures can be problematic for child data. Normalization techniques that rely on formant ratios, for example, are always inappropriate for data that spans childhood and adolescence as formants do not decline in a uniform fashion across childhood (see section 3). In particular, F3 reaches adult levels more rapidly than the other formants (Vorperian and Kent 2007), so normalization techniques such as the modified Bark Difference Metric (Thomas 2002, 2011) would underestimate the scaling necessary to account for maturational differences in F1 and F2 values.

Table 5.4 LOBANOV (1971) normalization algorithm.

$F_{norm}=(F_{ni-}\ \mu(F_n))/(\sigma(F_n))$	F_{ni} is formant n of token i. μ and σ are the grand mean and the standard deviation for formant n.

Similarly, normalization procedures such Fabricius et al. (2009) that use F1 values of high front vowels to estimate the back of the vowel space will be inappropriate due to non-uniform trajectories of change across formants.

As is true with all populations, normalization procedures should be checked for accuracy through visual examination. We argue that researchers who include child data in their analysis should measure corner vowels to inspect the effectiveness of normalization. In addition to spotting potential normalization errors, visual inspection may provide clues to how sound changes progress with age. Below, we present two case studies to illustrate the effectiveness of Lobanov normalization for childhood data as well as to highlight potential clues that normalization has failed. We focus on two speakers from the FPG corpus: Speaker 268, a female from Chapel Hill, NC (Figure 5.7); and speaker 1057, a male from Durham, NC (Figure 5.8). We

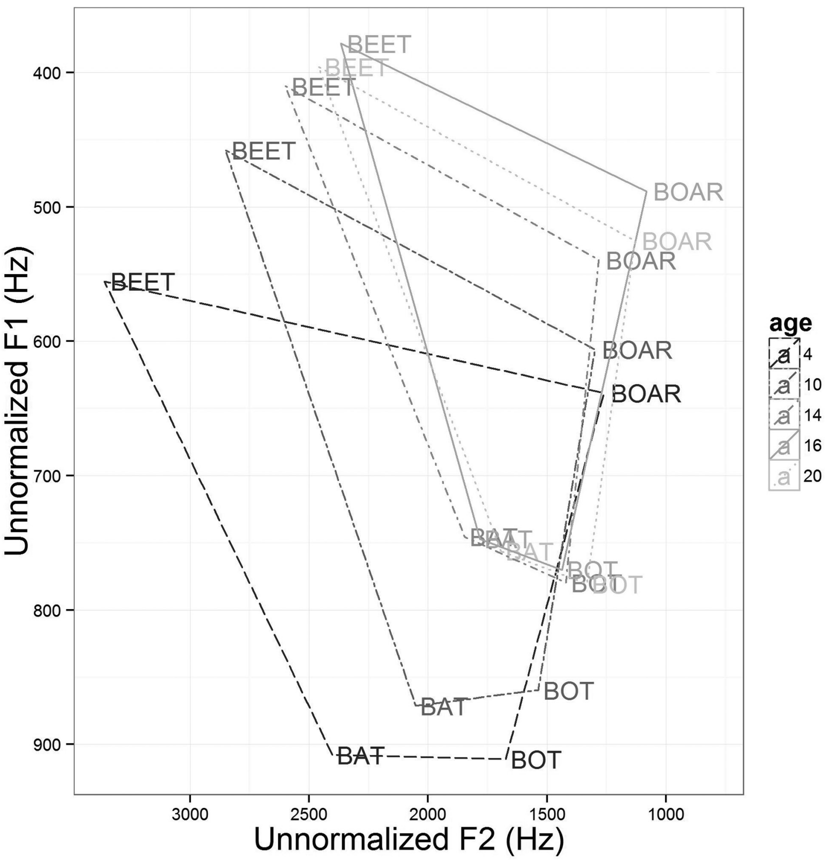

Figure 5.7 Unnormalized and Lobanov normalized vowel quadrilaterals for speaker 268.

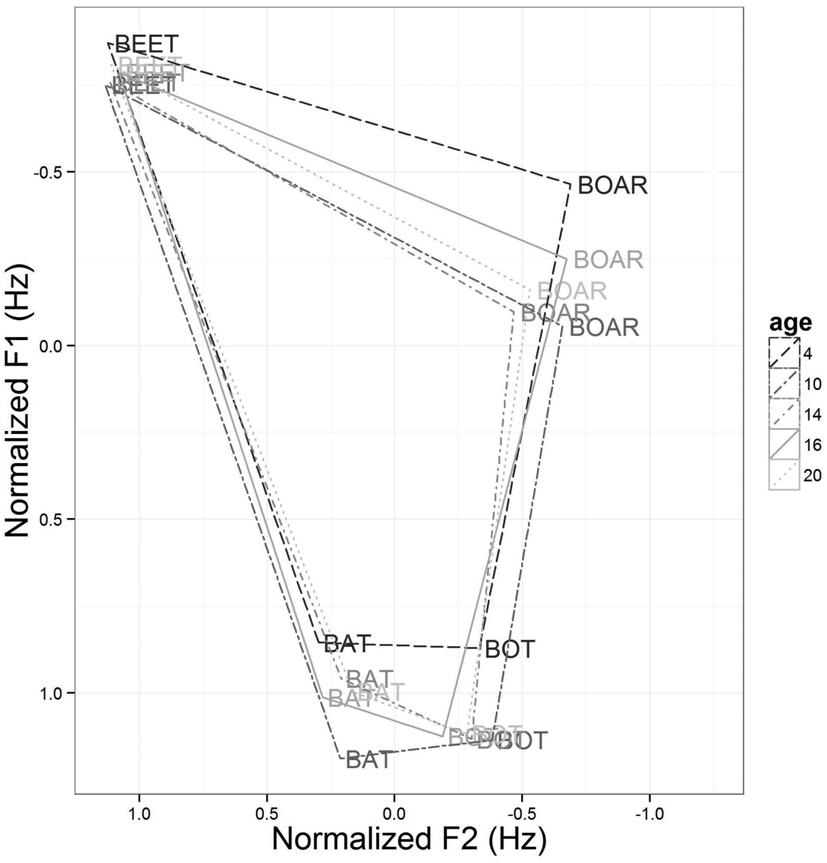

Figure 5.7 (Conttinued)

present both unnormalized and Lobanov-normalized vowel quadrilaterals across five time points for each speaker.

Visual comparison of speaker 268's vowel charts illustrates that Lobanov normalization improves overlap of VSA across time points. Statistical analysis of the normalized data confirms visual observation. In regression models with F1 and F2 as the dependent variable and age, corner vowels, preceding place of articulation, and duration as independent variables, only age four is significantly different from age 20, and only on the F1 dimension (p=.037). We can see that all of the vowels for the age four time point are more raised compared to other time points, strongly suggesting that normalization has failed. Without including corner vowels, it would be difficult to identify this shifted pattern for the age four time point. Note, however, that differences

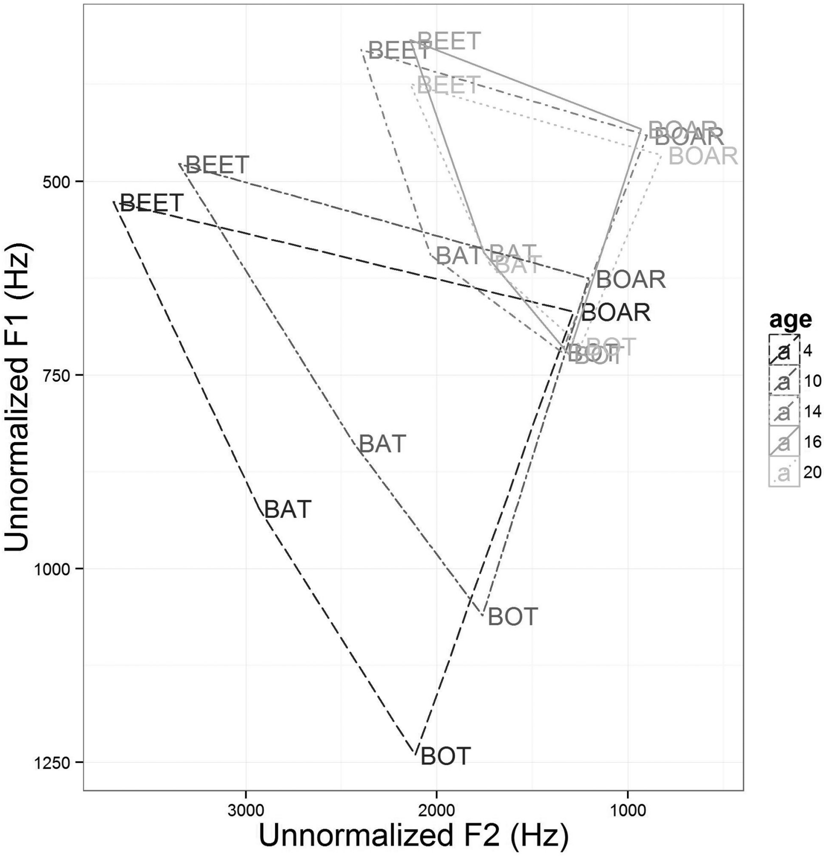

Figure 5.8 Unnormalized and Lobanov normalized vowel quadrilaterals for speaker 1057.

in alignment that do not appear related to age are preserved for all other time points. So, for example, BAT is lower than BOT for the age 10 time point in both the normalized and raw data. Because this pattern is unrelated to patterns associated with physical development, it is positive that the normalization technique maintains these differences.

A similar pattern is apparent for speaker 1057. VSA across time points are aligned, although the age four quadrilateral appears shifted up, again indicating a failure to normalize out acoustic correlates of size differences for this age. On the other hand, the fronted position of BAT at the age 14 time point has no ready developmental explanation and likely represents a difference in articulation between the time points. Once again, regression indicates that the age four time point is significantly different from age 20

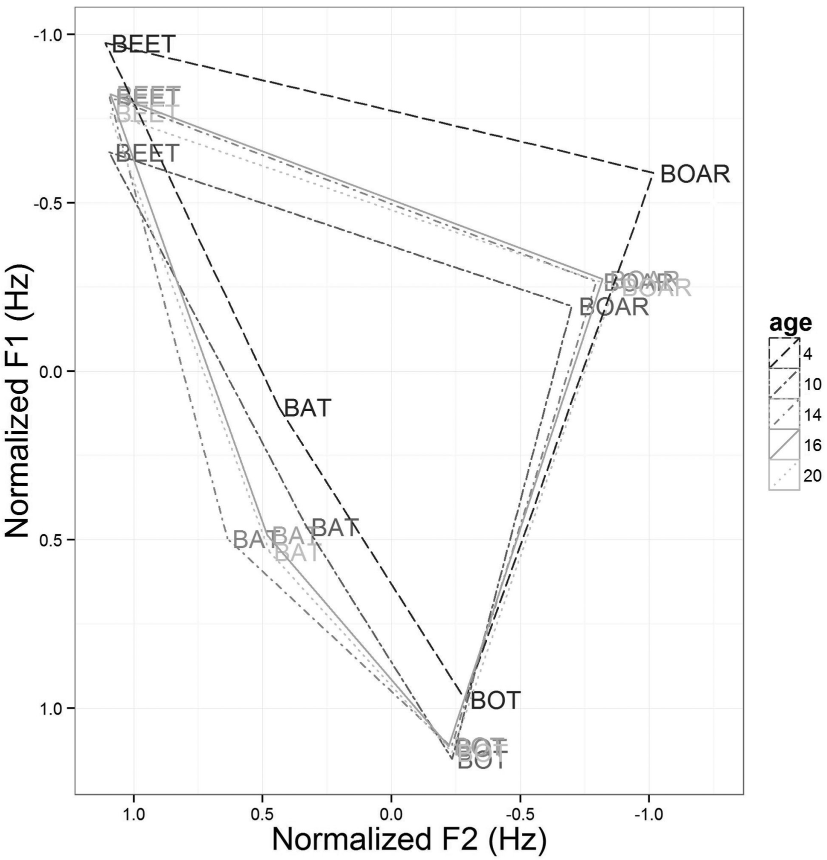

Figure 5.8 (Continued)

(p < .001). Age 14 is also significantly different from age 20, but only when BAT is included in the model.

Lobanov normalization appears effective at addressing physiological differences across time points (Kohn and Farrington 2012). However, we noted a systematic issue with normalizing the F1 dimension for the age four time point. This error may reflect variation attributable to higher F0 values that result in more widely spaced harmonics for children (see Van der Harst (2011) for a discussion of similar issues with women's speech, as well as suggestions for identifying measurement errors in data). Formants are difficult to discern in the spectrogram under these conditions. This is apparent in Figure 5.9, which depicts speaker 1057 at age four using extremely elevated pitch attributable to excitement (498 Hz). LPC accuracy decreases as pitch increases due to aliasing of the autocorrelation sequence (Rodríguez and

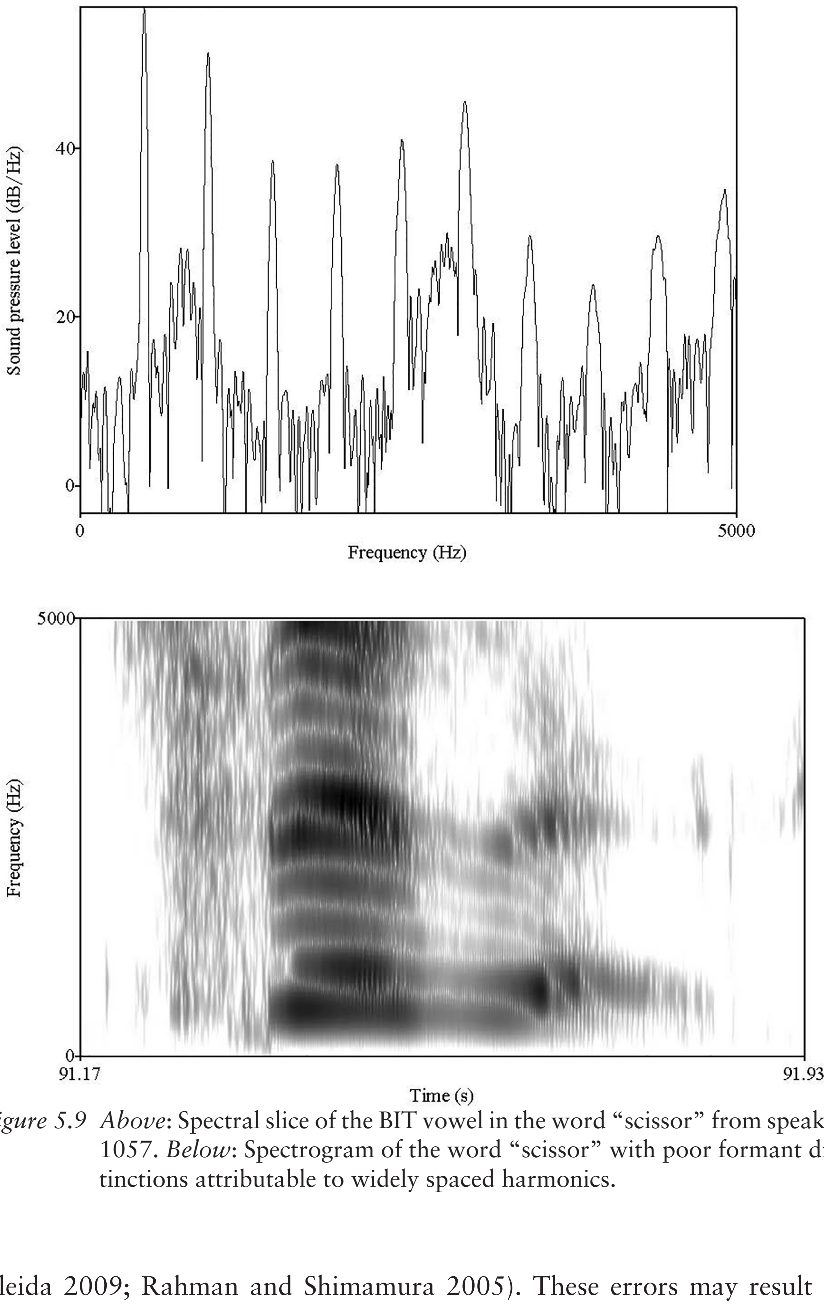

Figure 5.9 *Above*: Spectral slice of the BIT vowel in the word "scissor" from speaker 1057. *Below*: Spectrogram of the word "scissor" with poor formant distinctions attributable to widely spaced harmonics.

Lleida 2009; Rahman and Shimamura 2005). These errors may result in ineffective measurement for very young children, although see Rodríguez and Lleida (2009) and Rahman and Shimamura (2005) for some potential solutions to this issue using homomorphic analysis.

Measurement errors are not the only concern when studying early childhood speech. Low token counts also produce low statistical power. For example, we were only able to measure an average of 120 tokens per four-year-old, compared to an average of 204 tokens per ten-year-old. As noted by Roberts (2002) a researcher may need 8–14 hours of speech from a young child to capture similar token counts found in a 1–2 hour adult interview. In addition, a child's developing vocabulary has fewer word types than adults. In combination with increased intra-speaker variance, a researcher will likely have difficulties extracting enough data for statistical comparison. Indeed, although we analyzed the age four time periods for all speakers in this sample, we found that many of our speakers failed to produce enough speech for any meaningful analysis.

However, preschool time periods are vital for linguistic studies as the influence of school on social networks may be critical to the incrementation of language change (see Van Hofwegen and Wolfram, this volume). Therefore, time periods prior to the entrance of school, as well as studies of the increasing role of early education, are necessary to fully assess how social structures related to educational experience impact speech. Several researchers have included time points below the age of ten in vowel studies (Kerswill and Williams 2000; Roberts 2002; Smith, Durham, and Fortune 2007). These studies mostly rely on impressionistic coding, choosing to focus on phonemic differences salient within the community. For example, Smith et al. (2007) examined the speech of children age 2;10–3;6[11] acquiring Scottish English. Their focus on the monophthongal production of the /au/ diphthong (i.e. BOUT, or "the HOOSE variable") allowed for easy impressionistic coding. Kerswill and Williams (2000) similarly employed impressionistic coding when analyzing back vowel fronting for four-year-olds in Milton Keynes. When physical differences cannot be normalized, impressionistic coding may be preferable to acoustic analysis, as statistical comparisons can still be used to compare impressionistic scores across age groups.

5. Statistical Analysis

Appropriate selection of statistical techniques is just as crucial as appropriate selection of normalization techniques for longitudinal data. Best practices for longitudinal data indicate that mixed models should include speaker as a random effect, time as a random slope, and time as a fixed effect in regression analysis (Singer and Willett 2003). Researchers should include time as a random slope to recognize the non-independence of measurements within a speaker and across time. The incorporation of random effects also allows for statistical findings to be generalized to the group as any significant variation attributed to the fixed effect of time will represent group trends rather than individual variability.

It is important to visually inspect patterns of change as non-linear patterns are common in longitudinal studies. These patterns can be accommodated by either transforming the scale of the dependent variable or through modeling the fixed and random effect representing time as a polynomial function (see Singer and Willett (2003), Chapter 6 for additional details). Unique to acoustic data analysis, researchers may choose to model time as a categorical factor group. Because there are multiple measures per speaker per time point, rather than a single measure per time point as is common in most longitudinal studies, measures grouped by time point can be compared against each other using dummy coding in which a default factor group for one time point is compared to the remaining factor groups for time points in a dichotomous fashion. With this model, the researcher does not have to assume *a priori* that group change will follow a predictable pattern, such as a linear or curvilinear path. This approach is useful if researchers expect that one time point, such as an adolescent time point, will stand out as unique from the rest.

Some studies have used a random effect for speaker as an approach to normalization by allowing the random effect to control for between-speaker variation. Considering that this approach has been used in lieu of normalization in the past (e.g., Nycz 2011), we may ask whether normalization and the inclusion of a random speaker variable is overkill. However, relying on a random factor to account for physiological variability in child longitudinal data is not appropriate because group-level differences associated with maturation will be present in the data unless normalization occurs before statistical analysis. For example, the random speaker/slope component in the regression model will not control for the overall higher frequencies found at the age ten time point compared to the age 16 time point. Both normalization and a random factor are necessary for child data, unless the main focus of analysis is on developmental changes to acoustic correlates over time, rather than changes related to social variation.

6. Conclusions

Working with child data provides challenges for longitudinal studies. However, physical changes are predictable across childhood so researchers can control for acoustic correlates of physical development. Vowel extrinsic/formant intrinsic normalization like Lobanov (1971) are robust for adult and child data, although care should be taken to identify residual 'noise' from physical development. For this reason, researchers working with child data should plan to analyze corner vowels to identify systematic errors related to normalization. Additionally, caution must be taken to consider both unnormalized and normalized data to avoid "normalizing out" information which could potentially be relevant to sound change. Statistical analyses are only appropriate when normalization effectively eliminates acoustic correlates of physical differences. If normalization fails impressionistic coding may be an important alternative to acoustic analysis for very young children. As long as these steps are taken, the inclusion of child time points will enrich our understanding of how life stages influence language variation and change. Only research that includes these time points will answer questions central to the field of sociolinguistics about the role of childhood and adolescence in the speech community.

Notes

1 Ages represent mean age at the primary point of data collection for the 20 selected participants. Although age six was analyzed for ten participants in Kohn and Farrington (2012), many of the additional FPG participants did not have recordings of sufficient quality for acoustic analysis during this time period, so this time point was not included in the expanded sample.
2 While testing protocols allowed for consistent collection of specific words and phonetic contexts, the more naturalistic elicitation tasks were less controlled. We include phonetic environment in statistical analyses to control for unbalanced data collection. Additionally, we use grand means, instead of simple means, for normalization procedures so that unbalanced token counts across vowel classes do not skew normalization results.
3 This chapter follows Yaeger-Dror and Thomas (2009) in using Wellsian-style B_T vowel class frames. Actual phonetic production varies across the population.
4 See Reubold and Harrington, this volume, for acoustic changes related to aging beyond adolescence.
5 To allow for comparisons with Vorperian and Kent (2007), we focus on the mid-point measure of each vowel. An analysis of vowel trajectories for this population can be found in Risdal and Kohn (2014).
6 Using median values, rather than mean values, for each vowel class can prevent skewing caused by outliers.
7 We measured pre-rhotic /o/ as a back vowel measure due to the tendency for back vowels to front in many English dialects (Thomas 2011).
8 The Euclidean distance is the distance between two points on a plane. In this case, we calculate the distance between the central tendency (a) and the vowel token in question (b) on the F1 and F2 plane. For points a (a_{f2}, a_{f1}) and b (b_{f2}, b_{f1}), the distance is equal to.
9 The curve is calculated in a manner similar to a rolling average in which ordinary least squares regression is completed on subsections of the data. Coefficients from each localized regression are then used to estimate a predicted value for the evaluation point (see Jacoby (2000) for additional details). Loess curves were estimated using ggplot_smooth in the ggplot2 statistical package (Wickham 2009).
10 Lee et al. (1999) observe that duration reaches its minimum at age 15 before rebounding to adult levels, similar to the rebound observed for formant values.
11 Year; month.

References

Adank, Patti, Roel Smits and Roeland van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116: 3099–3107.

Assmann, Peter F., Terrance M. Nearey and Sneha V. Bharadwaj. 2013. Developmental patterns in children's speech: Patterns of spectral change in vowels. In Geoffrey Stewart Morrison and Peter F. Assmann (eds.) *Vowel Inherent Spectral Change*. Berlin and Heidelberg, Germany: Springer Verlag. 199–230.

Bradlow, Ann, Gina Torretta and David Pisoni. 1996. Intelligibility of normal speech I: Global and finegrained acoustic-phonetic talker characteristics. *Speech Communication* 20: 255–272.

Busby, Peter A. and G. L. Plant. 1995. Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America* 97: 2603–2606.

Clopper, Cynthia. 2009. Computational methods for normalization acoustic vowel data for talker differences. *Language and Linguistics Compass* 3: 1430–1442.

Eckert, Penelope. 1997. Age as a sociolinguistic variable. In Florian Coulmas (ed.) *The Handbook of Sociolinguistics*. Oxford, UK: Blackwell. 151–167.

Eguchi, Suco and Ira J. Hirsh. 1969. Development of speech sounds in children. *Acta Oto-Laryngologica. Supplementum* 257: 1–51.

Fabricius, Anne H., Dominic Watt and Daniel Ezra Johnson. 2009. A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. *Language Variation and Change* 21: 413–435.

Fant, Gunnar. 1975. Non-uniform vowel normalization. Speech Transmission Laboratory Quarterly Progress and Status Report 2–3: 1–19.

Fitch, W. Tecumseh and Jay Giedd. 1999. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America* 106: 1511–1522.

Flipsen, Peter and Sungbok Lee. 2012. Reference data for the American English acoustic vowel space. *Clinical Linguistics and Phonetics* 26: 926–933.

Green, Jordan, Christopher Moore and Kevin Reilly. 2002. The sequential development of jaw and lip control in speech. *Journal of Speech, Language, and Hearing Research* 45: 66–79.

Hindle, Donald. 1978. Approaches to vowel normalization in the study of natural speech. In David Sankoff (ed.) *Linguistic Variation: Models and Methods*. New York, NY: Academic Press. 161–171.

Hodge, Megan. 2013. Development of the vowel space in children: Anatomical and acoustic aspects. In Martin Ball and Fiona Gibbon (eds.) *Handbook of Vowels and Vowel Disorders*. Vol. 2. New York, NY: Psychology Press. 1–23.

Hollien, Harry. 2014. On pubescent voice change in males. *Journal of Voice* 26: 29–40.

Hubbard, Daniel J., Michael Kiefte, Shaikat Hossain and Peter Assmann. 2013. A developmental study of vowels spoken in syllables and in sentence context. *Proceedings of Meetings on Acoustics* 19: 060070. Retrieved from http://asa.scitation.org/doi/abs/10.1121/1.4799483.

Jacoby, William G. 2000. Loess: A nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies* 19: 577–613.

Kent, Ray. 1976. Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research* 19: 421–447.

Kerswill, Paul and Ann Williams. 2000. Creating a new town koine: Children and language change in Milton Keynes. *Language in Society* 29: 65–115.

Khattab, Ghada and Julie Roberts. 2011. Working with children. In Marianna Di Paolo and Malcah Yaeger-Dror (eds.) *Sociophonetics: A Student's Guide*. London, UK: Routledge. 163–177.

Kohn, Mary. 2014. *"The Way I Communicate Changes But How I Speak Don't": A Longitudinal Perspective on Adolescent Language Variation and Change*. Publication of the American Dialect Society 99. Durham, NC: Duke University Press.

Kohn, Mary and Charlie Farrington. 2012. Speaker normalization: Evidence from longitudinal child data. *Journal of the Acoustical Society of America* 131: 2237–2248.

Labov, William. 2001. *Principles of Linguistic Change*. Vol. 2: *Social Factors*. Oxford, UK: Blackwell.

Lee, Sungbok, Alexandros Potamianos and Shrikanth Narayanan. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America* 105: 1455–1468.

Lobanov, Boris M. 1971. Classification of Russian vowels spoken by different listeners. *Journal of the Acoustical Society of America* 49: 606–608.

Nearey, Terrance Michael. 1978. *Phonetic Feature Systems for Vowels*. Vol. 77. Bloomington, IN: Indiana University Linguistics Club.

Nittrouer, Susan. 1993. The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research* 36: 959–972.

Nordström, P.-E. and Björn Lindblom. 1975. A normalization procedure for vowel formant data. In *Proceedings of the 8th International Congress of Phonetic Sciences*. Leeds. 212.

Nycz, Jennifer. 2011. *Second Dialect Acquisition: Implications for Theories in Dialect Representation*. Unpublished Ph.D. Dissertation. New York University, New York.

Pettinato, Michèle, Outi Tuomainen, Sonia Granlund and Valerie Hazan. 2016. Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication. *Journal of Phonetics* 54: 1–14.

Pierrehumbert, Janet, Tessa Bent, Anne Bradlow, Benjamin Munson and J. Michael Bailey. 2004. The influence of sexual orientation on vowel production. *Journal of the Acoustic Society of America* 116: 1905–1908.

Rahman, M. Shahidur and Tetsuya Shimamura. 2005. Formant frequency estimation of high-pitched speech by homomorphic prediction. *Acoustical Science and Technology* 26: 502–510.

Risdal, Megan and Mary Kohn. 2014. Ethnolectal and generational differences in vowel trajectories. In Duna Gylfadottir (ed.) *University of Pennsylvania Working Papers in Linguistics: Selected Papers from NWAV 42*, 20.2. 139–148. Retrieved from http://repository.upenn.edu/pwpl/vol20/iss2/16/.

Roberts, Julie. 2002. Child language variation. In J. K. Chambers, Peter Trudgill and Natalie Schilling-Estes (eds.) *The Handbook of Language Variation and Change*. Oxford, UK: Blackwell. 333–348.

Rodríguez, William R. and Eduardo Lleida. 2009. Formant estimation in children's speech and its application for a Spanish speech therapy tool. In Martin Russell (ed.) *Proceedings of the 2009 Workshop on Speech and Language Technologies in Education (SLaTE)*. UK: Wroxall Abbey Estates.

Singer, Judith D. and John B. Willett. 2003. *Applied Longitudinal Data Analysis*. Oxford, UK: Oxford University Press.

Smith, Jennifer, Mercedes Durham and Liane Fortune. 2007. 'Mam, my trousers is fa'in doon!' Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change* 19: 63–99.

Thomas, Erik R. 2002. Instrumental phonetics. In J. K. Chambers, Peter Trudgill and Natalie Schilling-Estes (eds.) *The Handbook of Language Variation and Change*. Oxford, UK and Malden, MA: Blackwell. 168–200.

Thomas, Erik R. 2011. *Sociophonetics: An Introduction*. Basingstoke, UK: Palgrave Macmillan.

Van der Harst, Sander. 2011. *The Vowel Space Paradox: A Sociophonetic Study on Dutch*. Ph.D. Dissertation. Radboud University of Nijmegen.

Van Hofwegen, Janneke and Walt Wolfram. 2010. Coming of age in African American English: A longitudinal study. *Journal of Sociolinguistics* 14: 427–455.

Vorperian, Houri and Ray D. Kent. 2007. Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research* 50: 1510–1545.

Vorperian, Houri K., Shubing Wang, E. Michael Schimek, Reid B. Durtschi, Ray D. Kent, Lindell R. Gentry and Moo K. Chung. 2011. Developmental sexual dimorphism of the oral and pharyngeal portions of the vocal tract: An imaging study. *Journal of Speech, Language, and Hearing Research* 54: 995–1010.

Wickham, Hadley. 2009. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer Science and Business Media.

Yaeger-Dror, Malcah and Erik R. Thomas (eds.). 2009. *African American English Speakers and Their Participation in Local Sound Changes: A Comparative Study*. Publication of the American Dialect Society 94. Durham, NC: Duke University Press.