Vowel Acquisition in a Multidialectal Environment:

A Five-Year Longitudinal Case Study

Nikki Swift

PhD

University of York

Language and Linguistic Science

July 2023

Abstract

What happens when a child is exposed to multiple phonological systems while they are acquiring language? How do they resolve contradictory patterns in the accents around them in their own developing speech production? Do they acquire the accent of the local community, their parents' accent, or something in between? This thesis examines the acquisition of a subset of vowels in a child growing up in a multidialectal environment. The child's realisations of vowels in the lexical sets STRUT, FOOT, START, PALM and BATH are analysed between the ages of 2;01 and 6;11. Previous research has shown that while a child's accent is usually heavily influenced by their peers, having parents from outside the local area can prevent complete acquisition of an accent. Local cultural values, whether or not a parent's accent has more prestigious elements than the local one, a child's personality, and the complexity of the relationship between the home and local phonological systems have all been implicated in whether or not a child fully acquires a local accent. In the child studied here, a shift from the vowels used at home to local variants always happened at the level of articulatory feature, rather than at phonemic level, in the first instance, and vowels belonging to different lexical sets were acquired at different rates. This thesis demonstrates that acquisition of these vowels takes many years, as combinations of articulatory features stabilise. Moreover, even once a local variant has apparently been acquired, the variety of language spoken at home can leave a phonetic legacy in a child's accent. Naturalistic data collection combined with impressionistic and acoustic analysis in conjunction with a long and sustained data collection period reveals patterns in this child's phonological acquisition not seen in any previous research in this detail.

List of Tables

List of Figures

Acknowledgements

I am very grateful to my two supervisors, Professor Paul Foulkes and Professor Tamar Keren-Portnoy for their continual supportive feedback. Due to various life events, a pandemic and a full-time job, this thesis has been nine years in the making. Their ongoing enthusiasm and interest in my research has helped to sustain me over the years! Thanks also go to my Thesis Advisory Panel members, Professor Marilyn Vihman, Professor Paul Kerswill and Dr. Catherine Laing, whose comments and questions have helped to guide me. For their technical assistance, I must also extend my thanks to Dr. Vincent Hughes for his super-speedy help with Praat scripts when I couldn't get them to work on my Mac, and Dr. Jalal Al-Tamimi for his feedback on my LPC spectra interpretations. Any errors are of course, my own.

I must also thank my lovely colleagues at York St. John University for supporting me over the past nine years, especially for managing without me during the last six months, enabling me to finally finish! Special thanks go to Professor Rachel Wicaksono for supporting my request for research leave, Dr. Leesa Clarke for covering for me during my leave, Dr. Kate Whisker-Taylor for answering all my questions about Yorkshire dialect vowels and to Dr. Clare Cunningham for providing all the writing retreat opportunities over the years. Thanks also go to the students on 2LL200 for helping to transcribe my early data in PHON, particularly Joe Corfield, who produced a draft transcription of the conversation analysed in Chapter 6.

Finally, this thesis would not exist without the data-source himself. It has been the most incredible treat to be able to spend hours listening to recordings of you growing up. Pigpen, I dedicate this thesis to you.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other university. All sources are acknowledged as references.

Elements of chapter 5 were presented at the Building Linguistic Systems conference at the University of York June 2022 (poster presentation).

Chapter 1 Introduction

1.1 Research context

Historically, research in dialectology and sociolinguistics has focused on relatively straightforward cases of informants who have lived in one place for all or most of their lives. While this approach offers researchers a convenient way of conducting research, it is of course, limited in its ability to tell us about the much messier reality of mobility and dialect contact. Similarly, research in children's phonological acquisition also overwhelmingly tends to focus on the acquisition of stable patterns in an assumed homogenous speech community. In linguistically diverse environments such as that of the United Kingdom, however, the linguistic lives of many people are complex, with influences from geographically and linguistically disparate friends and colleagues, as well as family histories that may include geographical, educational and social mobility. With social networks that can extend over continents, as well as diverse family backgrounds, there is the opportunity to look more closely at the acquisition of phonology and to consider the multiplicity of influences on a child, and how these play out in language development. Thomas and Scobbie (2015) observe that the attention given to the phonological development of children raised in a multidialectal, monolingual environment is surprisingly rare, given how frequently families fit this profile.

The underlying assumption that children's acquisition of language is based on a stable target is so endemic that it is rarely acknowledged in the language acquisition literature. A small number of researchers, however, have challenged this assumption by focusing on the acquisition of variation (see for example, Foulkes et al., 1999; Foulkes et al., 2001), and language acquisition in a multi-dialectal environment (see for example, Khattab, 2002; Payne, 1980; Roberts, 1997).

In the idealised environments which dominate the phonological acquisition literature described above, the role of the idealised child is to acquire the apparently single phonological model present in their environment, with no interference. For a child whose phonological input is mixed, the challenge is more complex, and longitudinally uncharted. Multiple, competing linguistic input models mean that a child needs to sort out which sounds in the varieties they are exposed to are contrastive and which are not, and how these systems overlap with one other. At the same time, they need to produce sounds and figure out the allowable parameters for how they can sound and be understood. The small, existing body of research which addresses this tells us that children are more likely to end up sounding like their peers than their parents (see for example, Kerswill, 1996; Payne, 1980; Roberts, 1997), but that there may be complex elements of their peers' accents that they may never fully acquire. Foulkes and colleagues (1999: 1625) point out that while children necessarily develop their accent at the same time as their developing phonology, there has been little research conducted on how children acquire their accents over time. While they were writing over 20 years ago, this area remains under-researched. The existing literature mostly presents a snapshot of children's accents at a single point in time. By comparing the accents of a child's parents and peers, researchers determine where varieties are misaligned, and which sounds in a target accent have been problematic to acquire. What is missing is an understanding of how the transition from parental variety to the accent of a child's peers is navigated, when it happens and whether there is a point at which it can be said that the process is complete.

The research presented in this thesis aims to go some way to addressing this gap; a case study approach offers an opportunity to consider the development of one child's accent over a sustained and significant time-period, mapping this against variation in the child's exposure to different language varieties.

The subject of the study is a child, Henry, from North Yorkshire in the United Kingdom, whose parents speak southern varieties of English (different in many respects from the local variety). The child's accent development was tracked over a period of almost five years, when he was between the ages of 2 and 6 ½ years. A further sample at the age of 6;11 was analysed while he was playing with a friend, bringing the total data collection period to just short of five years. The continual development of the child's accent was analysed in the context of interactions with his mother (me), as he went through key changes such as starting nursery, moving house, starting school, and transitioning between friendship groups and classes at school.

The thesis investigates whether the child acquired his accent from his parents or from the local speech community, or whether aspects of both varieties were present in his speech, and whether this changed over time. The ways in which phonological or phonetic features of these accents present themselves are addressed, in conjunction with the point at which these features appear.

The phonological development of the child is considered with particular regard to vowels, the main source of differences between accents in England, (Wells, 1982). The subset of vowels analysed were selected as they are the site of salient differences between the parents' varieties and the local varieties, and the locus of notable variation and change in this child at sub-segmental level. This research investigates the range of variability in the acquisition of these vowels. Within this range of variability, I consider what can be attributed to expected variation in the child's variety, and what can be attributed to the range of linguistic inputs he has been exposed to. The analysis focusses on whether there is any evidence of the child's orientation to sub-segmental features in his multidialectal environment. Is there evidence in Henry's speech that his phonological acquisition is taking place at the level of the phonological feature rather than the phone, and if so, how does this add to our understanding of the somewhat controversial nature of phonological features? Can learning from a child's acquisition in a multi-dialectal environment inform models of phonological acquisition more broadly?

The research questions investigated are summarised as follows:

(RQ1) Does the child eventually acquire all of the vowel variants and patterns of his peers or does any parental influence remain?
(RQ2) Are vowels acquired at word level, phonemic level, phonetic level or is there evidence of acquisition at a more abstract level, for example, distinctive features?
(RQ3) Is each vowel acquired in the same way and at the same time/rate?
(RQ4) Is there evidence of the vowels being subject to accommodation?
(RQ5) Is the acquisition process complete by the age of 6;06?

1.2 Thesis structure

The thesis is divided into eight chapters. In Chapter 2, I present a review of literature relevant to this research. The chapter is divided into three main themes. In order to understand how Henry might comprehend or interpret the various accents in his environment, I begin by considering speech perception. This is followed by a look at phonological acquisition. As this research focusses on Henry's acquisition of vowels, I concentrate on vowel acquisition and the claims relating to how children break into the speech stream; in what order do children acquire linguistic structures such as word, segment, or phonological feature? Finally I review the evidence concerning variation in children's production. Who are the main influences on children's accents and what do we know about where children acquire their accent from: is it from their family or the community? At what age do they exhibit sociolinguistic variation, and how does their speech vary?

Chapter 3 presents a brief history of the methods of data collection used in the analysis of children's speech sounds. Here, I argue the importance of the case study, the method used in this thesis. This is followed by a rigorous review of the literature concerning the acoustic analysis of children's speech. I present evidence of the complexity of analysing children's speech acoustically, for example, the impact of physical growth of the vocal tract, and the challenges presented by some of the varied speech styles found in naturalistic data. I then outline best practices for the acoustic analysis of children's speech. In this chapter I also describe the existing published formant reference data available for children, against which Henry's speech can be compared.

Chapter 4 describes the methods employed in the collection and analysis of the data in this thesis. Here, Henry's home environment is introduced, including a description of the accents in his home and local community.

The main analysis of Henry's speech is presented in Chapter 5. The chapter is divided into an analysis of the STRUT and FOOT lexical sets followed by an analysis of PALM, START, BATH and TRAP, between the ages of 2;1 and 6;06. Data is analysed from intervals of roughly every two months, though there are some gaps during Henry's early years. The realisations of each lexical set are analysed impressionistically, followed by a supporting acoustic analysis for the ages 3, 4, 5 and 6 years.

In Chapter 6 an analysis of a conversation between Henry and a school friend, James, is presented when Henry is 6;11. This chapter establishes whether any changes in Henry's accent are evident when speaking to a friend rather than his mother. The same lexical sets are analysed as in Chapter 5 but with the addition of two additional lexical sets, GOAT and FACE, which are articulated as monophthongs in his friend's speech. In order to look for evidence of style shifting in this context, an analysis of linguistic features which are variable in Henry's speech such as glottal replacement of /t/ and /h/ dropping is also presented.

In Chapter 7, the discussion connects the analyses in Chapter 5 and Chapter 6 with the literature reviewed in Chapter 2 and Chapter 3. Henry's realisations of each of the lexical set vowels over the entire data collection period are tracked and compared against each other, drawing on the literature to look for an explanation for the distinctive behaviour of each lexical set. A detailed consideration of the acoustic data is made, before siting the results of this thesis in the context of existing research on phonological features and arguing their significance.

Chapter 8 forms the conclusion of the thesis. Here, the research questions raised in this chapter are revisited. Finally, a brief discussion of the opportunities for further research in this area is presented.

Chapter 2 Review of relevant research

I begin by presenting research on children's language perception, specifically the emergence of the ability to differentiate between accents in their environment (section 2.1). I then consider speech production (section 2.2), in particular, the development of vowels, as this is where most of the differences between English accents lie. Here, I also discuss what it is that children acquire and whether this changes over time; is it whole words, phonemes or sub-phonemic elements, for example, phonological features or articulatory gestures? Finally, in section 2.3, I look at structured variation in children's speech and what is known about how it is acquired.

2.1 Dialect awareness in children

An understanding of how children perceive different accents may be helpful in informing us of how they process the accents around them, and what they do with the various accent influences they are exposed to. At what stage in their development can children understand speech spoken in different accents, and at what stage do they notice accent differences? We can assume that a child's ability to correctly map unfamiliar accents to their own will be less developed than an adult's, as their own phonological system is not yet fully developed. Although there is some research on this skill in adults (see for example, Labov, 1989; Flege, 1992), its development in children is less well understood.

The following pages give an account of key pieces of research in the areas of dialect awareness in children. Some researchers focus on the development of accent perception including why and how they think development happens, while others explore the influence of the child's family on their abilities. Do children from multidialectal environments perform more or less well at identifying and differentiating accents than their peers from more homogeneous dialectal environments?

2.1.1 Do infants notice different languages?

Researchers have shown that even very young infants are capable of discriminating between different languages (Mehler et al., 1988). Nazzi and colleagues (2000) carried out a series of head turn preference procedure experiments with 5-month-old infants from the United States. The children were able to differentiate between languages from different rhythmic categories, for example, Italian and Japanese, even if neither were their native language. However, the 5-month-old infants could not discriminate between two languages of an unfamiliar rhythmic category, the syllable timed languages Italian and Spanish, but they did notice the difference between two languages which belonged to the rhythmic category of their own language, the stress-timed languages, English and Dutch (p. 11). Furthermore, the researchers were also able to demonstrate that 5-month-old infants could differentiate between US and British accents. The authors claim that this data suggests that 5-month-old infants are paying close attention to the organisation of sound in the language in their home environment and other languages which share the same rhythmic properties (p. 12). More specifically, the infants must be orienting to prosodic, phonetic and phonotactic aspects of speech (p. 15).

While the infants in Nazzi et al.'s study were found to be attending to a range of cues in the speech stream, van der Feest and Johnson (2016) point to some issues with children attaching importance to so much phonetic information. They claim that the results of previous studies suggest that infants pay attention to much more than is required to understand the language around them and consequently they can have problems in understanding the same word when it is accompanied by different prosodic features, is spoken by someone of a different gender, or is spoken in a different dialect. Infants, they claim, over-specify the detail of speech in the early stages of acquisition (see for example, Schmale et al., 2010); they store phonetic information which is not part of the phonological system, and it is only through exposure to more examples of speech that children begin to learn which aspects of phonetics they need to attend to, and which to disregard for the purposes of interpreting meaning (van der Feest & Johnson, 2016: 90). It is through this process, the authors claim, that children are able to develop a more competent understanding of different accents over time.

2.1.2 The development of comprehension over time

In an early study of accent perception, Nathan and colleagues (1998: 362) conclude that comprehension of speech in unfamiliar accents (henceforth 'comprehension'/'understanding' of accents) grows over time as children experience increased exposure to accent variation. Understanding different accents is a key component of sociolinguistic competence; adults usually have the ability to understand different accents of their own language, though these skills vary from person to person (p. 344). In order to better understand the development of the comprehension of unfamiliar accents in children, the authors collected data from forty-eight London children aged 4 and 7 years. The children were played words spoken by speakers of a familiar (London) and an unfamiliar (Glaswegian) accent in order to determine the children's ability to understand words spoken in an unfamiliar accent. The children were first exposed to a recorded extract of a Mister Men story (Hargreaves, 1971; 1976), followed by a word list comprising twenty individual words (Nathan et al., 1998: 353). The first story and word list were read by the London speaker and the second story and word list was read by the Glaswegian speaker. This served to enable the children to orient to each of the accents, allowing them the opportunity to familiarise themselves with each speaker's phonological system before hearing the individual words. The children were asked what word the speaker was saying – this was in an attempt to make clear that the child should repeat the word rather than imitate it. They were then asked to define it, in order to check their comprehension (p. 354). The children's responses were classified as phonological (the child had correctly mapped the unfamiliar accent on to their own phonological system), phonetic (the child produced an imitation of the unfamiliar speaker's articulation, lexical error (the child fails to map the new accent onto their own phonological system, for example, 'church' becomes 'touch'), and no response (p. 355).

Based on the results of the experiment, the authors concluded that 4-year-olds were less likely than 7-year-olds to understand words spoken in unfamiliar accents. The 4-year-old children were much more likely to give a phonetic response to the unfamiliar accent (44%) than the 7-year-old group (4.8%) (Nathan et al., 1998:357). This, according to the authors, demonstrates that they had not successfully mapped the phonology of the unfamiliar accent on to their own, but that this skill had developed by the age of 7. As an explanation for the younger children's higher likelihood of providing a phonetic response, the authors claim that younger children have not yet built up stable phonological representations, though it is also possible that this behaviour could indicate that the younger children did not understand what was expected of them.

2.1.3 Evidence from research on adults

Flege (1992) investigated the perception of English vowels spoken by adult Dutch speakers of English as a second language. His research on the perception of "non-native" accents provides some insights into the perception of unfamiliar accents that could extend to unfamiliar accents of a single language. He proposes that while a speaker may have tacit knowledge of prototypical speech sounds (see for example Oden & Massaro, 1978; Massaro & Oden, 1980; Samuel, 1977), there is a "tolerance region" around these; a range of articulation parameters that a sound must be produced within in order to be recognised. Vowels, Flege suggests, have a larger tolerance region than consonants, due to typically being more variable throughout their duration than consonants. He speculates that a speaker's representation or categorisation of what is a prototypical vowel or consonant may evolve over time as the range of variants they are exposed to expands. Consequently, they may also become better able to "gauge the degree of divergence" from those prototypes (Flege, 1992: 170).

Nathan and colleagues (1998: 346) claim that these "tolerance regions" will increasingly overlap as a child develops and is exposed to an expanding number of tokens. This, they suggest, could explain why adults sometimes misunderstand unfamiliar accents, but that context should help listeners to resolve ambiguous utterances. They provide an example: "bear", as spoken in Glaswegian is articulated as [ber], while in London it might be [bɛ]. The Glaswegian realisation sounds closer to "beer", pronounced as [biə] in London English, due to the closer vowel articulation — this could result in a London speaker misinterpreting a Glaswegian's pronunciation of "bear" as "beer", as the tolerance regions overlap (Nathan et al., 1998: 345).

Although hearing words in the context of an utterance rather than in isolation may help speakers to resolve misunderstandings such as those described above, Labov (1989) suggests that context does not always help listeners to understand unfamiliar accents. He investigated the perception of the pronunciation of "socks", [sæks], as spoken by Chicago speakers, and found that even when the word was presented within the context of a phrase, some non-Chicago listeners still interpreted the word as "sacks" (Labov, (1989) cited in Nathan et al., 1998: 345). In his experiment, even when a word was placed in the context of a sentence, some adults were unable to decode the sentence if the sound overlapped with another sound in the listener's phonemic inventory, causing the target word to be homophonous with some other word in their lexicon. Listeners were played the phrase "You had to wear socks". Only 20% of non-Chicago listeners were able to correctly identify the word as 'socks', and even Chicago listeners only correctly identified the word as 'socks' 40% of the time. When expanded to contain more contextual information, "You had to wear socks. No sandals", more listeners were able to correctly identify the target word as 'socks', but still only 60% (Labov, (1989) cited in Nathan et al., 1998: 345). Context can be helpful, therefore, even if it doesn't disambiguate a word, but surprisingly, more specific context is not a 'magic bullet'; adult speakers can still map words spoken in unfamiliar accents onto their own phonological system, even when we would expect the contextual semantic information to direct them to do otherwise.

Flege's tolerance regions offer an explanation of how exposure to different speakers shapes an individual's tacit understanding of variability, and how this over time, might explain their ability to interpret unfamiliar accents more readily. However, even though we have seen evidence that children get better at comprehending unfamiliar accents over time through exposure, even adults can still have problems with this skill. Most adults will presumably have been exposed to different varieties of US English through the media, and yet on hearing the Chicago variety, there seems to be a tendency, for some speakers at least, not to exploit that experience and assume that the sounds have the same meaning as those in their own variety, even in the face of reasonably unambiguous phrasing.

2.1.4 Can children recognise and differentiate between accents?

As seen above, part of the process of understanding a variety of accents is in understanding the parameters of variation. Children need to learn which phonetic information is phonologically contrastive, and which is down to phonetic variation (van der Feest & Johnson, 2016: 91). Mulak and colleagues call this ability to ignore non-contrastive phonetic variation, "phonological constancy" (2013: 2065). In their investigation of 15- and 19-month-olds exposed to familiar Australian pronunciations and unfamiliar Jamaican pronunciations, the authors claim that infants are able to ignore phonetic variation and develop phonological constancy by the age of 19 months (Mulak et al., 2013, see also Best et al., 2009). This was evidenced by the children being able to recognise words spoken in an unfamiliar accent. The authors compared the infants' vocabulary size with their ability to recognise words in both the familiar and unfamiliar accents. In the 15-month-olds, increased vocabulary size correlated with increased ability to recognise words in the unfamiliar accent, but by 19 months old there was no such correlation (Mulak et al., 2013: 2075). The authors claim that this correlation occurs because as children's vocabularies grow, so too does their ability to make phonological generalisations (see also Best et al., 2009; Mulak & Best, 2013; Swingley, 2003). At the start of the data collection period for this thesis, Henry is aged 2;1, which according to Mulak and colleagues (2013) should mean that he has phonological constancy and is therefore capable of recognising words spoken in the unfamiliar local accents heard at nursery, or in the local community while out and about.

Mulak and colleagues' study demonstrates that children are capable of recognising words in unfamiliar accents at a very young age; this indicates that children are capable of ignoring extraneous phonetic noise and focus on phonological contrasts while they are still young. But what of children's ability to orient to phonetic/phonological differences in accents and differentiate between them?

Girard and colleagues (2008) tested children's ability to differentiate between a regional French accent and their own variety. They collected data from 5-6-year-old French children from Besançon in the east of France for a series of experiments aiming to establish accent recognition skills. In their first experiment the children were asked to distinguish between adult speakers of their local variety and speakers from Toulouse in the South of France. The task involved grouping speakers into pairs according to colours, for example, voices from Toulouse being attributed to blue characters and local voices belonging to orange characters. The differences between the two varieties include a number of vowel differences, consonant gemination, the appearance of a nasal consonant after nasalised vowels, and the simplification of some consonant clusters intervocalically. The researchers found that the children were unable to reliably differentiate between the speakers (p. 412). In a second experiment, children were asked to listen to local speakers and foreign accented speakers; in this case, English speakers of French, selected due to their strong foreign accent. This time, the children *were* able to discriminate between the two varieties (p. 420). The researchers were unsure whether the results of the first experiment were due to the children not having the perceptual skills to be able to distinguish between the different varieties, or whether the task was too complex for them (p. 417). They had designed an experiment which was aimed at young children — some voices had been mixed up by a naughty computer and they needed to sort them out — but in spite of its child friendly intentions, the resulting experiment was somewhat complicated. The results of the second experiment, where children demonstrated an ability to notice foreign accents, revealed some of the reasons that children cited for their choices; they observed that the foreign accented speakers spoke more slowly than the local speakers, and that they did not have a "real" or "good" voice (p. 420).

Wagner and colleagues (2013) also investigated the ability of children to differentiate between accents. Like Girard and colleagues (2008), they were interested in a three-way distinction, whether the children could tell the difference between two regional accents, one local (Midland[1] US English), a British English dialect (Lancashire English) and a "foreign" accent, in this case, an Indian English dialect. The children's ages were also similar to those in Girard et al.'s (2008) experiments, though they included slightly older children as well; this time the children were between 5 and 8 years old. The authors classify Indian English as a second language dialect as the speakers had all learned another language before learning English. However, they acknowledge that Indian English is a regional variety in its own right, so consider it to be a geographical variety rather than the kind of foreign accented English that might be spoken by an adult learning English as a second language (Wagner et al., 2013: 1069) as was used in Girard et al.'s (2008) research. The researchers attempted to make their experiment more child-friendly than earlier research. Each task involved the children being asked to make a distinction between two of the three dialects. The children were presented with "monster" puppets of either green or purple, which were held up by the researcher while an adult voice was being played. The voice spoke in one of the two dialects, reading a passage from a children's book. The researcher told the children that this was what the puppet sounded like. The children were then told that the puppets had got muddled up and needed to be restored to their family (Wagner et al., 2013: 1073). As in Girard and colleagues (2008), the children were not able to successfully categorise speakers according to regional varieties of the same language, but they did notice that the second language speakers were different from their local dialect (Wagner et al., 2013: 1074). The children did not categorise their home dialect as being different from the regional dialect (Lancashire), and neither did they differentiate between the second language dialect and the regional dialect. The authors claim that

[1] A United States dialect region

the children must classify the regional dialect as being intermediate between the two (p. 1075). In a second experiment, a new group of children of the same age range were asked to link speakers to "cultural items". The children were played the same audio stimuli as in the first experiment and asked to choose between pictures of familiar and unfamiliar housing and clothing to match the voice they heard. For example, the children could choose between a North American ranch house and a mud hut, or between a woman in a business suit or a woman wearing a kimono (this method was borrowed from Hirschfeld & Gelman, 1997). As in experiment one, the children did not differentiate between the local and British English dialects but did associate unfamiliar cultural items with the second language dialect. The authors again note the specific nature of Indian English as spoken by the Maharashtran Indian speakers as having the status of being both a second language variety but also a variety of English in itself with its own grammar and phonology. This is relevant, they claim, in trying to understand why the children notice the difference between their own dialect and the Indian English dialect, but not the differences between the Lancashire dialect and their own (Wagner et al., 2013: 1080). The differences between the US and British English dialects are mostly vowel based, while the differences between the US and English dialects and the Indian English dialect are in both vowels and consonants. The authors argue that the significant variation in the vowel systems of US English may prime children to accept vowel variation more readily than consonant variation. They suggest that children whose first language contains extensive consonant variation may categorise varieties of the same language differently than the children in their research (p. 1081). In comparison to Girard and colleagues (2008), the Indian English speakers will have been fluent, unlike the foreign accented French speakers in their experiment. This makes it apparent that a "bad voice" or slower rate of speech is not the only factor in children's identification of what they perceive to be a foreign accent. In explaining their theory of "gradient dialect realisations", the authors claim that when children hear a regional dialect of their own language, they may be classifying it as a 'bad' example of that dialect – as a kind of unimportant "noise" (Wagner et al., 2013: 1080). This is somewhat similar to Flege's notion of "tolerance regions". In this case, the vowel variants are within the "tolerance region", or allowable levels of "noise", while the consonant differences found in the Indian English dialect might fall outside of those regions.

It is difficult to draw clear conclusions from Girard and colleagues and Wagner et al.'s research. Both indicate that children are capable of discriminating between some accents, but the exact parameters required to aid this differentiation are not clear. Although on the face of it, the two studies look quite similar, a direct comparison cannot be made. The differences between the dialects in each study make it hard to pinpoint exactly what it is that is makes the task easier or harder for the children. In each case, the children were unable to differentiate between two varieties of the same language. The two situations are not particularly easy to compare, however, as for Wagner and colleagues, the claim was that differences in consonants may be easier to spot than vowel variations, however, the two local accents that the children listened to in Girard et al.'s (2008) experiment varied in both vowels and consonants, and yet the children still could not tell these varieties apart – though those children were at the younger end of the range of ages in Wagner et al.'s study. Had they tested children at the older end of Wagner and colleagues' range, they may have found that they were able to distinguish between the two varieties of French. Similarly, in both cases, the children were able to differentiate the two "foreign" varieties, but these were completely different. In Wagner et al.'s research, the speakers were fluent speakers of a variety of English (Indian English) whereas Girard et al.'s foreign speakers of French were described as "non-fluent" and "foreign accented". Therefore, there is not enough information to confidently draw conclusions about what information children need in order to be able to differentiate between accents at this age. This may mean that Henry is not able to differentiate between speakers according to whether they speak the home and local varieties. However, being able to group speakers on the basis of their accent is a different skill from noticing the details of different accents. We saw above in section 2.1.1 evidence that children can tell the difference between languages (Nazzi et al., 2000) and accents (van der Feest and Johnson, 2016) from a very young age. Therefore, perhaps a child may notice the detailed differences between varieties but not attach any social significance to those differences. We will see some similar experiments to Girard and colleagues (2008) and Wagner and colleagues' (2013) work, below, where children have some prior experience of non-local accents, which will offer further insights into children's skills in this area.

Above, we have learned about the development of children's perception of unfamiliar accents and how it compares to adult perception, through looking at a range of experiments where children were exposed to different varieties of speech. However, most children are raised in environments where multiple varieties are present. We may therefore ask, what counts as unfamiliar? Perhaps the only accents familiar to a baby are the accents of their parents or guardians in early life, expanding to close family members who they live with or visit during their first months. With increasing exposure to different settings such as shops, parks and nurseries expanding their linguistic environment, an infant can expect to hear a growing range of voices, and what was once *un*familiar may soon become familiar. Chambers and colleagues (2003) observed that infants with an average age of 16.5 months old were capable of learning new phonotactic rules after only brief exposure. In two head turn experiments, the infants were exposed to new combinations of onsets and codas. The authors concluded from the results that infants learned new permitted consonant positions with the same level of exposure to the stimuli as adults doing a similar test. If we think about how phonotactics can differ between accents, such as rhotic versus non-rhotic accents, this could provide evidence that infants are capable of learning to recognise phonotactic features of different accents quickly.

As outlined in the introduction, language acquisition research has frequently ignored the multidialectal environments that children are often raised in. Even within a home where the parents and other family members have all been born within the same region, we can expect to hear variation in their voices; in some cases this may be due to age. We will learn below how children may be more or less affected by their parents or their peers, and how this can differ even within the same family (Payne, 1980) it may be influenced by gender (see Foulkes et al., 2005), or it may just be idiosyncratic variation (see Local 1983). All but the most unusual children are, to some extent therefore, being raised in multi-varietal environments. Our dialects may even be influenced to some extent by our exposure to media, though the extent of this is disputed; Trudgill (1986: 41) is representative of most linguists, who concede that watching television may result in the adoption of new words or "fashionable phrases" or even a "softening up" of speakers to make them more aware of, or more susceptible to linguistic change, (p. 55) but that media is not the cause of change. Some go further, however, claiming that watching a particular television programme can accelerate linguistic diffusion (Stuart-Smith, 2006; Stuart-Smith et al., 2013).

Of course, there is a difference between hearing variations in familiar accents in your linguistic environment and being exposed to more radically different dialects. There are a small number of researchers who have focused on this particular area: how children *living* in a multidialectal environment interpret accents outside of their community norms.

Above, I have looked at laboratory-based experiments designed to test how infants perceive unfamiliar accents. The following research focuses on children *living* in a multidialectal environment. In these cases there are substantial differences between the accents spoken in the home and those in the local community. Chambers (2002) proposes an "innate accent filter", which he refers to as "the Ethan Experience" after a case study involving one child of that name. He claims that children of immigrant parents filter out any accent differences between their parents and the local variety. His justification for this is that children of foreign sounding parents do not appear to acquire their parents' phonological systems. He cites the example of a boy named Ethan, a child of eastern European immigrants who settled in Toronto, Canada. Chambers provides no details of the child's age and very few details of the features that he develops and when; his claims are based on his own personal observations and reflections and those of Ethan's parents, with no supporting data. According to Chambers, Ethan did not acquire any of his parents' phonological features, for example the alveolar tap as a realisation of /r/, even before he attended school. Chambers suggests that the only explanation for this is that the child cannot hear the differences between his own speech and his parents'. This interpretation seems extremely problematic, however. While it is possible, or even probable that a young child does not notice the differences between their parents' accent and others, how would a child know which parts of the sounds that they are exposed to that they should filter out and which they should reproduce? Chambers claims that Ethan is representative of "countless" other children and that this linguistic behaviour is so typical that it is not noticed. This is in contrast however to many examples that will be presented in section 2.3.1, of children acquiring aspects of their parents' accents in their own speech, even after they attend school.

Nonetheless, there is other research which may at least partly substantiate Chambers' claim, though the children in their experiments are much younger than Ethan. Floccia and colleagues (2012) examined data from thirty-six 20-month-old infants from Plymouth, a rhotic accent area of England, whose parents spoke with either rhotic or non-rhotic English accents. Eighteen of the children had parents who were classified as mono-accentual (both parents spoke with rhotic accents) and another eighteen came from what were classified as bi-accentual families. In these bi-accentual families, one parent spoke with a non-rhotic accent (nine families) or both parents were classified as non-rhotic (another nine families) (p.96). The researchers used a preferential looking task to elicit responses from the infants when exposed to words with the potential for rhoticism for example ‘bird” when spoken in both rhotic and non-rhotic accents. The children were played 12 rhotic words, and 12 non-rhotic distractor words (e.g. bed), along with 14 pairs of non-rhotic control words (e.g. spoon, sock) spoken by two female speakers with rhotic accents and two female speakers with non-rhotic accents. The children heard the words spoken in the frame “Look! *Target*” and were simultaneously shown a picture of the object. Each infant spent longer looking at each picture for words spoken with a rhotic accent. The researchers conclude that these 20-month-old infants only recognised rhotic words, whether or not their parents spoke with rhotic accents. They interpret this as evidence that children’s phonological representations are determined by the accent of the local community rather than that of their parents (p. 98), that they only store one representation, that of the local community variety, and are unable to recognise other variants of the word. The authors observe that bi-accentual children may store more “accent-related” information than the infants from mono-accentual families, but that their research is evidence of “canonical” representations having “special status” in early representations of speech (p. 99).

Chambers (2002) claims that a child doesn’t notice their parent’s accent, while Floccia and colleagues (2012) suggest that the 20-month-olds in their experiment were “only able to recognise words spoken in the rhotic accent of their community, irrespective of the accent spoken by their parents”. While their conclusions may be similar to one another, they are substantiated and explained in quite different ways. Chambers does not make clear what exposure Ethan had to the community, through pre-school childcare etc., though it seems implausible that he would have had none if he had acquired a local accent. He does not make clear whether Ethan exhibits features of the local Toronto variety, only that he doesn’t share features of his parents’ varieties. Floccia et al., on the other hand, make clear that the children have had some exposure to the local variety. They do of course differ in many other ways – Chambers looked at Ethan’s own accent as the manifestation of his perception of features of his parents’ accents, while Floccia and colleagues used a preferential looking task; they do not consider the infants’ own nascent speech, as the children are much younger than Ethan. Floccia et al.’s experiment design is not without problems. Jeffries (2016) highlights the importance of treating their results with caution, as they grouped children with one parent with a non-local accent together with children with two non-local parents. This, she claims, is problematic, as it smoothed out the different extents to which children were exposed to the local variant and ignored the significant role of the primary caregiver and whether they were a rhotic speaker or not. The experiment also asked children to make a three-way distinction between two sets of non-rhotic forms (including the control words), which led the rhotic words to stand out and be potentially more interesting to the child. Furthermore, rhoticity is reported to be variable, or even absent in Plymouth (Wells, 1982: 341), which leads to further questions about the infants' level of exposure to rhoticity both at home and in the community.

Neither Chambers' accent filter, nor Floccia and colleagues' claim that the infants in their study had a "preference" for the local rhotic feature are convincing, but "the Ethan Experience" (Chambers, 2002) does present an interesting conundrum, even if the idea of a filter seems inadequate. This concept will be revisited in the context of parental and peer influence, in section 2.3 below.

Van der Feest and Johnson (2016) claim that research of the type that Floccia and colleagues (2012) conducted is very difficult to carry out due to the difficulty of finding large enough populations in which children are exposed to more than one variety consistently. This seems to ignore the range of varieties spoken in the UK, for example, there are many children who are exposed to both regional dialects and RP/Standard English, and diaspora communities. In spite of this, there do indeed seem to be few examples of published research in this area. We will, however, see a small number of experiments with children from multi-dialectal environments below. These studies consider the impact of coming from a multidialectal home.

In a range of accent awareness tasks, Beck (2014) found that monolingual children from Philadelphia, USA, with at least one parent from outside of the local area (she calls these "outsiders") were less likely to be able to identify different regional accents than a child with two local parents. Beck's experiment involved playing a recording of speakers with a local accent (Philadelphia, located in the northeastern United States) and a speaker of General Southern American English to 66 children aged 5-7 years. Out of 66 children, 28 had parents from outside of the local area – this included 13 speakers who were born overseas, ranging from the UK to India to Mexico (p. 27). The children were asked five questions by the experimenter; these included questions asking the children to point to where they lived on a map of the United States, and to point to any other places on the map that they knew. After playing the accent data, the children were asked explicitly whether they thought that the speaker came from the local area, and if not, where they thought they were from. The majority of the children were able to recognise an accent as local, but when asked to identify speech as *non*-local, 60% of children were unwilling to categorise the General Southern American accent in this way, and even less willing to identify where the speakers might be from. Beck speculates that this may be because they are unaware that accent can be related to geographical location (p. 47). Contrary to the author's presumption that increased exposure to different varieties of language would make children more likely to be able to identify an accent as non-local, children with at least one non-local parent were less successful at this (Beck, 2016: 113). Beck claims that this can be attributed to children being confused about what a local accent *is* (2014: 132). She suggests that many of the children who fell into this category could not yet differentiate between their parents' varieties and the local variety. She proposes that this may arise because as far as the children are aware, their parents *are* local (p. 52).

While Beck (2014) observes that children exposed to multiple varieties are less likely to be able to be able to identify that speakers came from a different region based on their accent than children from monodialectal environments, Jeffries (2016) reports that in her experiments, children of non-local parents *are* better at identifying and grouping speakers with different accents than children of local parents (see also van der Feest & Johnson, 2016). Jeffries' experiment focused on realisations of the vowels found in lexical sets including BATH and FACE (Wells, 1982) in 3- and 4-year-old children in York, Yorkshire, UK. These vowels were chosen in order to elicit differences between accents of northern and southern England. In northern varieties of English, the BATH lexical set is usually realised with an open front vowel [a] while southern varieties generally produce words from this lexical set with a longer open back vowel, [ɑ]. FACE may be realised as a monophthong, [e] in the north, while in southern varieties it will typically be realised as a diphthong, [eɪ].

In a departure from Beck's experiment design, which involved explicitly asking children whether a speaker came from the local area, Jeffries (2016; 2019) used a more child-oriented task to elicit children's perceptions of accent. Twenty pre-school children were played stimuli consisting of a sentence containing a word from one of the lexical sets above. In the first task, the children were asked to group speakers who pronounced words in the same way, for example, "This is my b[a]sket", or "Put me in a b[ɑː]sket". In a second task, designed to see whether children were able to notice whether a vowel was pronounced the same or differently in different words, they were played pairs of sentences such as "We need to walk on the p[a]th" and "I want to walk on the gr[ɑː]ss" (Jeffries, 2019: 16). The third task was the most complex — children were asked to group speakers according to accent features which belong together, for example, identifying that the following two sentences might be spoken by someone with the same accent "What did you br[eː]k?" and "It was a gl[a]ss". Like Wagner et al.'s (2013) "child-friendly" experiment with different coloured monster puppets, Jeffries presented the experiment as a game for the children. They were played a sound clip while seeing an image of a teddy bear wobbling. First, they heard the mummy bear "speak" and then the baby bears. The children were asked to choose which baby bears belonged with which mummy bear by pointing at the screen. The more complex tasks featured different characters, but the principle was the same.

On average, the children performed better than chance in all tasks, but 4-year-olds were significantly better at grouping the accent features than the 3-year-olds. The children with at least one parent from outside of Yorkshire performed significantly better than those who only had parents from Yorkshire (Jeffries, 2019: 21). Jeffries interprets this as evidence that both "maturational factors" (age + gender) and experience play a role in children's success in being able to group similar accents together (p. 28).

So, what of the implications of the research above for this thesis? On the basis of Chambers' Ethan Experience, we might expect Henry to filter non-local variants heard only at home out of his speech entirely, and therefore see only local variants in his speech. However, if Chambers' point is that Ethan only did this because his parents spoke with a foreign accent, how would a child know what is foreign and what is not? We learned from Beck (2014) that young children might not know what local and non-local means in terms of accents, so how could a child be expected to differentiate between a parent's accent being foreign and one that is just not local to the area?

Does age alone result in listeners being able to perfectly identify different accents? In Labov's experiment described above, we saw that even adults can have trouble understanding unfamiliar accents, and they can still have problems even where contextual information is available — though in Labov's experiment, exposure to the unfamiliar accent was brief — only a short utterance was played to the listeners. Flege (1992) told us about how experience could add to a listener's prototypical sounds over time, and consequently, the tolerance regions surrounding them could also adjust with experience. It seems obvious therefore, that previous exposure is important. The following laboratory studies consider the impact of exposure to an unfamiliar accent on an infant's recognition of words.

Van Heugten and Johnson (2014) performed an experiment in which 15-month-old infants acquiring Canadian English were exposed to words spoken with an Australian English accent. The infants were played recordings of both familiar words and "phonotactically legal" nonsense words as part of a headturn preference procedure (p. 343). The 32 infants were split into two groups of 16. The first group was played the lists spoken by a person local to the Toronto area (experiment 1a) and the second were played a recording of the words being spoken by an Australian English speaker (experiment 1b). The infants in experiment 1a spent longer listening to the familiar words than the nonsense words, indicating recognition of the known words, whereas in experiment 1b, there was less difference between average listening times, and only six out of 16 infants listened to the familiar words for longer, indicating that most infants did not recognise familiar words spoken in an unfamiliar accent. A second experiment repeated the process in slightly older children of between 17-18 months and 21.5-22.5 months. In keeping with previous research, the researchers concluded that the infants were able to start recognizing words spoken in an unfamiliar accent at some point between the ages of 15 months and 2 years, with the older group having the most success at recognising the words spoken by the Australian English speaker (van Heugten & Johnson, 2014: 344). In an attempt to discover whether the developing ability to recognise accents is as a result of developing abstract phonological representations, a third experiment was carried out. In adult speakers, brief exposure to an unfamiliar accent can assist in understanding future sentences spoken in the same accent, (see for example Clarke & Garrett, 2004; Kraljic & Samuel, 2005; Bradlow & Bent, 2008). In order to test whether exposure would have the same effect on the children, the researchers played 16 new 15-16-month-old children a two-minute video recording of the Australian English speaker reading from 'The Very Hungry Caterpillar' (Carle, 1969). They then listened to the same two-word list recordings played to the infants in experiment one. However, watching the video recording did not help the infants to recognise the stimuli spoken by the Australian speaker any better than before. The authors claim that this may be because the chosen text contained unfamiliar words such as "pickle" and "cocoon". Out of the 107 words in the story, the authors note that only seven are found in the 150 most frequently known words in Lexical Development Norms for English (Dale & Fenson, 1996) (van Heugten &

Johnson, 2014: 345). In a final experiment, experiment four, a new group of infants were read the text 'The Very Hungry Caterpillar' daily for two weeks by their parents before they were presented with the story in the unfamiliar accent, to enable them to map the phonological representations of the words across from their parents' speech to the Australian English speaker (p. 346). In this case, the 14.5-15.5-month-old infants were played the Australian English speaker reading the word list used in the previous experiments, having previously been played the 'Very Hungry Caterpillar' video read by the Australian speaker beforehand. This time, the infants preferred the real words over the nonsense words, demonstrating that, like adults, some exposure to an unfamiliar accent may prime infants to understand unfamiliar accents (p. 347). In this final experiment, the data collection occurred before the age where the children might be expected to be able to recognise the words without the priming video, as was the case in the authors' previous experiments one, two and three. The children were only able to recognise the new words after being familiarised with the text by their parents first.

In a subsequent series of similar experiments, this time using a preferential looking paradigm, van Heugten and colleagues (2015) established that by the age of 25 months, infants who were exposed to an unfamiliar accent had no advantage over infants who had not heard it before. This suggests, the authors claim, that the children's ability to map the phonetics to their mental lexicon has developed sufficiently well to enable them to tolerate some variability in accents by the age of 25 months. The authors propose that this is likely to be attributable to the increase in the child's vocabulary by this age (van Heugten et al, 2015: 59). They do acknowledge, however, that this does not mean that in a more complex task, children would not benefit from prior exposure to an unfamiliar accent. It has been well documented that adults can benefit from prior exposure to unfamiliar accents, as seen for example in Flege (1992). The authors also point to evidence that in some experiments, 20-month-olds have been shown to be able to recognise words spoken with an unfamiliar accent (see Mulak et al., 2013, above), while they were not able to in their own study. They conclude that this is because representations may still be "fragile" at this age (van Heugten et al., 2015: 60). In an attempt to reconcile the differences between their own results and Mulak et al.'s, they also explore the notion that it is difficult to determine how different an unfamiliar accent is to the local accent in various experiments. They claim that the accents in their study, Canadian and Australian, may be more distinct than the accents in Mulak et al.'s experiments, which were Australian and Jamaican Englishes. They do not offer any analysis of these pairs of accents to support this claim, however.

Further work on the effect of exposure on recognition of words includes that of White and Aslin (2011), who performed a preferential looking task experiment in which they manipulated vowel sounds to create an artificial accent. In this accent, vowels overlapped with vowels from a different phonological category. 18-20-month-old infants were played recordings of six familiar words (names of familiar objects) containing the vowel [a] in their local dialect. During an exposure phase of the experiment, the children heard words containing a "training vowel". They listened to recordings of a set of words being pronounced in the standard way, or with the vowel shifted to [æ], for example, the word dog was pronounced as either [dag] or [dæg] (US conventions) (White & Aslin, 2011: 375). During the test phase of the experiment, they listened to the same words as presented in the exposure phase, and further words, including *sock, block, bottle, ball* and *car*. Where the children had only had pre-exposure to standard pronunciations of the vowel, they were unable to recognise words containing the unfamiliar vowel. However, if they were exposed to the shifted vowel [æ] during the training phase, the children were able to recognise words produced with the unfamiliar vowel, providing they were exposed to another example first (p. 380). In other words, they were able to extend new phonetic information across analogous words. The authors claim that this is evidence that very young children quickly update their phonological representations in the face of new information (p. 382).

In a similar experiment, van der Feest and Johnson (2016) performed a comparison of speech perception between toddlers whose parents were both from the same dialect area and those with parents from a different dialect area. They ask whether young children are able to understand that speakers of the same language may organise their sounds in different ways. The authors note that children are assumed to learn their phonological system through exposure to the distribution of sounds in their environment. Where they hear overlapping distribution of sounds (what the authors term 'unimodal' distribution), they will learn that these are more likely to be allophonic variants, and where they do not overlap (bimodal distribution), a child will learn that these sounds are more likely to be contrastive in their language (see for example, Maye, Werker & Gerken, 2002). In the case of children who are exposed to more than one dialect where sounds have different distributions of phonological contrast, the authors ask how children resolve those contrasts. They argue that children who are exposed to more than one dialect are in a similar position to children learning more than one language, where children have to "track" more than one set of distributions (see for example Sundara & Scutellaro, 2011). On the other hand, children may "collapse" or simplify the two variants. If two sounds are contrastive in one variety but not in the other, a child may assume that the sounds are contrastive in neither variety. This could manifest itself in children treating different realisations as free variation (van der Feest & Johnson, 2016: 91). In most varieties of Dutch, contrast between voiced and voiceless fricatives has been neutralised in word initial position, while a small number of dialects have retained the contrast. Therefore, all word initial fricative realisations are voiceless for most speakers. Van der Feest and Johnson (2016) performed a preferential looking task featuring a mispronunciation detection paradigm. The children taking part in the study were born in and lived in the Dutch city of Nijmegen. The local variant in Nijmegen is the same as the Dutch Standard in that speakers do not distinguish between voiced and voiceless variants in word initial position. Sixty-four children with an average age of 24 months were selected for the study. The children were divided into two groups, a monodialectal input group of 32 children whose parents were from the local area, and a multidialectal input group of 32 children whose parents had relocated to Nijmegen from Limburg, a dialect area which has retained the contrast between voiced and voiced fricatives in the word initial environment. These children had been exposed to the non-contrasting local variety as well as their parents' fricative contrasting variety. The aim of the researchers was to determine the *expectations* of the children in each group. Would the multidialectal input infants notice when a speaker produced an 'incorrect' (i.e. apparently hypercorrected) variant for their variety? All groups listened to a speaker producing the word sock 'sok' as both [sɔk] and [zɔk] and the word soup 'soep' as both [sʊp] and [zʊp]. In both dialects, the target pronunciation was [s]. Half of the children within each of these groups were exposed to data produced by a speaker who naturally produced devoiced fricatives in word initial position. The other half listened to the voice of a speaker who naturally retained contrast of fricatives in word initial position (in words not included in the task). The researchers noted that in the group who had parents from out of the area, the multidialectal group, the children were more likely to notice mispronunciations in a speaker who shared their parents' fricative contrasting variety. They did not, however, notice mispronunciations by the speaker from the local area. In the monodialectal input group, the accent of the speaker did not affect the children's ability to identify the mispronunciations; the children did not notice the mispronunciations in either speaker. In this case, the children's exposure at home was to no contrast in fricatives in word initial position, as is the standard in Dutch accents. The authors claim that this is as they expected; the monodialectal group were still able to understand words spoken with initial fricatives, whether they were devoiced or not, (van der Feest & Johnson, 2016: 101). They suggest that the monodialectal input group may ignore fricative voicing because they are not familiar with other accents – they have no experience of hearing variation in this feature, so they do not notice it. The multidialectal input children on the other hand, do not assume that different realisations are an example of free variation, which would be indicated by ignoring the difference in pronunciation. Rather, they are able to alter their expectations and adapt to pronunciations depending on who the speaker is. The authors further note that while their results correspond with the conclusions drawn by van Heugten and Johnson (2014), Best and colleagues (2009) and White and Aslin (2011) above, they are at odds with Floccia and colleagues' (2012) claim that children recognise words spoken in the dominant local accent over the accent of their non-local parents.

2.1.5 Summary

In the work reviewed above, we have seen that very young children are able to notice differences between accents. In the case of Floccia and colleagues (2012) this manifested itself in a "preference" for the local accent over the accent of the parents. In that case, the children were too young for the researchers to consider the impact of this on their acquisition, but in Chambers (2002), Ethan apparently did not notice his parents' variety at all; the evidence for this was that he acquired no features of their accents. In both cases, the children paid more attention to the local variety, but is this because they prefer it over their home variety, or because they had more exposure to the local variety?

Unsurprisingly, most of the research points to children getting better at comprehending or recognising accents as they get older. It is interesting to note that even adults can struggle to comprehend unfamiliar accents, however, and van Heugten and Johnson's (2014) research offers a useful addition to our knowledge about how early children can orient to unfamiliar accents. Of particular interest is that they experience the same benefits of previous exposure to an accent as adults. However, we did see some examples in Beck (2014; 2016), Girard et al. (2008) and Wagner et al. (2013) of older children not being able to group speakers according to their different accents. There are a few plausible reasons why older children might perform less well in accent discrimination experiments than babies. Firstly, the data coming from babies may be flawed, as discussed in relation to Floccia et al.'s experiment above. Additionally, infants are only paying attention to phonological stimuli. 4- to 5-year-olds, on the other hand, are also processing meaning (Beck, 2016) (see also Werker & Fennell, 2004). Maye and colleagues (2002) observe that various studies show children to be extremely competent at distinguishing speech sounds from one another, while adults' ability to perceive phonetic differences is coloured by the phonetic organisation of their first language (Maye et al., 2002). For example, Werker and colleagues (1981) found that 6-month-old babies from an English-speaking environment with no prior exposure to Hindi could discriminate between two sounds found in Hindi, while adult speakers from an English-speaking environment could not (Werker et al., 1981: 354). This development in discrimination between speech sounds can mean that infants lose the ability to differentiate between sounds that they could previously distinguish (Werker & Tees, 1984: 50). Moreover, the difference in experiment design between infants and young children may be the key to these differences. For the tests on infants, the researchers determine discrimination via a head turn task, whereas the young children are being asked to group speakers. In some of the experiments, children had to infer that accent was a relevant factor. They may well be able to hear the differences between the accents but understanding the social and/or geographical significance of those phonetic differences is a different skill entirely.

The differing conclusions reached by Jeffries (2016; 2019) and Beck (2014; 2016) around the impact of having a non-local parent on the ability to identify different accents may be partially attributable to different experiment designs/differences between the accents in the experiment. Wagner et al.'s (2013) research aligns with Beck, in that they claimed that children classified regional dialects as being in some kind of in-between zone between a local and foreign accent (see also Girard et al., 2008). The salience of the features that differed between the accents may also be part of the explanation (see Floccia et al., 2009). However, Beck's explanation of children from multilingual or multidialectal environments being less willing to commit to a decision about what is local and what is not, is also plausible. There were of course, other differences between the experiments which make it difficult to make a direct comparison. The children in Jeffries' study were younger. She also focused on a more limited set of sounds. Beck's questions were overt, and thus less 'child friendly'. Did the children fully understand the questions posed in Beck's experiment? When the children were asked whether a speaker came from "here", what did they understand by that? Undoubtedly, the combination of these factors is at least partially responsible for the difference in results between Beck (2014; 2016) and Jeffries' (2016; 2019) results.

Overall, we have seen evidence here that in general, children seem to get better at recognising different accents as their own vocabularies grow, that some children can be primed by only brief exposure to a new accent, while others require more specific and targeted exposure to both their home dialect and the unfamiliar accent. Finally, some children seem to be better at identifying the differences between accents than others and having a parent from outside of the local area may have an impact on that ability.

In the case of this thesis, at the start of the data collection period, Henry was in between the ages of the infants who underwent some of the perception experiments (e.g. Mulak et al., 2013; Nazzi, 2000; van Heugten & Johnson, 2015) and the children involved in the grouping experiments (e.g. Girard et al., 2008; Jeffries, 2016; Wagner et al., 2013). According to the studies reviewed above, we can expect that he may have some 'phonological constancy' (see Mulak et al., 2013), and his exposure to the local dialect may mean that he is able to tolerate differences between different variants in the home and local varieties (see van Heugten & Johnson, 2015). While he may be able to group speakers of different dialects together as he has parents from outside the local area (see Jeffries, 2016), he may not understand the social significance of these accent differences (cf. Beck, 2014).

We now turn to literature concerning how developing perception manifests itself in the acquisition of the child's own accent.

2.2 Phonological acquisition

As we saw above in section 2.1, children are able to notice detail in the speech around them from the earliest age, including the organisation of sounds in their environment (see also Werker et al., 1981; Werker & Tees, 1984). The language input that infants are exposed to is extremely complex. Each time a child is exposed to language, it will be on some level, unique. Even in the shorter utterances or more restricted vocabulary typically found in child directed speech, children still have a hugely complex task in disentangling this input, sorting out the phonological structure, lexis and grammar of a language, which they have to do unaided (Ferguson & Farwell, 1975: 419).

Acquiring language includes a huge range of elements that the child must learn. Aside from the acquisition of lexis, they must learn the structural organisation of their language: phonology, morphology and syntax, alongside rules such as how much variability is allowable and in what contexts, when variability is optional and when it is not, how meaning works in explicit and implicit ways, and what is considered polite and impolite within their society and others. Here, I focus on phonological acquisition, and in particular, what is known about the typical development of vowels, as these are the sounds being examined in Henry's speech in this thesis. How much variability we can expect to see in the developing system, why variability occurs, and how that relates to children's organisation of language will also be explored.

2.2.1 Underlying representations

The input from adults, described above, can be complex, but it is the job of the child to internalise what they hear in their environment and start to construct their own phonological system. It makes sense therefore, that linguists should be interested in establishing the connections between an adult system and a child's. What does a child's phonological system look like and how is it established? Smith (1973) claims that a child's underlying phonological system is mapped from the adult system (see also Stampe, 1969). He supports his position with evidence from his case study of his own child, Amahl. Amahl's superior speech perception over his production (see discussion of perception in section 2.1 above) is evidenced through a series of tests of the child's perception, for example, by asking him to differentiate between "mouth" and "mouse" when these two words sound the same in his own speech. He also reports on the child's ability to reflect on his speech production and notice when he has acquired the ability to produce a word accurately, for example moving from pronouncing "yellow" as [lɛlo] to [jɛlo] (Smith, 1973: 137). As further evidence for the adult phonology as the underlying system, Smith points to the "across the board" nature of the language acquisition process" (p. 138-9). He suggests that it has often been claimed that when a child acquires a new sound or sound combination, they are able to instantly roll it out to all words containing that sound, demonstrating an understanding of the underlying system. They do not need to hear examples of the sound in all of those different environments in order to update their own pronunciation. Smith claims that the reality is somewhat different. Often children exhibit free variation between previous and new realisations for a few days, or in rare cases, weeks. Smith presents the child's ability to "restructure" his phonetic outputs based on updates to his understanding of a particular rule as further evidence of the child's internalisation of adult forms. We may expect, therefore that the subject of this thesis, Henry, will exhibit variation after being exposed to new local variants, but will the variation continue for only a few weeks, when he has a variety of dialects in his environment?

Macken (1980) re-analyses Smith's data and, specifically critiquing Smith's perception claims, suggests that some of the child's representations are incorrectly stored based on mis-hearing the adult (see also Braine, 1974; Braine, 1976). Macken highlights what has come to be known as the *puzzle-puddle-pickle* problem (Dinnsen et al., 2001). Smith proposed two rules which affected these words – a velarisation rule affecting *puddle* words, i.e. words which feature alveolar stops being velarised before /l/, and a stopping rule where affricates, stridents and non-sonorant continuants are realised as stops in *puzzle* words.

(1) Amahl (age 2;2-2;11)
(a) *Puzzle* words realized as *puddle* words (Stopping)
[pʌdl̩] 'puzzle'
[pɛntl̩] 'pencil'
[wɪtl̩] 'whistle'
(b) *Puddle* words realized as *pickle* words (Velarization)
[pʌgl̩] 'puddle'
[bɔkl̩] 'bottle'
[hæŋgl̩] 'handle'
(c) *Pickle* words realized target appropriately
[pɪkl̩] 'pickle'
[təːkl̩] 'circle'

Figure 1: Puzzle, puddle, pickle words (Smith, 1973) presented in Dinnsen et al. (2001)

Through an analysis of these and analogous words in the data, including a detailed look at exceptions, Macken (1980) concludes that the *puddle* words in Smith's data were mis-heard by the child and stored as velars in the first place, rather than being subject to a velarisation rule. The significance of this is that Macken disputes Smith's claim that a child's underlying representations are the same as an adult's and that it is only their articulatory limitations which result in their surface representations being different from adult speech. Macken's assertion is that that this proposal overlooks the possibility that the child's underlying representations may contain errors, which could affect their phonetic realisations.

Other authors have highlighted the impact of perception errors on a child's developing system. Vihman (1982) analysed her daughter, Virve's, speech, pointing to evidence of perception errors of the realisations of dental and labiodental phonemes, /θ, ð/ and /f, v/. Virve demonstrated these perceptual errors in her writing, representing "California" as KALATHORNJA, "stuff" as STUTH and "birthday" as BRVTEI. Vihman points out that while Smith later adjusted his model to account for the feedback from Braine (1976) and Macken (1980), he remained unconvinced that children's own realisations might be stored as part of their phonology. Vihman, on the other hand, suggests that if a child's surface realisations are lexified by the child, then they must be in some way stored by them, leading to a feedback loop from output lexicon to input lexicon. We will see in Chapter 5 evidence of Henry misallocating sounds he hears to an incorrect phoneme, and that he can also lexify his production 'errors'.

2.2.2 Variability in children's speech

So, the child's perception is ahead of their production, but their perception may contain errors which will be constantly overwritten as the child encounters new examples in their linguistic input. Meanwhile, the child's articulatory abilities are developing. A brief description of the development of infants' speech production follows.

Babbling typically begins at around 7-8 months and is characterised by vocalisations featuring a combination of phonation and articulatory gestures. These articulatory gestures have timings similar to those seen in adult speech (Davis & MacNeilage, 1995: 1199). It is generally agreed that babbling and early words have so much in common that it can be difficult to distinguish between the two (Vihman et al., 1986: 3).

It is well known that children adopt different approaches to phonological acquisition. Children develop at different rates (Sosa & Stoel-Gammon, 2006: 35) with different children favouring different strategies. For example, in Table 1 below, one child, Amahl, favours preserving the sonorant element of an initial consonant cluster, while the other, Gitanjali, retains the initial obstruent.

Amahl		Gitanjali	
snow	[no]	snow	[so]
small	[mɔ]	snookie	[sʊki]
slide	[laɪt]	sleep	[sip]
sleep	[wip]	smoke	[fok]
snake	[ŋeɪk]	sweater	[fɛrə]
smell	[mɛʊ]	smell	[fɛw]

Table 1: Children's differing strategies for consonant cluster reduction (Johnson & Reimers, 2010: 20)

Early articulations of words, typically beginning in the months following a child's first birthday (Clark, 2016: 87), are extremely variable, and infants often continue babbling alongside early word production (Vihman et al., 1986: 3). Children's production of vowels and consonants is much more variable than the variability found in adult speech (Clark, 2016: 108); so much so that children frequently produce multiple different realisations of the same word (Sosa and Stoel-Gammon, 2006: 35). For example, Ferguson & Farwell (1975) present data from a single child who pronounced the word 'pen' in 10 different ways within a single 30-minute recording, as she experiments with reorganising the features of bilabial and alveolar articulation and nasality. Donegan (2013: 106) reports on her daughter's varying pronunciations of her own name, 'Elizabeth'. These varied from what she describes as "hyperarticulate" [ɪ'jazəbɪs] to "hypoarticulate" ['miβbɨf]. Accuracy, she claims, may be better in imitated words than in words produced spontaneously. This could be because children are concentrating more on articulating, which might be reduced if they were simultaneously processing lexical and grammatical information (Donegan, 2013: 106). There is some evidence that this may not always be the case, however. Leonard and colleagues (1978) suggest that there aren't any substantial differences between imitative data and spontaneous speech, except for that in very young children (under 2 years), who may produce syllable shapes and some consonants not found in their spontaneous speech.

Sosa and Stoel-Gammon (2006: 35) note that there is little systematic research on variability in the speech of infants. They observe that there is also disagreement on how variability may progress throughout linguistic development. Variability may decrease over time, because as the child develops segmental phonological representations, their tacit knowledge of which features typically belong together in the phonemes of their language may increase. On the other hand, an increase in variability of realisations may represent instability of the system as the child shifts from whole word representations to segmental organisation. Sosa and Stoel-Gammon observed increased variability in infants' phonetic realisations, which peaked at around 150-200 words when they began combining words. This is the point at which, they suggest, a child is beginning to form phonemic representations (p. 48).

2.2.3 Vowels

Dodd and colleagues (2003: 618) report that the majority of research is on the acquisition of consonants rather than vowels (see also Bankson & Bernthal, 1998; Pollock, 1991). Historically, researchers have suggested that vowel acquisition is typically complete at an early stage in the acquisition process, and that errors are rare (Penney et al., 1994; Stoel-Gammon & Pollock, 2008). Donegan suggests that this may be because vowel production does not draw on as many features as consonant production does – jaw height, tongue advancement, tenseness and labiality or palatality are the main features to be mastered (Donegan, 2013: 71). Davis and MacNeilage (1990: 16) write that "vowels are the poor relations of child phonology" (see also Ball & Gibbon, 2002: xi). The authors qualify this by estimating that there is only one study on vowels for every twenty on the acquisition of consonants. Kent and Rountrey (2020) suggest that the situation has changed in the last thirty years; a significant development was the publication of Ball and Gibbon's 2013 book on vowels and vowel disorders. They acknowledge, however, that research on children's consonant articulations continues to dominate the literature (Kent & Rountrey, 2020: 1749). In contrast to the suggestion in the literature that vowel acquisition is unproblematic, Davis and MacNeilage (1990: 27) claim that children's vowel articulations demonstrate complex patterns. Examples of these patterns will be discussed below.

Vowels are defined by phoneticians as articulations without obstruction in the vocal tract, such as the narrowing of the articulators to the extent where turbulent airflow is produced, and they feature laminar (smooth) airflow. For phonologists, the definition relates to the position within the syllable. Vowels are found in the position of the nucleus. This results in some different classifications of vowels: for example, [n], [l], [j] and [w] may be classified as a vowel or consonant according to the position of the author (Ball & Gibbon, 2013).

Infants begin imitating specific vowels spoken by adults in their environment from around 12 weeks of age (Kuhl & Meltzoff, 1996). The production of vowels in early words, however, can bear little resemblance to the vowels found in the adult target. Clark (2016) gives the example of the word “squirrel” being realised as [ga] by an infant. Ignoring the reduction of syllables and elision of consonants, [a] seems unrelated to the vowels in the target. Vowels in early words, Clark notes, are usually produced without substantial narrowing of the vocal tract (p. 108). Vowels found in pre-speech babbling are typically front and central articulations (Kent & Bauer, 1985: 522).

The reported order of vowel acquisition in infants is inconsistent. Penney and colleagues (1994), report on the various orders that have been observed by researchers. After observing the phonological development of children in diary studies of children acquiring multiple languages, Jakobson (1968: 47) for example, claims that children universally acquire /a/ first, before contrasting with /i/ and /u/, while Paschall (1983) observed earlier acquisition of high vowels over low vowels. Penney and colleagues (1994: 48) note that it is unclear whether any inconsistencies in observations are due to true variability in children’s acquisition patterns or whether these differences are attributable to different methodologies. While some literature suggests that babbling infants typically favour low front vowels (see for example, Buhr, 1980; Kent & Murray, 1982), Davis and MacNeilage claim that there has been no suggestion that this carries into early word production (1990: 16).

James (2001: 460) notes that the justification for many linguists not including vowels in their data sampling is that they claim that the vowels have been acquired by the age of 3 years (see Smit et al., 1990 for an example of an experiment which uses this justification). Davis and MacNeilage (1990: 16) acknowledge that this is the dominant claim made in the literature but argue that vowel acquisition is more complex than is generally accepted and is thus worthy of study. In pursuit of evidence for this claim, they collected data from a single girl between the ages of 14 and 20 months. The authors analysed a randomly selected subset of data, noting similarities and differences between the vowels found in concurrent babbled vocalisations and recognisable words, and any patterns found in the relationship between vowel articulations and surrounding consonants, (see also Kent & Bauer, 1985; Stoel-Gammon & Pollock, 2008).

The child's vowel articulations were judged to be correct in fewer than 60% of instances, based on the adult target vowel. 'Correctness' was a judgement made by the primary transcriber, after their transcriptions were checked through a process of inter-rater reliability. Vowels were labelled as correct unless they sounded more like another vowel in the adult variety, in which case they were deemed to be incorrect (Davis & MacNeilage, 1990: 25). This means that vowel articulations which were close to the target articulation were still judged as incorrect, such as [ɪ] as a realisation of /i/, or [æ] for /a/. The authors note that in these cases, there was a tendency for the child to produce the vowel more open and front than the target sound. They also note the role of schwa; this vowel was substituted in a way which did not correspond with the more low/front pattern seen with other vowels and was substituted for vowels across the vowel space. This, the authors claim, suggests that the issue is not one of inadequate motor control, but that the child uses the neutral vowel as a default.

Davis and MacNeilage's case study concludes that the child was able to produce all vowels in monosyllables, demonstrating the child's full use of the vowel space. Surprisingly, a further analysis revealed that vowels typical in pre-speech babbling in other studies were produced correctly less frequently than vowels not typical of pre-speech babbling (Davis & MacNeilage, 1990: 21). The authors have no record of this particular child's pre-speech babbling for comparison, however. A relationship between vowel articulation and number of syllables, surrounding consonants, and stress was also revealed. The well-formed vowels observed in monosyllables led the authors to conclude that a single syllable utterance is an easy place for children to be able to produce vowels correctly (p. 24). Moreover, clear evidence was observed, the authors claim, of a complex relationship between the vowels articulated and the consonants. For example, front, close vowels tended to appear with alveolar consonants. Notably, it is the vowel which influences the realisation of the consonant, not the other way around. Open and mid vowels tended to appear alongside labial consonants, but there was weaker evidence to support a connection between velar consonants and back vowels (p. 26). A connection between vowels and stress was evident, in that the child showed an almost categorical tendency to produce a neutral vowel (schwa-like) in the second syllable of two syllable words (221 out of a possible 231 two syllable words). While an unstressed vowel is often found in the second syllable of English words and the child appeared to observe that trend, the authors suggest that she was not able produce the "specific vowel quality required" (p. 22). In fact, where schwa *was* the target, she only produced the vowel correctly in 55% of cases. The authors interpret the child's strong preference for schwa in second syllables as evidence that she knows that the second syllable is often unstressed, but that she is unable to produce vowels of the target quality (p. 23).

While the authors note the complex patterns evident in the child's developing mastery of vowels described above, they observe that the infant's vowel articulations did not show a clear and consistent trend towards more 'correct' realisations over the 6-month duration of the data collection period (Davis & MacNeilage, 1990: 20). Likewise, Donegan (2013: 108) notices that children "take many different paths to the mastery of the vowels of their languages".

Although Davis and MacNeilage's (1990) case study offers an insight into the development of vowels in an individual child, the authors' classification of 'correct' or 'incorrect' seems to ignore the range of variability of vowel realisations found in adult speech (see for example, Local, 1983; Keating & Huffman, 1984; and Veatch, 1991, discussed in section 2.3 and Chapter 3 below). There is evidence that variability is an aspect of acquisition which may alter over time, but the idea that adults produce vowels in the same way, consistently, is disputed. A distinction must be drawn, however, between the kind of variability discussed by Ferguson and Farwell (1975) where gestures such as articulatory closures, nasality and degrees of openness are reorganised, and the kind of variability seen in adult speech, as discussed by Veatch (1991, see below). These kinds of variability range from the extreme, which makes for unintelligible utterances unless given a clear context, and typical levels of variation, that is, variation in tongue advancement, jaw position and lip rounding which do not impact on communication and generally go unnoticed by interlocutors.

Donegan (2013: 108) comments that variation in vowel production may go unnoticed by researchers, resulting in an absence of variety appearing in published research. She also claims that what may be seen as atypical variation may not be atypical at all but is just a case of unusual substitutions continuing to be made at a point when most children have already resolved these developmental issues. Vowel variation in the early stages of language acquisition is likely to occur as a result of a child's lack of articulatory control. We will see in section 2.3 below however, that articulatory control is not the only factor in vowel variation, and there may not necessarily be a path, straight or otherwise, to consistency.

Above, we learned that vowels are generally overlooked in children's language acquisition research as they are assumed to be unproblematic, but on closer inspection, they present an opportunity to explore how children interpret and produce the vowels they hear around them. Vowels are where most accent variation occurs, for example, as we saw in section 2.1.4 on accent perception, above. We have seen that children (and adults) may map vowel categories incorrectly which can lead to errors in comprehension, but we have also seen that in some circumstances they may be capable of distinguishing vowel variation subconsciously, for example by grouping puppets who "belong together." Given that there is the potential for so much inter-speaker variability in adult vowels, how might this affect children's acquisition of vowels? In order to look more closely at the production of vowels, it is necessary to consider their composition.

2.2.4 Distinctive feature theory

Traditionally, there are three elements to a phonological system. 1) A system which represents contrasts (a phonematic system), 2) a set of rules determining well-formed syllables (phonotactics), and 3) the relationship between the phonological system and its phonetic realisations (including, for example, allophonic rules) (Ladefoged, 2005).

Relating to all three points above, distinctive feature theory aims to make connections between phonetic realisations and the phonological system. In its early stages, its primary purpose was to hypothesise a reduction in the number of phonological contrasts in a language, though over the course of the years its applications have been expanded to include the domains of phonotactic constraints and phonological rules (Mielke, 2011). Trubetzkoy's (1969) publication in 1936 was the earliest to propose phonological oppositions, but early distinctive feature theory is largely associated with Jakobson's work in the early 1940s. Jakobson claims that a large number of contrasts (as in the case of phonemic contrasts) is perceptually demanding, and therefore the number of theoretical contrasts should be reduced (Jakobson, 1942, cited in Mielke, 2011). Jakobson and colleagues (1963) distil these differences down to only 12 binary oppositions, which are mostly defined acoustically and, they claim, are underlying to all languages.

Jakobson et al.'s (1963) system proposed that phonemes comprised bundles of binary articulatory, acoustic, and perceptual features, which distinguish sounds from one another. These oppositions included, for example, distinctions such as vowel vs consonant, voiced vs voiceless, and acoustic features such as grave (low frequency energy) vs acute (high frequency energy). Jakobson argued that if phonemes were considered to be "primordial" oppositions, there are 28 binary distinctions differentiating the eight vowels of Turkish. With his distinctive features, on the other hand, this could be reduced to only three binary distinctions (Mielke, 2011). Over time, various researchers have developed Jakobson et al.'s distinctive feature theory. Notably, the original 12 binary oppositions have been expanded and have moved away from acoustic correlates towards articulatory parameters, for example in Chomsky and Halle's (1968) Sound Pattern of English (SPE), which utilised a much larger set of features. Chomsky and Halle's SPE continued with binary oppositions, and split these into five different areas: major class features (e.g. +/- consonantal, +/- sonorant), cavity features (e.g. +/- anterior, +/- high), manner of articulation features (e.g. +/- continuant, +/- tense), source features (e.g. +/- voice, +/- strident) and prosodic features, which they mentioned but did not report on in their 1968 publication, (Chomsky & Halle, 1968). Though Jakobson's aim had been to reduce the number of theoretical features, the number of distinctive features proposed in SPE was more than double the number in any previous model (McCawley, 1974). Honeybone (2009) refers to SPE as the 'standard model' of distinctive features, though they continue to be refined by phonologists.

2.2.4.1 The application of distinctive feature theory

Dodd and colleagues (2003) draw a distinction between the phonetic and phonological aspects of the system described above by Ladefoged (2005), in research on children's language development. The former, they suggest, relates to a child's ability to make particular articulations, for example, a [p], but the child may not be capable of producing this sound in all of the environments where it is found in adult speech. The latter relates to the child's ability to produce those sounds across different environments, reliably (Dodd et al., 2003: 618), as part of the child's transition to producing adult-like articulations.

Between the 1960s and the 1980s, a number of researchers were interested in analysing phonological development through the acquisition of distinctive features as proposed by Jakobson, Fant and Halle (1963), (see for example, Hodson & Paden, 1978; Menyuk, 1968; Wong & Irwin, 1983). These studies produced quite varied results. For example, Hodson and Paden (1978) reported that most children had acquired all features except [coronal] and [high] by the time they were 4 years old, while Wong and Irwin (1983) found that children had generally acquired all features apart from "linguadental" (their own feature) (quoted in Dodd et al., 2003). One possible explanation for these varied results is that these were large scale studies which did not follow the "sequential development" of individual children or account for individual differences. This is problematic as it can ignore the known issue of children apparently acquiring sounds and then seemingly regressing temporarily (Dodd et al., 2003). So these inconsistent results may not have been attributable to flaws in distinctive feature theory *per se*, but rather, other methodological issues.

In section 2.2.3 above, I pointed out the disparity between the limited research on vowels in comparison to the much more substantial work on consonants in phonological acquisition. Distinctive feature sets frequently focus on consonantal features, with vowels lacking the same attention. For example, in the SPE feature set, the binary oppositions of high and low are not particularly intuitive for describing vowel height, if the aim is to describe four different jaw positions. Donegan proposes a small set of phonological features for the analysis of vowels (Table 2). These, she claims, "are both binary and gradient" (p. 82). They are binary in that a vowel may either have the feature 'labial' or lack it, and gradient in that a feature may be present to a greater or lesser extent. A sound may have different levels of labiality, for example. This approach would seem to work particularly well for the analysis of Henry's vowels. In Chapter 5 we will see how he draws on articulatory aspects of both the home and local dialects and combines these in constrained, yet very variable ways.

	Palatal		*Non-Palatal*		
	Non-Labial			*Labial*	
	Tense	*Lax*			*Tense*
High	i	ɪ	ɨ	ʊ	u
Mid	e	ɛ	ʌ,ə	ɔ	o
Low	æ	a	ɑ		ɒ

Figure 2.1 Vowel Symbols and Features
In this framework, only three degrees of vowel height are phonologically relevant. Apparent four-height systems involve differences in tenseness (intensity of palatality or labiality) for a given vowel height. Non-palatal, non-labial vowels are lax. Note that the IPA has no symbol for a lax, low, labial vowel.

Table 2: Donegan's feature framework for the analysis of vowels (2013)

2.2.5 Whole-word representations

At the same time, during the 1960s, following on from early models of child phonology which presumed that the child's phonological system was similar to an adult's (Jakobson, 1968), linguists began to question the validity of segmental analysis, and proposed that each child might have their own linguistic system (Waterson, 1971: 179). Later research (see for example, Ferguson, 1986; Ferguson & Farwell, 1975) posited that the word or even phrase is the smallest meaningful unit in a child's phonological system (Sosa & Stoel-Gammon, 2006: 32). This hypothesis is supported by a child's lack of consistency in articulations in their early development (p. 32). Menn & Vihman (2011: 273) also reject the notion of features in early words where a small number of words in a child's lexicon vary by multiple contrasts. At this stage, they exhibit little evidence that they are capable of freely combining segments which could result in minimal pairs.

Liberman and colleagues (1974) claim to provide evidence of a lack of phonemic representations in young children in an experiment which demonstrates that they are unable to segment words into phonemes, or link words beginning with the same phoneme. 135 children (male and female) of nursery school (average age 59 months) kindergarten (average age 70 months) and primary school age (average age 83 months) took part in an experiment designed to test the ability of children to segment words and non-words by phoneme and by syllable. The investigators used an IQ test to perform statistical analyses confirming no significant difference between the IQ of children within their age group or across age groups. Following a training exercise which included modelling of the task, and a further training phase where they received explicit feedback on their responses, the children were required to use a wooden dowel to tap out the number of syllables or phonemes in a word spoken "in a natural manner" by the tester (for example, "popsicle", "cook", "holiday" for the syllable task, and sound out the phonemes in "mat", "cake" and "toys" for the phoneme task. The phoneme task also included non-words such as a single phoneme or combination of two phonemes). The authors report that none of the nursery school children and only 17% of the kindergarten age children could perform the phoneme segmentation task accurately. The older school-aged children were able to perform the phoneme segmentation accurately 70% of the time. On the other hand, 46% of nursery school children, 48% of kindergarteners and 90% of school-age children were able to perform the syllable task accurately. It is worth noting that this task required the children to count the number of syllables or phonemes rather than producing them orally. Counting phonemes seems like a particularly complex task in comparison to sounding them out, but this appears to be the point that the authors are making. In the absence of stable phonemic representations, younger children are unable to reliably perform this task. The authors explain that the identification of phonemes will necessarily be more difficult than identifying syllables, because consonants are "folded, at the acoustic level, into the vowel" (Liberman et al., 1974: 204), therefore there is not a single clear point at which a consonant ends and a vowel begins. Syllables on the other hand, all contain a more readily identifiable peak of acoustic energy in the nucleus. These do not necessarily aid the identification of the precise beginning and end of a syllable but help the child to count the number of syllables in a word (p. 204). A possible criticism of this method is that these tasks are meta-linguistic. The children are being asked to think about language rather than produce it. Whether the task can elicit responses which are directly representative of a child's phonemic representations is not certain. Children do generally improve at this phoneme counting skill with age, but there is evidence that suggests that speakers of some languages find the ability to count phonemes harder than others. Mann (1986) for example, observed that only 10% of Japanese first graders were able to pass a syllable counting test, in comparison to 70% of American children of the same age, a likely consequence of the differences between Japanese and English orthographic systems. There is also evidence to suggest that some children will never develop phonemic representations, through research on adults who are not literate in an alphabetic language (Morais et al., 1986).

Since Liberman et al.'s experiment, there has been further consideration of whether it is the size or the linguistic status of the segment that is important in a child's ability to segment it, with linguistic status being considered more important than size (Treiman & Zukowski, 1996), but the agreement is there that phonemic representations appear to be developed later than units such as words or syllables.

Vihman (2014) explains that the idea that childrens' phonological development begins with whole-word representations was slow to gain support after its appearance in the early 1970s (see for example, Macken, 1979; Menn 1978; Waterson, 1971). Waterson noted the divergence between her son P's phonetic realisations of words and their target, which she determined by their context. Following a detailed analysis, she concluded that the child's early word forms exhibited five different structural patterns (for example, articulations built around repetitions of labials), which accommodated a range of phonetic "differential features" (these included for example, rounding, labiodentality, friction etc.) (Waterson, 1971: 184).

Velleman and Vihman (2002) explain that a child is first exposed to speech input, after which they begin to produce their own vocalisations during the first year, practicing their articulations. They propose an "articulatory filter", through which the child perceives the input around them. After exposure to, and processing of particularly salient words, the child begins to internalise a range of patterns upon which they base their early attempts at words. Reflecting on the examples from Waterson (1971), Vihman (2014) describes a lack of "linear correspondence" between P's articulations and the target words and describes the relationship as "holistic". Segmentally, the targets and the child's realisations have little in common, yet the child produces words featuring a nasal in a stressed syllable with a common structure – NVNV; see Table 3 below.

P’s attempts	**Adult targets**
[ɲẽːɲẽ]	finger
[ɲeːɲeː]	window
[ɲaɲa]	another
[ɲaɲʷø]	Randall

Table 3: P’s articulations (Waterson, 1971: 181)

Patterns such as these are known by a variety of different names, including “articulatory routines”, “templates” (Menn, 1978; 1983) and “gestural routines” (Studdert-Kennedy & Goodell. 1992: 96).

Studdert-Kennedy and Goodell (1992) analysed the speech of a child between the ages of 1;9 and 2;2, called Emma, and note the similarities in phonetic organisation between these words.

Emma’s attempts	**Adult targets**
*[ˈbuːˈdiː]	berry, bird, booster
*[ˈbeːˈdə]	pillow, playdough
[ˈbeːˈdiː]	umbrella
[ˈpeːˈdə]	peanut
[ˈpəˈtə]	puppet
[ˈmeːˈnə]	tomato
[ˈmeːˈniː]	medicine
[ˈmuːˈniː]	money
[ˈweːˈdə]	playdough
[ˈweːˈdiː]	raisin
*[ˈɑˈmiːn]	Elephant, airplane
[ˈɑˈbiːn]	elephant
[ˈɑˈpiːn]	airplane
*[ˈɑˈbuːˈdiː]	Happy Birthday, cranberry, raspberry

Table 4: Emma’s articulations (Studdert-Kennedy & Goodell, 1992: 96)

*homonyms

Table 4 shows how in this range of words, Emma articulates a labial consonant followed by an alveolar consonant. Within this frame, she produces many words. The authors point out that once a child has settled on a gestural routine, they can use it as a route into pronouncing a wide variety of words. The authors describe how when presented with the new word “cranberry”, she produced the articulation [ˈbeːˈbiː]. She then, over several further attempts, refined the pronunciation by adding an additional syllable to get [ˈboːˈbeːˈbiː] before settling on the articulation in the table above, [ɑˈbuːˈdiː] (Studdert-Kennedy & Goodell, 1992: 96). Vihman and Croft (2007) provide examples of templates in children across multiple languages. Words, the authors claim, are a child’s first unit of acquisition, but children make generalisations based on their phonological structure in the form of templates. Children then select new words similar to these existing templates (see also Macken, 1979; Velleman & Vihman, 2002) before selecting words further from the structure of the original templates and adapting them to match (Vihman & Velleman, 2000).

2.2.5.1 The development of phonemic representations

Ferguson and Farwell (1975: 437) propose a four-stage process in phonological acquisition. They position this within the context of “de-emphasizing” the separation of the levels of phonetics and phonology, while maintaining the importance of contrast. The first stage, they claim, is that children acquire a “core” of lexical items and articulations. The second is that they begin to notice phonological patterns, building a model of the phonological organisation of their language; this seems similar to the templates described above. The third step is that they gradually become more aware of the phonological organisation of language in their environment. For the fourth and final stage, the authors propose that phonological development does not end in childhood but continues over the lifetime of an individual, though the level of detail noticed by adults will be necessarily different to those noticed by a child (Ferguson & Farwell, 1975: 438).

Similarly, Nittrouer and colleagues (1989) observe that as the child’s lexicon grows, so does their experience of phonetic patterns, both acoustic and articulatory. With this expanding experience, children are able to begin organising these phonetic features into phonological units (p. 131). While no definitive tipping point has been established, somewhere between the acquisition of 50-100 words has been posited as the point at which children begin to represent sound at a phonemic level (though we saw above that Sosa and Stoel-Gammon (2006) put this much higher at 150-200 words), with the process continuing until a child is perhaps as old as 8 (Fowler, 1991: 53).

Studdert-Kennedy (1987: 53) writes of how phonemes emerge as children "escape" consonant harmony. Once initial articulatory routines (see section 2.2.5 above) are well established, a child's growing vocabulary puts pressure on existing routines, and new ones emerge to accommodate the new words. New articulatory routines may "break up and redistribute" words previously articulated using an old routine. Studdert-Kennedy claims that the logical end to this is that the number of articulatory routines continues to grow until there is one for each phonetic segment. Like Fowler (1991), he places the termination of this process at around 50-100 words.

Waterson (1971: 181) suggests that as the child becomes more experienced in recognising and producing phonological forms, they may be able to detect - and eventually produce - more fine-grained detail. Menn and Vihman (2011) propose that the child's utterances may begin to show a more ordered relationship with adult speech between 30 and 70 words in a diary study, and after 25 spontaneous words in a 30-minute recorded session. At this point it is possible to begin mapping the differences between adult and child forms, looking for evidence of systematicity in the data. "We admit that it appears inhospitable, but this chaotic system, like many, does settle, albeit locally, lumpily, and gradually, into relatively stable and comprehensible systems at all levels" (p. 283).

In section 2.2.4 above, the move away from acoustically based distinctive features to an articulatory model was outlined, however, according to some linguists, features are problematic as they do not correspond directly with motor control, and are therefore unsuitable for children to imitate (Studdert-Kennedy, 1987). In the following sections, the problems of a featural account are raised and I consider some of the alternative proposals.

2.2.6 Articulatory gestures

Browman and Goldstein (1986) propose a phonological model based on articulatory gestures - Articulatory Phonology. Early concepts of phonological units were based on what Goldstein and Fowler (2003) describe as unsuccessful attempts to tie acoustic features to phonological units (see for example, Cooper et al., 1952; Harris, 1953; Jakobson, Fant & Halle, 1963). Most feature systems conflate both acoustic and articulatory properties (Browman & Goldstein 1989: 222), but articulatory phonology makes a connection between articulatory *gestures* (what they refer to as "constrictions of the vocal organs") and combinable phonological units. The authors make a distinction between the product of articulation, in other words, speech sounds, and the articulatory gestures that produce them.

Although there have since been developments (see for example, Byrd, 2000), at the time of its inception, articulatory phonology differed from most other phonological theories in that it does not require a fully developed phonological system in order to exist, as it makes use of the same gestures to analyse a nascent phonological system as a fully developed one, (Studdert-Kennedy & Goldstein, 2003: 241).

Articulations are divided into six types of constrictions of the organs of speech: lips, tongue tip, tongue body, tongue root, velum and larynx (Goldstein & Fowler, 2003: 4). Contrasts are defined where a combination of gestures contrasts with another combination, for example, the distinction between "pack" and "tack", where the only difference between the two articulations is the constriction of the lips in the sound at the beginning of "pack" versus the tongue tip constriction at the beginning of "tack" (p. 4). The authors note that these six organ constrictions do not account for all contrasts in all languages but suggest that they do account for the primary distinctions which occur in all languages: "within-organ contrasts (such as [p-f], or [t-θ] are not universal" (p. 5).

Browman and Goldstein (1992: 39) claim that early words are not divisible into phonemes, but that "segmental sized units" or a kind of "constellation" of gestures emerge over time. Multiple studies have proposed, the authors suggest, that children develop "higher level units" from "smaller units" during their development and that this could be accounted for by articulatory phonology, presumably in the form of phonemes or syllables being made up of constellations of gestures. The dynamic movements of multiple articulators with differing temporal overlaps means that constellations of gestures are unlikely to correspond exactly to traditional phonological units, however. Like phonological features, gestures may function as phonological "primitives", as changing a single gesture can lead to lexical contrast (p.24). There is evidence to support the claim that children are more likely to orient to sub-segmental features than adults. For example, Fowler and colleagues ((1991) cited in Browman & Goldstein, 1992) noticed that children are more likely to swap features rather than phonemes in speech errors than adults are. In the nonsense phrase example "pam dill", children were more likely to produce errors such as "bam till", allowing features to move from one segment to another (voicing switches between the initial phonemes), while adults were more likely to produce “dam pill”, where the whole initial phonemes have switched place, (Fowler et al. (1991) cited in Browman & Goldstein, 1992: 39-40). These examples of what the authors call “feature blends” decreased as children aged. Children of 4-5 years old blended features in 33% of speech errors, reducing to 18% at aged 8 and 8% of speech errors in adults. On the other hand, phoneme or onset exchanges increased with age, shifting from 33% in 4–5-year-olds to 44% in 8-year-olds and 74% in adult speakers (Fowler et al. (1991) cited in Browman & Goldstein (1992: 39-40). Browman and Goldstein claim that this is evidence of the shift from feature or gesture to an intermediate unit below the level of lexical unit as a child ages. One possible criticism of this interpretation is that Fowler’s results could be skewed by their choice of nonsense phrase in this experiment. In the case of “bam till”, switching the initial phonemes to “dam pill”, the version the adults were more likely to produce, leads to real words rather than a nonsense phrase. This could explain why adults were more likely to switch phonemes rather than features. Young children are unlikely to know the word “damn”, whereas adults may find it easy to reinterpret a nonsense phrase as a real one. However, this thesis provides further evidence of a shift from gestures to segments from the analysis of Henry’s speech in Chapter 7.

2.2.7 Criticisms of a featural analysis

Studdert-Kennedy and Goodell (1992) deem the phonological feature to be an inappropriate unit for the analysis of speech. One criticism, they claim, is that it does not exist independently, but in a circular relationship with the segment. A segment is defined by its features and a feature is defined by being a property of the segment it belongs to. This is evident, they claim, by the fact that features, for example “coronal, nasal, strident” are all adjectives rather than nouns – they are properties of a “larger unit”, and they do not have any temporal properties, which makes them inadequate for the description of children’s speech. Moreover, children cannot possibly orient to an abstract feature - “a relational property fulfilling the linguistic function of contrast across a phonological system”, as children don’t yet *have* a fully developed phonological system (p.90). An analytical framework for children’s speech must, they claim, be capable of handling the fact that gestures may move around beyond the segment where they could be found in a target adult utterance; they “slide along the timeline” (p. 97) (see also Browman & Goldstein, 1987). For example, they present data from ‘Emma’, introduced above, who had acquired around 100 words at the time recordings began. This ties in with the age at which Fowler (1991) suggests that phonemic levels have begun to be represented in infants. Emma produces the word "tomato" as ['meːnə]. Here, the velum is lowered for the alveolar consonant as well as for [m], as well as voicing continuing throughout the word. Studdert-Kennedy and Goodell describe this as gestural harmony due to the difficulty of planning and producing different gestures in close proximity to one another (p. 97). As described by Ferguson and Farwell (1975) above, as children attempt to articulate an adult target, aspects of the articulation may appear in various places in the word and are often not tied to a segment.

Waterson (1971) performed a non-segmental analysis of data that Leopold (1939) had struggled to explain. The child produced the articulation [lɔˑɪʃ], which Leopold had interpreted as "story", though the lack of similarity between the target and the child's articulation caused him to doubt this interpretation. Waterson notes that the target word has the features of friction, sibilance, liquid, continuance, open vowel followed by a closer vowel, rounding, backness in the first vowel and frontness in the second. The child's realisation of the target also features friction, sibilance, continuance, open vowel followed by a closer vowel, rounding, backness in the first vowel and frontness in the second. Thus, Waterson was able to conclude that the child's realisation had most of the features of the adult form, but these were organised in a different order (p. 86).

A further issue with a featural analysis of children's nascent language, Studdert-Kennedy and Goodell (1992) point out, is that it does not account for multiple realisations of the same underlying phoneme, while a gestural account can handle this kind of variability. They describe a typical process found in phonological acquisition in many children. Emma realises /r/ as [w] in initial position, (gliding), but as [d] in medial position (stopping). A gestural analysis recognises that the acquisition of speech sounds involves mastering the coordination of articulators, that this is a developmental process, and that they may be more or less difficult to articulate in different contexts (see also Menn 1983).

2.2.8 Do features actually exist?

In section 2.2.6 above, I presented some criticism of the concept of phonological features. Studdert-Kennedy and Goodell, (1992), for example, argue that traditional featural analyses require a fully functional phonological system, which, they claim, a child does not yet have. Taking a similar position to Browman and Goldstein's articulatory phonology, Ladefoged (2005) suggests that features are an artifact – a system imposed by linguists in order to explain the behaviour of language. The number of features required to account for all sounds across all of the world's languages, he argues, would be too large to be feasibly managed by a child. Ladefoged explains that while features can be helpful in explaining the patterns we observe in language, in his view, speakers do not adjust phonological features, they adjust articulatory parameters. "What speakers and listeners do may be better described in terms of articulatory phonology and direct perception" (p. 12).

On the other hand, there are still some who make the case for the existence of phonological features. Donegan (2013: 78) claims that the development of vowels in children suggests similar "features and processes" to those found in adult language. She proposes that phonological features are not abstract categories but simply links between gestures and the resulting sound. Children develop their awareness of phonological features by paying attention to the connections between articulatory gestures such as jaw height or lip rounding and speech sounds through their early articulations. The connections between these gestures and the resulting sounds establish the basis of the child's phonology (Donegan, 2013: 81).

Menn and Vihman (2011) argue that features are of limited assistance in describing children's early words due to the lack of systematicity in those utterances. However, they recognise that they offer a useful explanation of the relationship between targets and outputs later on in the child's development (p. 262). If features "reflect psychological reality", for example, by functioning systematically, the authors claim that they will be a valid analytical tool. The authors claim that psychological reality of features is related to the way they behave. If it "spreads, plays a role in generalisations, or divides sounds into classes that are treated consistently within class but differently across classes" then it can be deemed to be psychologically real (p. 265). They point out that as it is impossible to know what is inside a child's head, that when we say that a child "has a feature" what is actually meant is that the child is functioning as if they have a feature (p. 267). Menn and Vihman's position is that they will use the term "unit" to describe a feature if a child's speech is more usefully analysed in terms of features than segments, for example if sounds containing a feature behave in the same way.

2.2.9 Summary

Above, we have learned that phonological input is complex and infants have the complicated task of internalising multiple elements of their linguistic input in order to build their own phonological system. The question of what a child's underlying phonological system looks like is disputed; while some linguists believe that a child's underlying system is the same as the adult system, there is evidence that that there may be errors in a child's interpretation of that system. Consequently, the child's system is likely to be constantly under revision as they mature and are exposed to further linguistic input which resolves previously incorrectly stored representations.

Children's surface representations vary widely, both between children and within a single child's articulations. Much more variability is found in early speech than is found in adult speech (see section 2.3, below, for further discussion), and early speech bears little overt relationship to an adult system. This is seen, for example, in infants' features sliding "along the timeline" (Studdert-Kennedy & Goodell, 1992: 97). Variability may peak as children start to combine words and begin to form phonemic representations. Consonants are the focus of most phonological acquisition research, while vowels are generally overlooked. Researchers frequently claim that vowel acquisition is generally unproblematic, but others have found them to be rich in variation and argue that their acquisition reveals complex patterns.

We have seen how early phonological acquisition research made use of distinctive feature theory for the analysis of children's speech, and while these resulted in some inconsistent conclusions about the acquisition of features, these inconsistencies may be, at least in part, attributable to methodological problems. More significantly, features seem not to be helpful in analysing the early speech of children, where neither phonemes nor features appear to operate systematically. Over time, researchers have refined distinctive feature theory, and have moved away from the claim that features should be described in acoustic terms. Some researchers (e.g. Browman & Goldstein, 1989; Menn & Vihman, 2011) have rejected acoustically based features altogether in favour of gestures as a way of describing the organisation of early children's speech. Features still have currency, however, with proposals being made that they may be useful if considered as artefacts (Ladefoged, 2005) or that they may be "psychologically real" and useful if functioning systematically (Menn & Vihman, 2011).

2.3 Acquisition of systematic variation in children

As children are acquiring speech, they are of course acquiring the dialect and associated phonological system particular to their environment and any sociolinguistic variation within this variety. We now turn to the acquisition of systematic variation; the child's path to the variation they will continue to exhibit into adulthood.

Above, in section 2.1.1, we learned that children's perception of speech begins with infants noticing more phonetic variation than is necessary for the variety or varieties in the language that surrounds them (van der Feest & Johnson, 2016). Over time, this ability to notice difference reduces, as they pay closer attention to the phonetic organisation of their own language (Maye et al., 2002), and as meanings become attached to sound (Beck, 2016). In this section, we will consider variation in children's speech production.

Research focusing on the range of variation in children's speech represents a tiny fraction of work in the sociolinguistics literature (Foulkes et al., 1999; Nardy et al., 2013). In much of the child language acquisition literature, references to variation in children's input are simplified, fleeting or ignored. This approach often makes generalisations about the speech community at the expense of individual differences and situations.

Claims around the age at which children acquire systematic sociolinguistic variation range from 3 to preadolescence, depending on the author (Khattab, 2002). This range of interpretations may be because, as Foulkes and colleagues (1999: 1628) argue, the acquisition of sociolinguistic variables is difficult to separate from phonological acquisition more broadly. There is no research on the time period between phonological acquisition and the acquisition of sociolinguistic variables, because children acquire local variants at the same time as they acquire their "mother tongue" (Chambers, 1995: 158).

The earliest consideration of sociolinguistic variation in children was demonstrated in a research project in the late 1950s, in which variants of the -ing morpheme in present participles were linked to gender, style, mood, personality, formality and social class (Fischer, 1958: 51). Though this is recognised as being the earliest sociolinguistic study of children and highlights features later confirmed by others, it also receives some criticism for its lack of systematic methods (Nardy et al., 2013: 258).

Writing in 1964, Labov proposed a six-stage model describing the developing sociolinguistic competence of children. He claimed that under-fives acquire the basic grammar and lexis of language from their parents. This is followed by acquisition of the local dialect between the ages of five and 12, as children are exposed to their peer group. After age 12, while continuing to use the local dialect, children begin to attach social values to language use. It is not until after this stage, Labov claimed, that children begin to use more than one speech style. Once they notice the social significance of speech styles, they begin to apply them themselves in different social contexts (Labov, 1964: 91).

Nardy and colleagues (2013: 259) point out that Labov's model assumed that children are monodialectal and monostylistic until late adolescence, first, in the parents' dialect, and later in the dialect of their peers. It is not until late adolescence (16+), that they will be able to use different linguistic styles. They consider it unlikely, however, that this is an accurate representation of children's speech, as this model involves children switching from one dialect to another with no overlap or retention and use of their first dialect. Instead, Nardy and colleagues propose that the home and local dialects may co-exist.

An early example which recognises variation in both the adult and child's speech is Local's (1983) research on the realisations of one vowel in the speech of a single child recorded between the ages of 4;5 and 5;6 in Newcastle upon Tyne. Local analysed the fine phonetic variations of /i/ and the alignment of these realisations with the system of the child's parents. He showed that there was considerable variation in the child's articulations, demonstrating fluctuations in his developing system as he grasped the extent of possible systematic variation for that vowel. There was, he claims, far more variability than can be seen in the system of an adult speaker of the same variety (Local, 1983: 450). Some of the child's realisations occurred consistently in a particular environment, while others were "stylistic". Local gives some examples of these. In one case, the child adopted lip rounding for a stretch of talk – a "paralinguistic labial setting" (p. 450), while in another, his speech featured particularly large pitch movements which accompanied what the mother identified as "whingeing" (p. 451). Local found that the range of variants decreased as the child aged. Moreover, the child's variability began to align more closely with the adults' systems as he got older, in one case adopting centralisation of the vowel to [ɨ] in a phonological environment consistent with his father's idiolect (i.e. Local claims that this feature was not found in the wider speech community) (p. 450). Children have to sort out which aspects of the phonetics they hear are phonologically conditioned, which are socially conditioned, and which are neither. They then need to learn how much phonetic variability they can inject and still have articulated an acceptable variant – what Local terms a "hit" (p. 452).

Local (1983: 452) claims that not much is known about phonetic variability in vowels or their relationship to the phonological level, and that many writers ignore or "smooth out" phonetic detail to enable a better fit to their chosen phonological theories. He suggests that perhaps children's processing of the phonetic data they hear might be being treated somehow differently from the way that adults would process such data, and if we do not record the full extent of their variants we might misrepresent their language acquisition process.

Local (1983) is unconvinced by the widespread assumption evident in the literature (at that time) that the varieties children are exposed to are made up of stable patterns. For instance, he notes that the child's father realised /i/ as [ɨ] sporadically in stressed syllables in polysyllabic words. In spite of irregular patterns such as these, children appear able to work out what is structured variation and what is not, through their exposure to adult input (p. 452).

Labov's (2001: 437) discussion of children's sociolinguistic development has moved away from the idea that children do not notice stylistic differences until adolescence. In a model designed to explain how language changes over time, he sets out five principles of the transmission of linguistic change. The first three of these relates to the role of children in linguistic change, and map onto stages of linguistic awareness in children. The first principle, he claims, is that the model that children start from is usually that of their primary female caregiver. The second is that early examples of variation in child directed speech are based on differences in formality. Children will notice, for example, that formal language is for teaching and punishment, and informal language is associated with play and intimacy. The third principle is that children will then start to notice the social stratification of language. In other words, he now believes that children build on the input of their primary caregiver, rather than replacing it entirely with the dialect of their peers.

This approach does still seem to ignore complexities in children's home environments, certainly in the UK. Labov claims that the female caregiver provides the model, which is presumably based on the idea that the female is the caregiver staying at home during the day. This ignores families where parents may share caring responsibilities or a male caregiver, male single parent families or families where both parents are male. Equally, children frequently interact with multiple adults and may spend many or even most of their waking hours outside of the home. According to the Department of Education's 2021 survey of parents in England, 68%[2] of children aged 0-4 years spent time in some form of childcare – 64% in formal childcare). Therefore, the linguistic input children are typically exposed to is more varied than a single caregiver at home, and this range of different inputs could potentially affect a child's acquisition. There are also different cultural and economic contexts in which a primary caregiver may not be either parent and may be a paid employee (Ho-Cheong Leung, 2012). The issues of competing influences on a child's acquisition will be discussed in more detail below in section 2.3.1.

Foulkes and colleagues (1999) identified evidence of structural variation appearing in the speech of much younger children, aged 2;0- 4;0, who exhibited sociolinguistic patterns seen in the adult community. The data from two girls and two boys from Newcastle (collected as part of a larger study of sociophonetic development) were analysed acoustically to establish their realisations of /t/ across a range of phonological environments: initial position (e.g. "toy"), non-initial inter-sonorant position (e.g. "water", "bottle"), and pre-pausal position (e.g. "cat"). The authors report that the children showed distinctly different phonetic realisations in each of the three environments investigated. They conclude that the children were making good progress towards producing different allophones in each environment. They also note that the acoustic patterns were very similar to those seen in adults from the area. Finally, they observe that the children were acquiring pre-aspiration in pre-pausal position, a feature seen in their mothers, representative of an innovation seen in young

---

[2] Down from 76% in 2019 (Dept. of Education, 2019), possibly as a result of the effects of Covid-19 on working away from the home. Figures going back five years suggest that the 2019 percentage was stable (Dept. of Education, 2021).

women in the community. This, the authors suggest, supports claims, such as Labov's, that primary caregivers are the strongest influence on children's phonological development before peer influence begins (p. 1628). In this section we have seen that sociolinguistic and stylistic variation begins at a young age, so we should expect to see evidence of this in Henry's speech during the data collection period. The range of variation will be of particular interest due to the multiple models of variation in his environment. Local (1983) reported on the wide-ranging variation seen in naturalistic data in a child whose parents' dialects were consistent with the local area, so we may expect variability in Henry's vowels to be even more substantial.

2.3.1 The influence of parents and peers

Parents provide a model for children to adopt, but it is not predictable which features of variation the child will acquire, according to Hazen (2002). He claims that children will neither copy identically the patterns of the parents, nor will they vary from it hugely. He outlines the following possible outcomes for parental/peer influence on a child's linguistic development:

> 1. Children adopt their parents' variety
> 2. Children adopt a community variety
> 3. Children adopt a variety that is somewhere in between, perhaps becoming more like one or the other as they age. This could mean that the distribution of a child's sociolinguistic variables is somewhere between parental and community norms, or it could mean that children use features and/or styles from both varieties.
> 4. Different children in the family might behave in different ways.
>
> (Hazen 2002: 505).

In most cases, children and their parents belong to the same speech community, so it can be difficult tease out which features come from peers and which come from parents. The effects of parent or peer influence may be "masked" (Hazen, 2002: 504). However, when the parents belong to a different speech community, it is easier to trace which features are coming from where, (Chambers, 1995: 159).

In their seminal sociolinguistic work, Weinreich and colleagues (1968) consider the influence of parents and peers on a child's language development. They suggest that in only two situations are the parents the prime influence on the development of a child's dialect: where the child is isolated from other children, and in the transfer of prestige features. In other words, a child's peers are normally deeply influential in the acquisition of their mature dialect. An imperative to acquire dialect features from peers certainly bears out in subsequent research, though the relative influence of parents and peers depending on the context is heavily nuanced. For example, as we will see below, if geographical relocation is involved, the age at which this occurs, relative complexity of the local dialect and the influence of other dialects or languages spoken in the home are all influencing factors in whether a child will acquire a dialect from their peers.

That the early influence of parents on a child's speech is superseded by their peers is a widely supported view in the sociolinguistic literature. Above, in section 2.3, I wrote that Labov initially reported the transition as falling into two distinct phases – before and after the age of 5 (Labov, 1964). Since then, others (see for example, Chambers, 2009; Hazen, 2002) have examined the speech of children to explore the ways in which a shift from parental influence to peer influence manifests itself. A summary of this research is detailed below. There is evidence, however, both anecdotally and in the literature, that not all children make this shift, and that cultural differences (e.g. Kazazis, 1970; Stanford, 2008) or the prestige attached to dialect features (e.g. Surek-Clark, 2000) may also impact on a child's acquisition. Hazen claims that:

> If the family unit has an influence on language variation independent from other social factors e.g. gender or age, then we would expect the children in these families to align, in terms of dialect features, with their parents to some extent and not necessarily with their social categories or in the larger speech community. If the family has any influence on the children, the children would demonstrate language variation that would be unexplainable through any influence other than the family unit.
> (Hazen, 2002: 502).

This evidence of parental influence is reported below in the research of Surek-Clark (2000), who discusses the role of prestige in her research on the Curitiba dialect of Brazilian Portuguese. Here, she looks at the interaction between parental influence, and how prestigious a variable is. She claims that where at least one parent spoke a more prestigious standard dialect from out of the area, that the child would acquire (at least to some extent) these more standard features. She suggests that the home influence to use these more standard features is more powerful than the influence of the child's peers.

Specifically challenging Weinreich et al.'s assertion that parents can only affect children's acquisition of dialect in the two situations they specify, that is, in the case of isolated children and in the transfer of prestige features, Kazazis (1970) offers anecdotal evidence of a second-generation brother and sister in Athens whose parents were originally from Istanbul. He claims that in this case, the children acquired and retained Istanbul-Greek grammatical features in spite of teasing from their peers. These children were not isolated, and neither were the features retained by these children prestigious. He acknowledges the anecdotal status of his evidence but insists that his example is likely to be representative of others.

Mæhlum's (1992) research presents a further example of when parents might be the primary influence on a child's dialect. She focuses on the residents of Longyearbyen, a set of arctic islands between mainland Norway and the North Pole. This is a community without a stable dialect, due to regular turnover of residents. Families stay on average for around 10 years and spend their summers on the mainland. She found that children's varieties showed a greater affiliation to their parents in comparison to what would be expected elsewhere in Norway, because of the lack of a homogeneous variety outside of the home. Mæhlum claims that the only stable social unit in this community is that of the family.

The research described above provides evidence of exceptions to the dominant view that children are mostly influenced by their peers. More specifically, these are examples of where there is some evidence that children are subconsciously selecting the parental variety as desirable. The parental variety, or particular features of it, are selected for reasons of prestige. In the studies that follow, researchers note that many features in a child's speech correspond to the dialect of their peers, but parental influence remains in a range of different ways.

Hazen (2002: 500) claims that it is well established that adolescents do not have identical patterns of linguistic variation to their parents. There may be some overlapping variants, but children typically establish their own set of rules of variation based on wider influences than their parents alone. The family, he suggests, could be seen as some kind of other relevant grouping between that of the individual, and the wider speech community. The family is not typically included in models of the acquisition of language variation. However, Hazen argues that it has an embedded place within the speech community.

2.3.1.1 Incomplete acquisition

Payne (1980) considers what happens to the accents of children belonging to families who move to a new dialect area. Her analysis sheds light on not only the phonological acquisition of children who move to a new area, but also the development of dialect features of children born in the area to parents who speak a different variety. She performed a systematic analysis of the acquisition of a range of phonological variables by the children of 12 families who had moved to King of Prussia, Philadelphia, from another dialect area. Her aim was to investigate the ability of children to reorganise their grammars upon encountering new linguistic rules. She accounts for a range of phonetic variables, and a more complex rule, the short-*a* pattern, where the (æ) variable undergoes tensing and raising in certain environments. In addition to these phonological environments, some realisations in this set are lexically driven and there are also some lexical exceptions to this rule. The children were successful in acquiring the phonetic variables to varying extents. The simplest variables, for example, phonetic differences from their first dialect, were fully acquired by most of the younger children, but children who were 10-14 years old when they arrived in King of Prussia did not acquire most of the variables.

In her analysis of the much more complicated short-*a* pattern, Payne (1980: 175) found that its acquisition was "irregular, sporadic and incomplete". This variable is complicated in that short-*a* is split into tense and lax realisations, depending on a range of complex phonological and grammatical environments in complementary distribution, what Wells (1982) would characterise as a context-sensitive realisational difference between accents. A small set are also lexically conditioned; for example, *mad*, *bad* and *glad* are tensed and raised, but *sad* is not (Roberts, 1997: 250). Unlike the effect of age on the acquisition of the simpler phonological variables, there was no correlation between the age a child moved to the area and whether they learned the short-*a* pattern. "Unless a child's parents are locally born and raised, the possibility of his acquiring the short-*a* pattern is extremely slight, even if he were to be born and raised in King of Prussia" (Payne, 1980: 174). Her conclusion is that children are very susceptible to the influence of their peers and can reorganise their grammars to the extent that they can assimilate simple rules, but the acquisition of more complex rules may be confounded by exposure to non-local forms found in the home.

Like Payne, Trudgill (1986) considered the accents of people with non-local parents, but in a UK context. He claims that while these speakers, who were born and raised in Norwich, seemed to have local accents, they did not distinguish between /ʊu/ and /ou/. These are two, separate, lexically differentiated phonemes in Norwich, the first belonging to words such as *moan, rose, nose, sole*, and the second to *mown, knows, rows, soul*. This additional phoneme leads to a systemic difference between the accent of Norwich and other dialects (Wells, 1982). As it is found in a particular subset of words, these would need to be learned individually. While we cannot make a direct comparison with the short-*a* pattern of Payne's research, which was based on complex phonological environments and lexical conditioning, learning words by exception is plausibly more complex than a straightforward realisational difference between accents, where one sound is routinely replaced with another.

Deser (1989) analysed data collected as part of Shuy et al.'s 1966 Detroit dialect study. Participants were interviewed in family groups of one parent and two children (some of whom were adults) for around one hour. At that time, Deser points out, a dialect contact situation was ongoing as workers from the southern central United States had moved to Detroit for work, bringing their dialect with them. She analysed the data from six black families with a view to determining the dialect features present in the children and whether their influence was their parents or the local community. All children were born and raised in Detroit. Three speech and language therapists classified the parents and children as either northern in their dialect features, southern, or having mixed dialects. Deser then performed an acoustic analysis of two variables, /ai/ and /ae/[3] in the parents and children.

Variable	**Northern**	**Southern**
/ai/	[ai] (diphthongised)	[aː] (monophthongised)
/ae/	[E] (raised)	[ae] (unraised)

Table 5: Variables in Detroit (Deser, 1989)

[3] Bracketing and symbols as in original text.

In the northern families, the younger children (10-12 years old) broadly used their parents' variants and the older children (13-20 years old) used the community norm. In the southern families, the pattern was reversed. The older children from the southern families generally used the same variant as their parents, and the younger children used the variant found in the community (see Table 5). Deser claims that the older children who share their parents' dialect features are less rebellious than the other children and are trying to maintain their southern identities. This research demonstrates the complexity of individual differences, and how children in the same family may do different things. She argues, "an individual belongs simultaneously to a dialect community, a peer group, and a family structure and as such there must be a tension on that individual as s/he develops their identity vis-a-vis these various groups" (Deser, 1989: 120).

The influence of parents' varieties on a child's phonological acquisition is supported by Roberts' (1997) small-scale study of children's acquisition of a changing vowel system in Philadelphia. Roberts claims that having at least one non-local parent affected children's acquisition of more complex vowel systems. Of six children analysed (aged 3;4 to 4;10), one had a local father, but a mother from out of the area (Gia), and another had two Italian parents (Mike). The main language spoken in Mike's home was Italian. The other four children had "native" Philadelphian parents, though three of them lived with only one parent. Roberts looked at the acquisition of three vowel variables.

- Fronting and raising of (aw) in words such as "crown" and "south"
- Raising of (ey)[4] in words such as "cake" and "rate"
- Backing of (ay) before voiceless obstruents in words such as "fight" and "kite".

Roberts (1997: 251).

Roberts had previously examined the acquisition of the complex *short-a* pattern in the same children (Roberts and Labov, 1995). All six of the children in Roberts' study acquired at least one of the four local vowel variables investigated: (aw), a realisational difference from other accents. Gia acquired two of the three other vowels, including the complex *short-a* pattern (described in some detail above, in relation to Payne's King of Prussia research) to an extent, but Mike only acquired (ey) partially. He did not acquire the other two vowels, including *short-a.* Roberts notes, however, that although he did

[4] Bracketing and symbols as original: U.S. style notation.

not tense all tokens which could have been tensed, neither did he tense any tokens which should not have been tensed (1997: 260). So although Mike was not able to acquire *short-a* to the same extent as the other children, neither did he make any hypercorrections. Roberts surmises that Gia was more successful than Mike at acquiring the variables because her father was from Philadelphia, though her mother was not, while Mike's parents were both non-Philadelphians. Although she was more successful than Mike, she was not as successful as the children who had no non-local influence in the home. However, further nuance is evident in Roberts' analysis. None of the children in Roberts' research acquired the (ay) backing feature, including those with two local parents (1997: 260). Labov identified the centralisation of (ay) as a strong, male-led emergent new accent feature (1989). Roberts discusses the view put forward by Labov (1989), that caregivers are most typically female, and that this therefore makes it more likely that female-led sound changes will be acquired by children. She established that this feature was not present in any of the mothers' speech but was not able to obtain data from any of the fathers apart from Gia's, who she confirmed did have this feature. Roberts explains that Gia's mother worked, and that she has extended female family members, and spent time in childcare alongside workers from the local area. So although Gia's mother was not a Philadelphian local, she explains that as Gia was exposed to local Philadelphian female speech, the lack of acquisition of (ay) backing may mean that she learned more readily from female caregivers (Roberts, 1997: 263).

Back in the UK, Hewlett and colleagues (1999) investigated the presence of the Voicing Effect (VE) and Scottish Vowel Length Rule (SVLR) in seven children aged between 6 and 9 years from Edinburgh. The children were all observed to be acquiring the local accent. Of the seven children analysed, two had two Scottish English-speaking parents, two had one Scottish English-speaking parent and one parent who spoke a non-Scottish variety, while the other three had two parents who both spoke varieties other than Scottish English. A strong VE is found in many accents of English; vowels are longer before voiced consonants, for example, the vowel in "niece" is typically shorter than the vowel in "knees". The SVLR affects vowel length in Scottish English in certain phonological environments. Vowels are longer in open syllables, before /r/, before morpheme boundaries and before voiced fricatives (vowels in this position are also affected by the VE). Hewlett and colleagues (1999: 2157) claim that the relationship between the SVLR and the VE is contentious due to these potentially overlapping environments. The VE was found to have a minimal effect on vowel length before plosives in children with one or more Scottish English-speaking parent. In these speakers, vowel lengthening was much more strongly influenced by the SVLR. Children with only one Scottish English-speaking parent exhibited the same patterns in use of the VE and SVLR as those children with two Scottish English-speaking parents, while those with parents who spoke a non-Scottish variety showed evidence of the VE but little or no evidence of the SVLR. This is in direct contrast to Weinreich et al.'s observation that there were "regularly" no differences between the dialects of children of "first generation" parents to children of parents who had lived in their area for generations. They claim that this is the case even when the dialects of these "first generation" parents are substantially different from the local dialect (Weinreich et al., 1968: 145). These claims are also contradictory to Roberts' observations that a child with one parent from out of the local area is enough to interfere with the successful acquisition of a complex accent feature. Here, one non-local parent appears not to confound the successful acquisition of the complex SVLR rule, while in Roberts' data, it prevented Gia, the child with a local father and non-local mother, from fully acquiring the local accent. Admittedly, Gia did successfully acquire the complex *short-a* pattern, but did not fully acquire all of the features analysed in the study. Specifically, she did not learn the Philadelphian (ay) centralisation rule (Roberts, 1997: 264).

2.3.1.2 Compromise or variability in the sound system

Above, I discussed situations where children do not completely acquire the phonological system of a local dialect. Here, I describe circumstances where a child's system exhibits aspects of both the home and local varieties within a single segment.

The impact on children's accent development when their out-of-town parents moved to a "new town" in the south east of England, was investigated by Kerswill (1994). The precise phonetic realisations of the (ou) variable (related to the GOAT lexical set) were recorded in four children from different families under 4 years of age and compared to the realisations of their parents. The children's realisations were classified as being influenced by peers, similar to the mother or father, and in one case, a compromise between the parents. In that family, the father's typical realisation was [ɔ̝o], the mother's [æ̠ʉ̝], and the child's, [ɵ̠ʊ] Kerswill (1996:188). The compromises in vowel quality were evident in both parts of the diphthong.

Further evidence of impact on the phonological system of children of non-local parents comes from Scobbie (2006). Scobbie analysed Voice Onset Time (VOT) in 12 speakers from the Shetland Isles, Scotland, who had parents from Shetland, wider Scotland, or England. All had been born and raised in the same area of Shetland: Westside. Due to incomers connected with the oil industry, VOT in Shetland is mixed. In Scottish English, and most other varieties of (British) 'standard' English, /p/ has a long lag VOT (aspiration), while in vernacular Shetlandic, /p/ is articulated with a short lag VOT (Scobbie, 2006: 375). As expected, speakers whose parents were native Shetlanders exhibited the most vernacular VOT contrasts, but speakers whose parents were from wider Scotland had a long lag VOT. Those whose parents were English did not cluster together, but instead were found overlapping with both other groups and in between. The author suggests that this is evidence of speakers drawing on their exposure to varieties at home and in the local community but also demonstrating some arbitrary differences (p. 386).

Thomas and Scobbie (2015) add further detail to the picture of children's accent acquisition in an environment where they are exposed to multiple accents in the home. Two case studies are the focus of their research, one of a single child aged 3;1 in Glasgow, Scotland, with one Scottish parent (the father) and one English parent (the mother), and the other, of two pairs of siblings in Edinburgh, Scotland who were also exposed to mixed accents at home. The first case study focused on the FACE and GOAT vowels, as these lexical sets are the site of differences between the accent spoken by the boy's father (Scottish-accented Standard English, or SSE), where these vowels are monophthongs, and his mother's accent which is described as being close to Southern Standard British English (SSBE[5]), where they are realised as diphthongs. In the father's accent, FACE was realised as [e] while in the mother's accent it was realised as [eɪ]. The GOAT vowel, on the other hand, was realised as [o̞] in his father's speech and [ɤʊ] in his mother's variety. The child's realisations of these lexical sets showed the influence of both parents, though each lexical set exhibited a different influence, with realisations of FACE showing a stronger influence of SSBE (83% of realisations), and realisations of GOAT being more variable, but showing slightly more influence from the child's Scottish father (60%). In the second case study, data from two sibling groups was analysed,

[5] This variety is also sometimes called Contemporary Received Pronunciation (RP), BBC English, and more recently been called by a variety of different names: Non-Regional Pronunciation (NRP) (Collins & Mees, 2013) General British (Cruttenden, 2014) and Southern British English (SBE) (Wells & Colson, 1971).

each with one (broadly) SSE speaking parent and one (broadly) SSBE speaking parent (Thomas and Scobbie, 2015). As this data came from a pre-existing corpus (ULTRAX project), the authors examined realisations of FACE and rhoticity for signs of mixed accent influence. Family A contained an older male sibling of 8;7 and a younger female sibling of 6;8. Family B contained an older female sibling of 12;8 and a younger male sibling of 10;7. In family A, the older (male) child exhibited less influence from their Scottish parent (and presumably the community) than their younger (female) sibling and in family B, the younger (male) child exhibited less Scottish influence than their older (female) sibling. The number of tokens across both families was very small, but in each case, one sibling produced the SSE features 100% of the time, while the other sibling showed more influence from the SSBE speaking parent. Although based on case study research and small samples, these results highlight the stark variability that can be found within sibling groups in a way that no other research has shown to date. What is it that leads to one child adopting a feature completely while the other sibling does not? In one family it is the older sibling who shows no influence of the out of area parent, and in the other it is the younger sibling. In each case, it is the female child who exhibits the local features completely, but the authors acknowledge that the sample size is too small to draw any conclusions from that – they also point out that in the larger corpus that the data is drawn from, and in their casual observations, there are female children who do show influence of an SSBE speaking parent. They point to the potential for larger scale more detailed research which could investigate the variables affecting acquisition of local accent features, such as sex, birth order, age, personality and educational/childcare experiences.

The research discussed above reveals differences in the claims researchers make about the influence of parents and peers on the long-term acquisition of local accent features. Payne (1980), Trudgill (1986) and Roberts (1997) found evidence of incomplete acquisition of a sound system, and in Trudgill's and Roberts' research, we started to see evidence of the impact of multiple dialects in the home. Undoubtedly, children are able to acquire many local accent features unproblematically. It is perhaps not surprising that phonological complexity plays a role in determining the successful acquisition of features, though it would be difficult to rank these features meaningfully by level of complexity. Understanding the motivations for acquiring features is difficult to compare across different studies. Whether Gia's inability to acquire the (ay) variable can be attributed to her one non-local parent alone, and whether that would be replicated in other children with similar familial circumstances is unknown in that particular context. Hewlett and colleagues' investigations made the opposite claim – that one parent with a local accent is enough to ensure that the SVLR was acquired. These are of course, two different kinds of feature. One, a centralisation rule, the other a vowel length rule, but both are complex, and Gia did acquire the notoriously complicated *short-a* pattern in spite of the mixed dialect input at home. We then saw evidence of a different kind of effect; Kerswill (1994) reported on a phonetic compromise between the vowels in a mother and father's distinct dialects, while Thomas and Scobbie (2015) uncovered something similar in a Glaswegian child with mixed accents in their home. However, in their second case study, they saw variability in production rather than compromise, and preferred to attribute this to an unstable system rather than some kind of new or intermediate system. There are therefore multiple possibilities for Henry's vowel realisations. He may acquire the local accent, the home accent, or some kind of compromise between the two, and these outcomes may vary according to each lexical set depending on the complexity of the relationship between the phonological systems of the home and local varieties.

2.3.1.3 Social integration

Kerswill claims that peer groups become more influential during the pre-adolescent stage, between 6 and 12 years old. By the age of 6 or 7, children have typically acquired all their phonological rules (Kerswill, 1996: 192). His data demonstrated that the younger children in his 1994 study shared similarities with their parents' (ou) variable (GOAT lexical set), though, as in Local (1983) there was a great deal of phonetic variability in their realisations. By the pre-adolescent stage, Kerswill notes that children begin to move to new peer-oriented networks, when they change their language in "slight but systematic ways, accommodating to their peers and older children" (1996: 196). Commenting on the same data, Kerswill and Williams (2000) cite an example of one child of Scottish parents. Between recordings made 18 months apart., the child made the transition from sounding Scottish at age 4 to sounding local to the Milton Keynes area by age 6. That the child exhibited features of his parents' accents until at least age 4 chimes with the authors' claim that at this stage the child had not yet moved to a peer-oriented network. However, at age 6 the change to the local variety could be considered evidence of this shift to peer influence. The authors suggest that a child's *integration* into their social network influences their likelihood of adopting local features from their peers (Kerswill and Williams 2000: 94) and that slightly older children might be models for variants (Kerswill & Williams 2000: 107).

While Payne suggests that the major social variable in the acquisition of the Philadelphia dialect was age of arrival, Labov (2001: 430), re-analysed Payne's data and reached a different conclusion. Rather than finding age on arrival to be the key factor, Labov carried out a multiple regression analysis which, he claims, demonstrates that the most significant factor is the density of a child's social network; specifically, their connections to their peers. This was established in his analysis by how many times a child was mentioned by their peers in *their* interviews. Labov's measure then, appears to be a child's popularity, or social *belonging* within a group of children. It does seem likely, however, that social acceptance may be affected by how local a child sounds, in which case children who do not sound like their peers may be less likely to achieve social acceptance than those who do. It may be difficult to tease out whether children who have most successfully acquired the local accent have been accepted socially because they sound local, or that they sound local because they are a member of a close-knit group.

Kinzler and colleagues (2009) found evidence that children favour friendships with children with familiar accents. In a series of experiments, they presented a group of 32 children from the US (almost all white) with side-by-side images of a white child and a black child and asked the children which of the two children they would prefer to be friends with. 78% of the children selected the white face. The experiments then introduced matching the children's images with different languages (French and English), a different language (French) vs. French-accented English, and foreign vs. 'native' accented voices. In all cases the researchers found that children expressed a significant preference for friendship with the more familiar sounding accent, regardless of the race of the child. Although the experiment did not include choosing between a local US accent and an accent from a less familiar area of the US, the trend exhibited in the children's choices suggests that most children will select friends who sound most familiar. In a final experiment designed to test whether it is the familiarity of the facial features rather than skin colour which affected children's preferences, they manipulated the proportions of the white face from a 1:1 ratio to 2:3 to produce "novel" facial features. They found that when the faces were presented without voices the children chose the familiar features of the black child's face as the preferred friend, but once the distorted white face was presented with a familiar accent, 81% of the children chose that image as their preferred friend over a black child's face paired with French-accented English voice (Kinzler et al., 2009). Kinzler et al.'s research could therefore support Payne's original analysis. Labov argued that the children who had acquired the local accent features in King of Prussia were those who were most integrated into friendship groups. However, this could be because successful integration would be more likely in children who had acquired local accent features. Those who had retained their parents' features, on the other hand, would be less attractive as friends to local children and therefore be less likely to be mentioned by their peers in Labov's multiple regression analysis. Labov and Kerswill's arguments do appear plausible though — it does seem likely that if you are well integrated into a friendship group that you will be more likely to acquire the same accent features as your friends, but the relationship seems a circular one. Perhaps it is those children who attach importance to integration and therefore are open to picking up the local features more quickly that are those who are accepted, and more introverted children, less attuned to the significance of linguistic assimilation who will be less likely to develop the local features.

Above, I describe the age at which children may begin to be affected by peer influence and the role of accent in social integration. We now turn to an example of another cultural setting where different expectations and motivations for dialect choices exist. Stanford (2008) describes the linguistic behaviour of the children of exogamous Sui marriages in China. He collected data from three Sui clans in rice farming villages in Guizhou province in Southwest China. Here, women typically marry men from a different clan, and settle in their husband's village. Linguistic features frequently differ between these mutually intelligible dialects (Stanford called these "clanlects") thus the women often speak a clanlect distinct from their husband, and, Stanford claims, their adult children (2008: 569). Clanlect variations include tone, diphthongs and lexical items, but all clanlects share the same social status. Stanford notes that the women maintain their original clanlect permanently after their marriage, yet the adult children, according to local informants, always speak the clanlect of their father (p. 570). These claims formed the basis of Stanford's research question — how does this happen if women are the primary caregivers in these communities? Stanford took an ethnographic approach, by interviewing the adults, and he also analysed features of the children's dialects. Informants reported social pressure on the children to speak the father's variety — the patrilect — claiming that while in some cases children might use the mother's variety — the matrilect — when young, the children would be laughed at if they used lexical items associated with the matrilect once they were older. Another informant stated that even the youngest children would be admonished by members of the community for using matrilectal words (p. 571). The dominant forms in the community are the patrilectal forms; one informant explained that her children only heard her variety being spoken at home; all other speakers that the children came across would speak their father's variety, including other children (p. 572). Stanford's informants suggested that children have usually fully acquired the patrilect by 5-7 years old, though there were some differences between informants regarding whether very young children ever spoke the mother's variety (p. 573). In Stanford's interviews with the children, some reported that they could remember making the switch from matrilect to patrilect, and the transition was a conscious one, as they feared being laughed at. Like the adults, the children confirmed the age of transition as 5-7 years old (p. 573).

Above, we have seen that children's dialect acquisition may be more or less affected by their parents and peers depending on context. Typically, peer influence is strongest, but parental influences tend to confound the acquisition of particularly complex aspects of the peer/community/local variety. In Kinzler's research, we saw that children prefer to have friends with a familiar accent, which could present a strong motivator for children to acquire the variety of their peers in order to be accepted. On the other hand, in the community investigated by Stanford, the motivations appear to be driven by a more overt necessity to conform — children are laughed at if they don't use the patrilect. However, in both cases, the outcomes are similar; if a child does not acquire the appropriate dialect, they can face social exclusion. Henry's social integration and character may therefore play a role in his acquisition of the local variety, as he navigates the social pressures of school and friendship groups. It appears that there may be factors other than phonological complexity which can affect the completeness of the acquisition of the local dialect.

Stanford (2008) makes the case that patterns in which children align with their peers should be seen as culturally specific, and part of a more general pattern that is not specifically related to peer influence, but rather, is a manifestation of children's ability or desire to construct a linguistic identity from an early age. Within different cultures, this manifests itself in different ways. Rather than focusing on the split between parents and peers, he suggested that a focus on there being multiple competing groups that a child can acquire from would be a better way of accounting for dialect acquisition differences around the world. For example, this could be group one, group two, group three etc.. The populations of these groups would vary across different cultural settings (Stanford, 2008: 568).

2.3.1.4 How might children acquire structured variation?

Above, we saw how the phonetic input that a child is exposed to is extremely complex, even in a mono-dialectal environment, through Local's (1983) investigation of a child's acquisition of a single vowel sound. Local surmised that children have a range of phonetic variations in their input which they need to sort through and figure out what is ordered and what is not.

Local does not specifically comment on child directed speech; indeed, one can infer from his research that the child is doing all of the work in (subconsciously) working out the relevant phonological patterns in his environment. Smith and colleagues (2007) however, explicitly considered the role of the adult in passing their sociolinguistic knowledge on to children in their care. They explain that children acquire their understanding of how to use variables in language from their caregivers, though their acquisition of these variables is likely dependent on a range of factors such as the complexity of the feature (see also Kerswill, 1996). Smith and colleagues (2007) performed an analysis of two variables. A lexically conditioned variable, *hoose* (a Scots word meaning 'house') and a morphosyntactic variable, -s in 3rd person NP plural contexts were examined in in the speech of adult caregivers and children in Buckie in the north east of Scotland. The *hoose* variable is a subset of the MOUTH lexical set (Wells, 1982) which may be realised as [ʌʉ] or [u:] depending on the lexical item. The morphosyntactic variable requires –s not only for third person singular contexts (e.g. run*s*) but "when the subject is a noun, adjective, interrogative or relative pronoun, or when the verb and subject are separated by a clause, the verb takes the termination -s in all persons" (Murray, 1873: 211, quoted in Smith et al., 2007: 80). This manifests itself in Buckie in utterances such as:

"*Does* teachers have the video camera on?" (caregiver)
"Your feeties *is* cold as well" (caregiver)
(Smith et al., 2007: 80)

For the *hoose* variable, Smith and colleagues (2007) found that caregivers used fewer non-standard variants in child directed speech (CDS). Children acquired the standard variable before the non-standard variant, but after the non-standard (local) variant was acquired, they quickly learned the stylistic constraints governing its use. However, for the morphosyntactic variable, caregivers used a similar proportion of local variants in CDS to the proportion found in interactions in the adult speech community (p. 88). In this case, the children acquired both the standard and non-standard forms at the same time; they acquired complex grammatical constraints, but did not learn the stylistic constraints. That is, they did not learn to use the standard variant in more formal situations. The authors concluded that these variables are acquired in different ways due to their status as sociolinguistic markers or indicators[6] (Labov, 1973). The *hoose* variable is a sociolinguistic marker – the caregivers are aware that it is a socially salient variable and *teach* the children the stylistic rules of its use. The morphosyntactic variable is a sociolinguistic indicator – the caregivers are unaware of its social salience and therefore cannot pass this information on to the children through CDS (Smith et al, 2007: 91).

Foulkes and colleagues (2005) and Smith and colleagues (2007) demonstrated how adults might be modelling sociolinguistic variation through their interactions with children. Foulkes and colleagues (2005) found that variation in CDS directed at children between 2;0 and 4;0 differed from patterns of variation found in speech between adults. Specifically, mothers of girls used fewer local variants of /t/ than mothers of boys, and more broadly, corresponding with Smith et al.'s (2007) observations, local variants were less frequent in CDS than in inter-adult speech. Foulkes and colleagues prefer the terms 'local' (variants used in the local community) and 'supra-local' (variants in wider social and geographical use), (after Watt & Milroy, 1999) over 'non-standard' and 'standard'. The authors suggest that the mothers' lower use of local variants when speaking to girls is evidence of them modelling variants in line with their children's nascent gender identity (Foulkes et al., 2005: 198). A correlation was found between the children's production and the patterns of variation in their mothers' CDS in some variables, but this was strongest in children of 3;0 and above. The authors conclude that CDS influences children's acquisition of phonological variation, in particular, their understanding of sociolinguistic variants and their social significance (p. 200).

The account of the literature above illustrates a range of variation we can expect to see in children's language. This includes variation *per se*, how much of the variation that appears in a child's environment they might pay attention to or ignore, how children

[6] An explanation for this term is provided in section 6.4

might acquire stylistic awareness through CDS, and how they might orient to the dialects of parents or peers in different situations. We now turn to evidence of children's ability to accommodate to more than one variety.

2.3.2 Evidence of accommodation in children

Kobayashi (1981) analysed the speech of a child living in Osaka, Japan, whose parents were from Tokyo and spoke a standard Japanese dialect. The child, referred to as "C" was recorded at two points during her development, at 2;8, and again at 8;0 years old when Kobayashi expected her dialect acquisition to be complete. The two dialects vary in what Kobayashi describes as basic accentual patterns, and accentual rules. The first relates to Japanese pitch accent, where a syllable carries a different pitch associated with a lexical item, and the second, how the pitch accent interacts with the grammar, causing changes in the pitch accent. At 2;8, Kobayashi determined that the child's accent patterns were closely related to her parents' standard Tokyo dialect, due to her having limited interactions outside of the home, aligning with Labov's claim that a child's accent is at first influenced by their primary caregiver (2001). However, at age 8;0, C shows evidence of having acquired the local Osaka dialect (cf. Kerswill & Williams, 2000). Kobayashi recorded C in interaction with her mother and talking to a friend who spoke the local Osaka dialect. In conversation with her mother, features of the standard dialect dominated (seven non-standard to 34 standard variants), while in conversation with her friend, 30 non-standard variants were produced in comparison to only four standard variants. Kobayashi (1981: 19) claims that the results of her analysis reveal that C had restructured her system as a result of contact with the local dialect and was able to use different speech styles according to her interlocutor; she had developed a parent-code and a peer-code for use in specific social situations.

Like Kobayashi, Dyer (2007) also provides a case-study perspective on accommodation in a single child, but this time, her research focuses on the effect of multiple dialects being spoken in the home/family setting. She writes about her bilingual son's acquisition of two dialects, South-eastern British English and North American English (NAmE). The child (referred to as "J") also spoke Castilian Spanish at home with his father. Dyer was born in Britain and has a British accent, but was raising her son in Michigan, USA. J was exposed to NAmE outside the home. Dyer focuses on the acquisition of three phonological variables, the sites of particularly salient differences between NAmE and South Eastern British English. These are rhoticism, T-voicing and vowel backing. In the case of rhoticism, NAmE typically features /r/ in non-prevocalic contexts (a rhotic variety), while Dyer's British variety does not (a non-rhotic variety). In NAmE, /t/ is realised as a voiced alveolar tap between vowels, while in Dyer's British English accent it is realised as [t]. Finally, in NAmE, the vowel in the BATH lexical set is realised as a front [æ] and in Dyer's variety it is realised as the back vowel, [ɑː]. Dyer calls this "vowel backing" (Dyer, 2007: 3073). She explains that unless you are a first language learner, acquisition of the vowel in the BATH lexical set is typically difficult as it is lexically determined. She classifies her son's acquisition of NAmE as a second dialect, and British English as his first, therefore the acquisition of this vowel in the British variety was not problematic for him. She explains that up until the age of 3 there was no evidence of NAmE features in his speech, but at age 3½ he began to acquire the NAmE variants. As the vowel /ɑ/ is often accompanied by a following /r/ in NAmE, for example in words such as "cart", once J began to showing signs of NAmE features between the ages of 3;6 to 4;0, some hyper-corrections relating to the BATH lexical set started to appear. J pronounced the words "laugh" and "can't" as [lɑɻf] and [kɑɻnt].

Dyer's data analysis is based on J at 5;3. She recorded the child in play and in conversation with British English and NAmE speakers, though she does not detail the length or number of the recordings or over what time period the data was collected. It is also unclear what age the child's interlocutors were, though Grandma and Grandpa were listed, and the age of one other was given as 5 years old. There were five other interlocutors whose ages are not listed. As she mentions that an adult was often not in the room, I assume that aside from the grandparents, they were all children. Dyer focuses on J's ability to accommodate to his interlocutor by analysing his use of /t/ voicing, rhoticity and BATH vowel, and comparing it to the accent grouping of the person he is speaking to. This is complicated somewhat as one of the speakers listed as British English was identified as Scottish and a speaker of a rhotic variety.

J never used the NAmE variants in interactions with British English speakers. This is particularly interesting as he did not use the rhoticity present in NAmE when speaking to his rhotic grandfather. As a speaker of a Scottish variety, his grandfather may also have realised the BATH vowel as a front vowel, more similar to the NAmE vowel than the South Eastern British English one he heard at home, though Dyer does not mention this. One possible explanation could be that as Grandma and Grandpa are listed together, perhaps hearing Grandma's non-rhotic variety at the same time as Grandpa's rhotic one was enough to cause J to select the non-rhotic pronunciation. Alternatively, as J was likely in contact with his grandparents all of his life, before his significant exposure to NAmE, he may have recognised his grandparents' accents as part of a group of home accents rather than the community accent, NAmE. J's use of NAmE variants when speaking to NAmE speakers was very high but not 100%. His lowest use of the NAmE variant was in the vowel backing (BATH) category, perhaps reflecting its relative complexity. Rhoticity and T-voicing were present in more than 92% of cases (p. 3074).

Dyer (2007: 3077) concludes that her data demonstrates early sociolinguistic competence by a child in tailoring his variety to his interlocutor. At the time of writing, she suggests that little evidence has been produced of sociolinguistic competence in children under 9 years old. J's ability to notice the variety of his interlocutor and style shift accordingly is thus new evidence of an early ability to differentiate between different varieties in his home environment. In section 2.1.4 we learned that children from multidialectal environments could be more (Jeffries, 2016) or less (Beck, 2014) successful at recognising non-local accents than children from monodialectal environments. Here, the situation is slightly different as the accents that J demonstrates an awareness of were all familiar to him. J is also bilingual, while both Beck and Jeffries' research focuses on monolingual children. However, this case study still adds further evidence that children's accent awareness begins at a young age. In Kobayashi's (1981) research, the child was much older (8;0) when demonstrating competence in style shifting between the parent-code and peer-code.

Dyer's is not the only example of children being in control of more than one variety. Khattab (2013) collected data from three bilingual children in West Yorkshire, UK; one female aged five, and two brothers, aged 7 and 10 years old. The children were born in the UK to parents who had moved from Lebanon as adults. Khattab describes the parents' English fluency as "advanced" but with a noticeable foreign accent (2013: 448). The parents mainly spoke to their children in Arabic at home, while they were simultaneously exposed to English at nursery, which they attended from six months old. Khattab notes that the children were exposed to multiple varieties of English as their friends came from families who were from various areas of the UK. The children were recorded playing with friends of the same age, in conversation with the researcher (in English) and in conversations with their mothers (mainly in Arabic), though for the purposes of this analysis she focused on the interactions with the mothers, as this is where code-switching between Arabic and English occurred. Though

Khattab classifies Arabic as the first language (L1) and English as the second (L2), she describes the children as being English-dominant. Following an analysis of the children's accents, she classifies their accents as "native-like", containing a mix of Yorkshire features, more broadly northern features and "standard-like features". Typical Arabic phonetic features appearing in the data included trills and taps for /r/, clear /l/ across all environments, a lack of /h/-dropping and a lack of /t/-glottalisation as well as a range of alternative vocalic realisations of the following lexical sets: BATH, START, FACE, STRUT, GOAT THOUGHT, and GOOSE (Khattab 2013: 450).

Khattab (2013: 448) reports that some children for whom English was their dominant language, but whose parents spoke foreign-accented English, exhibited features of both local varieties of English and also features of their parents' varieties of English. She claims that the children managed multiple registers simultaneously, making use of aspects of a range of varieties. This included local features, features belonging to the wider North of England, and features of 'Standard English' as well as phonetic features of their parents' varieties.

Khattab (2013: 451) notes that the range of phonetic features that the children used in interactions with their mothers was much more wide-ranging than those seen in the peer interactions. Children were, she observes, switching between "native-like English phonetic features" and "Arabic-like phonetics", demonstrating that children can still have access to parental codes of speech even if they do not typically use them in interactions with their monolingual English-speaking friends. The proportion of code-switched utterances varied among the children, depending on their English and Arabic proficiency. The child most dominant in English made use of English-like phonetics in her code-switching to English, whereas the child with the highest Arabic proficiency used English less and was more likely to mix languages. The seven-year-old child's language skills were more evenly balanced across English and Arabic. His utterances contained the most examples of mixed languages, and his English utterances contained a mix of English-like and Arabic-like phonetics. After closer inspection, Khattab proposes that the children's choice of both language and phonetics were systematic and demonstrated awareness of the significance of their selections. For example, Khattab interprets some of the children's use of Arabic-like phonetics as a convergence strategy. In one interaction, for instance, the mother referred to a pair of glasses in English. The child explained, in Arabic, that the more precise word she was looking for was "sunglasses". In the child's pronunciation of "sunglasses", he articulated the STRUT vowel in the same way that his mother would and the [n] was unassimilated, as [sʌ̹nglæsəz]. Elsewhere in the data, he pronounced the word as [sʊŋglasəz] using local English-accented phonetics (p. 465).

The children also used phonetics as a divergence strategy from their mothers. In several examples, the mother produced an English word with Arabic-accented phonetics, which was repeated by the child in the same way, also with Arabic-accented phonetics. The child then went on to repeat the word with English accented phonetics. For example, the English word "pictures" was produced as [pɪktʃɚz], with an unaspirated word-initial voiceless plosive and a rhotacised vowel. After the mother pretended that she did not hear the word, the child looked irritated and repeated it with English-accented phonetics: [pʰɪktʃəz] (Khattab, 2013: 457).

Khattab also cites examples where the children switch from Arabic to English to "fill gaps" in their knowledge of Arabic. These gap-fillers were typically single word utterances and were produced with English-like phonetics. Similarly, she identified instances of the children producing English words as part of an Arabic conversation as part of a playful word substitution. For example, as part of a discussion about the seasons, the mother pointed to a picture of an umbrella and asks what it is. The child responded in Arabic-accented English, [ʔʌmbrɛlla]. Khattab notes that the child used gemination of /l/, an initial glottal, and a trill for /r/. When the mother asked for the word in Arabic, the child was able to produce it. This pattern appeared several times in the data. In one case, the child seemed to think that they had answered in Arabic, even though their response was Arabic-accented English (Khattab, 2013: 461).

In summary, the children in Khattab's study demonstrated highly skilled socio-phonetic behaviour. Even though these children were fluent English speakers with native-like accents, they also demonstrated a grasp of the phonetics of their heritage language and how to use this for a variety of communicative purposes. Khattab briefly mentions that the children were all "native-like", but detailed information about the analysis of their vowels is not provided. This suggests that these children had a different outcome to the children in Payne (1980), Scobbie (2006) or Roberts' (1997) (see section 2.3.1.2), whose research was focused on whether children had assimilated to the local variety when one or more parents came from out of the area, and in particular, the residual phonetics that betrayed their outsider status. Khattab, on the other hand, focuses on the benefits of having access to multiple phonetic codes, and how children are skilled and playful language users who deliberately choose their phonetics to achieve a desired communicative effect. It would appear that children are capable of using knowledge of multiple dialects if they are in a position where using them is of benefit to them.

Stanford's (2008) analysis of the children's data he collected (discussed in 2.3.1.3) presents a slightly more nuanced situation than the one reported in the interviews, revealing some evidence of accommodation. While all children were dominant in the patrilect, and even the under-fives showed dominance in patrilectal lexical items, there were some examples of matrilectal tokens appearing in the speech of older children. These represented a small proportion, however. Stanford notes that a 3-year-old girl used both lexical variants in the same recording. The matrilectal forms usually occurred after prompting from the mother, while the patrilectal variants spontaneously occurred a few minutes afterwards. This behaviour corresponds with Khattab's (2013) description of the children in her research using some phonetic variants to converge with their mothers and others to distance themselves from them, thus demonstrating sophisticated sociolinguistic skills. Stanford also reports an example of a phonetic compromise, which was made on tokens of the word "to transplant", by a ten-year-old girl and a ten-year-old boy, (2008: 582). This compromise aligns with the observations made by Kerswill (1996) (see section 2.3.1.2) of a child's realisations of the GOAT vowel in Milton Keynes UK, which was a compromise between the variants belonging to each of their parents. The perspectives of the two researchers reflect the children's motivations for speaking more than one dialect. In the community investigated by Khattab, the children's use of their mother's features was met with positive reinforcement; the children were encouraged to speak Arabic, and where they did not have the linguistic skills to do this, the Arabic-accented phonetics provided an achievable compromise. In Stanford's research, on the other hand, the children were actively discouraged from using their mother's dialect. Admittedly, there was no comparison of fathers' dialects in Khattab's research, but a distinction between male and female varieties was not being drawn as it is in the Sui clans, where there was a stark distinction between the matrilect and patrilect.

While the examples above involve children accommodating to members of their own families, Barbu and colleagues (2014) analysed the interactions of children from the Haute-Savoie department in the northern French Alps with their friends. The variable analysed was (Y), a French clitic pronoun. Standard forms of the variable are "le", "la" and “les”, while the stigmatised realisation “y” is stereotypical of the region. The authors analysed stylistic variation in 13 10–11-year-old children and their friends in 4 schools in French alpine villages (seven girls and six boys). The children’s friendships were categorised by duration and whether their friends were local to the area or not:

NL – Native children known for a long time
NNL – Non-native children known for a long time
NNS – Non-native children known for a short time

“Native” was the term used by the authors to describe children who had been born and raised in the local area. That is, they were local to the region and would therefore be familiar with the local dialect. The term “native” was not used to denote whether the children were native to France. Although the authors had intended to include a further group of friends, native children known for a short time, this group did not occur frequently enough to contribute to the dataset (Barbu et al., 2014: 4). Data analysis revealed that boys used non-standard variants in interaction with their native friends more than they did with non-native friends and that this reflected similar usage of the local variants in their friends. No such pattern was apparent in the girls’ speech, though they did use the local variant, it did not appear to be subject to the same systematic stylistic patterns as the boys’ variants. Barbu and colleagues (2014: 9). offer further analysis and explanation for the gender differences in their results, but this is not discussed here as gender is not of particular relevance to this thesis. The authors interpret their results as evidence of the male children being active participants in maintaining the regional variants in their social network. Their results, they claim, reflects similar patterns in the speech of children as are seen in the sociolinguistic literature for adults, but that “still little is known about the sociocognitive process by which children map language variation onto social group differences and situations”.

In the cases of Kobayashi (1981), Dyer (2007) and Khattab (2013), the children were exposed to substantially different varieties in their day to day lives. In the case of Kobayashi, the child was exposed to both Tokyo and Osaka dialects. In Dyer’s case, J heard British and American varieties of English, and in Khattab’s research, the children were exposed to local and supra-local varieties of English outside of the home, and heritage language-accented English inside the home alongside use of their heritage language. In these cases, the varieties in the home were from geographically distant places; these varieties were unlikely to be replicated in the wider community. This provides rather a stark contrast between the varieties, which can make it somewhat easier for sociolinguists to identify which patterns in a child's speech are coming from which influence.

Khattab's (2013) example of children making use of the phonetics of their heritage language for purposes of accommodation brings us back to "the Ethan Experience", which Chambers (2002) claims is seen in children of foreign-accented immigrant parents, discussed in section 2.1.4, above. These children, Chambers suggests, never acquire features of their parents' foreign accented English. Less categorically, other scholars have also observed that children do not *typically* acquire their parents' foreign accented English (DeJesus et al., 2017). The example Chambers gives is the absence of a tapped /r/ in Ethan, a child born in Canada, whose parents had that feature as part of their Eastern European accents. While I have never heard of a child born in the UK speaking with a French or German accent for example, anecdotally, I have come across examples of children with Scottish parents acquiring a Scottish accent. My own brother is an example of this. He left Scotland as a baby of 6 weeks old, moving to the south-east of England. As an infant, he acquired a Scottish accent from our parents, but as he began to develop a south-eastern English accent on starting school, he remembers being corrected by our parents and criticised for speaking with a "fake" accent. Our father, he recalls, was keen on encouraging him to value his Scottish heritage, so he learned to keep the two accents separate to please these two different audiences, his parents, and his peers. The English accent was only used at school, and a Scottish accent was used at home (similar to Kobayashi's research perhaps but in this case, no one was aware that these two varieties co-existed in this child). I grew up being unaware that he spoke with an English accent at school, and it was only after he left home, got married and brought his new family home that I recall hearing his English accent, as he found both of his audiences in the same room at the same time. This anecdotal example differs from Ethan in that while our parents were accented immigrants, their first language was (Scottish) English, and their ambition as recent arrivals to England was not to fully assimilate, but to preserve their Scottish heritage for their child. The case of speakers of another language coming to the UK may be different and is likely to be driven by whether their priority is for their children to assimilate or to strongly connect with their cultural heritage. It is well known that immigrant parents often choose not to teach their children their heritage language due to persistent myths about bilingualism, for example, that learning two languages simultaneously will confuse a child and that monolingual children will integrate better into mainstream culture and be more successful in school, (Genesee, 2009; 2015). Parents keen to support their child in fitting in to the community may therefore choose not to promote the child's heritage language at home. In the Kobayashi data, it is possible that the status of the Tokyo dialect is what contributed to its survival at home (see also Kazazis, 1970; Surek-Clark, 2000).

We have seen above that in spite of the early claim from Labov (1964) (see section 2.3) that children do not acquire sociolinguistic competence until early adolescence, there is evidence that not only are young children capable of perceiving differences in accents (see 2.1.4), but they are able to adapt their own accent to suit their interlocutor. They can make judgements about social situations and code-switch between varieties, even applying the phonetics of one language to another in pursuit of convergence or divergence with their interlocutor as shown by Khattab (2013). Young children might even be able to make sophisticated judgements about the group membership of a speaker based on more than their accent features alone, as seen in J's classification of his rhotic grandparent as a British English speaker as seen in Dyer (2007). In Barbu and colleagues' research, we see more evidence of children making use of the structured variation seen in adult speech (see also Foulkes et al., 1999) and actively contributing to the continuing status of local variants in their community. The evidence above indicates that it is possible that Henry may alter his accent depending on his interlocutor, therefore this should be considered when designing the data collection methods for this thesis (see section 4.3.4).

2.3.3 Summary

Above, we learned that variation in children's speech production has received little attention in the literature, but that those who have engaged in this research have found that children demonstrate evidence of structured variation. Early thinking was that this did not develop until early adolescence (Labov, 1964) but more recently, researchers have exposed structured variation in young children (e.g. Barbu et al., 2014; Khattab, 2013; Kobayashi, 1981; Smith et al., 2007).

Hazen (2002: 506) points out that in "almost all variationist research" since Weinreich and colleagues (1968), peer influence is seen as being the predominant influence on children's accents – though there is not yet evidence that it is the *only* influence. Chambers claims that where the varieties of one's parents and peers are different, "learners normally resolve the tension in favour of their peers" (2003: 185). This is so typical, he argues, that it could be considered abnormal for a child's variety to be more consistent with their parents than their peers. Researching the varieties of children with parents who have moved to a new dialect area presents an opportunity to test this claim.

In most of the research presented above, where complex rules were the site of differences between home and local varieties, their acquisition was affected (e.g. Hewlett et al., 1999; Kerswill, 1996; Payne, 1980; Roberts, 1997; Scobbie, 2006; Trudgill, 1986) by children having parents from out of the area. The effects of these differences in input on acquisition varied from partially acquired patterns (Hewlett et al., 1999; Payne, 1980) to arbitrary linguistic realisations (Scobbie, 2006), to phonetic compromises (Kerswill, 1996). However, there appear to be other factors contributing to a child's probability of acquiring a local variety. Kerswill and Williams (2000) suggested that a child's integration into social networks affects their dialect acquisition. This was also Labov's conclusion after re-analysing Payne's King of Prussia data (2001). This integration could also link to Kazazis' (1970: 118) claim that children may retain a home dialect. He claimed that Istanbul Greeks are typically proud of their heritage and that alongside characteristic Greek family orientation this could explain why the Istanbul features persist in the dialects of the second generation. Two of the studies above involved communities with no stable dialect – Longyearbyen (Mæhlum, 1992), and Milton Keynes (Kerswill & Williams, 2000). The outcomes for the children were different in each, however. In Longyearbyen the children retained their family variety, perhaps in recognition of their family's lack of investment in the community, seen in their long summer visits to their home area. In Milton Keynes however, where residents were making a permanent home, a new dialect emerged. Finally, Kobayashi (1981), Dyer (2007) and Khattab (2013) provide evidence of children acquiring the local dialect alongside the variety spoken one or more parents. In Kobayashi and Dyer's cases, the children retained both dialects, accommodating to the variety of their interlocutors, while Khattab's research revealed that children might draw on phonetic aspects of the parent's variety in selected interactions.

As we saw in Kerswill and Williams' (2000) and Labov's (2001) interpretation of the importance of social integration, there may be motivational factors at play in Roberts' (1997), Scobbie's (2006), and Hewlett et al.'s (1999) data that were not considered. Above, in section 2.3.1.2, Thomas and Scobbie (2015) noted the difference between the behaviour of siblings from the same family and called for larger scale research to attempt to expose other factors that could influence dialect acquisition, such as personality. This was cited as a factor in Deser's analysis of parent and peer influences in Detroit, where she described children who sounded like their parents as "less rebellious" than the children who exhibited the local variants, pointing to the effect of personality on linguistic allegiances.

Finally, taking Kerswill and Williams' example of the boy who switched from a Scottish accent to a local Milton Keynes accent as an example, and reflecting on both Kobayashi's research and on my own experience of having a brother who was secretly using different varieties for different audiences, what we do not know is whether a child's "new" accent is present in all interactions. We only know about those witnessed by the researcher. This raises questions about whether a child may retain some kind of latent phonology learned in their original acquisition, which may be accessed in some limited interactional circumstances, such as when speaking to their parents. This gives a clear advantage to linguists investigating the acquisition of their own children, as they are able to observe their linguistic behaviour in multiple settings, as in the case of this research.

2.4 Summary of the literature

This chapter has looked at literature covering children's language perception, production and their sociophonetic development. Through this review of existing research, I have attempted to uncover what aspects of speech sounds in a child's environment they orient to, and how this process manifests itself in a child's phonological development.

Children can differentiate between languages at a very early age, and there is evidence that they can differentiate between accents when they are under 2 years old. Vowels are where we see most difference between accents, at least in English, so this may be where we can find evidence of a child's perception of accents coming through into their production. These sounds are somewhat neglected in the phonological acquisition literature, however, as many linguists do not appear to see them as worthy of attention. Yet, due to their social salience, vowels can reveal important information about the influences on a child's phonological development. Vowels can carry information about region, education, social status or age, and in the case of children whose parents speak a non-local variety, a child's acquisition of vowels can tell us about who in their environment is influencing their linguistic development. At the same time, we need to try to tease out these complex patterns from the general variability children exhibit in their early speech and their acquisition of structured variation. These cases may unmask not only, as Hazen (2002) claimed, where the influences on their accent are coming from, but the level of linguistic structure at which children appear to be exhibiting these influences; word, phoneme or at the level of phonological feature/articulatory gesture.

Finally, with the exception of Kobayashi (1981) and Kerswill and Williams (2000), the research demonstrating the effect on children of having parents from another dialect area has all been synchronic. There is almost no longitudinal research which tracks the acquisition of a child's phonological system under these conditions. The challenges of frequent recording and close observation presents an opportunity for linguists to analyse the phonological development of their own children, as they have the advantage of being able to closely track their child's development at home. On that basis, this research aims to account for how a child's accent might move from the home variety to a local one over time though frequent analysis over a four-and-a-half-year period.

In the following chapter, I outline the history of and current practices for phonetic analysis of children's speech, before describing the methods used in this research in Chapter 4.

Chapter 3 Methodologies for the analysis of children's speech sound production

The following chapter addresses the techniques used for the analysis of children's speech and their history. This takes us from the earliest recorded examples of research on the acquisition of speech sounds, in the form of nineteenth century diary studies, to present day impressionistic and acoustic methods of analysis. These methods will be evaluated for their suitability to collect and analyse the data presented in the following chapters.

3.1 History of methodologies in children's language acquisition

The history of research on children's language acquisition falls into three distinct methodological and theoretical camps, beginning with diary studies in the mid-nineteenth century which have continued to the present day, followed by large cross-sectional studies beginning in the US in the nineteen thirties, and more linguistically rigorous studies appearing from the early nineteen sixties, (Local, 1978; Menn & Stoel-Gammon, 1995).

3.1.1 Diary studies and individual developmental studies

The diary method involves a parent keeping detailed records of a child's development and making observations. This method has both advantages and drawbacks. An expert parent is able to make observations of a child in a variety of settings over a long period of time in a way which would be impossible for any researcher external to the family (Menn & Stoel-Gammon, 1995), due to the time and intrusion of such an approach, though an obvious disadvantage is that the parent is the only observer (Khodareza et al., 2015: 4637). Samples in a diary approach may also be small and can be subject to parental bias about what they think is worthy of note. According to Ingram (1989) a common criticism is the lack of theoretical orientation; without a plan of what to observe, observations can be randomly selected. Diary studies are also limited to the behaviour of a single child or siblings, which may not be representative of the wider population. Furthermore, the reliability of transcriptions done live without a recording cannot be assured (Menn & Stoel-Gammon, 1995).

Claims vary of whose diary study was the first. According to McCarthy (1930) the earliest example of a diary study which included elements of language acquisition was Tiedemann, who kept a diary of his son's development in 1787. Though Tiedemann's diary study was mainly concerned with his son's psychological development, it included observations on his perception of language, the production of some speech sounds and the acquisition of some dialect features (Local, 1978). Ingram claims that the first was Taine's 1876 report on his daughter's linguistic development up to the age of two; this one was undoubtedly more focused on language acquisition than Tiedemann's, whose linguistic observations were only a small part of his diary. Other diary studies on language development followed in the second half of the nineteenth century and early twentieth century, mainly in Germany. Clara and Wilhelm Sterne carried out a diary study of the language development of their two children, Hilde and Gunter in 1907. This work is generally acknowledged to be the first comprehensive account of the stages of language acquisition, though it has never been translated into English (Ingram, 1989: 9). A frequently cited example of a diary study is Leopold's 1939 analysis of his daughter's phonological development as she learned both German and English, though Leopold's main focus was on her acquisition of English.

Diary data has uses beyond the original diarist; Jakobson (1968) produced some early theoretical work based on the existing diary studies available (e.g. Leopold, 1939) and anecdotal evidence. Stoel-Gammon and Sosa (2007: 244) point out that like most early theories, however, it was the unmodified application of adult phonological theory to children (what they call the "extension phase" of a theory (after Menn, 1980)). Historic diary data has gone on to be re-presented and reanalysed by many other linguists, though its usefulness is determined by the original detail and/or context captured by the diarist. This will be discussed further, below.

Vihman (2014) reports that only three studies of phonological development were published between 1938 and 1967: Leopold (1939), whose work is mentioned above, Velten, who carried out a case study on her daughter as she learned English, and Kolaric, who performed a study of two Slovenian children aged 0;6-2;0 and 0;11-3;0 (see Table 6). However, Vihman notes the increase in these individual and small group studies in the following years. These cover a range of languages, though English and German dominate. Both monolingual and bilingual acquisition are represented.

Table 1.1 Small group and case studies, 1938–2013

Years	*Studies*	*Languages*	*Children*
1938–1967	3	3	4
1968–1977	13	7	14
1978–1987	10	3	13
1988–1997	21	4	40
1998–2007	11	5	27
2008–2013	7	3	21
Total	**65**	**25**	**119**

Table 6: Small groups and case studies 1938-2013 (Vihman 2014: 14)

Some of these later diary studies have been extremely influential, for example, the work of Waterson (1971) and Smith (1973).

Waterson (1971) conducted a Firthian-influenced study of her child, “P”, based on diary data, but she does not present the whole dataset. Rather, she presents an analysis of a subset of data, categorising words according to particular structural patterns apparently favoured by the child. She then applies these structures to some of Leopold’s data where he struggled to find an explanation as evidence of the wider applicability of her hypothesis. This diary study signalled the beginning of whole word phonology and influenced the development of the concept of “templates” later described by Menn in 1983 (Vihman & Croft, 2007; Vihman & Keren-Portnoy, 2013).

Smith’s (1973) seminal diary study focuses on the phonological development of his son, Amahl. It has been described as “the first comprehensive longitudinal study of a child’s phonological development” (Demuth, 2014: 574). Before Smith, the oldest a child had been at the completion of the research was 3;0, and the largest lexical inventory was 500 words (Vihman, 2014). Smith’s diary study continued until his son was 3;9, though he does not specify the child’s lexical inventory at the completion of the research. Smith’s main claims are that at the point that a child starts talking, their lexical representations are equivalent to those of an adult in the target language, and that the child does not have their own phonological system (Smith, 2010: 19). One of the ways in which Smith’s work has had a lasting impact is that he published all of his data (Demuth, 2014).

Diary study methods have been the subject of frustration for some researchers, who complain about a general lack of systematicity (see Irwin, 1941; McCarthy, 1930). Criticisms include a lack of systematic research on infants under 6 months old, gaps in the data (Ingram, 1989; Irwin, 1941), and a lack of proper use of phonetic notation (Irwin, 1941; McCarthy, 1930). An early criticism of the diary method was that they were not representative of the larger community; the children appearing in diary studies are frequently either precocious or particularly slow in their language development (McCarthy, 1946: 494). Advantages of the method were, however, recognised. Leopold's case study on his daughter's bilingual development in particular, has been described as valuable (Irwin, 1941). Its value has continued to be recognised for decades after its publication due to Leopold's publication of his "raw data" which enables reanalysis by future researchers (e.g. Local, 1978).

Above we have seen that diary studies are a time-honoured method in language acquisition research (Vihman, 2014: 14), and in spite of some drawbacks they can offer detailed insights into children's linguistic behaviour. Diary studies are not unsystematic per-se, and some have provided a transparent, detailed account of a child's language, but their quality varies (Ingram, 1989). In some cases, this close analysis of individual children has provided ideas which have given rise to major theoretical developments. While diaries were the dominant research method for the analysis of language acquisition in the early twentieth century, some authors were making attempts to draw together diary data from a larger number of children (e.g. Lewis, 1999). These early attempts to combine diaries suffered from some methodological criticisms, however, due to the problems of comparing data which had been collected using different methods. At the same time, an interest in collecting data more systematically and from larger numbers of children was beginning to establish itself.

3.1.2 Cross-sectional studies

Large sample, cross-sectional studies (also called 'norming studies') started to appear in the early 1930s, though the first of its kind appeared in 1926 (Ingram, 1989). This method originated from a new approach to psychological study, behaviourism. In a departure from individual or small group studies, these investigations attempted to respond to what were thought of as weaknesses of the diary study approach by following a precise systematic method that was quantitative rather than qualitative. Data was collected from large numbers of children across multiple societal groups, for example, sex, age and socio-economic status. These studies tended to use techniques such as picture naming tasks to elicit data from subjects rather than collecting examples of naturally occurring talk. The intention with these studies was to establish the typical ages (norms) at which children acquired phonemes. Large sample studies provide a systematicity absent in some of the earlier diary studies, and they have a practical use in establishing normal ranges of development as is evidenced by their continued use by speech and language therapists. They can also provide a wealth of data for reanalysis, but data must be used cautiously as explained below. They do have some significant drawbacks, however. Unlike diary studies, data from cross-sectional studies is not longitudinal, though collecting data from children of different ages does provide an apparent time construct. This can offer a view of a typical child at age 3 or 4 for example, but it does not give the more detailed information about how an *individual* child develops. These studies tend to be superficial and descriptive, for example, counting the number of words in an utterance, rather than looking for underlying patterns in the child's grammar. Even though the main period of popularity of cross-sectional studies (circumscribed by Ingram as 1926-1957) was later than the period of early diary studies, data was usually not recorded, but was transcribed live, with the experimenter quickly noting down a child's responses. Templin (1957), author of the last large sample study during their period of peak popularity, explains why she does not use recording equipment in her data collection: "The use of recording equipment is not efficient when recording must be done in many places under varying and often unsatisfactory acoustic conditions," (Templin, 1957: 19). Ingram (1989) feels that we should therefore be worried about interpreting data collected under these conditions. Finally, large sample study data is experimental rather than naturalistic and generally provides information on a small set of words. Studies of this type, therefore, are best looked at in tandem with individual or small group longitudinal data such as that which comes from diary studies. "Taken together they complement one another and provide a fuller picture of the phenomena at hand" (Menn & Stoel-Gammon, 1995: 336), though Vihman (2014) claims that this approach is rare.

Large, cross-sectional studies continue to be carried out, filling gaps in the data, as they remain an important resource for speech and language therapists, though they have been fewer in number since 1957, however, when interest in a new kind of methodology began to emerge (Ingram, 1989).

3.1.3 Small group experimental studies and naturalistic studies

The next major trend in language acquisition research was longitudinal sampling, a natural progression from diary studies and cross-sectional studies, taking the best of both approaches (Ingram, 1989). The major contributors at this time were based at different U.S. universities: Roger Brown and associates, Susan Ervin and Wick Miller, and Martin Braine. These linguists each developed their own longitudinal sampling method, independently (Ingram, 1989). Brown (1973) claims that at that time, researchers had had their interest stimulated by the work of transformational linguists such as Chomsky's *Syntactic Structures* in 1957. Until then, the focus had generally been on description of the child's language output in terms of describing phonemic or lexical inventories. The influence of *Syntactic Structures* led to an interest in explaining the rules underlying language development (Local, 1978: 5), (see for example, Braine, 1963; Brown & Bellugi, 1964; Miller & Ervin, 1964). Brown claims that although transformational linguists had inspired this new approach to language acquisition research, they did not approve of it, as they were not convinced that it was possible to "discover constructional knowledge" from the "mere performance" of children (Brown, 1973: 99).

These studies differ from diary studies in that the subjects are not usually children of the researchers, and the children are chosen specifically to meet the needs of the study, for example, being at a particular stage in their language development. This method is similar to diary studies in that data is collected longitudinally, so the progress of an individual child can be tracked. In all but one of the studies in Table 7, the data was made up of samples from three children. Ingram explains that more than two children are selected so that the researcher can compare development across a number of children. The smallest number of children needed to support the author's claims would be three: with a single child it is not possible to know if development is atypical; if data is collected from two children, they may behave differently, and how does the researcher know which one is typical; but with three children there can be a majority (Ingram, 1989: 21).

Table 2.5 *General information on four major studies using longitudinal language sampling*

Investigator	Children (age range in months)	Sampling schedule
Braine (1963a)	Andrew (19–23)	parental diary of all multi-word utterances produced. For Steven, there were tape-recordings for four hours over a four-week period (12 sessions.)
	Gregory (19–22)	
	Steven (23–24)	
Miller & Ervin (1964)	Susan (21–)	initially weekly in 45-minute sessions; later every two months for 2 or 3 sessions for 4–5 hours. Sampling over a two-year period.
	Lisa (24–)	
	Christy (24–)	
	Harlan (24–)	
	Carl (24–)	
Brown (1973)	Adam (27–44)	two hours every two weeks; two observers present.
	Eve (18–27)	
	Sarah (27–44)	half-hour every week.
Bloom (1970)	Eric (19–26)	eight hours over three or four days, every six weeks.
	Gia (19–27)	
	Kathryn (21–24)	

Table 7: General information on four major studies using longitudinal language sampling (Ingram, 1989: 22)

As in the cross-sectional studies, data is collected systematically at predetermined intervals for a fixed length of time. In a significant departure from earlier studies, much more data is collected from each child, and all sessions are recorded and transcribed at a later time. Aside from these commonalities, the methods used by individual investigators can vary; some of these early studies complemented their recordings with parental diaries (for example, Braine, 1963) and adjusted their recording sessions to coincide with key developments reported by the parents, while others stuck to a more rigid schedule. Ingram suggests that a more flexible approach such as Braine's is sensible as regimented data collection points might miss out on the emergence of important developments, but such an approach is rarely seen (Ingram, 1989: 23). Following on from this move towards longitudinal sampling, linguists began to introduce experimental methods to support naturalistic data collection; this approach allowed linguists to test hypotheses (Menn & Stoel-Gammon, 1995: 337). The combination of naturalistic and experimental methods emerging in the mid to late twentieth century are generally accepted as the most rigorous methods and continue to be used today; "Together, naturalistic and experimental approaches provide data essential for hypothesis testing and theory construction," (Menn & Stoel-Gammon, 1995: 337).

3.1.4 Summary

We have seen above that over the past century, the methods used to investigate children's language acquisition have evolved. The earliest known method, the diary study, has in some cases, provided excellent data which continues to be retrospectively analysed. In response to some of the early, unsystematic diaries, methods in data collection have become more systematic over the past century, though some of these large sample cross-sectional studies offer a superficial view of language development. A range of data collection methods continue to be used in phonological acquisition research, though diary studies are now generally supplemented by systematic recordings thanks to advances in technology. The combination of naturalistic and experimental data collection tends to be considered the most rigorous approach, but case studies continue to offer the opportunity to follow the development of a single child in detail, generating new ideas to be investigated in a larger group. As recording technology has advanced, leading to more reliable transcription of children's speech and corpora which may be reanalysed, there are further methodological opportunities to add rigour to the analysis of data collected from children. I now move on to an investigation of acoustic analysis of children's speech.

3.2 Acoustic analysis

The vast majority of acoustic research has been performed on men. There has been little attention given to the acoustics of women and children's speech for 'technical and social reasons'. The choice of a 300 Hz analysing bandwidth in early spectrographic analysis worked well for the analysis of male speech, but less so for women and children's voices. Kent and Read explain that this 'probably' had the effect of discouraging the acoustic analysis of women and children's speech (2002: 189). Some limited analysis of children's voices does exist, however. In this section I explain why acoustic analysis of children's speech is important, before presenting an overview of the reference data available, as this will be used to compare my acoustic analysis of Henry's vowels to existing acoustic reference points. An explanation of the challenges of performing acoustic analysis on children then follows, before a roundup of best practices for analysing the acoustics of children's speech.

3.2.1 The limitations of impressionistic transcription

In section 3.1 above, I proposed that while the quality of diary studies can be variable, collecting longitudinal data from a single child remains an excellent way to understand the path a particular child takes through their acquisition of language, and can expose phenomena worthy of study in larger populations. Keeping a detailed and accurate record of a child's productions is, however, essential. Indicating a lack of access to recording equipment at the time, McCarthy (1930: 24) writes that a common perception is that once a child reaches 5 years old, they are so fluent that it is impossible to keep an accurate record. Yet, above, we saw how the use of audio-recording in language acquisition studies was fairly slow to start, given the limitations of early equipment (e.g. Templin, 1957), but tape-recording started to appear alongside the small-group longitudinal and experimental studies described in section 3.1.3, above. Tape recorders have of course been replaced by other equipment in the last few decades, most recently by solid-state digital recorders, whose use is now standard practice. Recordings ensure that time can be taken over a precise transcription. However, even in the case of a careful impressionistic transcription, there can be questions about its reliability. Oller and Ramsdell (2006: 1392) explain that when the speech under investigation is different from well-formed adult speech, inter-rater reliability can decrease. Others have found that impressionistic transcriptions often differ from acoustic records (Shriberg et al., 1984). Shriberg and Lof (1991) found that the reliability of transcription varied from 20-100% depending on the context. They carried out a review of the transcription literature and concluded that it was not possible to make generalisations about the reliability of transcriptions, as the methodologies and reliability results were so variable for each study. Even broad claims such as intra-transcriber reliability is higher than inter-transcriber reliability were not held up across the studies they reviewed. When looking at individual studies, there is certainly evidence that it can be more difficult to achieve agreement on the transcription of vowels than consonants. Davis and colleagues (2002) provide detail of reliability across five transcribers in their research which analysed the babble and first words of infants from 12-25 months. Consonants were more reliably transcribed than vowels, though this varied according to place of articulation: 79% agreement for labials; 79.2% for coronals; and 69.6% for dorsals. This compared to 77.5% agreement for back vowels; 66.3% for front vowels; and 55% for central vowels (Davis et al., 2002: 83). However, in another study, Norris et al.'s (1980), analysis of listener agreement between the transcriptions of 4–5-year-old children, reanalysed by Shriberg and Lof (1991), broad transcriptions of consonants were found to be less reliable than transcriptions of vowels. On the other hand, when looking at the reliability of narrow transcriptions, the agreement on vowel transcriptions was lower. Listeners achieved 91% agreement in 16 out of a possible seventeen vowel phonemes on broad transcriptions, but when performing narrow transcriptions of vowels, that agreement dropped to 53%, or nine out of seventeen vowels. Shriberg and Lof attribute this drop in agreement to "confusing acoustic cues" contained in children's speech (1991: 255). Variability in reliability seems dependent on multiple factors: "transcribers, subjects, sampling modes, sounds, error types and target contexts" (Shriberg & Lof, 1991: 230) and ultimately, they conclude that "multiple sources of evidence should be presented to support each claim" (Shriberg & Lof, 1991: 273). This leads us to look to acoustic methods of analysis to validate impressionistic analysis.

The first instance of children appearing in a study of acoustic features of speech was in Peterson and Barney's (1952) seminal work. This research provides a frequently cited baseline for vowel formant frequencies in American English; no previous research had provided a systematic study of the frequencies of speech, and the data is still frequently cited today. The study reports on data collected from 76 speakers (33 men, 28 women and 15 children), speaking from a list of words containing 10 vowels between the consonants /h/ and /d/: *heed, hid, head, had, hard, hoard, hood, who'd, hud,* and *heard*. The list was read twice by each speaker. Peterson and Barney (1952: 183) reported that there was a bigger difference in inter-speaker formant values than intra-speaker values. They conclude that inter-speaker differences are not due to vocal tract length alone but are also due to the different ways in which speakers articulate the vowels.

In Table 8, children's averages by vowel are presented in the same order as Peterson and Barney (1952). Data was not separated by age or sex.

	i	**ɪ**	**ɛ**	**æ**	**ɑ**	**ɔ**	**ʊ**	**u**	**ʌ**	**ɝ**
f0	272	269	260	251	256	263	276	274	261	261
F1	370	530	690	1010	1030	680	560	430	850	560
F2	3200	2730	2610	2320	1370	1060	1410	1170	1590	1820
F3	3730	3600	3570	3320	3170	3180	3310	3260	3360	2160

Table 8: Average formant frequencies (Hz) by vowel (Peterson & Barney 1952)

While Peterson and Barney do not specify the ages of the children in their informant group, Lee and colleagues (1997: 1468) estimate that the children were probably around 8 years old based on comparison with their own data. A comparison will be made against values from children in the UK, in section 3.2.4 below. Informants in Peterson and Barney's research were mostly from the "mid Atlantic speech area" in the US, though two were born overseas and an unspecified number spoke a language other than English. The male speakers represented a wider geographical range than the women and children, though most were speakers of General American. Based on the differences in production and perception in Peterson and Barney's research, Watt (1998) concludes that there may be considerable accent variation in the group.

3.2.2 Why is acoustic analysis of children's speech important?

As discussed above in section 3.1, early research on children's language development was largely unsystematic (for example, diary studies and observations without recordings, alphabetic notation in place of phonetic script etc.) and there were few attempts to establish the reliability of data (Irwin, 1941; Local, 1978). In conjunction with evidence that inter-rater reliability can be lower in the analysis of children's speech in comparison to adult speech (Oller & Ramsdell, 2006; Shriberg & Lof, 1991), it seems clear that employing additional methods to support an auditory analysis is an important step in establishing the veracity of an analysis of children's speech. This is routine for adults, and as transcribing children's speech is notoriously difficult it seems appropriate to adopt this additional measure here. Yang and Fox (2013) discuss the limitations of using auditory analysis alone, pointing out that it can be subjective. For example, analyses can be affected by the child's personality, their age, intelligibility and even the child's physical characteristics. They explain that while some researchers might claim that a certain sound has been acquired by a particular age, other studies argue that the acquisition process continues for some time until adult like patterns are reached (see for example, Sander, 1972). In some cases, different conclusions may be drawn from the same data by different individuals. Di Paolo and colleagues (2011: 87) also insist that auditory analysis alone is not sufficient in the sociophonetic analysis of vowels. Drawing on acoustic techniques in addition to impressionistic analysis adds a further layer of robustness to the analysis, as it offers a way to capture phonetic detail through more objective measurements (Yang & Fox, 2013). While some elements of acoustic analysis of children's speech are relatively straightforward however, such as duration or VOT, the analysis of vowels can be challenging (Khattab & Roberts, 2011: 170). The issues presented by the acoustic effects of a growing vocal tract, the lack of dialect specific data for the purposes of comparison, and a range of additional factors in the analysis of children's acoustic data will be explored below.

3.2.3 Existing reference data

Vorperian & Kent (2007) carried out a review of formant frequencies in 14 separate studies which took place over 50 years, beginning with Peterson and Barney's 1952 data. Eleven of the studies contained data from children between the ages of 3 months and 13 years; the remaining three studies contained adult data only. Using the technique of plotting F1 and F2 on to a scattergram, values of the four "corner" vowels were extracted from 12 of the 14 studies. F1 and F2 values were plotted to show the vowel space of each age group in relation to one another (Figure 2). Joos (1948) was the first to publish a plot of F1 and F2 frequencies in graphic form. Through inverse scaling of the axes and deliberate manipulation of the scales, he made an acoustic chart that was designed to resemble the IPA's vowel quadrilateral, and thus permit relatively straightforward interpretation of formant data as correlates of tongue position in a theoretical two-dimensional space. F1 corresponds inversely with tongue height (lower vowels have higher frequency F1), while F2 frequency corresponds with tongue advancement or retraction; a front tongue position having higher frequency than a more back tongue position. Formants are also affected by many other factors, including lip rounding, aspects of vocal setting, speech rate, stress, the Lombard effect, emotional stress and a range of technical factors relating to medium, transmission and analysis methods.

The formant frequencies of males are presented here as an example of the impact of age on formant frequencies.

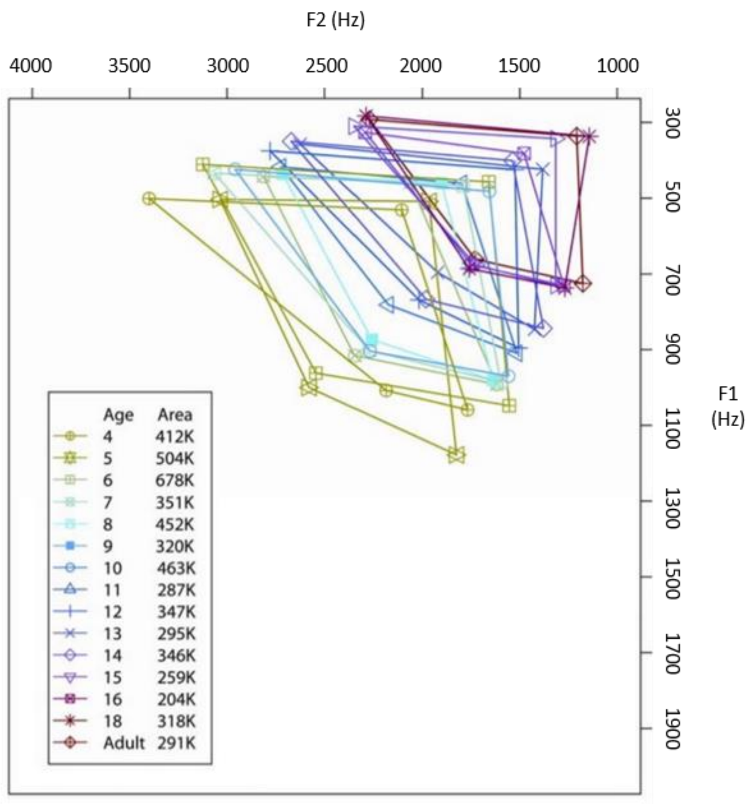


Figure 2: F1 /F2 frequencies of males by age (Vorperian & Kent 2007) rotated 90° clockwise and inverted

Broadly speaking, Figure 2 shows a clear lowering of F1 and F2 over the age range, and reduction of the overall acoustic space. F2, however, (corresponding to tongue advancement) drops much further than F1 (corresponding to tongue height). For example, in close front vowels, F2 drops from around 3400 Hz at age 4 to approximately 2300 Hz in adults. F1 on the other hand drops from around 500 Hz to approximately 350 Hz. This represents an approximate drop of 1200 Hz in F2 in compared to a drop of only 150 Hz in F1. This difference in the proportions of change in F1/F2 during the maturation of the vocal tract has been previously noted by Eguchi and Hirsh (1969). While Figure 2 shows the combination of all data, Vorperian and Kent consider inter-speaker variability, and identify that F2 is particularly variable in the high back vowel, /u/, while F1 values are less variable (Vorperian & Kent, 2007). They suggest that this might be attributable to a number of factors including non-uniform vocal tract growth (Fant, 1975), changes in articulatory gestures (Nittrouer, 1993), along with dialect differences (Vorperian & Kent, 2007). This may reflect variability in GOOSE fronting across different accents (see for example, Sóskuthy et al., 2015). These are outlined below in section 3.2.4.3.

It is important to note that the studies included in Figure 2 incorporate data from children in different age ranges. Each study did not include every age range represented in the diagram. The ages most relevant to this research are 3-6 years. For those ages, in the male category, data from the following studies (Table 9) were included:

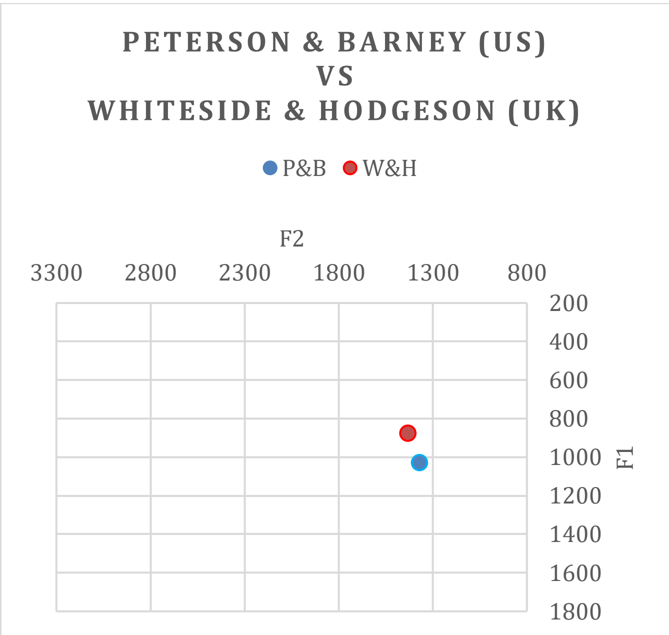
Age	Number of studies	Authors	Dialects
4	1	Perry, Ohde and Ashmead (2001)	US English
5	2	Busby and Plant (1995) Lee, Potamianos and Narayanan (1999)	Australian English US English
6	1	Lee, Potamianos and Narayanan (1999)	US English

Table 9: Composition of Vorperian & Kent data by age and dialect

Table 9 indicates that the number of studies contributing to the formant values at each age is much smaller than the number of studies included in the overall dataset, and that although US English is present in each age category, the presence of Australian English in the data for age 5 may be skewing the formant values for this vowel space. The close front vowel for age 5 appears more retracted than for ages 4 and 6, and the low back vowel more open which may be attributable to this influence. Vorperian and Kent acknowledge that dialectal influences are most likely present but have not been accounted for in any way, (2007: 16). While this data acts as a useful reference point, the mix of dialects, lack of representation of British dialects, and uneven distribution of dialects at each age range may affect its worth as a point of comparison.

3.2.3.1 Reference data dialects

In order to compare US and UK dialect formant data in children, the values from Whiteside and Hodgson (2000), the only study of the formants in UK based children, are compared to Peterson and Barney's data from children aged 8 years in Figure 3. This data has been chosen due to Lee et al.'s (1997) claim (referred to above) that the Peterson and Barney data corresponded best with children aged 8. Boys and girls have been combined as there was no separation by sex in the Peterson and Barney data. The formant values from the Peterson and Barney data are for the low back vowel in the word "hod", which corresponds with the lexical set LOT, alongside values for the words "bar", "jar" and "car" which align with the UK PALM/START lexical set. The vowels in these lexical sets are realised as a low back vowel in each dialect. The children taking part in Whiteside and Hodgson's research were speakers of a dialect from the north east of England, while Peterson and Barney did not specify the dialect spoken by their informants, only that they were mostly born in the "middle Atlantic speech area" (1952: 177). The data was collected 48 years apart, though as we are comparing dialects on two different continents, this seems less relevant than it might be if we were comparing data from the same dialect area at two different points in time, as we would expect there to be differences between these dialects in any case. Note that F1 (vowel height) places the UK low back vowel 151 Hz lower (higher tongue position) than the US equivalent, and F2, (front/back) places the North East UK vowel as 133 Hz higher (more advanced tongue position), thus more centralised than the US vowel described in Peterson and Barney (1952) (see Figure 3).



	P&B 1952 (US)	W&H 2000 (UK)
	ɑ	ɑ
f0	256	255
F1	1030	879
F2	1370	1428
F3	3170	3448

Figure 3: Comparison of Whiteside & Hodgson (age 8 boys and girls) (2000) (UK) with Peterson & Barney (1952) (US) Formant values in Hz for low back vowel /a/ (scale consistent with other vowel plots in this thesis)

Within the Whiteside and Hodgson data, unlike the gradual decrease in F2 by age seen in Vorperian and Kent (2007), F2 values for 8-year-olds were slightly higher than for the 6-year-olds (Whiteside & Hodgson, 2000). This anomaly may be explained by inter-group differences, as the rest of the data followed the expected patterns. As outlined above, F2 variability is typically more than F1 variability (Vorperian & Kent, 2007), so overlap of these values due to inter/intra-speaker variability is perhaps more likely than overlap of F1.

3.2.4 Factors affecting formant frequencies

3.2.4.1 Sex differences

As outlined above, the size and shape of the vocal tract is closely connected to F1 and F2 values; as we saw in Figure 2, as the vocal tract grows in size, these values reduce. Even within children's data, there are differences between the sexes from at least the age of 6-7 years (Bennett, 1981).

Bennett analysed five vowels in 42 seven- and eight-year-old boys and girls. The children all shared a common US dialect. The children were recorded producing the vowels /i, ɪ, ɛ, æ, ʌ / within the frame d_d, (i.e. *deed, did, dead, dad, dud*) inside the phrase "I will say d_d again" (Bennett, 1981: 231). She reports that sexual dimorphism is clearly evident in the formants of a "large majority" of girls and boys of this age. On average, girls' formant frequencies were 10% higher than those of boys. In Figure 4 below, the lines dividing each ellipse are designed to show that most tokens from each gender clustered together. As a proxy for detailed measurements of the vocal tract, Bennett recorded body size (standing height, sitting height, weight, and neck circumference), which she suggests has a direct impact on formant values. Fitch & Giedd (1999) later confirmed a strong correlation between vocal tract length and body length.

Bennett notes that the difference in frequencies between male and female children is less than the difference between male and female adults. While there was limited data available on vocal tract size in pre-pubescent males and females at that time, there was some evidence that the only differences between the sexes in children is that the boys' pharynxes are longer (see for example, Mol, 1963). Boys were an average of 7cm taller than the girls, and 4kg heavier. Based on formant values, Bennett concludes that the larger physical size of the male children leads to them having a larger vocal tract. She claims that children with a taller sitting height, which she suggests would include a longer neck, are strongly correlated with lower formant values. However, the correlation between physical size and F1 values was weak.

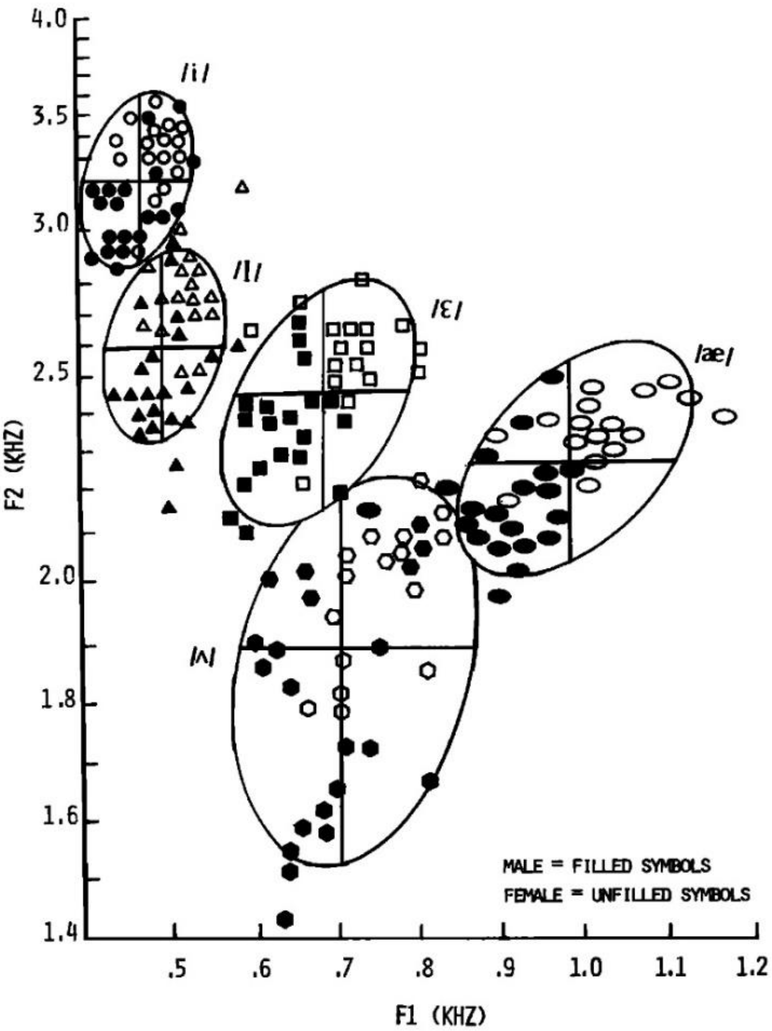


FIG. 1. ***F*1–*F*2 plot of male and female children's vowels.**

Figure 4: F1 -F2 plot of male and female children's vowels (Bennett, 1981)

Bennett claims that the differences in formant frequencies between boys and girls are primarily attributable to the size of their vocal tracts, but there may be other factors which contribute to these differences, for example, if girls were to use a wider mouth opening in open vowels than boys, or if boys were to use more lip rounding (see also Lieberman, 1984; Lindblom & Sundberg, 1971). Formant lowering can also be achieved by lowering the larynx by 10mm (Lindblom & Sundberg, 1971).

Like Bennett, Busby and Plant's (1995) study investigated formant frequencies in children, differentiating between boys and girls. They did not collect size data from the children, but they did collect data on a wider range of vowels than Bennett (11 vowels to Bennett's 5). They tested whether formant frequency differences between girls and boys varied according to which vowel was being articulated. Data was collected from 40 Australian children of 5, 7, 9 and 11 years. Five boys and five girls were recorded in each age group. Eleven vowels were elicited in the form of the test words: ‘sheep’, ‘ship’, ‘bed’, ‘cat’, ‘cart’, ‘cut’, ‘four’, ‘dog’, ‘shoe’, ‘book’, and ‘bird’. The recordings were made under quiet classroom conditions, pronounced within the context of “I can see a ____” (p. 2603). As expected, F1 and F2 decreased as age increased for both genders, though there was an exception in the case of F1 in /ɪ/, /ɒ/ and /u/. See Figure 5 and Figure 6 below.

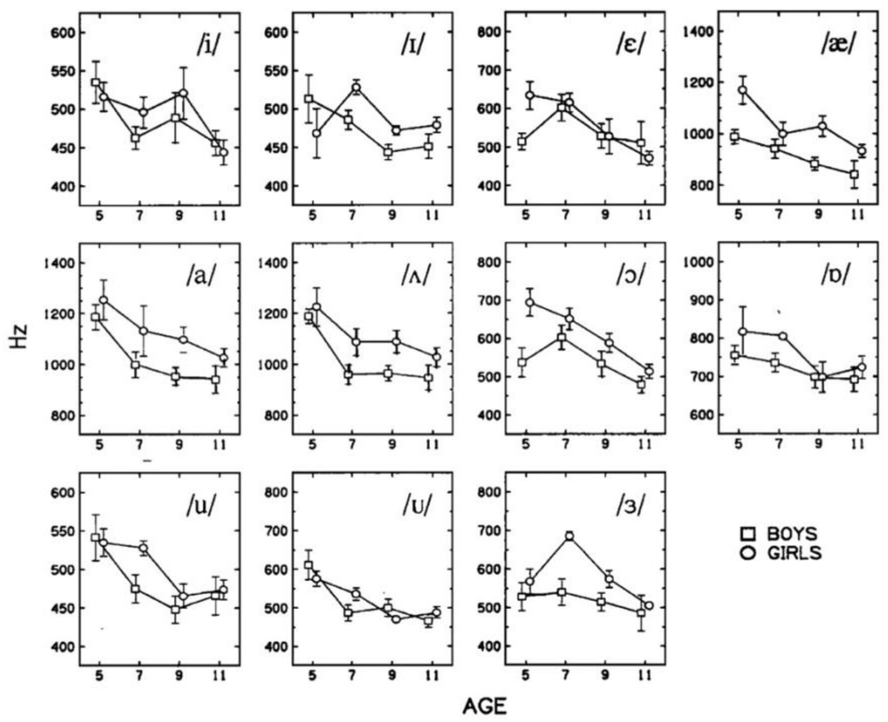


FIG. 2. The mean *F*1 values for the different vowels according to the age and gender of the children. Error bars show the standard error.

Figure 5: Mean F1 values for the different vowels according to the age and gender of the children (Busby & Plant, 1995)

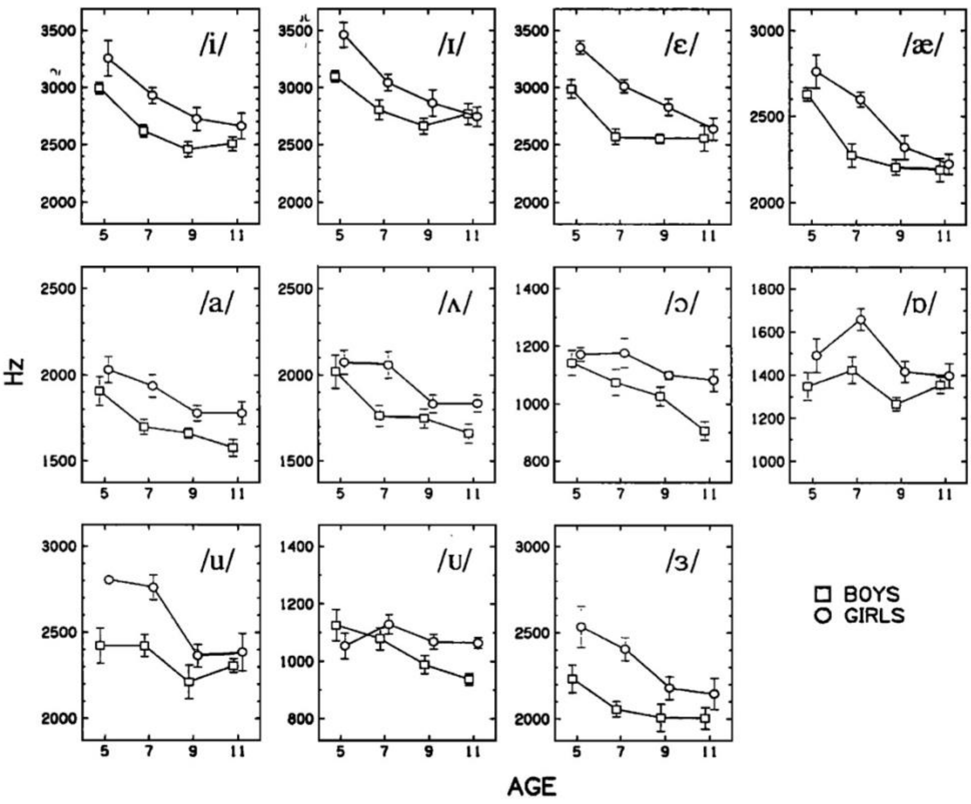


FIG. 3. The mean $F2$ values for the different vowels according to the age and gender of the children. Error bars show the standard error.

Figure 6: Mean F2 values for the different vowels according to the age and gender of the children (Busby & Plant, 1995)

The authors claim that their results are consistent with Bennett (1981) in that she also reported that F2 was lower in boys across all vowels, and F1 was lower in low vowels. A detailed discussion of vocal tract size and its connection to formant values follows in section 3.2.4.3.

3.2.4.2 Phonological environment

Whiteside and Hodgson (2000) collected data from twenty children from the North East of England, with a view to determining the impact of phonological environment on the formant frequencies of a single vowel, /ɑ/. The children were classified into age groups: 6, 8 and 10 years old (ten girls and ten boys). The vowel was measured in two phonological environments (after /b/ in 'bar' and following /dʒ/ in 'jar' in phrase final position), (see Table 10). Almost all average formant values for girls were significantly higher than those of boys in all age groups (Whiteside & Hodgson, 2000: 125).

Although the dataset in this case was rather small, others have reported the same conclusions (see for example, Busby & Plant, 1995; White, 1999).

Age	6		8		10		Adult	
	M	F	M	F	M	F	M	F
F1	1065	997	879	1113	712	921	618	718
F2	1398	1583	1428	1578	1315	1498	1073	1274

Table 10: Mean formant values (Hz) for /a/ (Whiteside & Hodgson, 2000)

The authors report that the phonetic environment of 'jar' versus 'bar' and 'car' had a statistically significant effect on the formant values of F2 (combined in Table 10). The authors interpret this as a palatal coarticulatory effect, F2 being an indication of tongue advancement or retraction. The effect of the bilabial in 'bar' and velar in 'car' however, were not significant (Whiteside & Hodgson, 2000: 130).

3.2.4.3 The physiological effects of growth on acoustic information

Peterson and Barney's (1952: 183) data (Table 11) demonstrates that formant frequencies decrease with age. Formant frequencies are consistently lowest in adult males.

TABLE II. Averages of fundamental and formant frequencies and formant amplitudes of vowels by 76 speakers.

		i	ɪ	ɛ	æ	ɑ	ɔ	ʊ	u	ʌ	ɝ
Fundamental frequencies (cps)	*M*	136	135	130	127	124	129	137	141	130	133
	W	235	232	223	210	212	216	232	231	221	218
	Ch	272	269	260	251	256	263	276	274	261	261
Formant frequencies (cps)											
F_1	*M*	270	390	530	660	730	570	440	300	640	490
	W	310	430	610	860	850	590	470	370	760	500
	Ch	370	530	690	1010	1030	680	560	430	850	560
F_2	*M*	2290	1990	1840	1720	1090	840	1020	870	1190	1350
	W	2790	2480	2330	2050	1220	920	1160	950	1400	1640
	Ch	3200	2730	2610	2320	1370	1060	1410	1170	1590	1820
F_3	*M*	3010	2550	2480	2410	2440	2410	2240	2240	2390	1690
	W	3310	3070	2990	2850	2810	2710	2680	2670	2780	1960
	Ch	3730	3600	3570	3320	3170	3180	3310	3260	3360	2160
Formant amplitudes (db)	L_1	−4	−3	−2	−1	−1	0	−1	−3	−1	−5
	L_2	−24	−23	−17	−12	−5	−7	−12	−19	−10	−15
	L_3	−28	−27	−24	−22	−28	−34	−34	−43	−27	−20

Table 11: Peterson & Barney (1952) Average fundamental and formant frequencies and formant amplitudes of vowels by 76 speakers (divided into men, women and children)

Fant (1966) measured the formant frequencies of 7 male and 7 female Swedish speakers and analysed them together with measurements from Peterson and Barney

(1952). Based on this data, he claims that female frequencies are an average of 20% higher than those of male speakers. That a larger vocal tract will produce formant frequencies lower than a smaller one may seem obvious, and some linguists, for example, Mol (1963), do assign this as the reason for differences in formant values between men, women and children. However, multiple researchers claim that there is no straightforward correlation between vocal tract size and formant frequencies. For example, Fant, (1966), Mattingly, (1966), Eguchi and Hirsh (1969) and Hillenbrand and colleagues (1995) all claim that the relationship between vocal tract length and formant values cannot be attributed to vocal tract length alone. Fant's (1966: 29) research looks at more detailed proportions of the pharynx and oral cavity. His measurements reveal that the proportional differences between male and female formant values vary according to which vowel is being articulated, but the relationship between women's and children's formant values is consistent regardless of vowel class. Öhman[7], (cited by Fant, 1966) suggests that proportionally larger mouth openings in women may also be a factor contributing to differences in women's formant values (this was also suggested by Bennett (1981) in section 3.2.4.1) but no evidence is provided for this. Mattingly (1966), whose analysis is also based on Peterson and Barney's data, agrees that vocal tract size is responsible for differences in formant frequencies in men, women and children, but claims that proportional distribution of formants across all vowels in the Peterson and Barney data is not consistent. "The separation between male and female distributions for some vowel formants is much sharper than variation in individual vocal tract size can reasonably explain" (Mattingly, 1966: 1219). The differences between these vowels, he claims, can probably be explained as stylistic or related to linguistic convention. Further detail on the changes in vocal tract morphology during maturation, their relationship to formant values and possible stylistic explanations will be discussed below.

Performing acoustic analysis on children is more problematic than the analysis of adult speech, due to a number of factors. The child's growing vocal tract, children's speech habits, and a relative lack of data for the purposes of comparison are among them. However, the advantages of developing detailed acoustic knowledge of children's speech is of benefit to many areas, for example, the adding evidence to support impressionistic data in academic research, supporting appropriate speech and language therapy interventions, and more recently, to support the development of

[7] No date provided by Fant.

speech recognition systems that will reliably recognise children's voices (Vorperian & Kent, 2007). The increase in Automatic Speech Recognition systems (ASRs) in recent years, found in particular in popular home 'smart speakers' such as Amazon's Alexa device, might lead one to believe that manufacturers would have invested in research to support accurate recognition of children's speech. These systems have, however, generally been designed for adult speech and can therefore make errors with children's voices (Ureta et al., 2020). Children's vocal tracts are shorter, their vocal folds are smaller, their speech is more variable, and research in this area is expensive (Chen et al., 2020). This means that research on the acoustic features of children's speech is limited in comparison to research on adults, and consequently there is a lack of reference data for formant frequencies of sounds in different languages and dialects. Capturing comprehensive data on children's acoustics is a challenging task, however. As well as changes to the length of their vocal tracts, their faces are also growing and changing, along with elements of the oropharynx, such as the tonsils and adenoids (Vorperian & Kent, 2007). Velopharyngeal closure also undergoes change during development. These developmental changes have an impact on resonance properties. Furthermore, children's motor skills are developing; some researchers have claimed that a child's vowels may be more variable than adults' (Yang & Fox, 2013), or may over or under "shoot" in terms of duration (Lee et al, 1997). This claim will receive scrutiny below.

A child's vocal tract grows from around 6-8cm in neonates to 15cm in an adult female and 18cm in an adult male (Vorperian & Kent, 2007: 3). The vocal tract increases in length by around 1.5-2cm during the first two years of life, and another 1cm before age 3 (Vorperian et al., 1999). At puberty, hormones cause the male larynx to descend one whole vertebra lower than the female larynx, and their vocal folds are 60% larger (Pisanski et al., 2016).

Whether there is a difference in vocal tract size and morphology between the sexes before puberty has been the subject of debate for decades. The complexity of measuring vocal tracts accurately has limited the research in this area, and has led to using body size as a proxy for x-ray/Magnetic Resonance Imaging (MRI) based vocal tract measurements, as seen for example, in Bennett (1981). Fitch and Giedd (1999: 1517) report that pre-pubescent sex-based differences in formant frequencies are most likely due to issues of style rather than vocal tract morphology. The authors collected vocal tract measurements of 129 people, including 53 females and 76 males from 2.8

years to 25 years old. The males had a mean age of 11.5 years while the females had a mean age of 11.6 years old. Through MRI technology, they were able to establish that children's vocal tracts do not differ significantly between the sexes before puberty. This observation has also been made in relation to the differences between adult male and female formant frequencies, which, it is claimed, cannot be attributed to vocal tract size and shape alone (see for example Fant, 1975; Mattingly, 1966; Sachs et al., 1973). Sachs and colleagues characterise this style as men making themselves sound bigger than they actually are, and women making themselves sound smaller, (1973: 75) (see also Pisanski et al., 2016). The proposal that children might be altering the resonance characteristics of their vocal tract based on identity presents some interesting questions around whether a child's projection of their identity might shift in different social situations. For example, could a child behave in a more masculine manner when talking to his male peers, but shift his largyngeal settings to suit a more child-like identity when in the company of family such as their mother or grandmother? This could present issues of interspeaker variation within different settings.

We saw above that formant frequencies decrease with age, and that female voices have higher formant frequencies than males. This carries through to differences between formant frequencies in girls and boys as young as 4 years old (Vorperian & Kent, 2007). Vorperian and colleagues (2005) investigated the vocal tract measurements of 37 subjects including 25 children and considered whether there is evidence of differences in the size of children's vocal tracts and if so, at what age these changes emerge. They report that contrary to earlier work by King (1952), who claimed sexual dimorphism was present in the pharynx length of 1 year olds, there is no sign of differences in vocal tract measurements between the sexes at the age of 6;9. King's data was based on x-ray measurements of the hard palate to the hyoid bone, but Vorperian and colleagues (2005: 348) claim that the full length of the vocal tract should include the naso-pharynx. A much larger study by Vorperian and colleagues (2009) extends the age ranges being studied up to 19 years old. MRI/CT scan data from 327 males and 278 females was analysed for multiple, detailed measures of the vocal tract. The authors conclude that most variables in vocal tract growth appear to diverge in males and females after age 12, though some variables, such as the measurement of the lips to the posterior pharyngeal wall diverge at age 4. Other measures, such as vocal tract length and posterior cavity length appear to fluctuate slightly before puberty, with some differences being temporary. Vorperian and colleagues (2011) reanalysed the data from Vorperian and colleagues (2009), refining their methodology to account for more granular age differences. In using this approach, they identified significant pre-pubertal differences between the sexes. These emerge in the oral cavity first (what they call the horizontal plane) between 3 and 7 years old, followed by the nasopharynx (the vertical plane) at around 8 years old, which is longer in pre-pubertal females than males. The authors claim that these differences were masked by growth rates in the previous study in 2009. It is important to note that the changes in vocal tract growth vary during development, and differences do not necessarily persist beyond puberty. For example, the longer naso-pharynx in pre-pubertal girls is overtaken by boys' growing pharynxes in puberty. After this age, males have a proportionally larger pharynx to oral cavity than females. Finally, growth does not stop at puberty; changes continue to a lesser degree throughout adulthood, as the skull and associated craniofacial structures continue to grow (see, for example Israel, 1968; 1973; Linville & Rens, 2001; Petrosino, & Squibb, 1991; Rastatter et al.1997; Scukanec et al., 2003).

There are some differences of opinion in the literature around the impact of vocal tract growth on formant values – for example, Fant (1966) claims that F2 values correspond with the length of the pharyngeal cavity, and F3 with oral cavity length, while Lieberman and colleagues (2001) claim that the correspondence is between pharyngeal width rather than length. Martland and colleagues (1996) on the other hand, suggest that F3 relates to pharyngeal cavity length in under 2s.

Vorperian and colleagues (2009) and the subsequent re-analysis of their data by Vorperian and colleagues (2011) remain the largest and most detailed studies of vocal tract measurements to date, and while the 2011 analysis revealed pre-pubertal sexual dimorphism in vocal tract growth, the results were unexpected; in their analysis, girls have a longer naso-pharynx than boys before puberty. Coming back to the data presented by Vorperian and Kent (2007) which shows that pre-pubertal boys have lower formant frequencies than girls, Vorperian et al.'s (2011) vocal tract measurements do not explain this observation. The authors acknowledge this body of work but claim that there are still not enough refined measurements of the oro-pharynx available to be sure whether style is the reason for these pre-pubertal lower frequencies in boys.

3.2.4.4 Articulatory control (intra-speaker variability)

It is obvious that in typically developing children, motor skills develop over time, their articulatory control being perfected as they develop physically and cognitively.

Researchers have made varying claims about the nature and timing of these developments.

Nittrouer (1993: 969) collected data from 10 adults and 30 children aged 3, 5 and 7 (10 children of each age, split equally between boys and girls). The subjects were asked to produce 15 syllables with the structure CV, inside the carrier phrase "It's a _____ Bob". The syllables were made up of combinations of five consonants (voiceless fricatives and plosives with one voiced plosive for comparison) and three vowels. All but one of the syllables was a real word, e.g. 'tea', 'two', 'shoe', elicited by showing the subject pictures. Ten samples of each syllable were analysed from each speaker. The subjects' dialects and geographical origins are not disclosed. Through an acoustic analysis, Nittrouer noticed that children's consonant articulations were typically slower than those of the adults, and also varied temporally more than the adults' articulations. While Nittrouer (1993) reports that children were capable of producing the articulatory gestures required for the sounds under investigation, she notes that these developed at different rates. She reports that F1 appeared to be less variable than F2 within her dataset, which she attributes to "adult-like" jaw movement skills, while F2 continued to vary until at least age 7, reflecting a slower acquisition of tongue movements. Nittrouer's claims regarding the invariability of F1 are supported by other researchers, for example, Green and colleagues (2002), who noted that lip movements are more variable and slower to mature in children compared to jaw movements. However, there have also been reports that although lip movements in children vary more than jaw movements, variability in jaw and lip movements decrease equally as the child develops (Walsh & Smith, 2002).

Lee and colleagues (1997) measured formants, duration and pitch in 10 monophthongs and 5 diphthongs from 536 children aged 5-18 and 56 adults in the United States. Most of the subjects had been born in Missouri or Illinois (Lee et al, 1999), but no further information about dialect is provided. The research was carried out in a laboratory setting, with the children being asked to produce each word within the frame "I say *uh* _____ again" (*uh* represents a schwa inside a word such as 'a', 'the' etc. and was intended to put the vocal tract in a neutral position before the articulation of the target word). Formant values were extracted using an automatic segmentation and formant extraction system. A sample was measured by hand to evaluate the accuracy of the automatic system, and clearly erroneous measurements were discarded. The authors remain cautious that some erroneous measurements may have remained in the dataset

(Lee et al., 1997: 474). 5-year-old children were found to have a statistically significantly longer vowel duration than other age groups. The authors claim that while children of between 5 and 7 years share vowel patterns with adults, they have a tendency to "overshoot" or "undershoot" vowel duration in comparison (Figure 7, 10, 11). They claim that this could be evidence that children may have a larger dynamic range for vowels than adult speakers of the same dialect. They also found that children under 10 years old exhibited wider ranges of spectral variability than adults. The authors attribute this to excessive tongue movements connected to lack of articulatory skill in coarticulation, which corresponds with Nittrouer's claim that F2 varies much more than F1 in children. This variability diminishes to adult levels at around 11-12 years (Lee et al., 1997: 475). Teenagers were found to display less variation, which the authors attribute to fast speech rates.

In a series of laboratory experiments, Eguchi and Hirsh (1969) collected data from 84 children from 3-13 years old. Between five and six children in each age group repeated two test sentences ("I am tall" and ''He has a blue pen") five times. Children aged 3-6 years repeated the utterances after they were produced by one of the researchers, while children of 7 years and above read the same utterances from a card. The youngest children's repetitions were much more variable than the older children, with changes in both F1 and F2. This variability decreased progressively until age 11 at which point they demonstrated much more precision in repetitions of the same vowel in the same sentence (Eguchi & Hirsh, 1969: 257).

In a more recent experiment, similar to Eguchi and Hirsh's (1969), Yang and Fox (2013) collected laboratory speech data from 15 children in Columbus, Ohio (7 girls and 8 boys), and six mothers of the children. The subjects were asked to produce an unspecified number of repetitions of 20 monosyllabic/disyllabic words containing 10 vowels from a word list, after receiving an audio prompt of each word. Their method differed slightly from Eguchi and Hirsh's in that they considered more vowels (10, compared to Eguchi and Hirsh's six), they included adults in their sample, and rather than using a separate method for the younger children, they asked all subjects to repeat a word spoken by the researcher rather than older subjects reading the word from a card. While Eguchi and Hirsh used test sentences, Yang and Fox (2013: 1264) preferred to use a word list. They found that the high back vowels in particular were subject to continuing refinement as the child develops, demonstrating an increase in articulatory precision over time (Figure 7 - 10). Adults also showed variation in production of vowels, though the ellipses (which contain 95% of the articulations) show a tighter distribution in the adults performing the same task (Figure 9). The authors note that /u/ has the most dispersed set of articulations, owing to typical fronting of this vowel found in Ohio dialects.

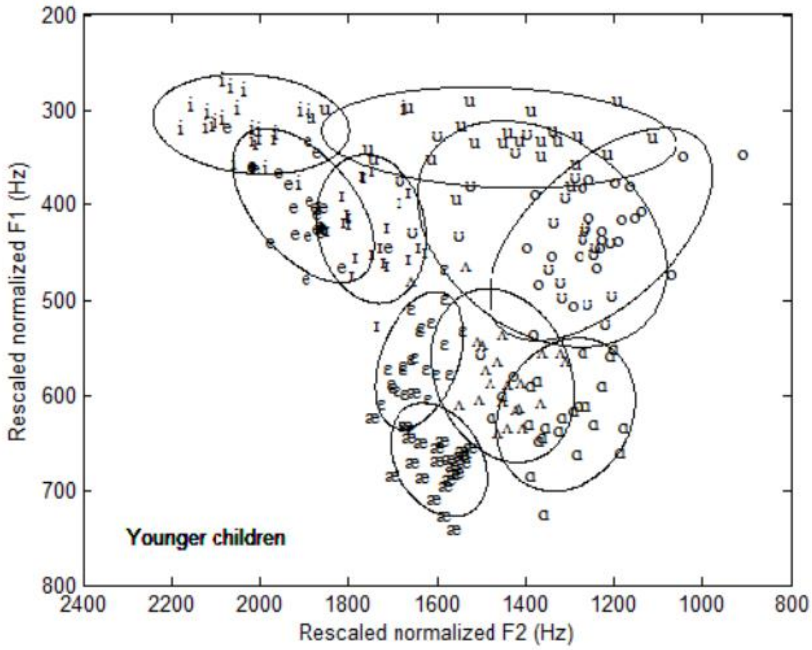


Figure 7: Vowel ellipses in younger children 3-5 years (Yang & Fox, 2013)

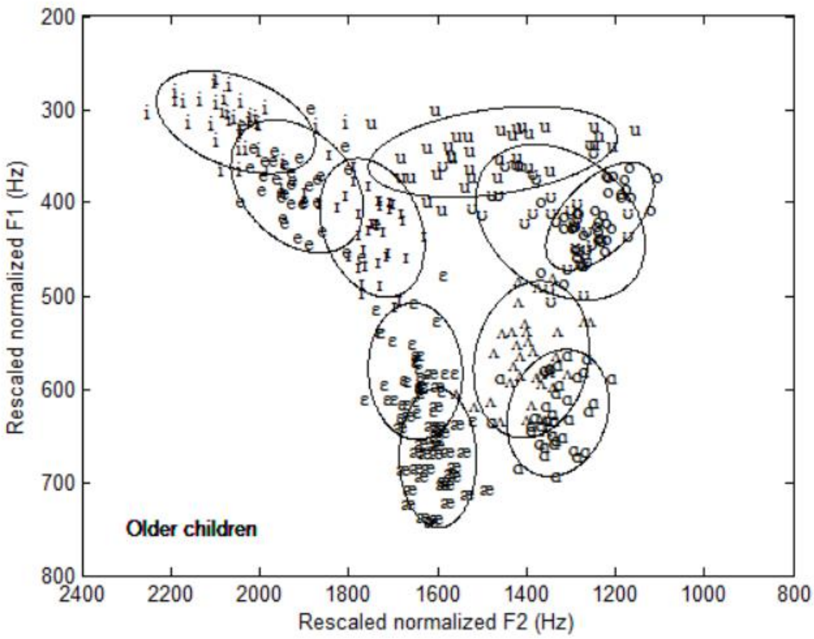


Figure 8: Vowel ellipses in older children 5-7 years (Yang & Fox, 2013)

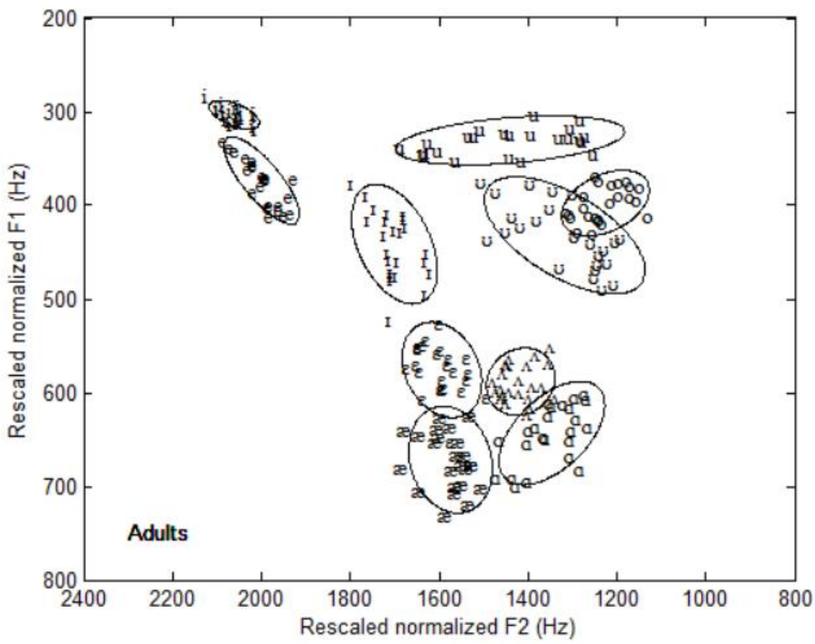


Figure 9: Vowel ellipses in adults (Yang & Fox, 2013)

Younger children exhibited overlap between the acoustic spaces of back vowels, but less so for front vowels. More centralised vowels were apparent in the younger children (see Figure 7). Between the ages of 5-7 the acoustic space becomes more stable (see Figure 8) (Yang and Fox, 2013: 1266). The authors attribute this to the development of motor skills controlling the lips, as the back vowels in English are mostly rounded. They claim that motor development of lip shape is slow and gradual, and that this is borne out by their results. Of course, Yang and Fox's data is based on single style laboratory speech, and yet there is evidence that speakers use the vowel space in quite different ways across a range of speech styles. A description of some key research which demonstrates this, follows.

While researching adult articulation rather than children, Keating and Huffman (1984) reported on the use of the vowel space by seven speakers of Tokyo Japanese. Their intention was to discover whether a language with a small number of vowel contrasts uses only a small portion of the available acoustic vowel space, or whether these languages allow greater phonetic variation of each phoneme. They established that in word list style, phonetic realisations formed discrete clusters within the vowel space, but in reading style, speakers filled the vowel space.

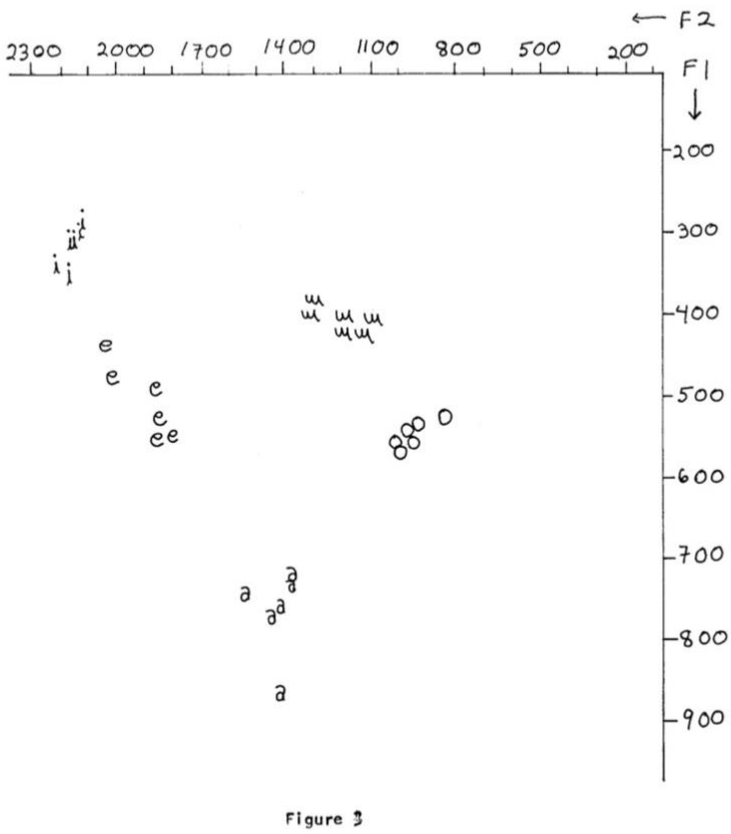


Figure 10: Word list style Japanese vowels (Keating & Huffman 1984)

Figure 10 shows the 5 vowels of a Japanese speaker's reading of a word list. The tokens are tightly clustered and occupy discrete acoustic spaces. Figure 11 shows the same speaker's much larger, overlapping space occupied by tokens from a reading passage (called "prose texts" by the authors).

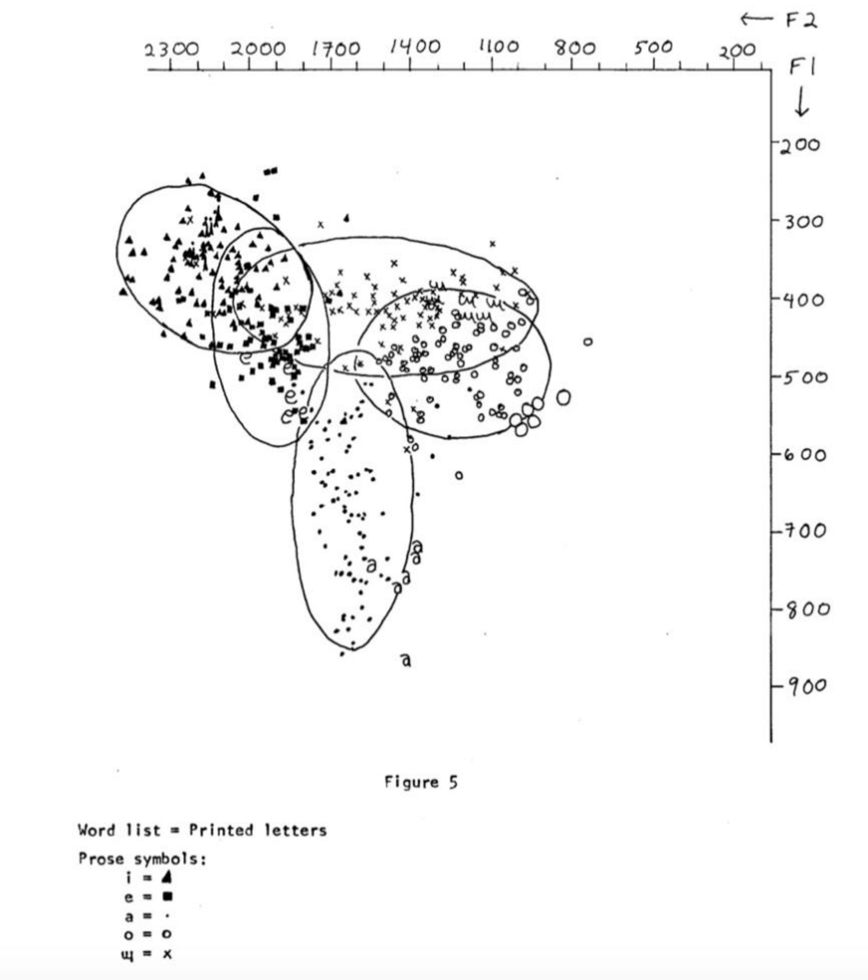


Figure 11: Reading style Japanese vowels (Keating & Huffman 1984)

Veatch (1991) makes an even bolder claim based on his analysis of the range of realisations of a single vowel in an individual speaker. He analysed vowel formants from three speakers of different US dialects, Alabama, Chicago and Los Angeles, and a Jamaican Creole speaker, all collected by sociolinguistic interview. If all tokens of a vowel are analysed in naturalistic speech, including stressed and unstressed vowels, Veatch claims, realisations of a *single* vowel phoneme can not only overlap the spaces of other vowels but can *fill* the vowel space. The most extreme example of this is of realisations of the vowel /ɛ/ in a single speaker of white Chicago English, Rita (Figure 12).

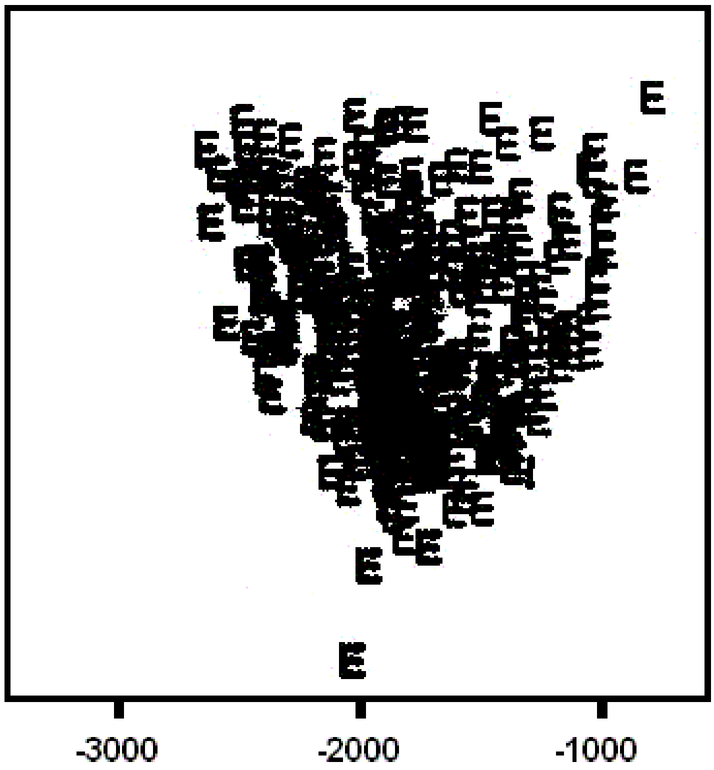


Figure 12: Rita's /ɛ/ in naturalistic speech. Veatch (1991: 202)

Although the evidence from Keating and Huffman (1984) and Veatch (1991) comes from adults, we can expect that children will also exhibit variability in formant values depending on the task/context, though the extent of this is unknown. We saw in Local (1983) in section 2.2.2 that children's speech styles can be even more varied than those of adults, due to singing, whingeing, pulling faces etc., and yet all of the studies we have looked at so far in this chapter have come from laboratory speech, where, for example, children are asked to produce a word carefully, within a given carrier phrase (though see McGowan et al. (2014) for a hybrid method which includes conversational speech augmented with word lists and repetitions of words spoken by a researcher). There is of course a sensible motivation to perform acoustic analysis away from the home setting, and particularly on laboratory speech. Naturalistic speech is messy; there are overlaps and background noise which means that some tokens need to be disregarded, and there may be few tokens of the features to be studied present in the data. Moreover, the acoustic analysis of children's vowels is already more challenging then for adult speech, as described above. However, in making the decision to collect data in a laboratory setting, we must accept that it will not be representative of the variation found in truly naturalistic speech.

3.2.4.5 Amplitude

Naturalistic data collection can include all kinds of variation in a child's voice, which can cause problems for acoustic analysis. Non-modal voice qualities can, for example, make interpretations of formant values more difficult due to the introduction of noise into the spectrogram (Kent, 1976: 422). This appears to have received little attention in the literature, but the effects of amplitude have received some scrutiny. Huge variations in amplitude can of course be common in children's language in naturalistic settings. Local (1983) describes a range of vocal styles a child used in recordings made for an impressionistic analysis (see section 2.3 above). Children may cry, shout, whine or whisper among other speech styles. The vast majority of research on the acoustic effects of increased amplitude on speech has been confined to adults, however, there have been a few studies which have included children in their dataset.

Huber et al. (1999) investigated the impact of vocal intensity on formant frequencies in children between the ages of four and eighteen and adults, collecting data from children of ages 4, 6, 8, 10, 12, 14, 16 and 18. 10 males and 10 females were included in each age group. Subjects were asked to produce the vowel /ɑ/ at three different intensities, classified as "comfortable", "high" (10dB above "comfortable") and "low" (5dB below "comfortable"). Younger children were shown a bar representing vocal intensity with a sticker placed at the target level for each production to guide them on how loudly to produce the sound (Huber et al., 1999:1534). /ɑ/ was chosen, as, according to the researchers, it is easily replicable, and because it has a high F1, making it less likely to be confused with the fundamental frequency on the Linear Predictive Coding (LPC) analysis. The authors hypothesised that F1 and F2 would increase with increased vocal intensity. They predicted that increased vocal intensity would increase F1 values as vocal intensity is often accompanied by a more open jaw position (Schulman, 1989), which in turn affects tongue height and F1 values (Fant, 1971). Kent and Read (2002) claim that a high F1 leads to increases in the frequencies of all other formants due to the higher amplitude of the tail of F1. Huber and colleagues therefore hypothesised that F2 would also raise with amplitude. Their results confirmed all hypotheses with the exception of the raising of F2 as a consequence of a higher F1. While F1 did increase with amplitude, the authors reported no significant effect on the frequency of F2 (Huber et al., 1999: 1539). The fundamental frequency was also found to increase with vocal effort. The difference between the "comfortable" and "high" levels of intensity and "high" and "low" levels were significant, but the difference between "comfortable" and "low" intensity was not significant (Huber et al., 1999:

1538). Liénard and Di Benedetto (1999) reported similar results in adults. They also identified a strong correlation between vocal effort and f0 and F1, but they too found no significant statistical effect on the relationship between amplitude and F2 or F3.

Traunmüller and Eriksson (2000) included 7-year-old boys and girls in their experiment on the effect of vocal intensity on formants. Unlike Huber et al., they used increasing physical distances between interlocutors as a means of increasing vocal intensity, rather than asking subjects to produce a vocalisation while monitoring a bar representing intensity for feedback. As in Huber et al.'s (1999) study, they found that f0 and F1 were affected by amplitude. They also reported an effect of increased vowel length correlating with increased vocal intensity.

More recent research on the acoustic correlates of higher amplitude speech in adults has looked at individual differences between speakers. Koenig and Fuchs (2019) concluded that the effects of increased vocal effort on F1 vary according to speaker and open vowels are more affected than close front vowels. Subjects performed three tasks designed to elicit different speech styles, communicating through a glass window in order to elicit louder than normal speech. Vocal intensity was therefore elicited in a naturalistic way rather than the more precise technique used by Huber et al., which forced all subjects to produce the same vocal intensity. Koenig and Fuchs found that there were extreme differences between the vocal intensity of the speakers, which determined the inter-speaker variation uncovered by the authors. The authors make the point that speakers use multiple strategies to increase the loudness of their speech These vary both according to the task and to the vowel being articulated (Koenig & Fuchs, 2019: 1293). While this research was performed on adults, there is no reason to believe that children will behave differently.

3.2.4.6 Additional challenges in the analysis of children's speech

Above, we saw how children's speech style can affect the acoustic signal available for analysis. Here we discuss other aspects of infants' speech which can cause problems for acoustic analysis. The following section will explore the impact of resonances (and their absence) found in children's speech which can present challenges for the researcher.

In section 3.2.4.3, we learned about the impact of the growth of the vocal tract on formant values, in particular, F1 and F2. A smaller larynx and shorter vocal folds leads to children typically having higher pitched voices than adults. Infants, having the smallest vocal tracts, have the highest fundamental frequency (f0) of around 400 Hz. This is around three to four times higher than an adult male, (Kent and Read, 2002: 196), whose average pitch is around 120 Hz. This high fundamental frequency results in widely spaced harmonics, which can make it difficult to read formants (Huggins, 1980). Widely spaced harmonics can result in interference with individual formant peaks rather than the more general influence of closely spaced harmonics (White, 1999; Story and Bunton, 2015). We will see some examples of how these situations have been resolved in this study in Chapter 4. High f0 values can also have other effects. Although their average can be 300-400 Hz, depending on speaker, they can range from the adult male range up to 1000 Hz, making it difficult to identify (Kent and Read, 2002); it may sometimes be mistakenly analysed as a formant, and can result in formant estimation errors, (Lindblom, 1962). The higher the fundamental frequency, the less likely it is that harmonics will coincide with a formant peak. Strong harmonics may therefore be misinterpreted as formants, (Kent, 1976: 422). Lindblom (1962) claims that the lower the fundamental frequency, the more accurate the spectral envelope, while Vorperian and Kent (2007) agree that more closely spaced harmonics result in better definition of the peaks of the vowel spectrum.

Adding further complexity, Kent and Read (2002: 197) point out that in some cases, researchers have observed “harmonic doubling” in children’s speech, resulting in the appearance (and disappearance) of harmonics at half of the frequency of f0. They also note that unusual phonatory features, such as biphonation, can result in a double series of f0 and vocal tremor. This also contributes to the complexity of interpreting formants. In experimental settings, shifting pitch by singing has been used as a technique to expose formant measurements (White, 1999), but in naturalistic speech, it can be more challenging to reliably determine the formants accurately.

In addition to the interference to formants seen above, nasality can also produce unwanted effects on a spectrogram. A child’s developing motor control includes their control over the velum; a typical infant shifts from almost consistently nasalised cries in its early months to an ability to close the velo-pharnygeal port during oral sounds by the age of 7-9 months (Vorperian & Kent, 2007). Nevertheless, Kent (1976: 422) claims that children often exhibit “inappropriate” nasalisation, which can cause problems in the identification of formant frequencies due to the introduction of "unexpected resonances and anti-resonances". Di Paolo and colleagues (2011: 94) explain that the combination of amplitude and harmonics which don't correspond with formants can lead to an additional pole appearing in the spectrogram.

3.2.5 Extracting and interpreting acoustic data in children

Given the complexities described above, achieving accurate formant measurements from children's data may be difficult, but accuracy is of paramount importance, as incorrect measurements can be mistaken for articulatory imprecision (Vorperian & Kent, 2007).

Commonly used techniques apparent in the literature include spectrographic analysis incorporating automatic formant tracking, and/or Linear Predictive Coding (LPC). These techniques extract formant values at a particular point in a vowel. While formant frequencies are usually taken from the mid-point of the vowel, spectral changes can be evident in the speech related to dialect, generation, and importantly, motor control. Yang and Fox (2013: 1263) point out that a slight movement of the articulators can affect resonance, and consequently, formant frequencies. The development of motor control, therefore, has an impact on formant frequencies. The precise point at which formant frequencies are measured is crucial to the results; these decisions should therefore be made carefully and on an individual basis, using visual and auditory cues to determine a stable portion of the vowel, free from coarticulatory effects wherever possible (see also Khattab & Roberts, 2011). As described above, widely spaced harmonics due to the child's high f0 can make formant values difficult to read (Kent, 1976; Story & Bunton, 2015). Huber and colleagues (1999: 1535) used Linear Predictive Coding (LPC) to establish formant frequencies, but noted that F2 was often missed, and consequently, following formant frequencies were incorrectly attributed to a lower formant. In these cases, the LPC was cross referenced with the wide band spectrogram and formant values reported in the literature to help locate the second formant. Therefore, as with the careful consideration of which point in the vowel the frequencies should be extracted from, LPC values should also be compared to the spectrogram, to ensure that the correct F2 values have been extracted. Khattab and Roberts (2011: 172) suggest that the researcher's first choice should be to extract formant values from the spectrogram using the burg method, which should then be checked visually against the LPC autocorrelation and FFT graph to ensure that the correct formants have been identified. Vorperian and Kent (2007: 15) suggest that formant frequencies beyond F1/F2 should be reviewed. F3 values can help to determine the accurate placement of F2 and eliminate anomalous readings. In circumstances where the acoustic measurements produce outliers, it is also recommended to cross check with an auditory analysis (Di Paolo et al., 2011; Watt et al., 2011). As an additional measure, Khattab and Roberts (2011: 170) recommend that Praat may be set to display up to 8 kHz rather than the standard 5 kHz to account for the higher frequencies in children's speech, and that if only looking at the first two formants, reducing the number of formants displayed from five to three may help formant tracker to avoid picking up harmonics instead of formants.

3.2.6 Summary

The vast majority of acoustic research has been done on adults, but there are some examples of research on children, or of children being included in larger datasets. The lack of existing reference data available can cause problems for researchers, as coverage of different dialects is patchy, with little data for UK accents available. Vorperian and Kent's excellent 2007 review of the existing body of data available for children is useful, but its broad coverage of children's dialects across three continents means that their corner vowel values may lack precision for comparison against local dialects. In spite of large-scale work on vocal tract growth and differences between the sexes, it is still unknown whether lower frequency formants in boys compared to girls is down to physical differences alone, or whether style plays a role.

We should expect there to be intra-speaker variability in children's articulations, as is seen in adult speech. In lab-based experiments, variability in children's articulations of a single vowel appear to become more precise over time (Yang & Fox, 2013), but if we compare to studies of adult speech, we see that where different speech styles are analysed, the more informal the speech style, the more likely adults are to produce the same vowel in overlap with the acoustic space used for other vowels (Keating & Huffman, 1984). In the most extreme cases, in truly naturalistic speech, a speaker's articulations of a single vowel may fill the entire vowel space (Veatch, 1991). Evidence of this phenomenon in children's speech is absent from the literature but should be considered as possible.

The extraction of formant values from children's speech can be problematic, particularly where naturalistic data collection is concerned. Their varied speech styles may contain huge variations in amplitude, which can have the effect of raising F1, and temporary vocal settings such as nasalisation may appear in unexpected places, which can result in incorrect formant tracking. The higher pitch of children's speech can also cause issues; widely spaced harmonics can be mistaken for formants more easily than in adult speech, where their narrow spacing tends to have a less noticeable effect.

In spite of the challenges of working on children's data, the inclusion of acoustic analysis can help to reassure readers that the subjectivity which may be present in impressionistic transcription may be minimised. While the complexities described above can make the analysis of children's formants difficult, and perhaps time-consuming, careful methods can help to offset these issues. Careful selection of a stable portion of the vowel, not relying on automatic formant trackers alone, and employing visual inspection of multiple spectrographic displays such as the LPC are essential to ensure the accuracy of readings.

3.3 Summary of methodologies literature

In this chapter, we started off by considering diary studies, the earliest recorded method for analysing children's speech. Following on from the introduction of large sample cross-sectional studies in the 1930s, and later, small group naturalistic studies or experimental studies from the 1960s onwards, diary studies have received some criticism for being unsystematic and/or subjective (Ingram, 1989; Irwin, 1941). In some cases, these criticisms are justified. Some early examples did lack systematicity, and, in some cases, language development was treated inadequately, as an add-on in a study of more generalised development. Some of the researchers who carried out this work were not primarily linguists, for example, alphabetic systems were sometimes used in place of phonetic transcription. However, there have been diary studies which have stood the test of time and have provided data for later re-analysis by other researchers (Demuth, 2014). The development of audio recording technology has of course added to the reliability of diary study data. Those early diary studies of a single child or small number of children were so-called as they were reliant on written records, whereas the data in more recent studies of that nature are now primarily audio or video recorded and may be more accurately described as case studies. With careful consideration of sampling and methods of analysis, the study of a single child can provide a unique opportunity to access longitudinal data in a naturalistic setting. The emergence of the small group study method recognises the value of a detailed analysis of a small number of children above small amounts of data from a large group.

A number of children fulfilling exactly the same criteria, however, is not always available to make a small group study possible, and the unique benefit of a case study is that the linguist has access to a child (usually their own) across a range of different settings. Even when the recording equipment is switched off, they are able to observe a child and take note of linguistic behaviour. Such unrestricted access to a number of children, who fit the profile that the linguist seeks, is rare. Case studies have often provided the ideas for larger scale studies to investigate in other populations, so they can go on to influence our understanding of the way that language functions, even if in the first instance, those observations are made of a single child (see for example, Waterson (1971) later developed by Menn, 1983).

The quality of case study data can be further assured through inter-rater reliability checks (see section 4.3.2 below) and acoustic analysis techniques which support the impressionistic analysis by comparing acoustic correlates against reference data, or through plotting vowels on to the acoustic space. Unfortunately, there is not enough data from comparable UK dialects available for comparison to provide a reliable benchmark against which to measure vowels of different dialects. However, Yang and Fox's (2013) evidence of children's vowels being articulated differently each time, even in a laboratory task, gives us pause for thought. They noted that the children's articulations became more precise over time, as they developed their motor skills, but we must remember that this is *elicited lab-speech* – requiring children and young adults to insert a chosen lexical item into a framing sentence, prompting a careful speech style. In contrast, the data presented in the following chapters is truly naturalistic. The child is recorded at home, at different times of the day, going about his normal daily activities. It appears that a formant analysis of naturalistic vowels in children is currently absent from the literature, thus this study provides an opportunity to learn more about this under-studied area.

In line with the work on adults done by Keating and Huffman, (1984) and Veatch (1991), we might expect to see even more variability than is reported in Yang and Fox's study. In which case, the most important acoustic work will be to confirm whether the acoustic values correspond with where the impressionistic analysis places a vowel in the vowel space, to see how a child's vowels vary within their own vowel space, and to analyse how that variation changes over time. Further explanation of this process follows in Chapter 4.

Chapter 4 Method

4.1 The child and his environment

4.1.1 The home environment

Henry[8] is a boy born in 2010, who lives with his parents in a rural location in North Yorkshire, UK. Neither of Henry's parents speak with local accents. Both of Henry's parents work full time, so he has typically spent around 8-9 hours each day away from home from the age of 2 years. Until starting school, he spent one day each week with his Scottish grandmother.

4.1.2 Nursery

Henry started at a nursery (pre-school/kindergarten) in York at the age of 2;0, initially for four full days per week. All staff at the nursery spoke with an accent local to the York area. Henry attended the nursery for two years and two months. The nursery was small, and Henry mainly played with two other boys, who were both from the north of England. One child's parents were from the local area and the other boy's were from Middlesbrough, approximately fifty miles north east of York.

4.1.3 School

Henry started school at 4;4 at a village school in rural North Yorkshire. The school was small, and he was in a smaller than average class of around 20 children. Most children in Henry's class spoke with a local accent, though one child arrived at the end of the reception year (the first school year – in the UK, children usually start school in the September following their fourth birthday) from Brighton (Sussex, southern England), and another child had parents who spoke Southern regional varieties. The UK school year runs from the start of September to the middle of July, and is split into three terms, September to December, January to Easter (as the dates of Easter vary between 22nd March and 25th April, so do the dates of the transition between terms), and Easter to July. In the first term of the reception year, the class was taught by a female teacher who spoke with a Southern Standard British English (SSBE) accent, and two female teaching assistants speaking varieties local to North Yorkshire. In term two, another female teacher joined the class having returned from parental leave. The first teacher

[8] All names given in this thesis are pseudonyms.

remained in the class for two days per week and the second, who spoke a local variety, taught on the other three days. The classroom assistants remained as before. In years one and two, the school years following the reception year, Henry's teacher (again, female) spoke with a local accent. Therefore, the accents Henry heard at school were largely northern, and most were specifically from the local market town and surrounding villages.

4.1.4 Integration

Henry took a long time to settle into the class. All but two children in the class had already been to nursery together – a different setting from the one that Henry had attended in York – so friendships had already been formed. Henry is a very cheerful child and said that he enjoyed school, but he did not make close alliances easily. He made two close friends, however, James and George. Both children have what I would characterise as particularly strong local accents and used monophthongal GOAT and FACE vowels (typical monophthong and diphthong variants of these lexical sets are provided in the following section).

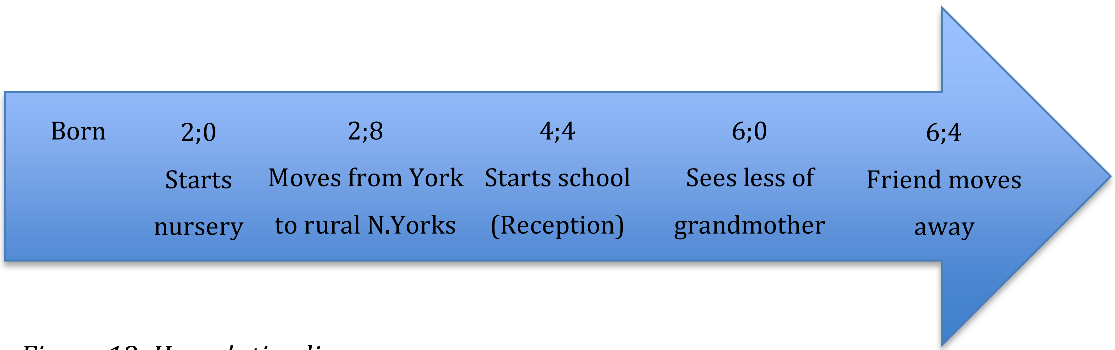


Figure 13: Henry's timeline

4.1.5 Accents in the child's environment

As outlined above, Henry's mother (me) was born in the south of England, to Scottish parents. As a child I lived in Buckinghamshire, Oxfordshire, and the Middle East until moving to the North of England, within 15 miles of York, UK at the age of 15. My accent consists mostly of features of SSBE, though my articulations of the STRUT vowel can approach [ə], and BATH is very occasionally articulated as [a] (see Wells (1982) for an explanation of lexical sets).

Henry's father was born in London but moved to a village in Kent at the age of 2 and lived there until he moved to York at the age of 18 to attend university. His accent exhibits some southern regional features, but realisations of the lexical sets under investigation here are largely consistent between both parents.

Henry's grandmother was born in Glasgow and moved to the south of England aged 21. Her accent is still noticeably Scottish, but her Scottish accent features have (by her own account and that of others who know her) diminished over the five decades that she has been in England. In terms of the lexical sets under examination here, only the realisation of START differs noticeably from that used by Henry's parents, where some *r*-colouring is present.

4.1.5.1 Accents in the home

Mother

As outlined above, I have Scottish parents and moved around a lot as a child, including three years overseas where I attended an international school, though I spent most of my years in the south east of England until I was 15, when I moved to northern England. At this point, I had no regional accent, and spoke a version of SSBE. Although the move north came at 15 years, it has had an impact on my accent. I recall trying to fit in when I arrived, and adopted some local accent features, though these were, in hindsight, not acquired, but imitated. As the initial desire to fit in reduced, so did my attempts to sound like I was from the area I had moved to. However, I have now spent over thirty years in Yorkshire, and my accent now does exhibit slight differences from SSBE, particularly in the STRUT vowel, as will be seen below. In addition to the move north, my own home environment as a child contained more than one accent. My parents were both from Glasgow but had moved to England as a young married couple in their early 20s. My father's dialect included more Glasgow vernacular elements than my mother's, but both spoke with noticeable Scottish accents. My mother's accent will be described below. As discussed in section 2.3.2 above, my eldest brother, who is 10 years older than me, was born in Scotland but left at 6 weeks old. In spite of not growing up in Scotland he grew up speaking with a Scottish accent, though he developed a regional London accent after leaving home to go to university at 18. The Scottish accent is now reserved for speaking to other Scots. My other older brother is five years older than me. He was born in Buckinghamshire, as I was, and grew up speaking Southern Standard British English. His accent also developed some regional features after leaving home as he settled in the south east of England. Looking back on the dialectal environment I grew up in, I attribute some idiosyncrasies in my own variety to this multi-dialectal environment, probably both the home environment and the move to the north while I was still quite young. Most particularly this is exhibited in some members of the BATH/PALM lexical set. For example, BATH, would be pronounced as /baθ/ in my parents' varieties, and PALM as /pam/. In SSBE and other southern varieties I was exposed to, BATH would be /bɑθ/ and PALM, /pɑm/. The differences between BATH and PALM are, however, complex. Wells describes a sound change from /a/ to /ɑ/ in R.P.'s history. The vowel changed in words before /nt#/, so words such as 'plant', 'can't' and 'slant' belong to the BATH set. However, the sound change stopped before reaching words such as 'rant' and 'pant' (Wells, 1982: 100), which belong to TRAP. Most speakers will have no trouble distinguishing between these, however, for children growing up in multidialectal environments, acquiring these exceptional lexical items may be problematical, as seen for example in the research of Payne (1980) and Roberts (1997) (discussed in section 2.3). As a child I was exposed to different systems – one in which TRAP, BATH and PALM were distributed in a particular way – mostly outside of the home, and one in which they were all pronounced the same – inside the home. To take 'rant' and 'pant' as examples, I would only ever have heard these as /rant/ and /pant/, both inside and outside of the home. However, I would have heard 'plant' and analogous examples as both /plant/ (at home, mostly with the exception of one brother) and /plɑnt/ outside of the home, for example at school. The relatively high frequency word 'pant' was unproblematic, but I pronounced the lower frequency 'rant' as /rɑnt/ until it was pointed out to me (by a linguist) as an adult. As I have lived in the north since my teenage years, many speakers who I have associated with may not have noticed (or at least, not mentioned) my idiosyncratic pronunciation, which meant that it continued until I was in my thirties. This example is not totally isolated. Other words in the BATH set continue to cause me problems, for example, I tend to pronounce 'stance' with a TRAP vowel while I pronounce 'France' and other BATH words as /ɑ/.

Father

Henry's father had a more stable linguistic background. His parents were both born in London and lived most of their lives there until adulthood. They both have regional London accents. The family moved from Lewisham in southeast London to a large village 25 miles away in Kent when their children were small and still live in the same house today. Henry's father moved to York at age 18. He has retained his southern regional accent.

Grandmother

Henry's Grandmother was born in Park House, Glasgow, Scotland, the youngest of four children. Her father was from Glasgow, and her mother was from Limerick in Ireland. She lived there until she married my father (from Clydebank, a town within the Greater Glasgow area), and moved to Buckinghamshire, aged 21. Aside from three years in the Middle East in her early forties, and a few years in Oxfordshire in her thirties, she lived in Buckinghamshire until moving to the East Riding of Yorkshire in her mid-forties. As a child, she reports that she was aware of her accent and did not sound like either her siblings or her peers. She won a high school diction competition when she was 11 years old, which involved reading aloud. She remains very aware of her accent and has a tendency to move her vowels towards SSBE when asked to read a word list and exhibits more Scottish vowels when she is not aware of being observed.

Wells' lexical set	**Laver's keyword**	**Mother**	**Father**	**Grandmother**
KIT	bid	ɪ	ɪ	ɪ
DRESS	bed	ɛ	ɛ	ɛ ~ e
TRAP	sam	a	æ	a ~ aː
LOT	cot	ɒ	ɒ	ɒ
STRUT	mud	ʌ ~ ə	ʌ	ʌ
FOOT	pull	ʊ̟	ʊ̟	ʉ ~ uʊ
BATH		ɑ	ɑ	a
NURSE	bird	ɜ	ɜ	ɝ
NURSE	word	ɜ	ɜ	ɝ
NURSE	heard	ɜ	ɜ	ɚ
FLEECE	bead	i	i	i
FACE	bay	eɪ	eɪ	eɪ
PALM	psalm	ɑ	ɑ	ɑ ~ a
THOUGHT	caught	ɔ	ɔʊ	ɒ
GOAT	go	əʊ	əʊ	əʊ
GOOSE	pool	ʉ	ʉ	ʉ ~ uʊ
PRICE	side	aɪ	aɪ	aɪ
PRICE	sighed	aɪ	aɪ	ae
CHOICE	boy	ɔɪ	ɔɪ	ɔɪ
MOUTH	cow	aʊ	aʊ	aʊ

NEAR	beer	ɪə ~ iə	ɪə	ɪɚ
SQUARE	bare	eə	eə	eɚ
START		ɑ	ɑ	ɐ˞
NORTH		ɔ	ɔ	ɔ˞
CURE	poor	ʊə ~ ɔ	ʊə	ʊɚ

Table 12: Home accents, compared (Based on Wells' Lexical sets (1982) combined with Laver's key words (1994))[9]

The data in Table 12 was transcribed from recordings of a word list and augmented by listening to conversation and taking transcribed notes in an attempt to capture variability. This was easily done due to the high contact with these family members in informal circumstances. In order to record my own usage, I transcribed my own word list read-though and added my own reflections on the variability of my pronunciations. These have been informed by listening to recordings of myself in conversation with Henry throughout the duration of this project.

Notable differences between Henry's mother (me) and his father are that Henry's father differentiates between THOUGHT and NORTH while I do not, and his father has a more consistent realisation of STRUT and NEAR than I do.

Henry's grandmother exhibits rhoticity in the expected places for Scottish English, though rhoticity is slight. 'Heard' differs from NURSE. The realisations of the vowel in 'pull' and 'pool' are affected by l-vocalisation. The vowels in 'side' and 'sighed' are affected by the Scottish Vowel Length Rule (Aitken, 1984). The BATH vowel is the same as TRAP, though TRAP shows some variation, being long in 'Sam'. PALM and psalm were differentiated, with PALM being articulated with /a/ and psalm being /ɑ/. In separate word list recording made five years previously, PALM was articulated as /ɑ/. I attribute some of these changes to inconsistent style-shifting. As mentioned above, Henry's grandmother is very aware of her accent and sometimes minimises her Scottish accent features when she is being recorded. THOUGHT is produced with the same vowel as LOT as would be expected for Scottish varieties.

[9] Length markers ':' are used here and in the analysis to indicate contrastive length distinctions here such as the difference between the vowels in TRAP and PALM in the local accent, or noteworthy extra-long phonetic realisations.

The accents described above will be known as the 'home' accents. Aspects of accents found outside of the home, in the local environment, are described below.

4.1.5.2 Accents outside of the home

From the age of 2 years, Henry attended a local childminder for four days each week. The child minder was from York, and she employed other assistants who were also from York. All spoke with local accents. One worker in particular had a very strong local accent. At age 4 Henry started school at a community primary school (approximately 150 children on roll) in a village around 20 miles north east of York.

<table>
<tr><th>Wells’ lexical set</th><th>Mother</th><th>North Yorkshire</th></tr>
<tr><td>KIT</td><td>ɪ</td><td>ɪ</td></tr>
<tr><td>DRESS</td><td>ɛ</td><td>ɛ</td></tr>
<tr><td>TRAP</td><td>a</td><td rowspan="2">a</td></tr>
<tr><td>BATH</td><td rowspan="3">ɑ</td></tr>
<tr><td>PALM</td><td rowspan="2">aː</td></tr>
<tr><td>START</td></tr>
<tr><td>LOT</td><td>ɒ</td><td>ɒ</td></tr>
<tr><td>STRUT</td><td>ʌ ~ ə</td><td rowspan="2">ʊ</td></tr>
<tr><td>FOOT</td><td>ʊ</td></tr>
<tr><td>NURSE</td><td>ɜ</td><td>ɜ</td></tr>
<tr><td>FLEECE</td><td>i</td><td>i</td></tr>
<tr><td>FACE</td><td>eɪ</td><td>eɪ ~ e</td></tr>
<tr><td>GOAT</td><td>əʊ</td><td>əʊ ~ o</td></tr>
<tr><td>THOUGHT</td><td rowspan="2">ɔ</td><td rowspan="2">ɔ</td></tr>
<tr><td>NORTH</td></tr>
<tr><td>GOOSE</td><td>ʉ</td><td>u</td></tr>
<tr><td>PRICE</td><td>aɪ</td><td>aɪ</td></tr>
<tr><td>CHOICE</td><td>ɔɪ</td><td>ɔɪ</td></tr>
<tr><td>MOUTH</td><td>aʊ</td><td>aʊ</td></tr>
<tr><td>NEAR</td><td>ɪə ~ iə</td><td>ɪə</td></tr>
<tr><td>SQUARE</td><td>eə</td><td>eə</td></tr>
<tr><td>CURE</td><td>ʊə ~ ɔ</td><td>ʊə</td></tr>
</table>

Table 13: Comparison of Henry's mother with a typical representation of the local variety

Wells (1982: 350) outlines the main features of northern English accents. He places York in the “middle North”. Table 13 lays out typical realisations of one of the varieties Henry was exposed to at home, alongside a typical representation of a local variety. Further details on the local accents are explained below.

Broadly, Wells describes northern accents as having a system of five short strong vowels; what he calls a “five term system” as opposed to the “six term system” found in Received Pronunciation (RP). This is due to the lack of a STRUT/FOOT split – Wells describes this as a failure of northern accents to make this split into two phonemes during the Middle English period (1982: 351). He claims that all five tend to be produced with a more open articulation than in RP (Wells, 1982: 356). The TRAP vowel is realised fully open, unlike RP's more typical [æ]. Wells is however, writing in 1982, and the RP TRAP vowel has been moving to a more open articulation in the intervening years (Cruttenden, 2014; Upton, 2004; Wells, 2001). Consequently, Cruttenden (2014) began using /a/ rather than /æ/ to symbolise this vowel in 2014.

The long open vowel found in START and PALM is typically a front [aː], sharing the quality of the TRAP vowel. This vowel quality tends to be found in the Middle North (Wells, 1982: 360).

STRUT also varies in its articulation, varying from a complete alignment with the FOOT vowel, to speakers who observe a split. However, Wells notes that the lack of split extends further up the social scale the further north you go (1982: 352).
He points out that the STRUT/FOOT vowel may be unrounded in some northern near RP speakers, for example a mid-unrounded vowel [ə].

While FACE words in RP are realised as a diphthong, in northern varieties, they are often articulated as a monophthong in the vicinity of cardinal vowel 2, though he claims that the influence of the RP diphthong may be found, particularly in urban northern accents (Wells, 1982:357).

Like FACE, GOAT may also be realised as the more traditional monophthong, typically somewhere around cardinal vowel 7, [o], or diphthongs of [oʊ] or [əʊ] may also be found (Wells, 1982: 358).

The THOUGHT vowel may be articulated as an open back vowel [ɑ] in the Middle North (Wells, 1982: 360). This can sometimes be heard in the pronunciation of the name of the city of York as [jɑk] by some speakers.

Of course, there is a range of sociolinguistic variation that will be found in northern accents, and of particular interest here, in the STRUT and BATH vowels. BATH vowels are typically short, i.e. the same vowel as TRAP, /a/. TRAP and BATH vowels are pronounced the same "much further up the social scale" than "unsplit /ʊ/" (Wells, 1982: 354). Wells claims that some northerners would find the pronunciation of STRUT as /ʊ/ "vulgar" but consider the pronunciation of BATH with the short /a/ as an inherent part of their northern identity.

4.2 Data collection

Data was collected in the form of video and audio recordings of Henry from the age of 1;01. Early recordings were all video recorded and were very short, though frequent, with many containing only babble or no language data at all. The mean recorded video length up to this time was only 89 seconds. The first recording containing a recognisable word (*wool*) was at 1;03. Audio recordings began at 1;09. Between that time and the age of 6;06, 127 recordings were made.

Early recordings (to the age of 3;07) were recorded in .mp3 format (64kbps), because they were initially collected for memories rather than specifically for linguistic analysis. Once the research purpose of the materials became clear, later recordings were made in .wav format to ensure better quality (44.1 kHz 16 bit). Up to age 4;08, recordings were made on an iPhone 4, and later using an iPhone 6, both using NCH Software's 'Wavepad' application. Video files were recorded on the same devices in .mov format. A range of types of recording equipment and formats were tested by De Decker and Nycz (2011) with a view to determining whether recordings made on these devices are suitable for acoustic analysis. They tested the consistency of F1 and F2 values in recordings of the same speakers recorded on 4 different digital devices. These were a Roland Edirol R-09 wav recorder (44.1 kHz, 24bit), an Apple iPhone (first generation) recording in lossless m4a through a voice recorder application, 'Voice Memo', a Macbook Pro running Praat 5.1 (recording in WAV), and a Mino Flip video camera (MPEG 4 AVI converted to AIFF). The Mino recording was uploaded to YouTube, downloaded again and converted to WAV. The authors' discussion was around the effects of compression on the Mino recording, and the efficacy of the built-in microphones on the phone and in the laptop. Unfortunately there is no consideration of variance between different file formats on the same device. Predictably, the audio collected on the flip camera which went through the upload and download process proved to be the most problematic in terms of the effect on F1 and F2. Recordings that had been downloaded from YouTube exhibited inconsistent differences in F1 or F2 across the two speakers analysed. The authors caution against using data from different informants recorded on multiple devices. However, they claim that recordings made on an iPhone or Macbook Pro may be suitable for acoustic analysis, particularly of F1 and F2.

In this case, a comparison of F1 and F2 between different speakers is only made in Chapter 6; in the main analysis of Henry's speech in Chapter 5 comparisons of F1 and F2 are only being made within the same speaker over time. It is unfortunate that early recordings were made in MP3 format. However, by the age of 3;08 all recordings were made in WAV format. Careful consideration will be given to the acoustic analysis of the data from the age of 3, where it is possible that there may be a small effect on F1/F2 values when the recording method shifts.

De Decker and Nycz's research was published in 2011, but the first-generation iPhone they used in their experiment was released in 2007. The iPhone 4 (fourth generation) used in the early recordings in this research was released in 2010, and the iPhone 6 (eighth generation) in 2014 (Montgomery & Mingis, 2022). Little information on the built-in microphones is available, but it seems likely that over this seven-year period, improvements will have been made to the technology.

For the purposes of this research, analysis begins at 2;01, as recognisable words began to appear regularly in the recordings. From this time, recordings were made on a fairly frequent basis (see Figure 14), though there was a period between the ages of 2;11 to 3;08 where few recordings were made, and these were typically short in length. The focus of this research emerged at around the age of 4;0. From this time, recordings became more regular, being made on a monthly basis for periods of around 10-15 minutes. Some recordings were slightly shorter, but on one occasion the recording length was around two hours. Although the recordings were short, they captured naturalistic data while Henry played, ate, got dressed, or chatted with me about his day. Some recordings contain his early reading, and any change in style is discussed where appropriate.

Analysis of these recordings has been carried out on data from approximately every two months, though where only shorter recordings were available, this frequency has been increased to maximise analysable data (see Figure 15). The length of these recordings varies according to what data is available from that month. Analysis has been typically limited to a maximum of 15 minutes of data per recorded interaction. In total, just over six hours of recordings were systematically transcribed and analysed.

The recordings mostly consist of interactions between Henry and me, and situations range from getting ready for school to singing songs, building Lego, playing with other toys and reading together.

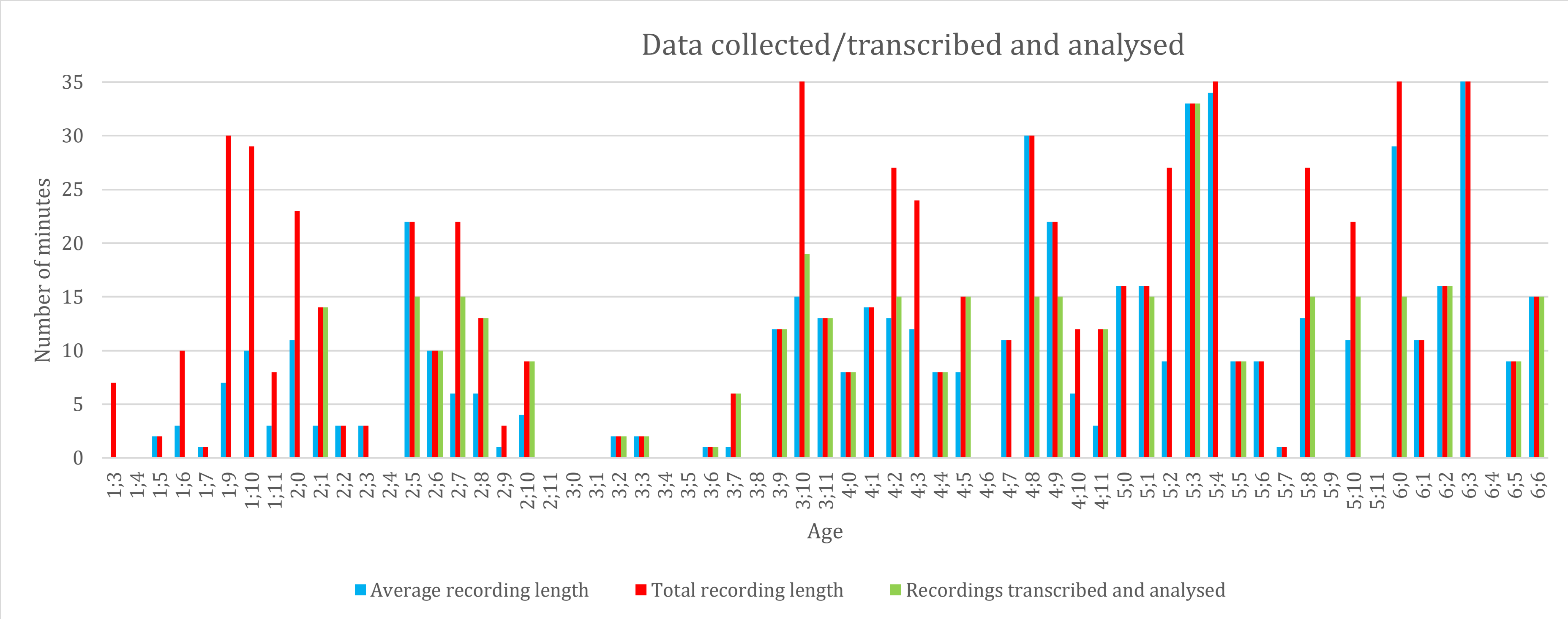


Figure 14: Data collected/transcribed and analysed

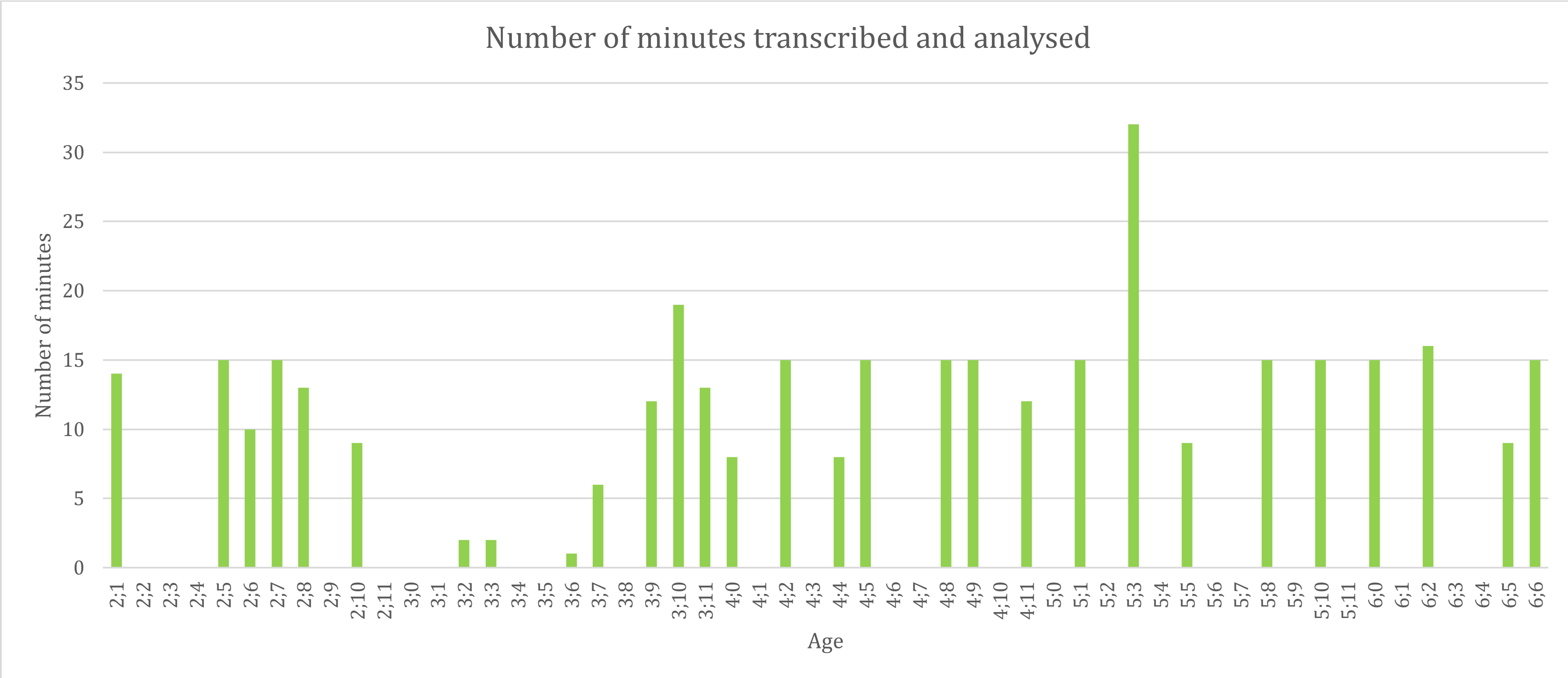


Figure 15: Number of minutes transcribed and analysed

The data was segmented and transcribed orthographically and phonetically in PHON (Rose et al., 2006) version 2.1. All non-WAV files, including video, were converted to 16-bit WAV for the purposes of importing data into PHON. See Table 14 for details of all original recording formats.

Codec	**Sample rate**	**Data rate**	**Audio channels**
.mp3	44.1	64 kbps	mono
.wav	44.1	16	mono
.mov (AAC, H.264)	44.1	7.41 mbit/s	mono

Table 14: Original recording format

4.3 Analysis

Henry's vowels were analysed according to target word membership of key lexical sets (Wells, 1982). These lexical sets were selected as the site of potential accent-based differences between Henry's pronunciation and the accents spoken both at home and in the local area. The STRUT and FOOT vowels, for example, are distinct from each other in the accents of both parents and his grandmother. However, in local varieties, STRUT does not contrast with FOOT, representing a systemic difference between the home and local accents (Wells, 1982: 76-78) (see Table 15).

<table>
<tr><th></th><th>Home</th><th>Local</th></tr>
<tr><td>STRUT</td><td>ʌ</td><td rowspan="2">ʊ</td></tr>
<tr><td>FOOT</td><td>ʊ</td></tr>
</table>

Table 15: Distribution of STRUT/FOOT

The typical realisation and distribution of TRAP, START, PALM and BATH is more complex. Both home and local varieties have a contrast between long and short vowels, for example, in the home varieties, BATH, PALM and START words are articulated as a low back vowel /ɑ/, for example, 'grass' /grɑs/, 'can't' /kɑnt/ and 'star' /stɑ/ all share the same phoneme, while TRAP words are articulated with the front vowel /a/, for example, 'gas' /gas/.

In local varieties however, BATH and TRAP share the same front articulation, /a/, as in 'grass' /gras/ and 'gas' /gas/. There is no long back vowel in these accents, but a long front vowel /aː/ features in PALM or START words, for example, 'can't' /kaːnt/ and 'star' /staː/. The distinction between TRAP/BATH/PALM/START in local varieties is one of length (see Table 16).

<table>
<tr><th></th><th>Home</th><th>Local</th></tr>
<tr><td>TRAP</td><td>a</td><td rowspan="2">a</td></tr>
<tr><td>BATH</td><td rowspan="2">ɑ</td></tr>
<tr><td>PALM/START</td><td>aː</td></tr>
</table>

Table 16: Distribution of TRAP/BATH/PALM/START

While the BATH/PALM/START vowel may typically be longer in the home varieties than TRAP, length alone is not phonologically meaningful. Wells describes the distribution found in the local varieties as flat-BATH accents and the home varieties as broad-BATH accents (1982: 134). These, he explains, are terms used to differentiate between accents where TRAP and BATH vowels sound the same (flat-BATH accents) and those where BATH and PALM words share the same vowel (broad-BATH accents). The origin of the terms 'flat' and 'broad' is not clear, though it can be seen as early as Kenyon's 1930 essay on the distribution of "flat-a and broad-a" in Standard British English and its status in the United States. Kenyon refers to the popularity of the terms with "teachers of speech", who valued the "intrinsic beauty" of broad-a above flat-a, before they later turned to "the science of phonetics" (Kenyon, 1930: 323). Wells, of course, attaches no differential values to the variants, but finds the terms a helpful way of distinguishing between groups of accents which behave in a particular way. He describes this distinction as a lexical distributional difference between accents (1982: 79). These are differences where a particular vowel is attached to a set of lexical items, which may not form a predictable pattern and if not acquired, may need to be learned individually. Wells claims that a phonemic split between TRAP and BATH vowels occurred in 18th century English, leading to BATH words being produced with a long, back vowel, /ɑ/ in the accent that would eventually become RP. Some words may belong to the BATH or TRAP sets depending on the speaker, and in some cases are variable within the same speaker, for example "transfer, Glasgow, stance, exasperate" (Wells, 1982: 134). Others, which may apparently seem to follow the pattern for BATH words, such as 'rant' and 'pant' are categorically members of TRAP. See section 4.1.5.1 above for details of how the BATH vowel is articulated in Henry's home environment.

There are further lexical sets which are articulated differently in the home and local accents, such as FACE and GOAT. While the STRUT/FOOT sets and the BATH/PALM/START sets are not differentiated by *most* northern speakers, a monophthongal pronunciation of FACE and GOAT tend to be used by *more non-standard* speakers. Many of Henry's friends have exhibited all these features, but Henry has never shown any signs of acquiring local variants of the FACE and GOAT lexical sets. Therefore, the main analysis in Chapter 5 will focus on the STRUT/FOOT sets and the BATH/PALM/START sets. Further discussion and an analysis of the FACE and GOAT lexical sets when Henry is in conversation with a friend will be addressed in Chapter 6.

The number of tokens for each lexical set varies in each recording depending on the activity being recorded, what happened to be spoken about that day and the number of repetitions of a single word. A low or high number of tokens is not solely due to the length of the recording, though that is a consideration. The natural variation in frequency of phonemes occurring in the language is also a contributing factor. The vowel /ɑ/ appears as only 0.68% of all English phonemes (see Table 17) and is further divided into membership of the BATH, PALM and START sets. FOOT occurs much less frequently than STRUT and TRAP. This distribution is apparent in Henry's speech.

Phoneme	% of total English phonemes
a	1.62
ʌ	1.56
ɑ	0.68
ʊ	0.62

Table 17: Relative distribution of English phonemes (Cruttenden, 2014)

Cruttenden (2014: xvi) outlines the relative frequency of phonemes in what he calls "GB", his term for General British English. While he acknowledges that linguists describe Received Pronunciation as a current and flexible standard variety, he claims that the media continue to prescribe RP as the "posh" accent of elites. General British, he claims, is a term designed to diverge from this to a working standard accent. Cruttenden does not make explicit the source of data or size of corpus, aside from stating that it is text based (p. 159).

The process for the analysis of the data collected is described below.

4.3.1 Impressionistic analysis

Following the segmentation and transcription of the data in PHON described above, members of the six lexical sets STRUT, FOOT, PALM, START, BATH and TRAP above were tagged for analysis. Tagged data were assigned to a category impressionistically, according to whether they most closely matched the home variant, the local variant, or 'other', i.e. something not found in the home or local varieties. The 'other' category was usually an articulation with features belonging to both the home and local variants (categorised as 'blends'), but in early recordings, this was sometimes a different vowel altogether. Only tokens judged to be stressed were included in the analysis. Unstressed, unclear, overlapping or particularly rapid tokens were excluded as they were sometimes difficult to judge. While PALM and START were initially kept separate, they were later merged, as aside from in Henry's grandmother, these two lexical sets behave as one in both the home and local accents, and there were no signs of her influence on this lexical set.

Following the initial allocation to the categories above, I performed two further full auditory checks on my initial analysis, separated by several months each time. A PHON report was generated for each lexical set on each occasion, and individual tokens were checked against my original classification. On both occasions, tokens were reassigned or were excluded based on the criteria above after careful reconsideration. A fuller picture of which tokens were suitable for inclusion emerged during the course of the initial analysis. For example, judgements of which tokens were stressed and which were unstressed developed over time based on the experience of the full dataset. The margins between home or local articulations and blends also became clearer as my experience of listening to these vowels grew. Details of the inter-rater reliability checks follow in section 4.3.2, below.

4.3.2 Inter-rater reliability

Inter-rater agreement in the transcriptions of children's language can be low, particularly in the case of vowels. For example, Davis and MacNeilage (1995) reported on their analysis of a 6659-utterance corpus from six infants between the onset of babbling and age 3:06. While reliability of consonant transcription varied between 63-83% with an average of 76.8%, vowel reliability was much lower, ranging from 33-69% with an average of 44.8%. Agreement of individual vowels varied widely, for example, /u/ achieved 0% agreement, while /a/ achieved 60% agreement, and back vowels were found to have agreement of only 15%. When the researchers limited the possible analysis to a 9-way classification through the use of the labels high, mid, low and front, mid, back, agreement improved to 49.9%. This is still lower than the inter-rater agreement of the consonants, however. The authors were able to improve this agreement to 80% if they accepted sounds allocated to neighbouring cells as a match. Broadly speaking then, vowels are typically more difficult to reach agreement on than consonants, and while higher levels of agreement are achievable through the allocation of vowels to predefined categories, inter-rater agreement remains much lower than for consonants.

A professional phonetician (my supervisor) performed a categorisation exercise on 20% of the corpus in order to determine an inter-rater reliability score. The 20% sample included recordings from ages 2;10, 3;09, 4;02, 5;05 and 6;06.

Some tokens provided were inaccessible to the second listener for technical reasons (for example, some sound files would not open in the software). Taking these into account, 15% of the overall dataset was included in the inter-rater agreement process (see Table 18).

The sample size of each lexical set varies in proportion to their overall distribution in the data. This distribution follows that described by Cruttenden (2014) in terms of rank order, though here I have split what he characterised as a single phoneme /ɑ/ into three separate lexical sets, START, PALM and BATH.

	TRAP	STRUT	FOOT	START	PALM	BATH	PALM/START /BATH total	Overall total
Total no. of tokens analysed	584	387	117	90	51	47	188	1276
Selected for inter-rater ag.	117	77	23	18	10	9	37	254
Analysed for inter-rater ag.	97	53	18	10	7	9	26	194
% of total lexical set	17%	14%	15%	11%	14%	19%	14%	15%

Table 18: Tokens selected for inter-rater reliability as a percentage of total

For the STRUT lexical set, tokens were allocated to the classifications of rounded (local), unrounded (home) and blend. For FOOT, tokens were assigned to the categories of front, mid, or back. PALM, START and BATH were split into front, mid/long mid or back. TRAP tokens were divided into categories of short (typical articulations), and long (hypercorrections).

	Rounded	**Blend**	**Unrounded**
STRUT	ʊ	ʌ̹ ~ ɒ ~ ɒ̜ ~ ə̹ ~ ʊ̜	ʌ ~ ə
	Front	Mid	Back
FOOT	ʏ	ʊ̟	ʊ
	Front	Long/long-mid	Back
PALM, START	aˑ ~ aː	ɐˑ	ɑ
	Front	Long/long-mid	Back
BATH	a	aː	ɑ
	Typical	Hypercorrection	
TRAP	a	aː	

Table 19: Indicative realisations of lexical sets (not exhaustive)

Initial inter-rater agreement for the FOOT lexical set was 61% and for STRUT, 66%. For PALM/START/BATH, initial inter-rater agreement was 77%. 100% agreement was achieved in the classification of TRAP hypercorrections. Many disagreements were around the classification of blends, which exist on a continuum. For example, the listener needs to decide where the dividing line between front and mid articulations or back and mid articulations lies. It is extremely difficult for two independent listeners to agree perfectly and consistently where this tipping point lies, when it is inevitable that there may even be some slight intra-listener differences, as evidenced by my own adjustments to categorisation on reflection following second and third listenings.

Cohen's Kappa (Cohen, 1960) statistical analysis was performed as a metric of inter-rater reliability. Cohen's Kappa provides a rating of reliability between judges, accounting for the agreements that can be predicted by chance. A Cohen's Kappa score of above 0 indicates that inter-rater agreement is above the rate that could happen by chance, up to a score of 1.0 as the maximum agreement (where agreement between judges is perfect). A score of less than 0 is an indicator of a number of agreements being lower than chance. An interpretation of the strength of the Kappa score is suggested by Landis and Koch (1977), (see Table 20).

Kappa Statistic	**Strength of Agreement**
0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Table 20: Cohen's Kappa strength of agreement scale (Landis & Koch, 1977)

There is, however, no agreement on the reliability of these interpretations; Landis and Koch themselves admit that the categories are arbitrary. Bakeman and Gottman (1997: 66) suggest that a rating of less than 0.7 should be treated with caution, but they also admit that this is an informal claim.

Weighted Kappa (Cohen, 1968) enables close disagreements to be rated as less problematic than distant ones. For example, in this data, there is necessarily a closer relationship between a blend and a home STRUT vowel than there is between home and local variants as the blends are, in articulatory terms, midway between the other two categories. A weighted Kappa score takes these closer relationships into account.

Lexical set	**Kappa**	**Weighted Kappa**	**Strength of agreement**
STRUT	0.256	0.278	fair
FOOT	0.417	0.423	moderate
PALM/START/BATH	0.598	0.693	Moderate/substantial
TRAP	1.0	1.0	Almost perfect

Table 21: Cohen's Kappa strength of agreement

As seen in Table 21, all results were well above chance. The strength of agreement for STRUT was fair on the established scale, compared to better results (moderate) for the FOOT lexical set, and much stronger agreement for PALM/START/BATH and the TRAP lexical sets.

Following the initial inter-rater agreement process, I provided a more detailed explanation of my classification process to the second listener. Tokens where we had initially disagreed were reclassified independently.

	STRUT		FOOT		START/ PALM/ BATH		TRAP		Average agreement
	N	%	N	%	N	%	N	%	
Tokens selected for analysis	77	20%	23	20%	38	20%	116	20%	
Tokens analysed (after exclusions)	53	14%	18	15%	26	15%	96	17%	
Initial agreement	35	66%	11	61%	20	77%	96	100%	71%
Final agreement	49	92%	17	95%	26	100%	96	100%	97%

Table 22: Resolution of differing classifications

This time, independent classifications of STRUT were closer (66% agreement) than the initial inter-rater test (55%). Where we continued to allocate these tokens to different classifications, we discussed individual examples in detail in an attempt to discover the exact point of variance. Through these discussions all but a few tokens were resolved (Table 22).

The overwhelming majority of shifts in listener perceptions were small, for example, in the case of STRUT realisations, changing from an initial classification of [ʊ] (local) to blend, which could be any of the following [ʊ̜ ~ ə̜ ~ ʌ̜] (and vice versa) or from an initial classification of blend to unrounded (home) [ə ~ʌ] and vice versa. In only two out of 14 cases was an initial classification changed from rounded to unrounded, and there were no examples moving in the other direction, from an unrounded classification to a rounded one. The tipping point between a home or local articulation and a blend is on a continuum, and as such, the point at which it becomes classified as belonging to one category or the other is subjective. The high number of resolutions achieved through discussing these marginal examples in detail is evidence of this. Detailed discussions were also had around coarticulation and how much the influence of other consonants could or should be ignored.

The second inter-rater test also resulted in better agreements for the other lexical sets. Table 23 shows how on second independent listening, after further explanation of the criteria, much better levels of agreement were achieved.

	STRUT	**FOOT**	**PALM/START/BATH**	**TRAP**
Initial inter-rater agreement	55%	61%	76%	100%
Initial agreement after further explanation	66 %	61%	77%	100%
Final agreement after token-by-token discussion	92%	95%	100%	100%

Table 23: Improvements to agreement through second inter-rater test

Further, more detailed data from the inter-rater reliability process is provided in the Appendices.

4.3.3 Acoustic analysis

Following the inter-rater tests, a subset of lexical items from the impressionistic analysis were analysed acoustically using Praat (Boersma & Weenink, 2018). The purpose of the acoustic analysis is twofold: to ascertain the validity of the impressionistic analysis, specifically by considering F1 values (inversely related to tongue height) and F2 values (lower values equate to retraction and higher values to advancement of the tongue), and to consider the range of variation in each vowel, based on distribution and overlap (if any) within the vowel space, as seen in Yang and Fox (2013) and Keating and Huffman (1984). Plotting the vowels using F1 and F2 values enabled confirmation of the impressionistic analysis, based on relative tongue retraction and height. Formant values were recorded in Hertz (Hz) to enable comparison with the acoustic reference data seen in the literature.

Tokens were selected for acoustic analysis from the age of 3;0 upwards (Table 24). In each year of life, the first 10 tokens representing unique lexical items produced in each category (i.e. home, local and blend, based on the impressionistic analysis) were selected for acoustic analysis in each of the six lexical sets. Naturally occurring speech was prioritised, but where not enough tokens of this type were present, read or recited tokens were permitted. Where 10 tokens of unique lexical items were not present, different tokens of the same word were selected.

Due to the naturalistic data collection methods, many tokens which had been considered suitable for impressionistic analysis were found to be unsuitable for acoustic analysis. Background noise and overlapping talk were frequent features of recordings as well as wide-ranging vocal styles, such as shouting, singing, whining and whispering. As some speech styles affected formant values, as discussed in section 3.2.4 above, this reduced the number of tokens suitable for analysis.

The rigorous selection process described above meant that in some cases, 10 tokens suitable for acoustic analysis were not available, particularly in the case of the less frequent lexical sets such as BATH. As Henry's accent changed over time, in some cases, fewer tokens were available in each of the home, local and blend categories. As will be seen in the analysis, sometimes he favoured home realisations, and sometimes local realisations dominated his speech. A lack of examples in some of these categories reflects his changing accent. A full list of the examples analysed acoustically is provided in the Appendices.

Age	3	4	5	6
STRUT				
Home	10	9	9	3
Blend	10	10	10	6
Local	2	9	8	10
Total N	22	28	27	19
FOOT				
Front	10	9	10	5
Mid	0	3	8	6
Back	1	9	4	10
Total N	11	21	22	21
PALM/START				
Home	1	2	2	2
Blend	2	8	0	2
Local	10	6	7	7
Total N	13	16	9	11
BATH				
Home	0	0	0	0
Blend	0	3	5	2
Local	0	4	4	8
Total N	0	7	9	10
TRAP				
Home	10	10	10	10
Blend	0	7	3	0
Total N	10	17	13	10

Table 24: Number of tokens available and suitable for acoustic analysis

For the initial impressionistic analysis, FOOT tokens were divided into rounded and unrounded (hypercorrections). The acoustic analysis revealed an interesting distribution of tokens between back and front realisations (see section 5.1.3), thus the categories for classification were revised to front, central and back. The impressionistic analysis was repeated at this point, using the new categories. Inter-rater reliability checks were performed on the new categories.

Each vowel within this subset of the data was selected and annotated in Praat using the textgrids function. The vowel portion of the file was selected by close listening in the first instance, followed by a visual appraisal of the formants.

Surrounding liquids, nasals and velars can impact on formant values, and some authors argue that this needs to be addressed in the selection of tokens suitable for a sociophonetic acoustic analysis (Di Paolo et al., 2011: 87). Given the naturalistic data collection methods used in this study, however, and the young age of the subject, once rapid tokens or those overlapping with other speech or background noise were discounted, few tokens suitable for acoustic analysis remained. It was not therefore feasible to apply any further filters. Instead, formant plots were examined to look for evidence of these effects and are raised as appropriate in the analysis below.

Options are available to the researcher for the selection of the point at which formant values are measured. The mid-point of the vowel, between the vowel onset and offset is a simple choice, but it may ignore coarticulatory effects. The 'maximal displacement method' requires the researcher to select the first point in the vowel where it appears to be free of coarticulatory effects. Alternatively, multiple values throughout the duration of the vowel may be selected manually, in order to provide the most detailed information on formant movement (Di Paolo et al., 2011: 90-91). Here, I have chosen a single point measurement which is somewhere between the mid-point method and the point of maximal displacement method. The most stable portion of the vowel free from cooarticulation (Khattab & Roberts 2011: 170) was selected using the texgrids functionality in Praat (Figure 16). Praat then extracted values from the mid-point of this selection. This method was chosen as there was considerable movement of formants, and in some shorter vowels, this could extend towards the centre point. Its benefit over the maximal displacement method is that the process can be automated.

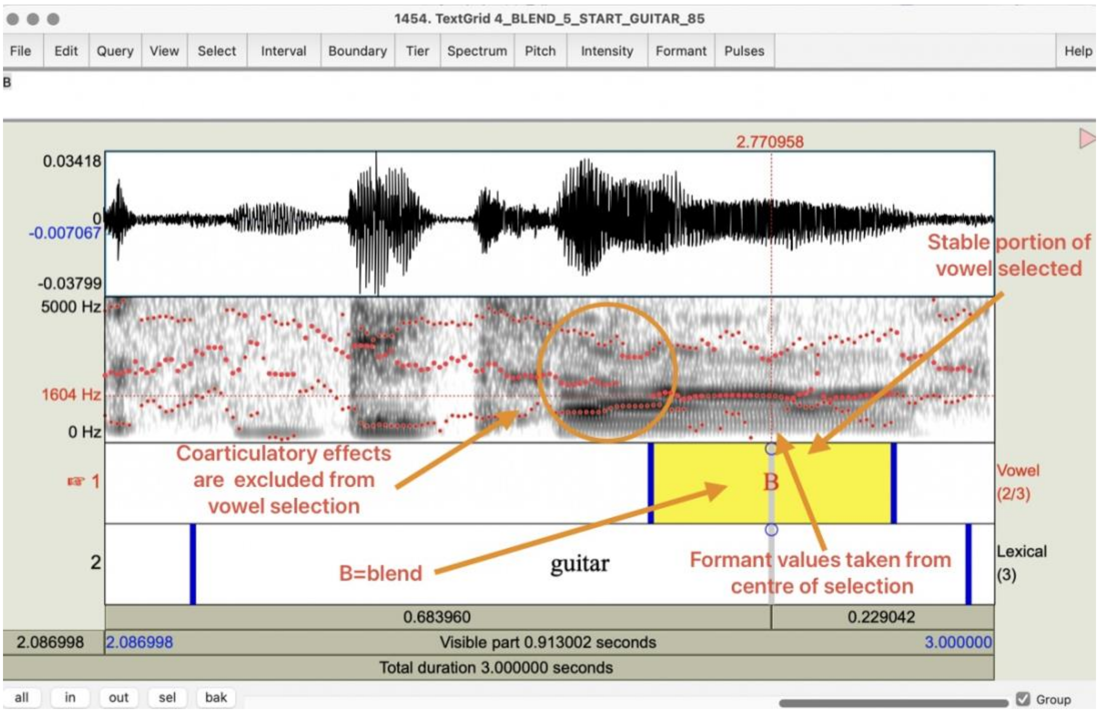


Figure 16: Textgrid showing selection of stable vowel portion

Praat was set to display three formants in 5 kHz rather than the standard five for adults, to reduce the problem of harmonics being incorrectly identified as formants (Khattab & Roberts, 2011: 170). Formant values (F1, F2 and F3) were extracted from the mid-point of the selection using a batch Praat script (Remijsen, 2019) which was modified to also extract F3 values to assist in accurate selection of the first two formants. Linear Predictive Coding (LPC) spectra (autocorrelation algorithm) were also produced for analysis. The F1 reference value was set to 750 Hz, F2 to 2250 Hz and F3 to 2475 Hz, in reference to the Remijsen script adapted for Khattab (2011) by Al Tamimi (2011). Harrison (2013: 235) found that altering reference values to match the average of a particular vowel can improve the accuracy of tracking. However, given that in this data the formant values can vary substantially according to home, blended or local variants, I did not feel that altering them would be useful. For example, at age 4, F1 for PALM/START varies from 764 to 1619 Hz, and F2 varies from 1545 to 2309 Hz.

Formant values were plotted inversely onto scattergraphs (after Joos, 1948) to both check the reliability of the automatic readings and create a visual representation of vowel variation. Following an approach recommended by Khattab and Roberts (2011: 170), each LPC spectrum was then examined individually. Formant values were adjusted manually after comparing the LPC spectrum with the spectrogram, as well as the impressionistic auditory analysis, as recommended by Di Paolo et al., (2011: 94). A detailed example of this process is outlined below.

At this stage, outliers were scrutinised (Di Paolo et al., 2011: 94). Some were rejected on the basis of formant values which placed the F1/F2 plot in an unfeasible place in the vowel space, in line with the process adopted by Lee and colleagues (1999). For example, Harrison (2013) explains that errors in F1 and F2 measures are more frequent in higher frequency speech, such as is found in women and children. He also points out that errors occur more frequently at the edges of a speaker's vowel space, so in this case, vowels which I perceive as particularly front, back, high or low, are at more risk of having erroneous values from the script's automatic tracking algorithm.

Where an alternative reading of F1/F2 placed the vowel plots in a cluster with other tokens judged to be similar on an impressionistic basis, these were retained. A very small number of tokens were reassigned to a different category at this stage. As outlined in 3.2.4, formant peaks can be difficult to read, due to the influence of the widely spaced harmonics attributed to the child's high f0 (Kent, 1976; Story & Bunton, 2015).

For example, compare Figure 17 and Figure 18 below, of the child's PALM/START vowels from age 4.

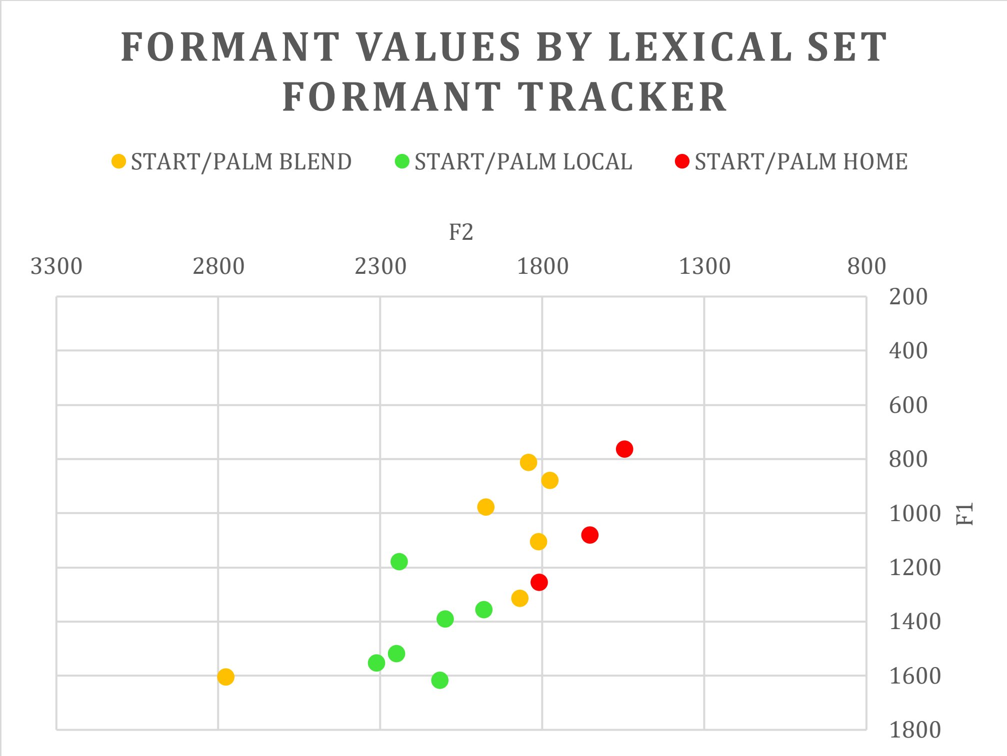


Figure 17: F1 /F2 values (formant tracker) by lexical set age 4

Figure 17 was populated using values generated by the automatic tracker function of the script, which used the 'to formant (burg)' algorithm (recommended by Khattab & Roberts, 2011), and the tracker (Remijsen, 2019). Note the outlying blend example in the lower left of the figure. This is the START vowel in 'guitar'. The automatic script identified F1 as 1604 and F2 as 2774. Impressionistically I judged this token to be a blend, that is, somewhere between front and back – definitely not fully front. The formant tracker and burg algorithm place this example as much more front than any other example including all of the tokens judged as front. All other examples have minimal differences between the formant tracker values and the values extracted from the peaks in the LPC spectrum (autocorrelation algorithm), which do not affect the overall distribution of the variants.

To check whether altering the reference values would have had an effect on the values for this vowel, I re-ran the script for the 'blend' tokens at age 4, altering the reference values to 1100 for F1, and 1850 for F2, however, the script algorithm continued to extract the original formant values. While this was not a systematic experiment on whether altering the reference values had an affect across the dataset, it appears that in this case, the soft constraints of the reference values are not enough to over-ride the burg algorithm's initial reading.

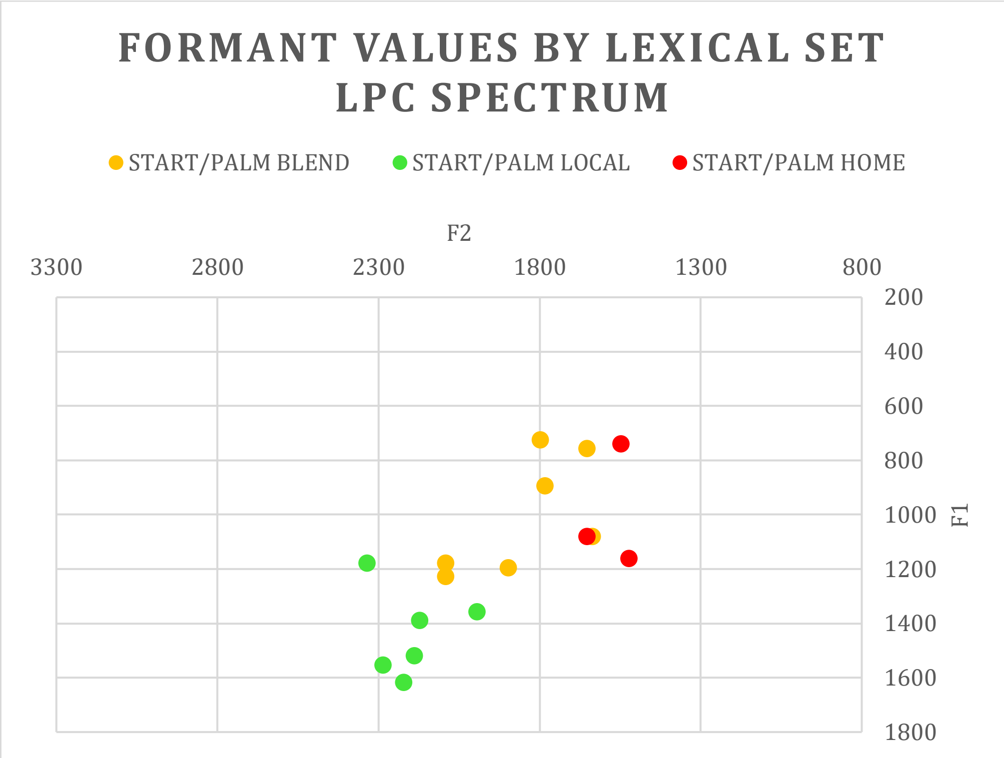


Figure 18: F1 /F2 values by lexical set (LPC spectrum) age 4

On examining the spectrogram (Figure 19), we can see that as picked up by the automatic tracker, F1 does indeed appear to be at around 1688 Hz. There is, however, another band of energy lower in the spectrogram. f0 at the centre point of this vowel is 211 Hz, therefore the 8th harmonic is at around 1688 Hz, which could be intensifying the appearance of the energy at this point.

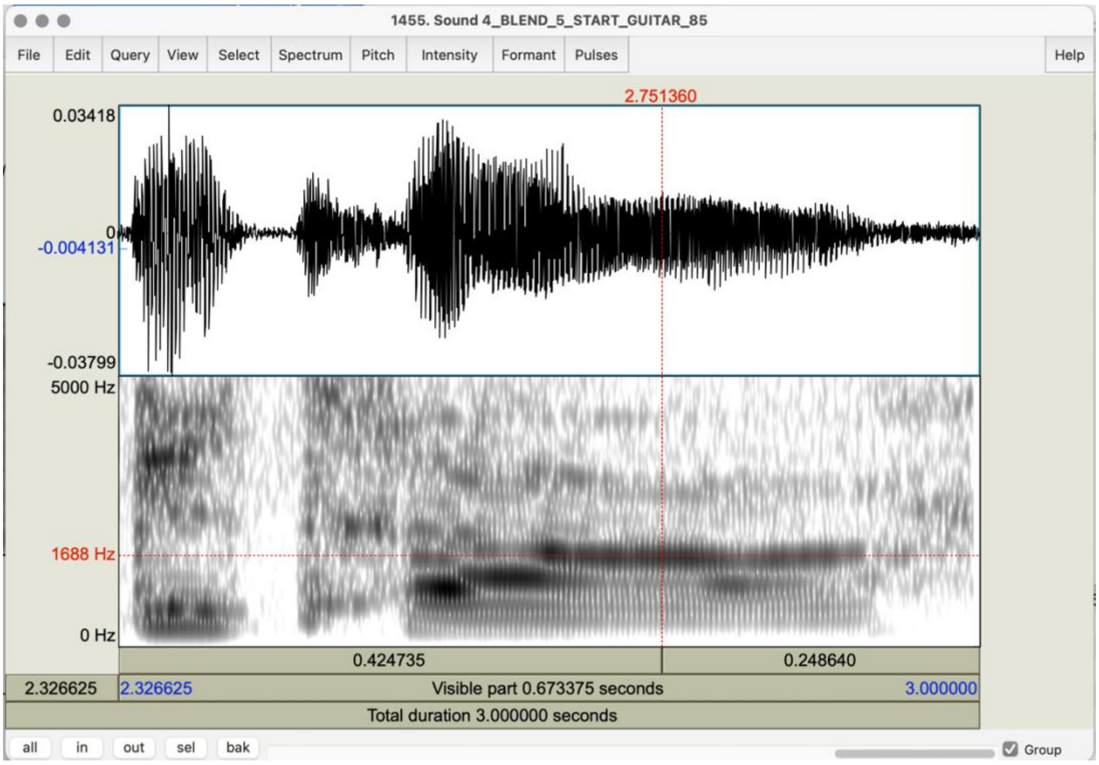


Figure 19: Spectrogram 'guitar'

An inspection of the LPC (Figure 20) reveals the first peak to be at 757 Hz, and if we treat 1651 Hz (the LPC reading, as opposed to 1604 which was the formant tracker reading) as F2, this places the vowel in a plausible space in the scatterplot, alongside other tokens which had also been judged to be blends. It is possible therefore that the intensification of the energy at F2 caused by the presence of the 8th harmonic may be fooling the formant tracker into thinking that it is F1, as the energy around the real F1 is comparatively less intense. The final values for this vowel were therefore taken from the LPC.

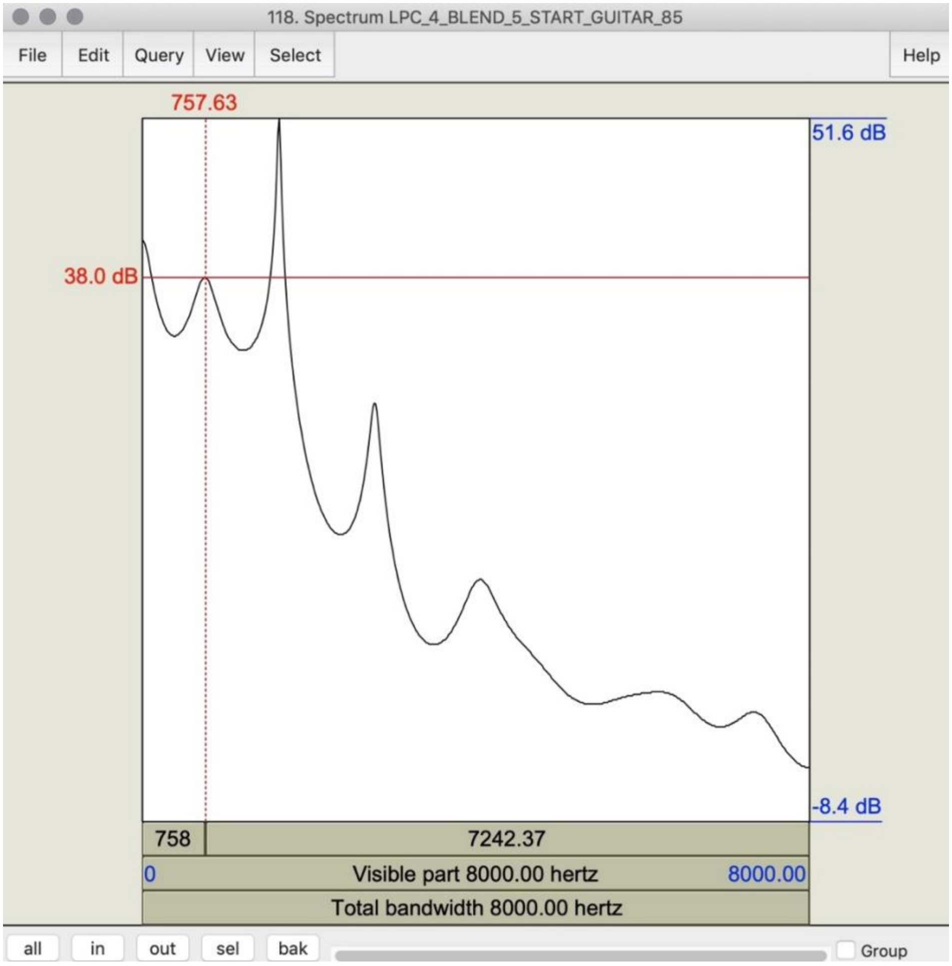


Figure 20: LPC 'guitar'

However, below is another example from age 4. If taken without further scrutiny, the LPC spectrum gives an implausible F1 reading. While in the spectrogram below, (Figure 21) the automatic formant tracker identifies F1 at 1199 Hz and F2 at 2248 Hz, a constant path of F1 during the course of the vowel is difficult to follow – there appears to be a break in the energy intensity which corresponds with the centre of the vowel, right where the formant tracker is reading its information from.

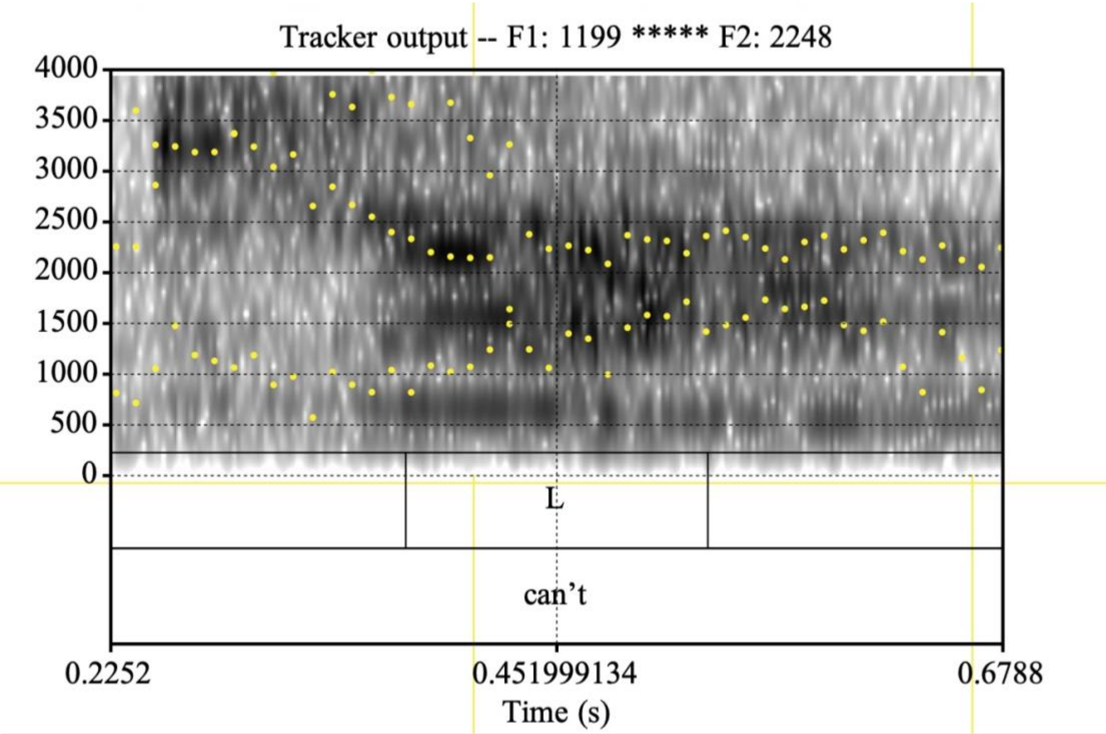


Figure 21: Spectrogram 'can't'

In the associated LPC spectrum shown below (Figure 22) however, peaks appear at 660, 1554, 2220, 3910, 4800 Hz. I perceive this vowel as a low, very front vowel. If F1 is taken as the first peak of the LPC spectrum, this places the token as a high back vowel. However, if the second peak is taken to be F1, the vowel clusters with low front vowels. The LPC spectrum's first peak may be influenced by the first harmonic, as f0 is around 310 Hz. Anticipatory nasality as a result of the adjacent nasal consonant may also be influencing the formant value. As we learned in 3.2.4, nasality can affect formant frequencies, for example, by creating extra poles or zeroes in the speech signal. This example demonstrates the necessity of scrutinising each LPC alongside the spectrogram in conjunction with impressionistic listening. Here, I resolved this example by ignoring the first peak of the LPC based on a combination of scrutinising the spectrogram, taking into account impressionistic listening, considering F3 values and possible coarticulatory interference.

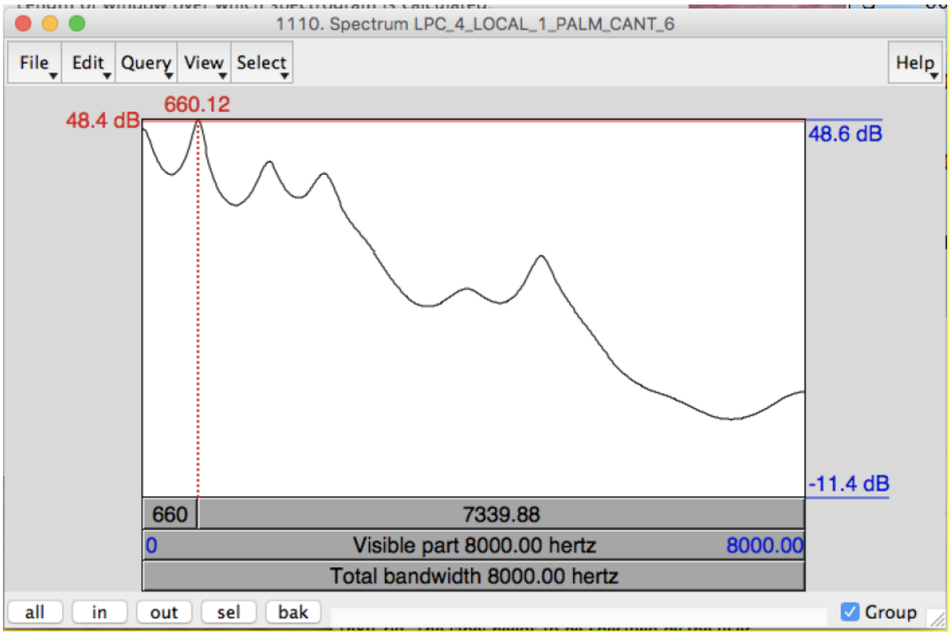


Figure 22: LPC 'can't'

4.3.4 Accommodation analysis

Throughout this project, Henry has been recorded in conversation with me, his mother. In order to establish whether Henry changes his accent when talking to friends, he was recorded playing with a school friend, James, and James' younger brother. Henry was aged 6;11 at the time of this recording, five months after the last recording in the main corpus. The boys were recorded using the same equipment and settings as the main data collection described above. During the recording, the children were left alone to play with Lego in Henry's home, while James's mother and I were out of earshot in a different room. The boys were aware that they were being recorded, though Henry was more used to this than James was, as evidenced by the following exchange 16 minutes into the recording:

James: I think I think we're getting listened to.
Henry: I know. It's in the phone.

James thought that our voices in the next room had stopped so that we could listen in on them, while Henry was aware that he was being recorded all along by the phone on the table, as he has grown used to being recorded.

The children were recorded playing by themselves for 32 minutes. As was the case for the main corpus, the recording was transcribed orthographically and phonetically in PHON (Rose et al., 2006) version 2.1.

The same six lexical sets were tagged as for the main corpus: STRUT, FOOT, PALM, START, BATH and TRAP. James realises GOAT and FACE vowels as monophthongs, while Henry has shown no sign of acquiring these monophthongal variants in his speech. As such, these lexical sets have not been analysed in the main corpus, but as the GOAT and FACE vowels are salient features of James' dialect, for the purposes of the accommodation analysis, these lexical sets were tagged, transcribed and analysed in both Henry and James' speech. An analysis of these lexical sets will determine whether Henry showed any signs of shifting towards these vowels in conversation with his friend. The younger brother's speech was not tagged, transcribed or analysed in any way.

In the main corpus, the beginnings of recordings were not excluded; this is because Henry has grown up being recorded on a frequent basis, and I do not usually draw his attention to the recording device when recording starts. He sometimes 'shows-off' for the recording, for example, by using silly voices, but I considered the full range of his speech productions to be worthy of analysis. In the case of the friends playing, however, James was unfamiliar with being recorded, so my preference was to select examples from further into the recording where possible, in case he should be nervous about the recording. Although he made reference to being listened to, he seemed otherwise unconcerned, and his realisations were very consistent throughout the recording, showing much less variation than Henry's speech.

4.3.4.1 Acoustic analysis of accommodation recording

The acoustic analysis of this recording follows the same pattern as the acoustic analysis of the main dataset. Ten tokens from each lexical set for each child were selected for acoustic analysis. The tokens were selected from a starting point of 10 minutes into the recording, to enable the children to relax and become less conscious of the recording device. If fewer than 10 examples were found, the first 10 minutes of the recording was used to find additional tokens. If fewer than 10 unique lexical items were present, repetitions of the same lexical item were included. In some lexical sets, (STRUT, FOOT, START, PALM and BATH) fewer than 10 tokens of each lexical set were present for each child. Unstressed and very rapid realisations were excluded, as were ambiguous utterances.

Chapter 5 Analysis

In this chapter I present an impressionistic analysis of Henry's lexical sets STRUT, FOOT, START, PALM, BATH and TRAP, supported by an acoustic analysis. The acoustic analysis is designed to support the impressionistic classifications, to offer a visual representation of how these vowels are distributed in Henry's vowel space and how this distribution changes between the ages of 3 and 6 years old.

5.1 STRUT/FOOT

5.1.1 STRUT

There are very few tokens of STRUT in the early recordings, corresponding with few intelligible words being spoken (see Figure 23). The number of tokens increases gradually over time. The dip in number of tokens at 3;06 is an anomaly, as only a short recording was made of Henry singing 'Little Donkey' at Christmas time. Another dip at 4;04 corresponds with a shorter recording, but in the recordings at 5;01, 5;08 and 6;04 the small number of tokens is not related to a shorter recording time.

In the figures throughout the analysis, the time between recordings is varied, but recorded sessions are compressed here due to limited space. Please see Figure 15 for a visual representation of the gaps between the recordings.

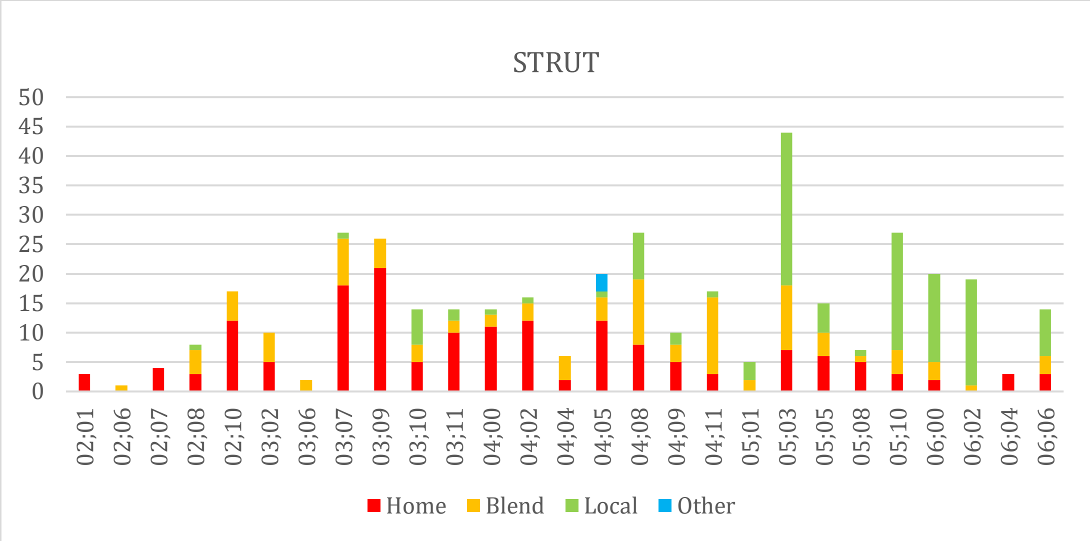


Figure 23: Realisations of STRUT [10]

The earliest recorded examples of STRUT are articulated with a vowel close to Henry's parents' STRUT vowels, i.e. [ʌ], as seen at age 2;01:

Example (1) 'Buzz'

[ᵐbʌz̥]

The first example illustrating some influence of the local vowel [ʊ] appearing in the recordings appears at 2;06 in the form of lip rounding of STRUT:

Example (2) 'upstairs'

[ˌʌ̹p'stɛəz]

Here, the STRUT vowel shares its height with the home variant (open-mid), but impressionistically it is articulated with some lip rounding. Articulations such as these were categorised as a 'blend'. These blends took various forms, for example, [ʌ̹], [ɒ̝] and [ʊ̞] as can be seen in Table 25.

[10] *Excludes 'one', 'once' etc. as they belong to the* LOT *lexical set. This will be discussed later in this chapter.*

	Home	Blend	Local	Other
02;01	[ˈ**bʌz̥**]			
	'Buzz'			
02;06		[ˌ**ʌ̹pˈstɛəz**]		
		'upstairs'		
02;08	[ˈtaɪ **ˈʌp**]	[ˌ**ʌ̹pˈzɛəz**]	[ˈti **ˈmʊ̹mi**]	
	'tidy up'	*'upstairs'*	*'tea, Mummy'*	
02;10	[ˈpɪk ˈmi **ˈʌp** ˈpwiz]	[ˈaɪ əm **ˈʊ̹nɪn** əˈʊaʊnd]		
	'pick me up, please'	*'I am running around'*		
04;04.	[dɪd ˈjɔz ˈfɔl ɒf **ˈsʌmtaɪm**]	[ˈɪz ˈbəʊ **ˈkʌ̹mɪn**]		
	'did yours fall off sometime'	*'is Beau coming?'*		
04;05	[ˈwiəʊ **ˈtɹʌmˈbəʊn**]	[ˈwɒt ˈdɪt ˈdadi ˈhav ˈɪn ˈhɪdən **ˈmɒ̹ŋki**]	[ˈhiz **ˈdʒʊstɛə**]	[ˈfaɪv **ˈplas** ˈfɔ]
	'wheel trombone'	*'what did Daddy have in Hidden Monkey?'*	*'he's just there'*	*'five plus four'*
04;08	[ˈaɪ **ˈlʌv** ˈju]	[a **ˈlɒv** ˈdaʔ]	[**ˈtʊnz**]	
	'I love you'	*'I love that'*	*'tons'*	
04;09	[**ˈsʌkəz** an ˈsʌ̹kəz]	[ˈsʌkəz an **ˈsʌ̹kəz**]	[**ˈsʊmdɪ** ˈslap mi ˈwaɪʔ ɒn də ˈhand]	
	'suckers and suckers'	*'suckers and suckers'*	*'somebody slapped me right on the hand'*	
05;03	[lɛʔs ˈdu ðaʔ **ˈfʌni** ˈθɪŋ]	[ˈpak **ˈɒp**]	[ˈaɪm ˈʊɹəli **ˈhʊŋgɹi** ˈmʌmi]	

	Home	**Blend**	**Local**	**Other**
	'let's do that funny thing'	*'pack up'*	*'I'm really hungry Mummy'*	
05;08	[hav ˈaɪ **ˈθəŋk** ˈju]	[ˈwaɪ dɪd aɪ **ˈd̹ɒs** seɪ ˈjɛs]	[ˈaɪ hav ˈaɪ ˈaɪ ɒˈɹədi **ˈdʊn** eɪ ˈtɛn]	
	'have I sunk you?'	*'why did I just say yes?'*	*'I have I- I- already done A ten'*	
06;00	[ˈaɪ **ˈlʌ̹v** ˈju ˈmʌm]	[əʊ ˈðɛəz ði **ˈʊ̹ðə** ˈwɒn]	[ˈjɛə ˈðɛ ˈɡəʊɪn ˈɒn nə **ˈfʊʊnt**]	
	'I love you Mum'	*'oh there's the other one'*	*'yeah they're going on the front'*	
06;02		[s **ˈɹə̹n** aʊʔ ə ˈbaʔəʊi]	[**ˈɡʊmz**]	
		'it's run out of battery'	*'gums'*	
06;06	[ˈjəp ɪʔ **ˈdəz** lʊk ˈsɔ]	[ˈwɛn ˈwɛn ˈhi ɡʊu **ˈʊ̹p**]	[ˈakəli **ˈkʊʊʃt** wɒn]	
	'yup it does look sore'	*'when when he grew up'*	*'actually crushed one'*	

Table 25: Examples of STRUT *realisations*

Figure 23 above shows that Henry's early articulations of STRUT are typically home or blended variants. Local variants appear gradually, usually alongside both home and blended variants, coming eventually to dominate the articulations over time (from 5;03). Notably, home and blended variants reduce but do not disappear altogether.

Rounded variants appear alongside the unrounded home variants throughout the recording period. That is, articulations of the home, unrounded STRUT vowels are consistently present, even though the local influence is increasingly evident. At 2;08, the first example of a local, back rounded STRUT vowel is recorded, though its articulation sounds more open than is typically found in local varieties:

Example (3) ‘tea Mummy’
['ti **'mʊ̞mi**]

Examples of the local, rounded variant start off as sporadic, with home realisations dominating the data until 5;03. The recording made at 5;03 is more than twice as long as most of the other recordings, at 32 minutes; however, the dominance of local vowels is apparent from the beginning of the recording. At this point, Henry has been at primary school for a full academic year. Following on from this, local variants of STRUT are more common, although this is not a consistent pattern. At 5;05 and 5;08 local vowels are fewer (40% and 15% respectively) and at 6;04 they do not occur at all, but the number of tokens of STRUT in this recording is very small (N=3). Peaks in the number of tokens in this lexical set are related to the activity being recorded. For example, at 4;08, we are baking, and there are several repetitions of ‘lumps’ and ‘Mummy’; at age 5;03 many repetitions are of the word ‘hungry’ as Henry is wheedling for something to eat. At 6;04, where we see only three tokens of STRUT, Henry is not particularly talkative, but there are also no topics which elicit STRUT tokens readily. Neither the home nor blended variants appear to have diminished completely (see Figure 23).

Figure 24 expresses Henry’s realisations as a local/non-local percentage of the total number of articulations in any recorded sessions. Blended variants have been grouped alongside home variants as they still exhibit influence of the home varieties. Home and blended variants dominate the recordings with local vowels only appearing sporadically until 5;01 when they begin to appear more regularly. Even then, recordings at 5;05 and 5;08 mainly feature home variants, and at 6;04 Henry exhibits *only* home variants, though it should be noted that this represents only two tokens.

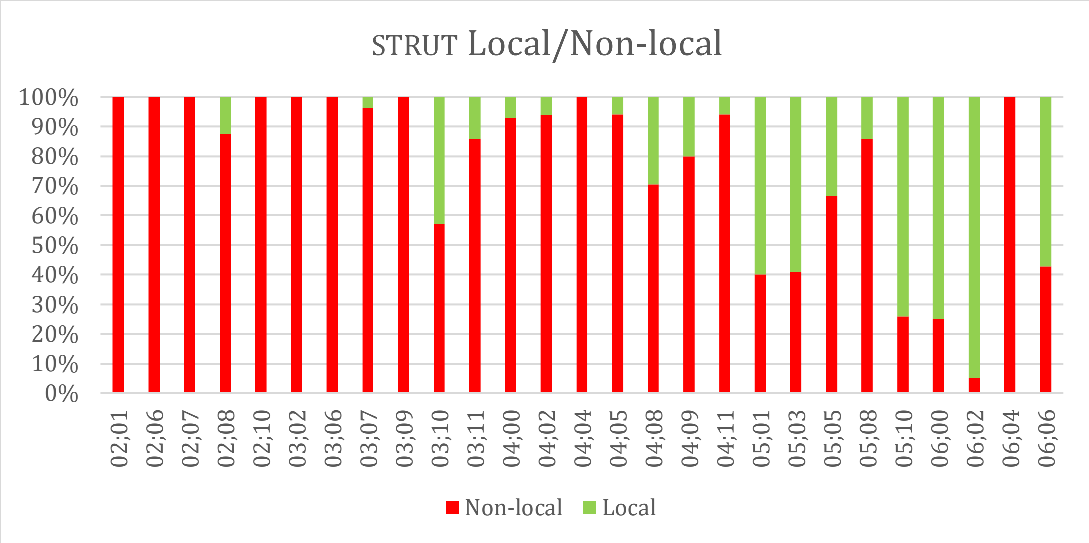


Figure 24: Acquisition of the STRUT *lexical set (including Mum/Mummy)*

Home variants vary from an open unrounded vowel as in the home variety, to a central unrounded vowel, also found in the home variety in stressed positions, for example at 6;06:

Example (4) ‘Yup, it does look sore’
['jəp ɪʔ **'dəz** lʊk 'sɔ]

Blended variants vary from a little bit of rounding on an open back vowel [ʌ̹], to full rounding, analogous to LOT words, [ɒ]. For example, at 2;10:

Example (5) ‘I’m running around’
['aɪ əm **'ʋɒ̠nɪn** ə'ʋaʊnd]

From age 3;09 onwards, we see some blended variants taking on a closer articulation but with less rounding than the local variant [ə̹] and [ʊ̜], as seen here at 6;0 and 6;02:

Example (6) ‘Oh there’s the other one’
[əʊ 'ðɛəz ði **'ʊ̜ðə** 'wɒn]

Example (7) ‘It’s run out of battery’
[s **'ɹə̹n** aʊʔ ə 'baʔəʊi]

As outlined in 4.3.3, an acoustic analysis was performed to support the impressionistic analysis and provide a visual representation of the distribution of Henry's realisations. The first and second formants have been plotted on to a scattergram (see Figure 25 for example) to show the relationship between vowel height and tongue advancement/retraction of a sample of realisations between the ages of 3–6 years. Axis ranges have been set consistently to represent the whole of Henry's vowel space throughout the thesis so that vowel plots may be compared. Though lip rounding cannot be seen here, we would expect home vowels to feature a more open jaw position than local variants of STRUT.

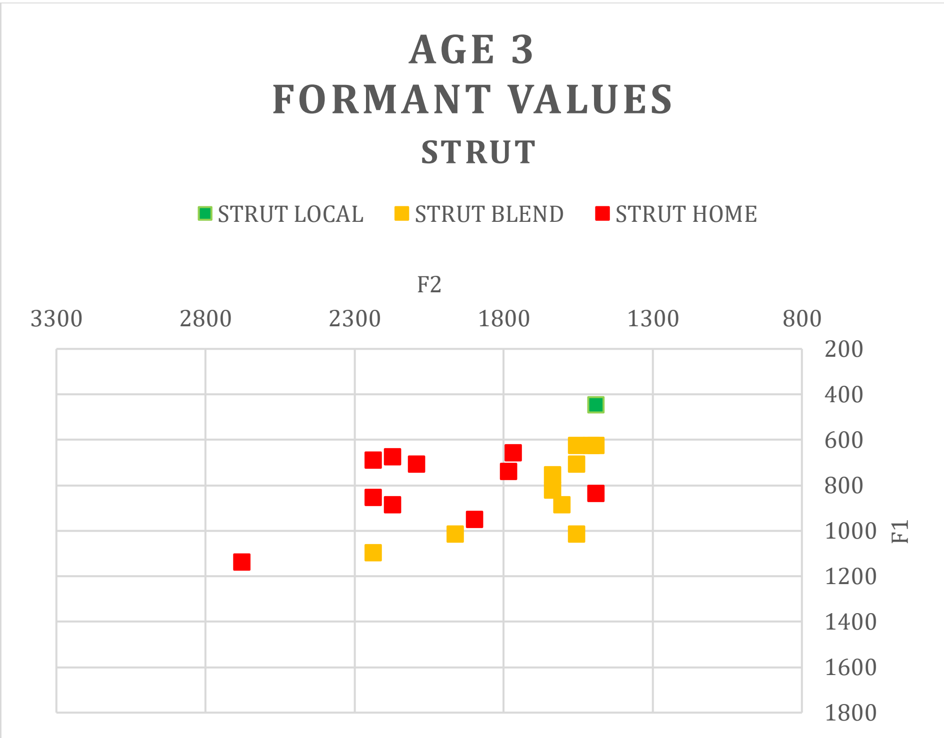


Figure 25: Formant values STRUT *lexical set at age 3*

At the age of 3, a single local realisation of STRUT was available for acoustic analysis as few local tokens appeared altogether. Figure 23 shows that local realisations at this age were rare, and all but one of these were unsuitable for acoustic analysis due to overlapping speech or background noise. This example can be seen to have a lower F1 than the variants classified as home or blends (Figure 25), indicating its close jaw position. The F2 value is at the low end of values seen in blends and home realisations, placing it at the back of Henry's vowel space. Home variants mostly feature higher F2 values and are rather widely dispersed. We will see the placement of these realisations in relation to FOOT variants, below, where some interesting patterns emerge.

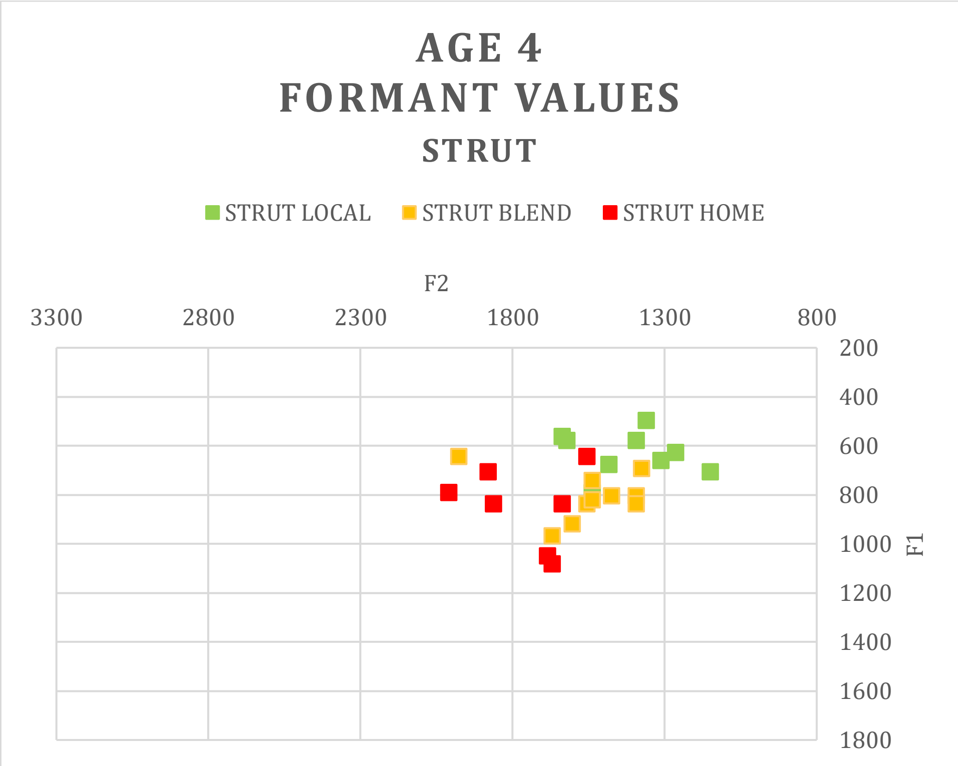


Figure 26: Formant values STRUT *lexical sets at age 4*

At age 4 (Figure 26) we see many more local realisations of STRUT, which cluster together as relatively high, back vowels. The general pattern seen at age 3 continues; the home and blended realisations overlap each other in the vowel space, with the home realisations generally featuring higher F2 values and no rounding. Broadly speaking, realisations at this age appear to be less widely dispersed than at age 3, and there is less variation in F2 than was seen at age 3. In comparison to the reference values seen in Vorperian and Kent (2007), the F1 value of Henry's local realisations is in the region of the average F1 of their high back vowel in 4-year-olds (approximately 500 Hz), but F2 in Henry's speech is much lower than the approximate 2100 Hz average seen in their averages for back vowels. His values are much closer to Busby and Plant's (1995) values (Figure 6) for F2 of /ʊ/ in 5-year-old Australian children (approximately 1100 Hz).

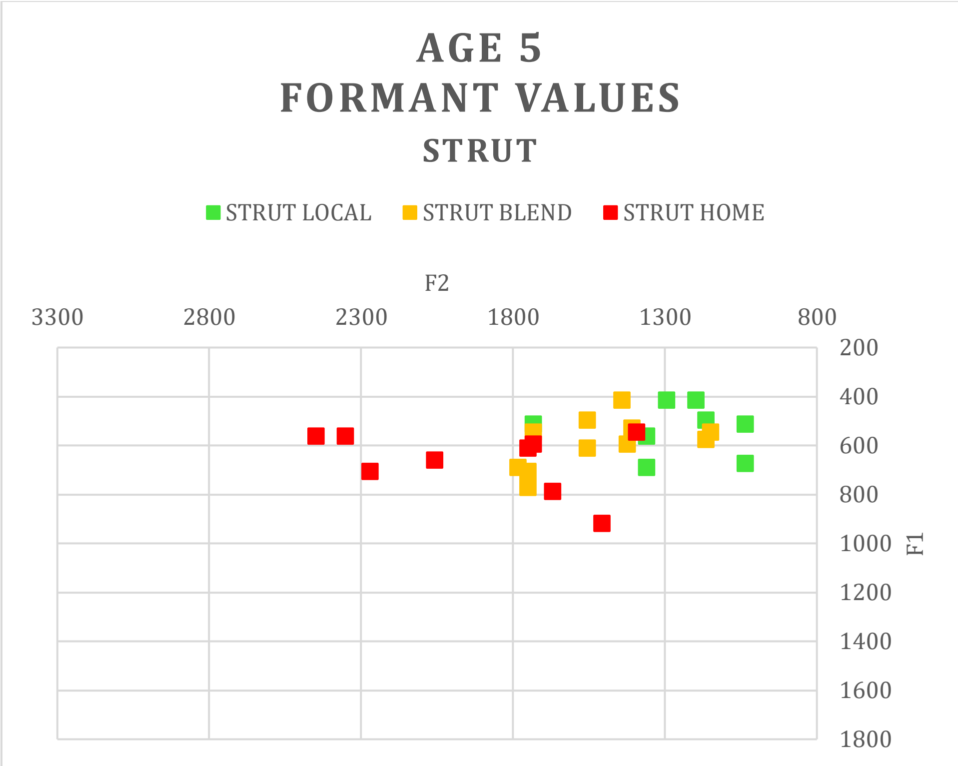


Figure 27: Formant values STRUT *lexical sets at age 5*

At age 5 (Figure 27) a lowering of F1 is evident in the more open, home variants. There are also several home variants which feature very high F2 values. These include higher F2 values than Busby and Plant's (1995) averages for 5 years olds, which place /ʌ/ at around 2100 Hz for F2, and their F1 average of 1200 Hz reflects a much more open vowel than seen here, however, their data was based on a laboratory experiment. Henry's data is naturalistic, and therefore we should consider Keating and Huffman (1984) or Veatch's (1991) work which showed much more variability in vowel realisations in the read/naturalistic speech of adults. Impressionistically, the most close, front example of STRUT here sounds like a raised and advanced schwa. The F1 values of the closer vowels and F2 of the most back vowels are also lowering in comparison to the data from ages 3 and 4 as would be expected due to his growing vocal tract.

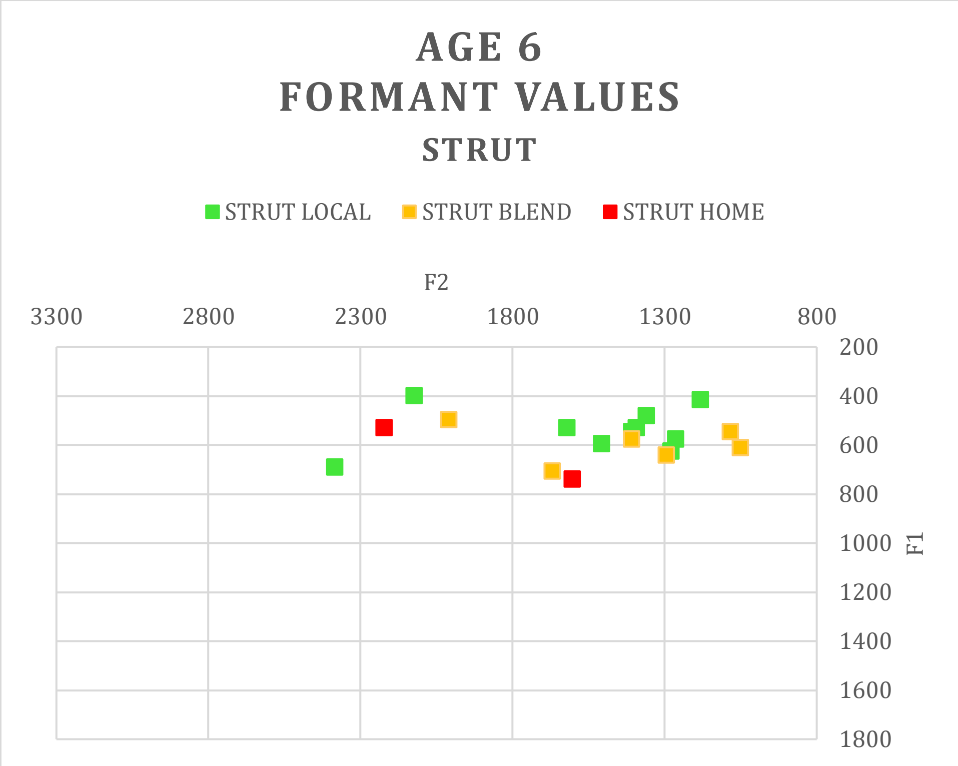


Figure 28: Formant values STRUT *lexical sets at age 6*

Finally, at age 6, we see fewer home variants suitable for acoustic analysis (Figure 28), reflecting the reduction in appearance of home variants seen in Figure 23. In comparison with the data from age 3, where the lowest F2 value across all the variants was 1488 Hz, the lowest F2 seen at age 6 is 1050 Hz. Again, these values are much lower than those seen for F2 in Vorperian and Kent's (2007) corner vowels but are more similar to Busby and Plant's Australian /ʊ/ formant values. In fact, Busby and Plant's data shows an interesting difference between girls and boys for this vowel, where girls' F2 values were lower than boys. This was not the case for any other vowel they tested (Figure 6). Variability in jaw movement appears to have decreased over time, as seen in the lower range of F1 values. At age 3, F1 values for STRUT vary from 488 Hz to 1017 Hz, while at age 6 this has reduced to between 416 Hz and 741 Hz. These reduced values tie in with the impressionistic analysis, where closer articulations were observed over time, with blends tending to be closer vowels such as an unrounded close vowel, and home variants being realised as schwa. Close vowel F1 values are lower than those seen in Busby and Plant (1995) but are similar to Vorperian and Kent's (2007) averages. As discussed above, the lack of British English dialects in the literature and the dominance of laboratory data makes differences between Henry's formant values and any reference data inevitable.

Having addressed the broad patterns in the development of the STRUT vowel up to the age of 6, we now turn to some specific examples of interest. One month after starting school, at age 4, Henry was introduced to addition in maths. As his teacher's accent differentiated STRUT and FOOT, he interpreted the 'new' word 'plus' as ['plas], which he appeared to categorise as a TRAP word. I noted that this lasted no more than a few days before he realised that 'plus' belongs with STRUT, though the word did not occur in a recording again.

Example (8) 'five plus four'
['faɪv **'plas** 'fɔ]

The formant values for 'plus' can be seen alongside formant values for the STRUT and TRAP lexical sets in Figure 29. Its placement on the other side of the TRAP variants clearly demonstrates its distance from the rest of the STRUT tokens (Figure 29).

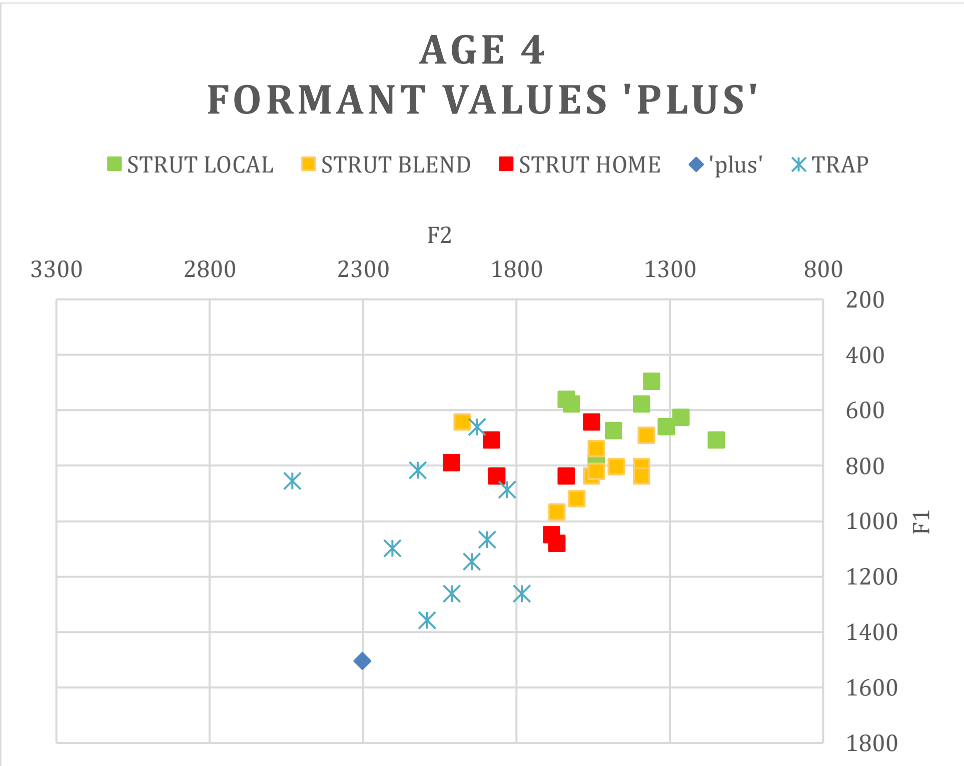


Figure 29: Formant values 'plus' compared to STRUT *and* TRAP *lexical sets at age 4*

At 5;03, during the school holidays, Henry wrote on his lunch box, ‘pac oq’ (he frequently but inconsistently reversed p, b and d in writing at that time (see Figure 30). In the local dialect, a packed lunch is known as a ‘pack up’, and though this is not a lexical item used in Henry’s home, it is a term he has learned at nursery or school. When asked to read what he had written, Henry read ['pak 'ʌ̜p], and in repetition, ['pak 'ʌ̜p] and ['pak 'ɒp]. This was not an isolated example; in Henry’s early spellings, he frequently spelled STRUT words with ‘o’ rather than ‘u’, demonstrating some kind of connection in his mind between some realisations of STRUT and the Roman character ‘o’.

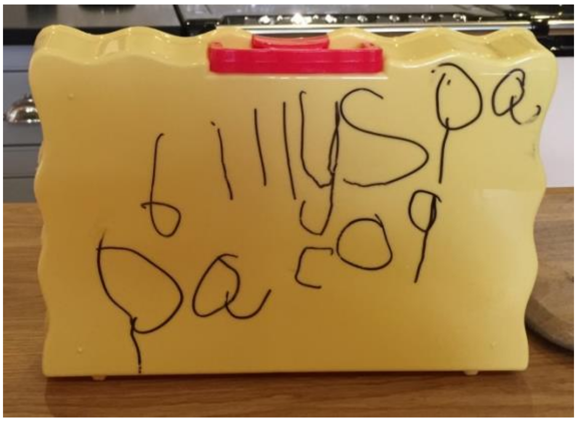


Figure 30: Graphological representation of ‘pack up’

At 6;11 evidence of this spelling idiosyncrasy continued to present itself. Figure 31 shows a picture of the ‘Smurf’ character, ‘Clumsy’, who Henry drew as a gift for a friend (also called Henry).

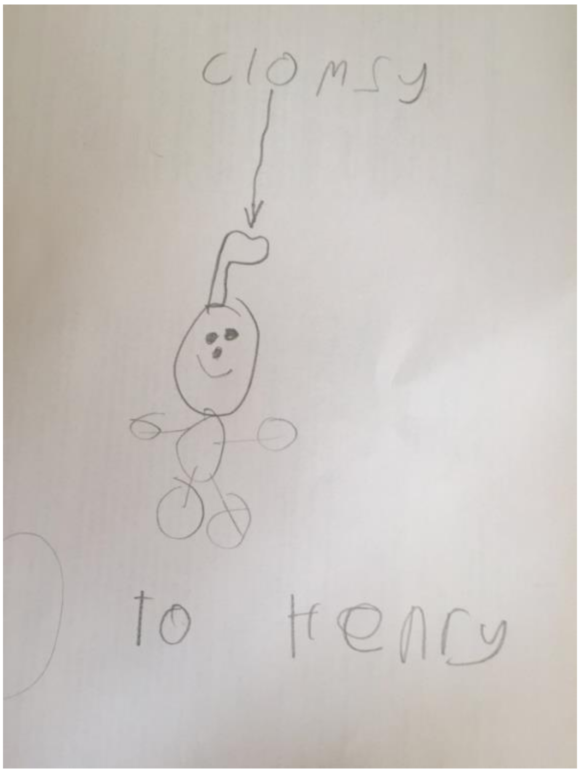


Figure 31: Graphological representation of 'Clumsy'

Further, more detailed analysis was performed on high frequency lexical items to look for evidence of patterns, as they have the potential to behave differently to the rest of the set (Di Paolo et al., 2011).

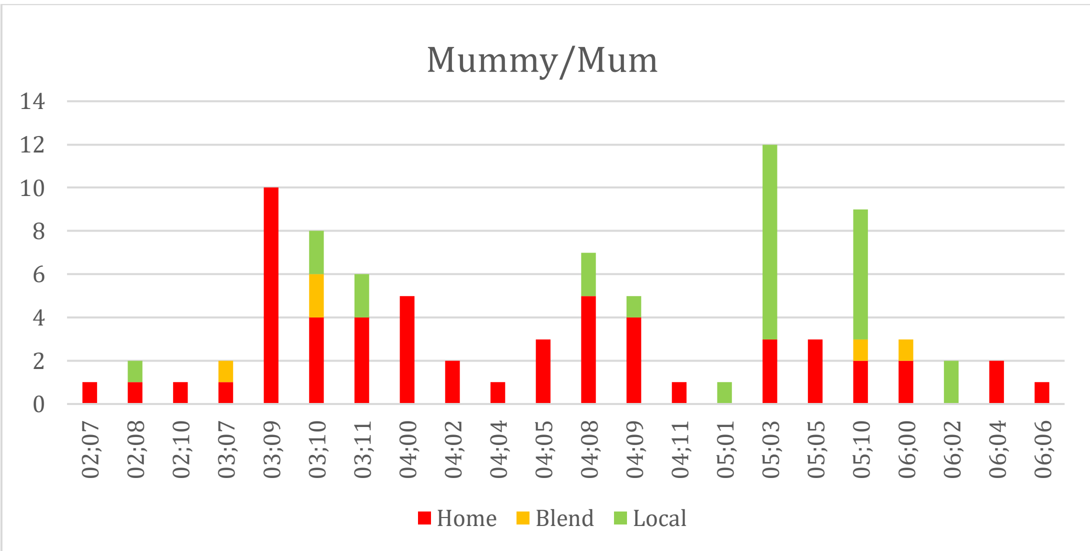


Figure 32: Realisation of Mummy/Mum

‘Mum’ and ‘Mummy’ appear in 22 out of 27 recordings (Figure 32). Vowel realisations in these lexical items are more likely to be realised as a home vowel: 64%, compared to 35% of STRUT tokens with ‘Mum’ or ‘Mummy’ excluded (see Table 26), though more local realisations are found in later recordings, home variants do not disappear altogether. Blended variants are very rare, only appearing in 5 tokens in total (6% of all tokens). This compares to 33% of other STRUT words being realised as a blend.

Tokens	Home		Blend		Local		Other		Total Ns
	Ns	**%**	**Ns**	**%**	**Ns**	**%**	**Ns**	**%**	
ALL STRUT	163	42	104	26	120	31	3	1	390
STRUT (without Mummy/Mum)	107	35	101	33	96	31	3	1	307
Mummy/Mum	56	64	5	6	26	30	0		87

Table 26: Balance of realisations between STRUT *and ‘Mummy/Mum’*

Other high frequency words include ‘just’ and ‘does’. As with the rest of the data, only stressed tokens are considered here. While these are much less frequent than ‘Mummy’, they offer some insight into variations within a single word. Table 27 shows all the variations (N=15) of stressed ‘just’ found in the recordings. The earliest example appears at 3;09 and is a blended articulation, with the vowel height of the home variant and rounding associated with the local vowel. From 3;10 local realisations appear, but not consistently, with blends appearing regularly, including very rounded open vowels. The pattern here appears to be broadly the same as the rest of the STRUT dataset, moving from home to local variants over time, though there are few home variants as this word does not appear in the recordings until rather late on.

Age	**Stressed**	**Ns**
03;09	ʤʌ̹st	2
03;10	d̪ʊs ~ʤʊs	3
04;02	dʊst ~ ʤʊ̜s ~ dʌ̹st ~ ʤɒs ~ dˡʊs	5
04;05	ʤʊst	1
04;11	jɒst ~ ʤʌs	2
05;08	d̪ɒs	1
06;00	d̪ʊs	1

Table 27: Realisations of ‘just’

In comparison with 'just', stressed examples of 'does' appear much earlier in the recordings, and occur more frequently (N=25) (Table 28). The first articulations are unrounded, though at 3.09 centralised realisations appear, showing some movement in height towards the local vowel. Like 'just', local vowels appear at 3;10 but unrounded variants and blends are apparent throughout.

Age	**stressed**	**Ns**
02;08	dʌzː	1
03;09	dəz ~ də̹z	2
03;10	dʊz	1
04;02	dɒz	1
04;05	dəz ~ dɒz	3
04;08	dʌ̹z ~ dʊz	3
04;09	dɒz	1
04;11	dʊzn	1
05;05	dɒz	4
05;08	dʌznt	1
05;10	dʊn ~ dʊz	4
06;02	dʊz	2
06;06	dəz	1

Table 28: Realisations of 'does'

Articulations of 'one', 'ones', 'anyone', 'everyone' and 'once' have also been analysed separately, as in local varieties they can belong to the lexical set LOT rather than STRUT/FOOT. For some speakers, 'one' belongs to the LOT lexical set, and 'once' belongs to the FOOT lexical set, while for others, both 'one' and 'once' pattern with LOT (Wells 1982: 300). In the home varieties, all of these 'one'-related lexical items belong to the STRUT lexical set. Wells classifies this type of difference between the home and local varieties as 'lexical-incidential'. The distribution of the vowels is tied to specific lexical items rather than being part of a larger systemic pronunciation difference. In his discussion of the regional distribution of 'one', Wells specifically mentions Liverpool, Manchester and Sheffield, but not North Yorkshire. However, my own informal investigations[11] suggest that most speakers of North Yorkshire varieties produce 'one' as a member of the LOT set. A smaller subset of northern varieties also produce 'once', 'among', 'none' and 'nothing' with a LOT vowel (Wells, 1982: 362). Again, personal observations include North Yorkshire varieties in this set.

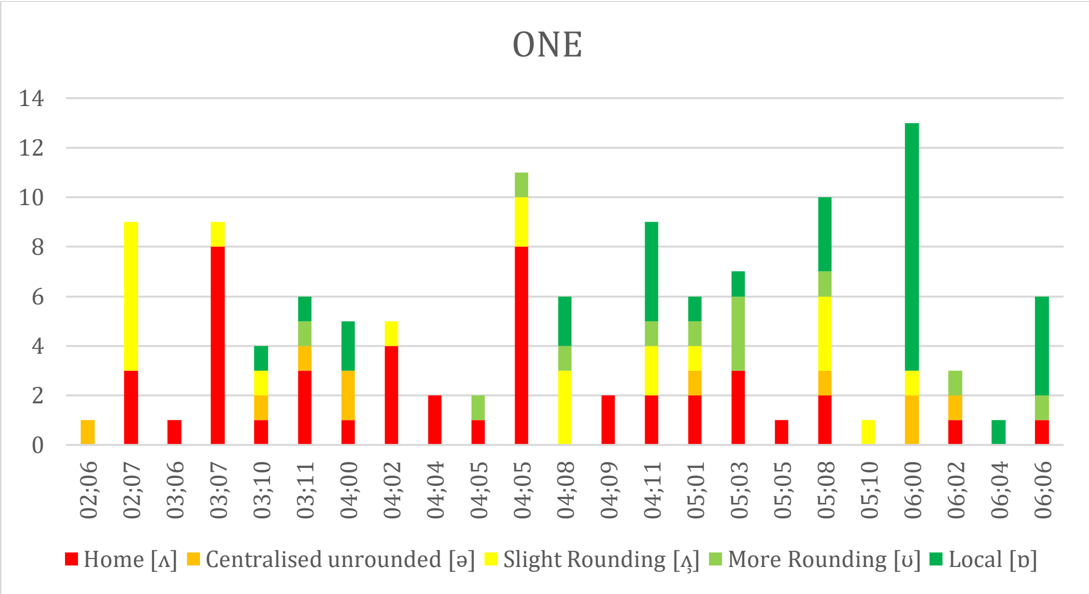


Figure 33: Realisations of 'one'

An analysis of the realisations of 'one' revealed a range of articulations which have been classified according to vowel height and rounding (see Figure 33). The earliest recorded realisations were of a central unrounded vowel; these were followed by realisations which would have been heard at home, but slightly rounded variants also appeared early at 2;07. Local LOT variants of these lexical items do not start to appear until 3;10 — around the same time as local vowels are beginning to appear regularly in the STRUT lexical set (see Figure 23). Like STRUT, however, while the local LOT variant becomes more prevalent over time, the home, blended and local STRUT variants continue to be articulated alongside it.

Realisations of 'one' were divided into pronouns and numerals to check whether there is any systematic variability in their realisations. While there is some variance in the

[11] Including personal communication with Dr Kate Whisker-Taylor, a socio-phonetician local to the North Yorkshire area.

realisations in each category, all articulation categories are represented in both word classes (Table 29), showing no overall pattern.

	Home [ʌ]		**Centralised unrounded [ə]**		**Slight Rounding [ʌ̹]**		**More rounding [ʊ]**		**Local [ɒ]**		**Total tokens**
	Ns	%	Ns	%	Ns	%	Ns	%	Ns	%	
Pronoun	32	38%	6	7%	12	14%	11	13%	24	28%	85
Numeral	14	40%	4	11%	10	29%	1	3%	6	17%	35
ALL	46	39%	10	8%	22	18%	12	10%	30	25%	120

Table 29: Realisations of 'one' by word class

5.1.2 FOOT

While there are relatively few words in the FOOT lexical set, several occur frequently (Wells, 1982: 133), such as 'put', 'look' and 'good'. Overall, the distribution of this vowel is less than half as frequent as STRUT, as we saw in Gimson's analysis of the distribution of English vowels (Cruttenden, 2014) in section 4.3.

FOOT and STRUT do not contrast in local varieties, but they do in the home varieties, with STRUT being realised as [ʌ ~ ə] and FOOT typically realised as [ʊ̟], though unrounded variants occur occasionally. Wells (1982: 133) notes that the back rounded variant tends to be linked to rural speech, while more centralised and unrounded variants may be linked to urban speech. The FOOT set is typically realised as [ʊ] in local varieties, but more front realisations may be heard, depending on the speaker.

Auditory impressionistic analysis of the FOOT lexical set reveals a range of articulations from front rounded to back rounded vowels, with some vowels being articulated centrally (see Figure 34).

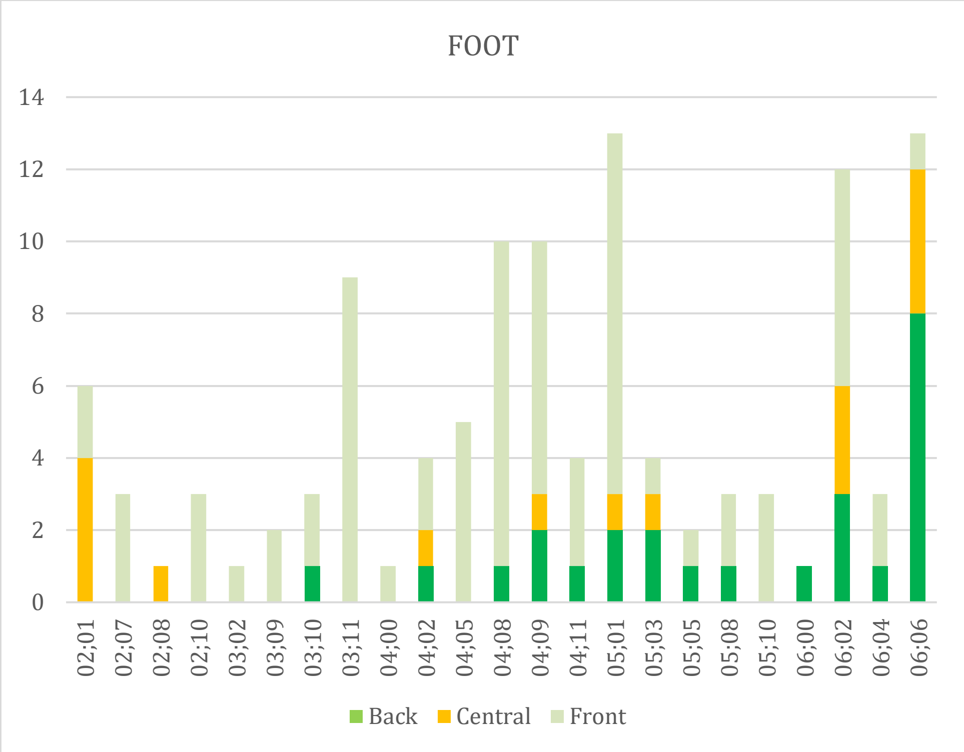


Figure 34: Realisations of the FOOT *lexical set*

Early articulations were mostly very front and rounded:

Example (1) ‘my book’

['maɪ **'bʏːk**]

Example (2) ‘put it down’

[**'pʏt** ɪt 'daʊn]

	Front rounded	Central rounded	Back rounded	Unrounded
02;01	[**'wʏdi**]			[**'wədi**]
	‘Woody’			*‘Woody’*
02;07	['maɪ **'bʏːk**]			
	‘my book’			
02;08		[**'kʊ̟kɜ**]		
		‘cooker’		
02;10	[**'pʏt** ɪt 'daʊn]			

	Front rounded	Central rounded	Back rounded	Unrounded
	‘put it down’			
03;02	[ˈstɛːt ən **ˈlʏk** ən ˈstɛə]			
	start and look and stare’		‘	
03;09	[ˈdəz ˈbeɪbiz **ˈpʏt** ˈtɔɪz ɪn ˈnɛə ˈmaʊf]			
	‘does babies put toys in their mouth?’			
03;10	[ˈpliz ˈkan ju **ˈpʏt** ɪt ˈdaʊn]		[ˈpliz ˈkan ju **ˈpʊt** dɛm ˈdaʊn]	
	‘please can you put it down’		*‘please can you put them down‘*	
03;11	[ˈas ə **ˈwʏmᵇən**]			
	‘that’s a woman’			
04;00	[**ˈwʏʃ**]			
	‘whoosh’			
04;02	[hav ə **ˈlʏk** ɪn ˈmaɪ ˈbɛd]	[ˈaɪ ˈhav **ˈwə̹di** an ˈbʌ̹z]	[**ˈlʊk**]	
	‘have a look in my bed’	*‘I have Woody and Buzz’*	*‘look’*	
04;05	[ˈfɔ ˈju **ˈgʏdi**]			
	‘for you goody’			
04;08	[ɪz ˈaʔ ˈpaːstə **ˈʃʏgə**]		[**ˈʃʊgə**]	
	‘is that pasta (caster) sugar?’		*‘sugar’*	

	Front rounded	Central rounded	Back rounded	Unrounded
04;09	[ˈjɛp **ˈlʏk** ˈteɪk ˈdɛm ˈɒf]	[ˈ**lʊ̟k** ˈjɛə ˈfɪʔs mi]	[ˈ**lʊk**]	[**ˈlək**]
	'yep, look, take them off'	*'look, yeah, fits me!'*	*'look'*	*'look'*
04;11	[ˈkan ju ˈʧu maɪ **ˈbʏk** aʊʔ]	[maɪ ˈmɒnstə **ˈbʊ̟k**]		
	'can you choose my book out'	*'my monster book'*		
05;01	[**ˈpʏʃ**]	[**ˈpʊ̟ʃ**]	[**ˈpʊs** ɪn ˈbuts]	
	'push'	*'pushed'*	*'puss in boots'*	
05;03	[**ˈpʏt** ɪʔ ˈbak ˈɒn ˈðɛn]	[ɒm ˈmaɪ **ˈfʊ̟t**]	[**ˈlʊk**]	
	'put it back on then'	*'on my foot'*	*'look'*	
05;05	[**ˈlʏk**]	[ˈbʌʔ **ˈlʊ̟k**]		
	'look'	*'but look'*		
05;08	[**ˈlʏk**]		[**ˈgʊd**]	
	'look'		*'good'*	
05;10	[jʊ **ˈkʏd** ˈiʔ ˈðəm]			
	'you could eat them'			
06;00			[wɛl **ˈlʊk** aʊ ˈfaːʊ aɪ ˈam]	
			'well look how far I am'	
06;02	[**ˈfʏʊ** ˈstɒp]	[**ˈgʊ̟d** ˈəʊld ˈkɪpə]	[ˈkən ˈju **ˈpʊt** ˈɪt ˌaʊtˈsaɪd]	
	'full stop'	*'good old Kipper'*	*'can you put it outside?'*	
06;04	ˈwɛə ʃəl a **ˈpʏʔ** ˈðɪs		[si **ˈlʊk**]	

	Front rounded	Central rounded	Back rounded	Unrounded
	'where shall I put this'		*'see, look'*	
06;06	[ˈ**lʏk** əʔ ˈɛvɹiwɒnz ˈaɪz]	[ˈa **ˈgə̹d** ˈʤɒb nadim ˈɪz ˈwɪð ˈʊz ˈsɛd ˈbɪf]	[ˈhaʊ ˈkan ə ˈtɜm mə ˈfʊʔ əˈʋaʊnd]	
	'look at everyone's eyes'	*'a good job Nadim is with us, said Biff'*	*'how can I turn my foot around?'*	

Table 30: Examples of FOOT *realisations*

At age 3, front variants continue to dominate. One back rounded realisation was captured at this age. Two separate realisations of the same word, 'put', were articulated with different degrees of tongue advancement within the same turn at talk (Figure 35).

Example (4) 'please can you put it down'
[ˈpliz ˈkan ju **ˈpʏt** ɪt ˈdaʊn]

Example (5) 'please can you put them down'
[ˈpliz ˈkan ju **ˈpʊ̈t** dɛm ˈdaʊn]

Example 5 was the most retracted token at this age (F2=1797 Hz), while Example 4 featured a more advanced tongue position (F2=2269 Hz).

Vowels with a central quality also begin to appear.

Example (5) 'Woody'
[ˈaɪ ˈhav **ˈwə̹di** an ˈbʌ̹z]

Example (6) 'My monster book'
[maɪ ˈmɒnstə **ˈbʊ̟k**]

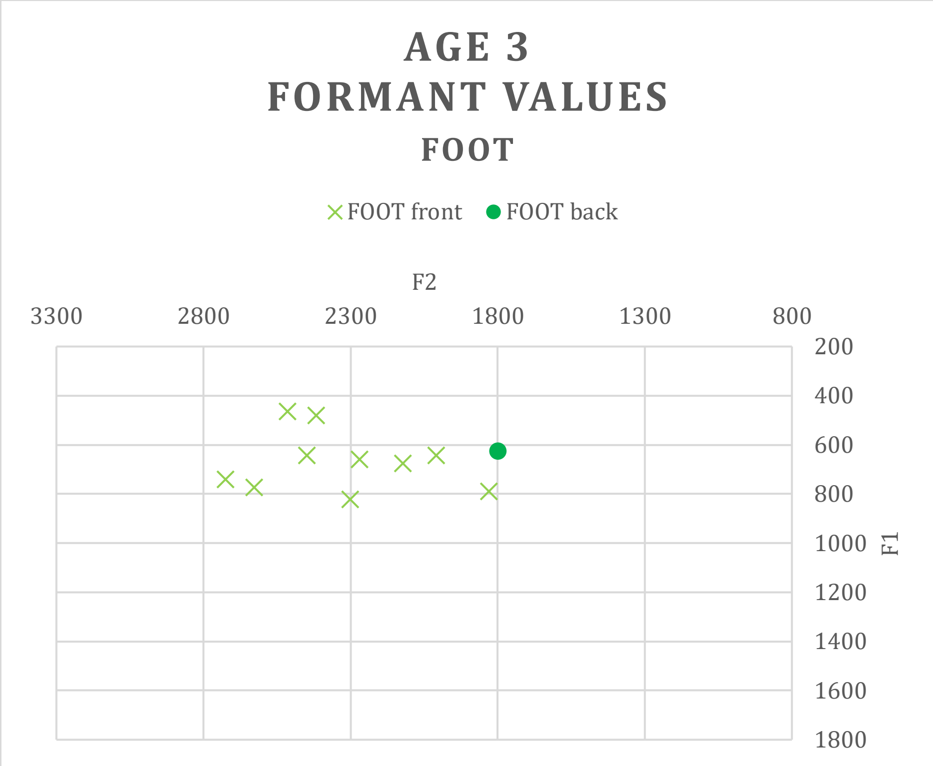


Figure 35: Formant values FOOT *lexical set at age 3*

At age 3, most tokens of FOOT suitable for acoustic analysis were impressionistically categorised as front, with only one token judged to be back, though the F2 of this back token was only marginally lower than those considered to be front. The range of F2 values of this vowel are remarkably varied. As discussed in Chapter 4, formant values were verified via an LPC spectrum as this is considered to be the most accurate way of determining formant values in children. Figure 36 shows the LPC spectrum for the most advanced token of FOOT at age 3. There is a clear second peak at 2724 Hz which corresponds with the impressionistic analysis of a front rounded vowel [ʏ].

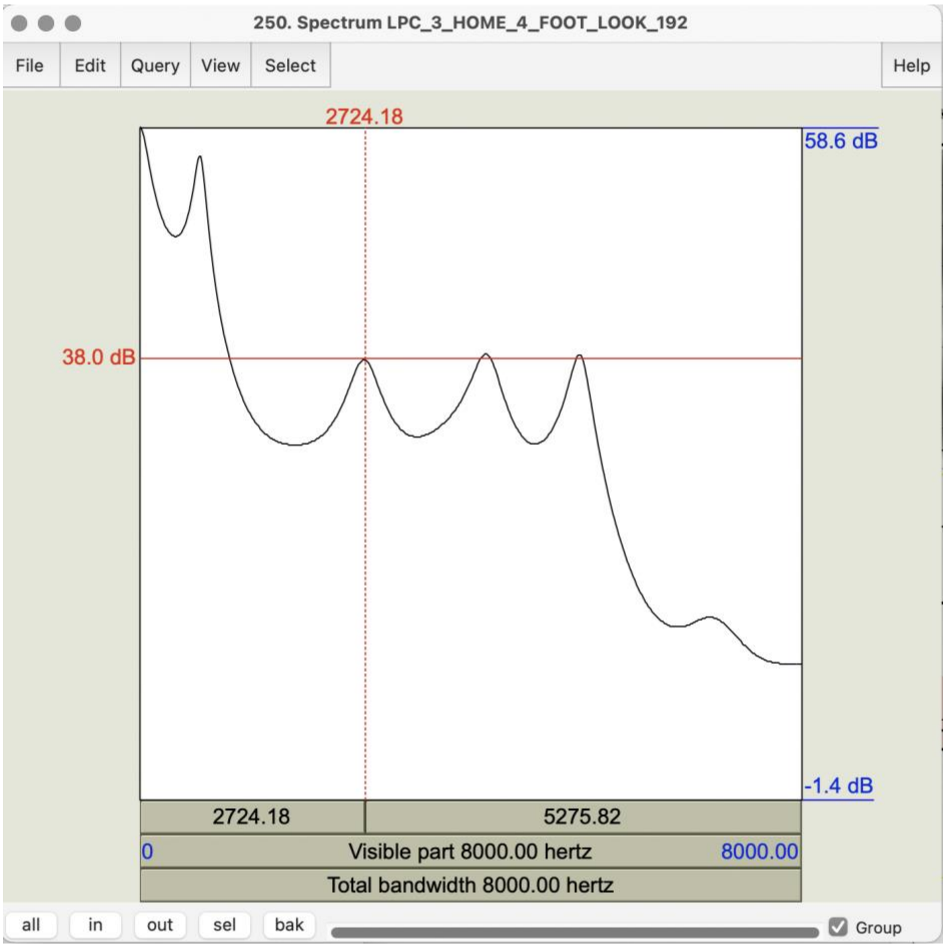


Figure 36: F2 value of front FOOT *vowel*

Figure 37 shows the LPC for the most retracted of the tokens judged to be front. Again, this is a clear picture of F2, demonstrating an 894 Hz range in front tokens alone.

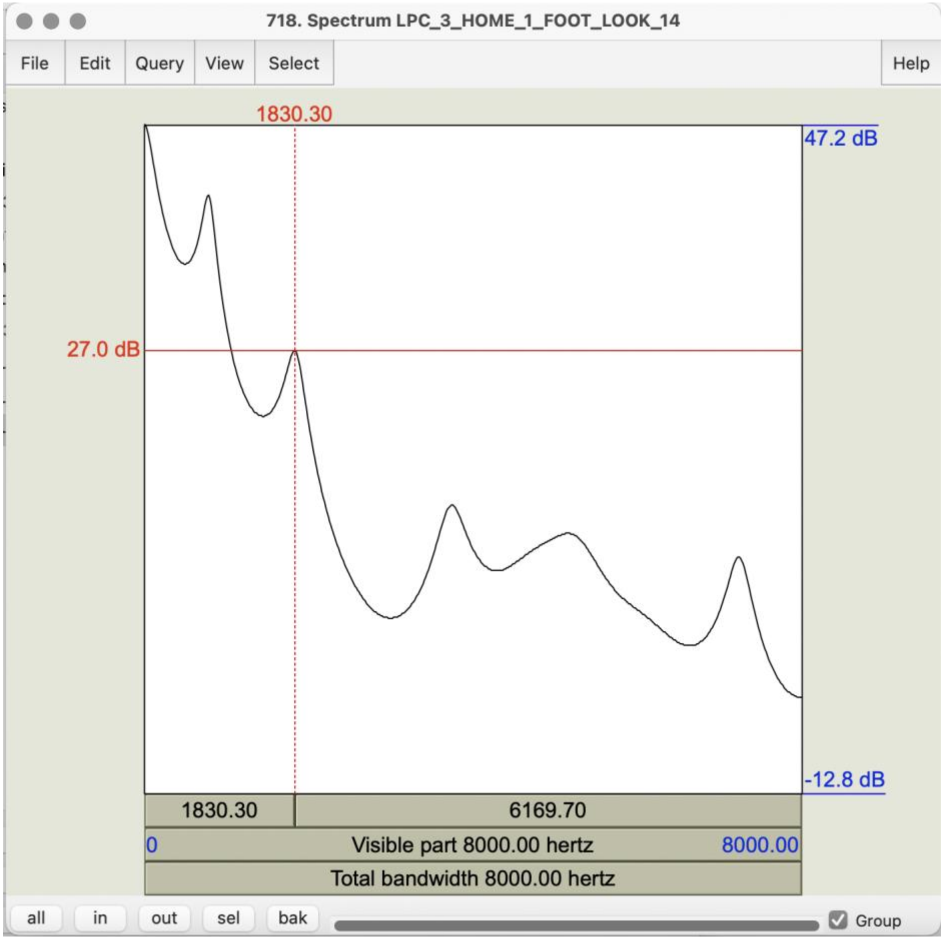


Figure 37: F2 value of retracted front FOOT vowel

At age 4 (see Figure 38) the back vowel [ʊ] is still rare in realisations of FOOT. Front realisations do not feature such an advanced tongue position as at age 3, and more jaw movement is evident, as seen in the wider range of F1 values.

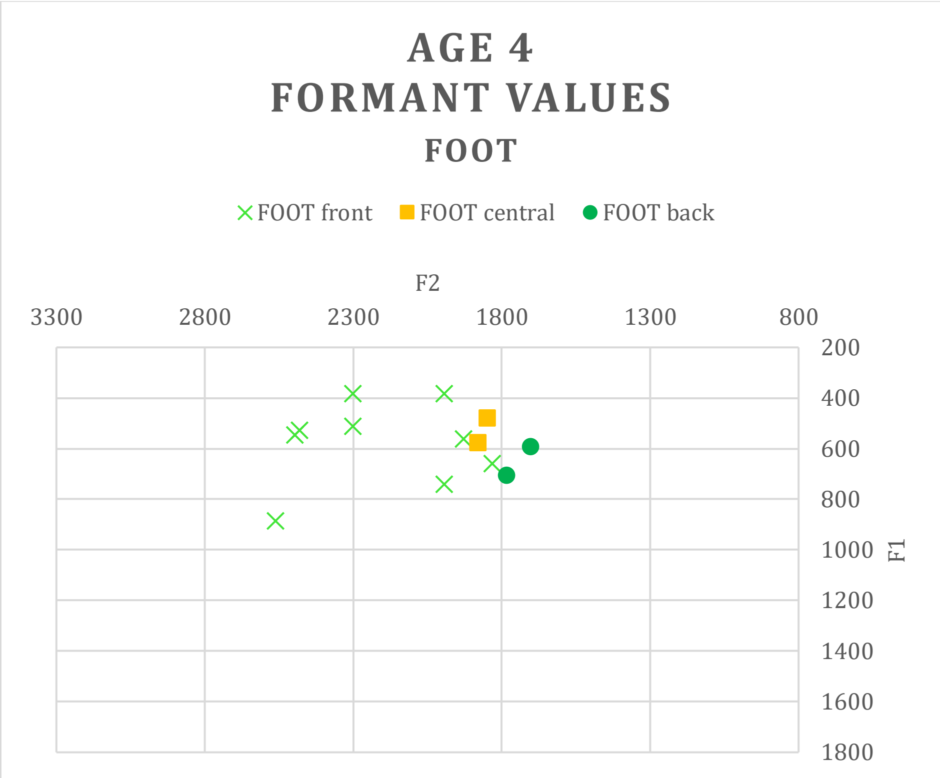


Figure 38: Formant values FOOT *lexical set at age 4*

At age 5, we continue to see a few examples of back rounded and central vowels, but front rounded vowels still appear most frequently (Figure 39). The front vowels seem to have very little difference in jaw position as seen in the tightly clustered F1 values at this age. As with previous examples, there is some small overlap between the F2 values of vowels perceived as front and central or central and back, as the vowels were classified impressionistically.

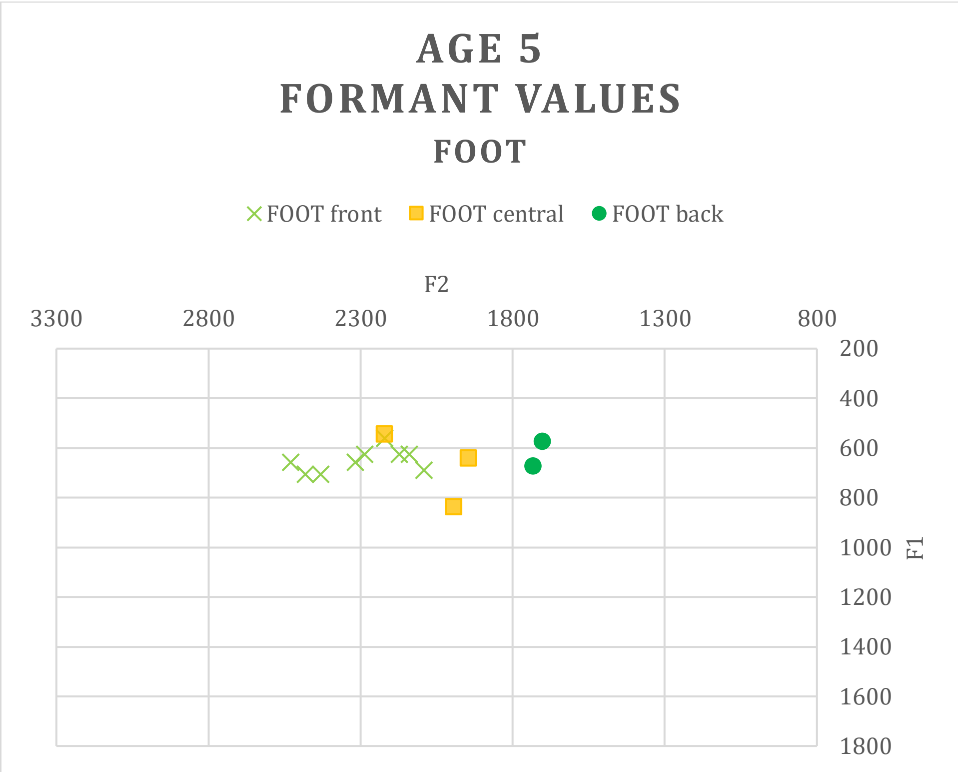


Figure 39: Formant values FOOT *lexical set at age 5*

Age 6 brings an increase in back rounded variants, and a substantial drop in some F1 values (Figure 40). F2 remains similar to previous ages, and there is still a great deal of variation in tongue advancement.

As, gradually, more back tokens emerge as realisations of FOOT, Henry appears to be slowly increasing the distribution of tokens from front to back, with the central tokens representing a stage in that journey.

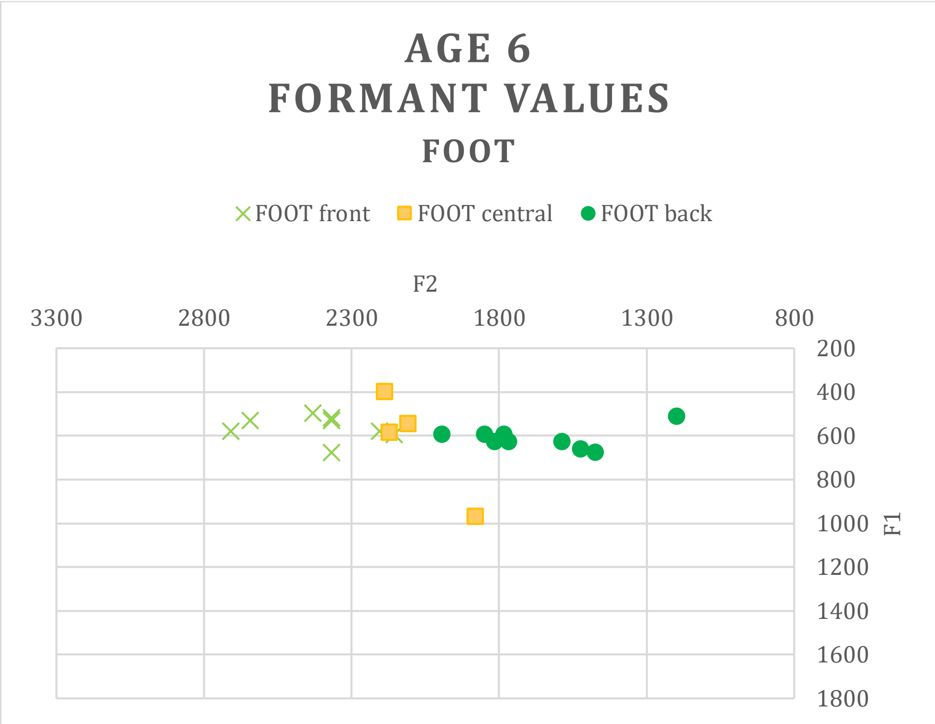


Figure 40: Formant values FOOT *lexical set at age 6*

Henry's F1 is in a similar range to Vorperian and Kent's (2007) average F1 value for the close corner vowels at age 4 (around 500 Hz), though he exhibits variation including both higher and lower frequencies (Figure 38). Vorperian and Kent's (2007) average F2 values at this age for close vowels range from around 2000 Hz at the back corner to about 3400 Hz at the front corner (see Figure 2). Henry's lowest F2 value at this age is lower than this by around 200 Hz. This was confirmed by the F2 shown in the LPC spectrum in Figure 41. His most front variants of FOOT, [ʏ], appear at around 2800 Hz at age 6, by which time Vorperian and Kent's (2007) average front corner vowel value has dropped to around 3000 Hz. Henry's F1 is a good match for Busby and Plant's (1995) F1 values of [ʊ] in Australian children (600 Hz) (Figure 5). Most back variants, articulated in recordings when he was 6 years old, are approaching F2 in the Busby and Plant data (around 1120 Hz), but most of his earlier realisations are much further forward, approaching Busby and Plant's (1995) average F2 value for [i] in boys - 3000 Hz (Figure 6).

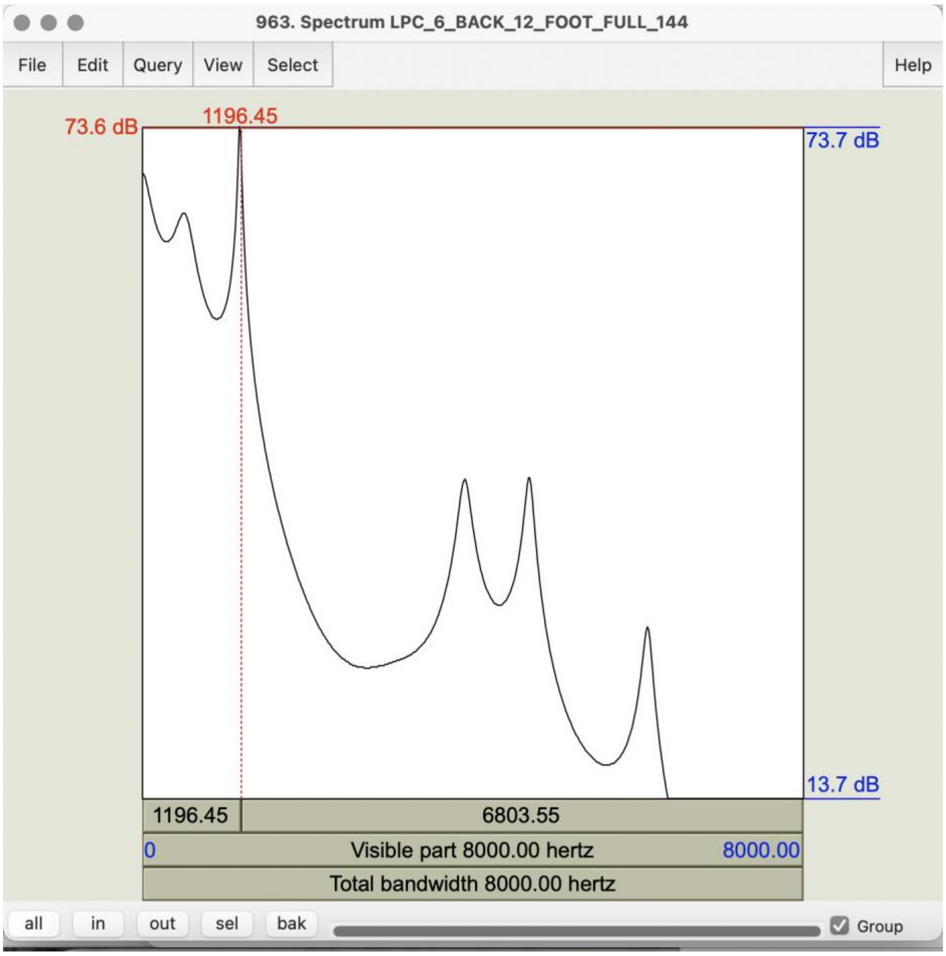


Figure 41: F2 back FOOT *value age 6*

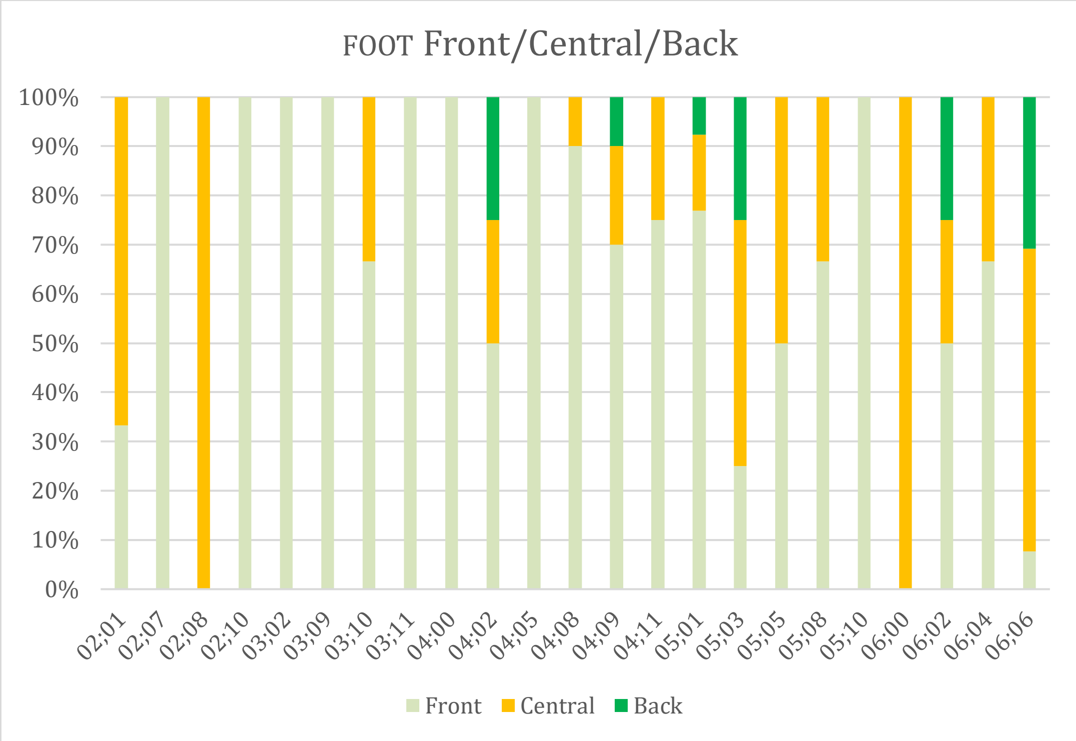


Figure 42: FOOT *front/central/back*

Representing a percentage of total realisations, Figure 42 shows the dominance of front variants of FOOT during most of the recordings. However, as seen in the STRUT analysis above, Figure 43, below, shows the front tokens of FOOT compared to the back and central realisations combined, which are most consistent with the local variety. This allows us to see more clearly the movement of this vowel towards a local articulation over time. Months where 100% of one token occur are misleading as these may represent only one token. If we ignore these, we can see that the front vowels are gradually being replaced by central and back vowels by the age of 6;06. Most of the early articulations are short front close vowels in the region of [ʏ], though isolated examples of the back rounded vowel [ʊ] appear occasionally. From 04;08 onwards, a steadier pattern shifting away from front realisations appears.

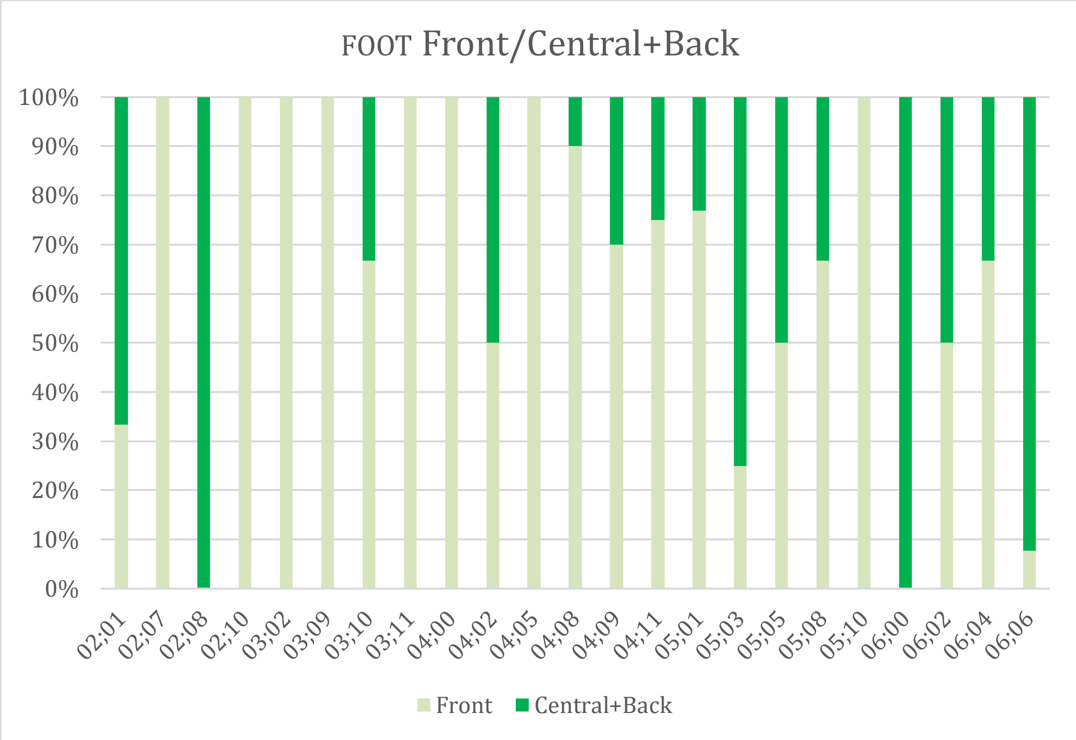


Figure 43: FOOT *front/central back*

Completely unrounded variants appear rarely – at age 2;01 Henry says 'Woody', the name of his toy (a character in the Disney film 'Toy Story') six times (five tokens suitable for acoustic analysis appear in Figure 44). In four out of six utterances, the name is realised with an unrounded vowel ['wədi], whereas in the other two, he is named as ['wʏdi], with a very front close rounded vowel. Much later, at 4;09, the same word appears featuring a similar unrounded vowel.

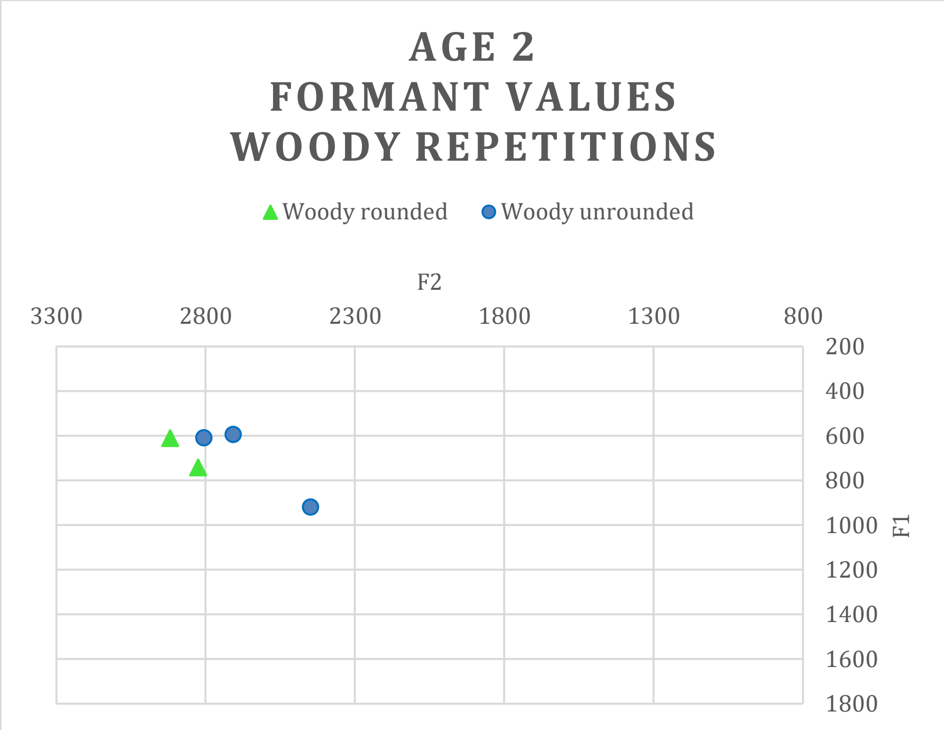


Figure 44: Repetitions of 'Woody' formant values

As seen in Figure 44, unrounded variants may appear with a very advanced tongue position, overlapping with the front rounded variants; they need not be central. In the analysis above, we saw that aged 5, Henry produced several unrounded schwa-like realisations of STRUT in this acoustic space. In the STRUT analysis, we saw one case where Henry mis-categorised a STRUT word, ('plus'), where he appeared to have perceived a phonetic overlap and assigned the word to the wrong phonemic category, which aligned with the TRAP lexical set. Could these unrounded variants of FOOT also be a hypercorrection in which Henry is assigning members of the FOOT lexical set to STRUT? This is possible, as in the local variety, FOOT and STRUT are produced with the same vowel, and the division is on a lexical basis. As Henry's system is in flux as he moves from the home dialect to the local dialect, it is possible that he may assign FOOT members to STRUT. As he produces STRUT with both rounded and unrounded variants at this time, it is plausible that he might do the same with FOOT. There are complicating factors however; as we saw above, Wells (1982) claims that unrounded FOOT vowels can occur in some speakers anyway, so it is likely that Henry has been exposed to this. Also, Henry produces the word 'Woody' six times, very close together, with both rounded and unrounded variants appearing. It seems unlikely that he would change his mind about which category the vowel was assigned to and change it back again in articulations so close together.

5.1.3 STRUT/FOOT comparison

Above, we took a detailed look at each of the STRUT and FOOT lexical sets, considering the realisations of their members in Henry's speech, whether these were closer to the home or local realisations of these vowels, and how these changed as he aged from 3 to 6 years old. Below, the interaction of these two lexical sets is investigated.

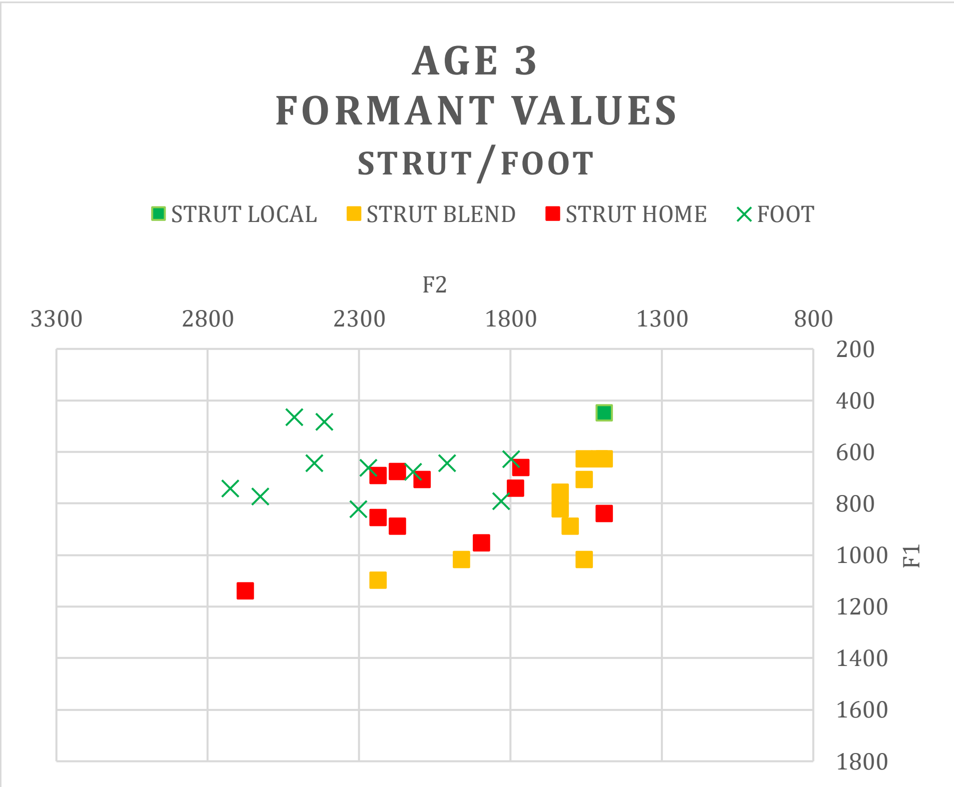


Figure 45: Formant values STRUT/FOOT *lexical sets at age 3*

As seen in Figure 45, at age 3, FOOT tokens are generally more front and close than STRUT tokens, but less so than might be expected based on canonical STRUT realisations. There is some overlap between home realisations of STRUT and the FOOT realisations, but these are usually distinguished by rounding (though we did see some examples of unrounded FOOT, above). STRUT blends occupy a different space from the home variants, and there is only one local realisation of STRUT at this age, which has a more retracted tongue position than the FOOT variants.

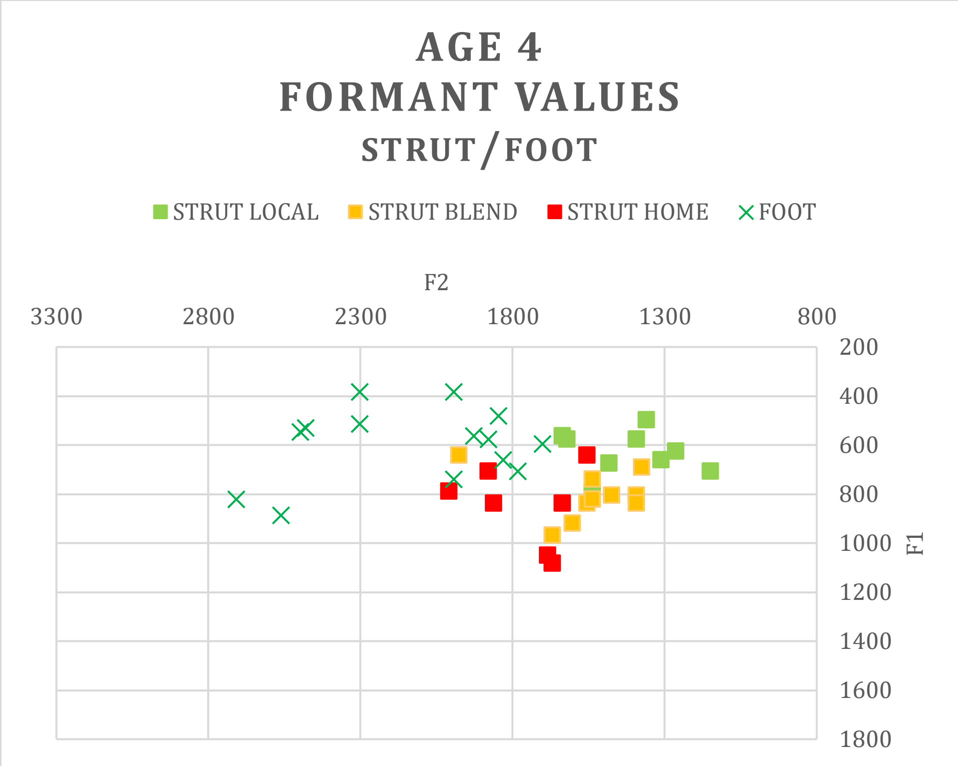


Figure 46: Formant values STRUT/FOOT *lexical sets at age 4*

At age 4, more local STRUT vowels are evident, and these continue to exhibit a difference in tongue retraction to the FOOT vowels, which are often realised as [ʏ]. In the local variety, these vowels are not differentiated. However, it appears that at some level, Henry is differentiating between the articulations of STRUT and FOOT by means of fronting (see Figure 46).

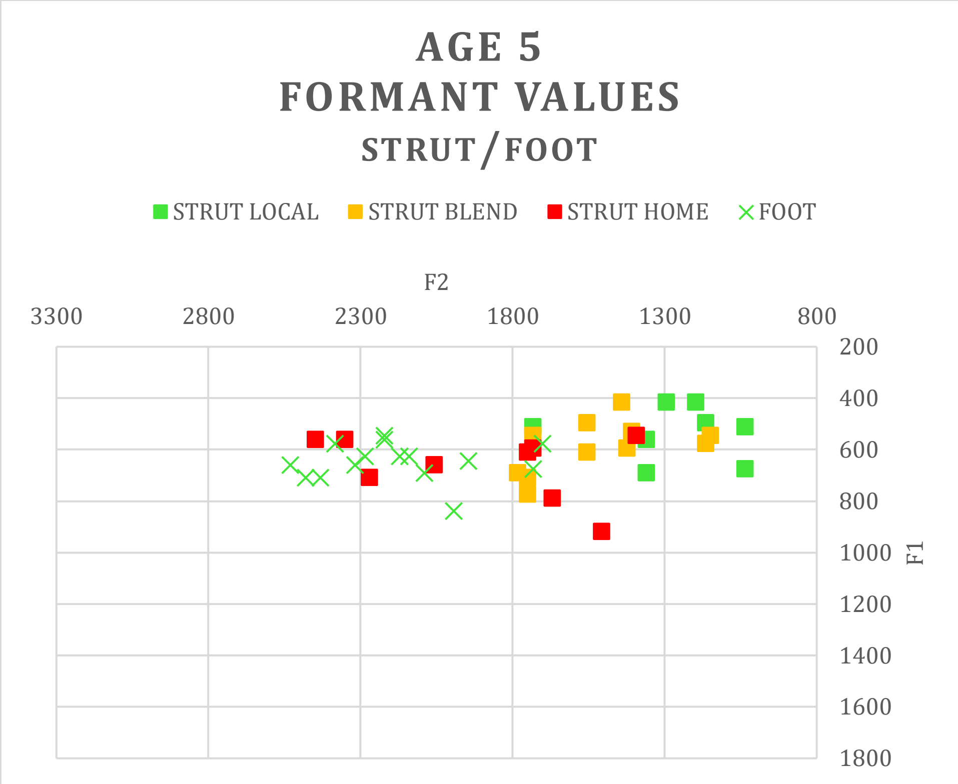


Figure 47: Formant values STRUT/FOOT *lexical sets at age 5*

At age 5, the distinctive distribution of FOOT variants and local STRUT variants continues (see Figure 47). Again, there continues to be some overlap in tongue advancement between the home STRUT variants and FOOT, but these continue to be differentiated by lip rounding.

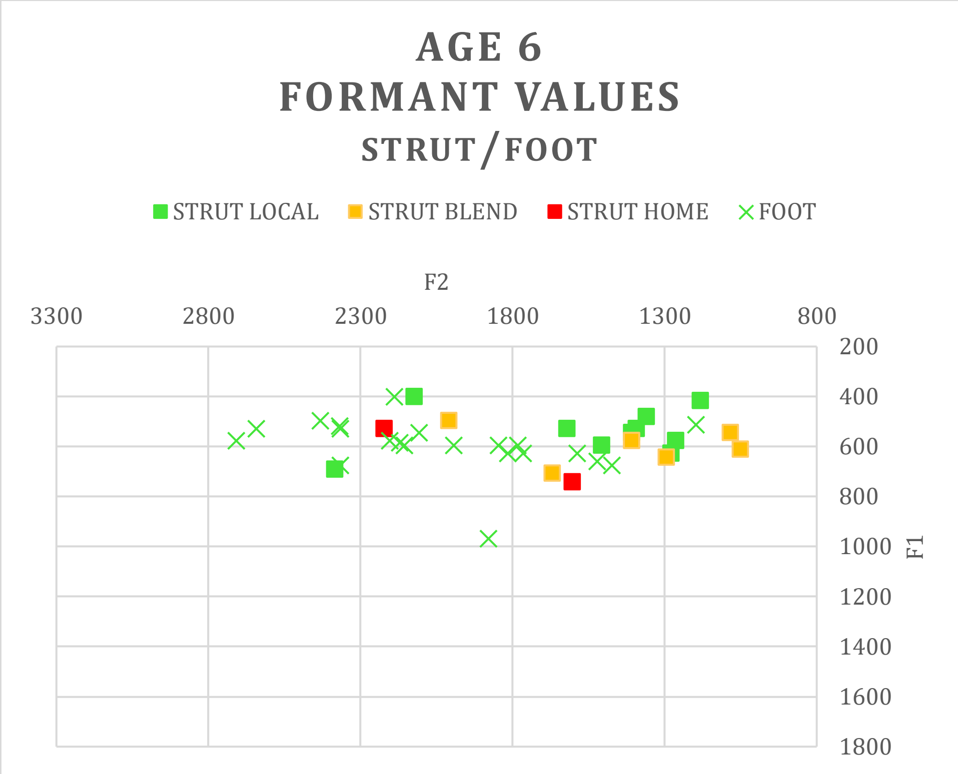


Figure 48: Formant values STRUT/FOOT *lexical sets at age 6*

While many of the FOOT and STRUT variants at age 6 remain distributed as seen between the ages of 3 and 5, there is now also some overlap between FOOT and local variants of STRUT (see Figure 48). FOOT realisations are also articulated further back in the oral cavity than seen at age 5 and below. This may be interpreted as the beginnings of a wider range of articulations being permissible for these categories and the end to the apparent contrast between these tokens. Only a few home STRUT realisations remain as well as a small number of blended tokens, as these are replaced by local articulations. Less variation is also seen in F1 values, as STRUT loses its open articulation, moving towards the close vowel found in local varieties.

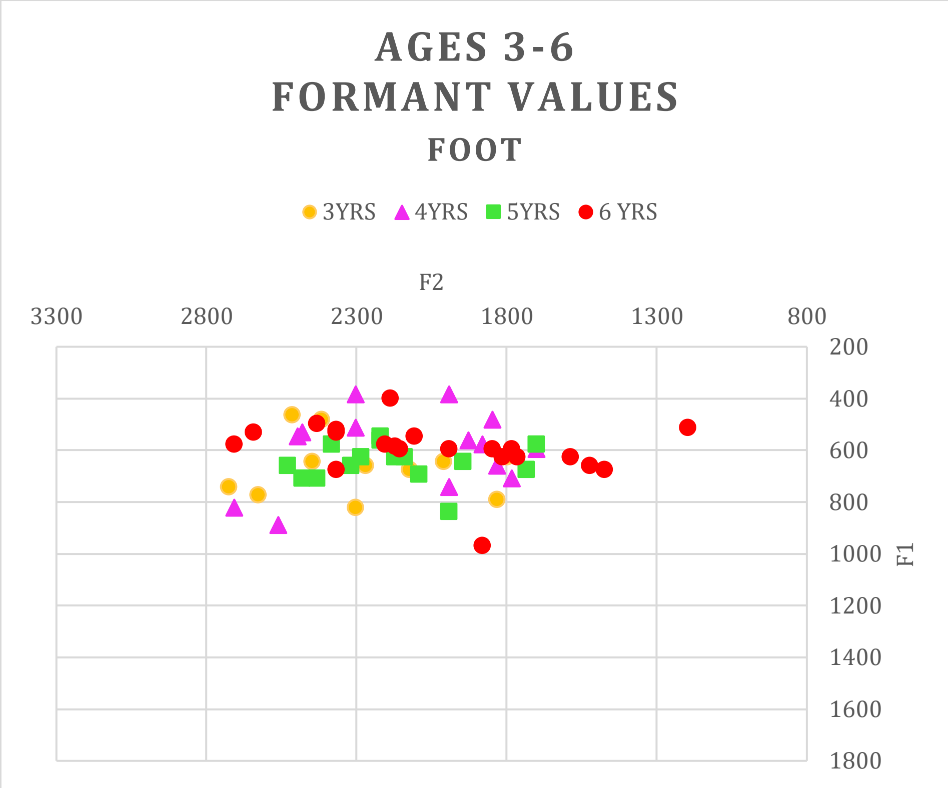


Figure 49: FOOT *Formant values between ages 3-6 years*

Figure 49 shows all F1/F2 mappings for FOOT between the ages of 3 and 6. There is evidence of a slight decrease in formant values over time as would be expected as Henry's vocal tract grows. There is of course overlapping distribution of the vowels, but it is notable that the highest F2 values are seen at age 3 and the lowest at age 6 (see Figure 50).

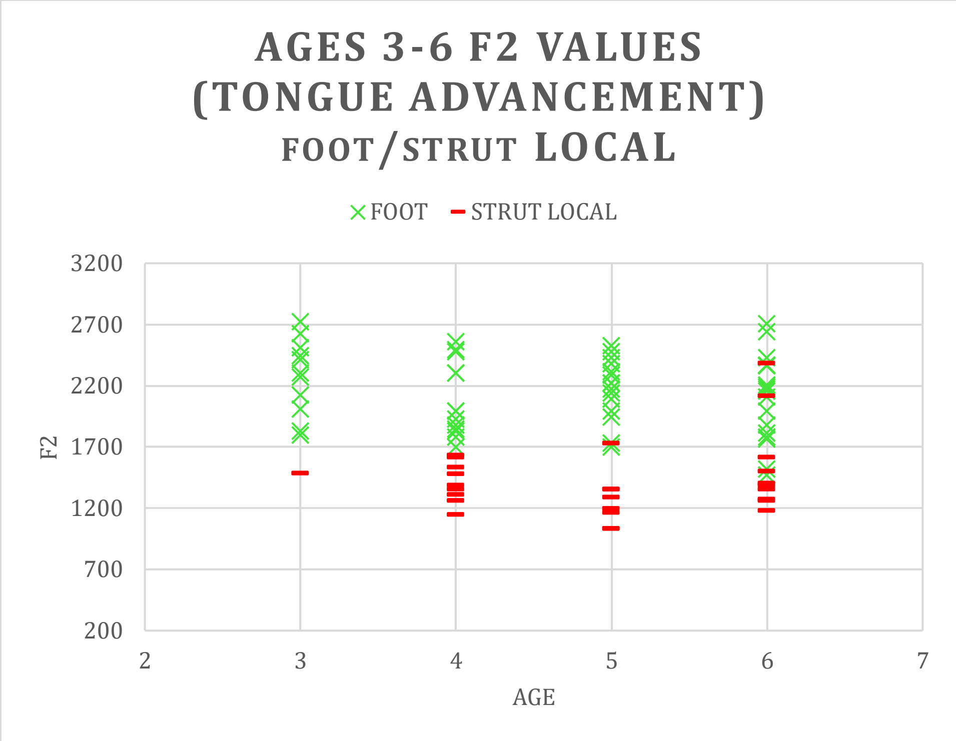


Figure 50: F2 values FOOT/STRUT *local*

Figure 50 shows the initial separation followed by an increasing overlap between F2 values of FOOT vowels and the local realisations of STRUT as Henry ages. Although there is only a single local token of STRUT at age 3, a clear pattern emerges over time, as the F2 values start off completely separate from one another, until at age 5, when they begin to overlap slightly. By the age of 6, there appears to be total overlap between the values, representing a loss of contrast between these two vowels for Henry. No contrast between these vowels was predicted, as while FOOT and STRUT are contrastive in the home varieties, the most salient difference between these vowels is their rounding. In Henry's speech, both local realisations of STRUT and FOOT are rounded; the distinction is made by tongue advancement. He appears to have introduced his own distinction between these vowels. Membership of FOOT and STRUT is not predictable, so the tacit knowledge of which lexical items belong to which lexical set must be acquired token by token. Henry initially organises these two groups according to the phonemic categories found in the home dialects but, surprisingly, realises them in a phonetically distinctive way (Table 31).

	Home dialect	Local dialect	Henry's dialect
FOOT	ʊ	ʊ	ʏ
STRUT	ʌ	ʊ	ʊ

Table 31: A possible STRUT/FOOT *system for Henry at age 3-4*

This is a simplification, of course – Henry's STRUT vowel is gradually moving from the home realisation of [ʌ] to the local [ʊ] at the same time — but the table above is indicative of what is happening in Henry's dialect at a particular point in time. It is noteworthy that the F2 values of Henry's *home* realisations of STRUT overlapped with his FOOT realisations, but these were differentiated by rounding (see Figure 45 and Figure 46).

There is not such a clear decrease in F1 by age as was seen in Figure 50 for the F2 values, as jaw movement appears to be most variable at age 4, with (aside from an outlier at age 6) the highest and lowest F1 values appearing at this age (Figure 51) (see Nittrouer, 1993).

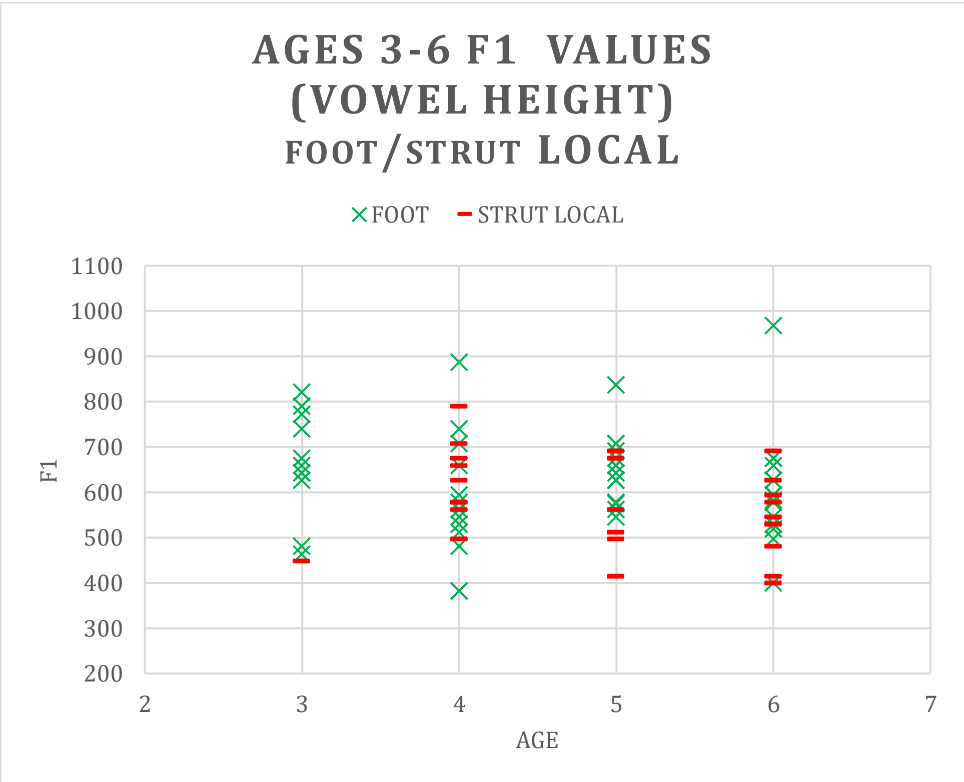


Figure 51: F1 values FOOT/STRUT *local*

A reduction in F1 would be expected over time as the STRUT vowel moves from the more open realisation in the home dialects to the closer local articulation, alongside a decrease due to his growing vocal tract. There is a general trend showing a decrease in F1 over time, though there are some outliers. At age 4, there appears to be a wider range of F1 values than at age 3 or 5. We saw in section 3.2 that amplitude can have the effect of raising F1, but this is not the case, here. This wider range of variability does correspond with Local's (1983) observations, however, that as children acquire language, they have to learn how much variability is acceptable. It is also consistent with Yang and Fox's (2013) observations of wider variability in the vowels of young children which started to become more stable between 5-7 years old.

5.2 PALM/START/BATH/TRAP

The relationship between the lexical sets PALM, START, BATH and TRAP is more complex than the relationship between STRUT and FOOT. In the local variety, there is no contrast between BATH and TRAP, which are both /a/, or between PALM and START /aː/. In the home variety, however, PALM, START and BATH are all realised as variants of /ɑ/, while TRAP is realised as variants of /a/. Of these sets, TRAP is the most frequently occurring, followed by START, BATH and then PALM.

As with STRUT, blended variants are also evident in Henry's realisation of these lexical sets. In the case of START and PALM, the local variant is a long open front vowel, while the home variant is an open back vowel. Henry's blended variants manifest themselves as [ɐ] and [ɑ̟], approaching the backness of [ɑ].

5.2.1 PALM

Many of the PALM tokens are realisations of 'can't'. From the earliest recordings, articulations are mainly local and blended variants, though home variants persist sporadically (Figure 52).

Example (1) 'I can't see it though'
['aɪ '**kɐːnt** 'si 'ɪt 'dəʊ]

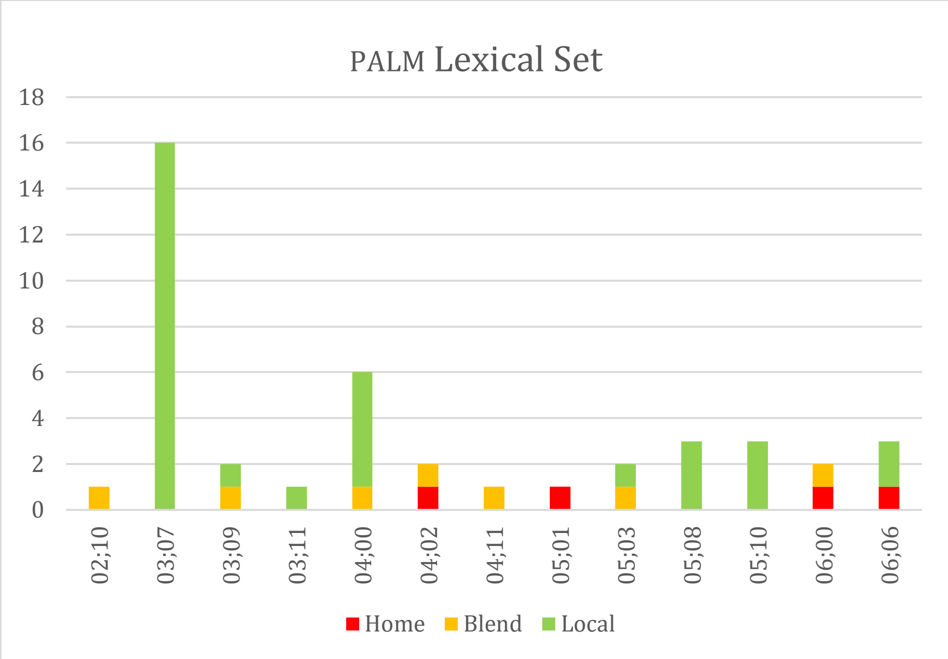


Figure 52: Realisations of the PALM *lexical set*

At age 4;0, Henry produced a typical blended articulation. The PALM vowel in 'can't' is articulated in a mid-position rather than the back position of the home variety, or the front articulation typical of local varieties. Earlier, at age 2;10, we see a different kind of blend, one where the vowel is back, but rounded rather than unrounded:

Example (2) '*can't remember it*'
['**kɒnt** 'ɪmɛmbʊ 'ɪt]

This is an isolated example; no other rounded variants of PALM appear in the data. At this time, Henry was using open, rounded, blended variants for STRUT, so this could be related to that, or possibly rounding could have moved from the /r/ at the start of 'remember' to this vowel (see Ferguson and Farwell, 1975; Studdert-Kennedy and Goodell, 1992).

	Home	Blend	Local
02;10		[**'kɒnt** 'ɪmɛmbʊ 'ɪt]	
		'can't remember it'	
03;07			[**'faːðə** 'kwɪsməs]
			'Father Christmas'
03;09		['ju **'kɑ̟nt** 'katʃ 'mi]	['aɪ **'kaːnt** 'si 'dadiz 'kaː]
		'you can't catch me!'	*'I can't see Daddy's car'*
03;10			['ju **'kaːŋ** 'geʊ 'hɪə]
			'You can't go here'
04;00		['aɪ **'kɐːnt** 'si 'ɪt 'dəʊ]	['aɪ **'kaːnt** bi'kəz]
		'I can't see it though'	*'I can't, because'*
04;02	['aɪ **'kɑnt**]	[**'kɐːnt** 'si]	
	'I can't'	*'can't see'*	
04;11		['ɪ'**ka̠ŋ** 'gəʊ 'ðɛə]	
		'He can't go there'	
05;01	[**'kɑnt** 'si 'ɪʔ]		
	'can't see it'		
05;03		['haʊ 'mʌ̥tʃ 'ɪz **'hɐːf** ən 'aʊə]	[**'kaːnt** 'faɪn nə 'ʊðə 'hɛd]
		'how much is half an hour'	*'can't find the other head'*
05;08			['aɪ **'kaːnʔ** 'kaʊnʔ 'ðɛm 'ɔl]
			'I can't count them all'
05;10			['aɪ **'kaːnt**]
			'I can't'
06;00	[ə'**kɑnt** 'faɪnd 'ɪʔ]	['lɪp **'bɐm**]	
	'I can't find it'	*'lip balm'*	
06;06	['aɪ **'kɑnt** 'iv̞ən 'faɪnd 'ðɛm]		['ju **'kaːnt** 'sɛd 'bɪf]
	'I can't even find them'		*'you can't, said Biff'*

Table 32: Examples of PALM *realisations*

Fully local variants appear frequently around Christmastime at 3;07 in 'Father Christmas', as Henry learnt a song at nursery, where the staff all spoke with local accents:

Example (3) 'Father Christmas'
[**'faːðə** 'kwɪsməs]

The influence of the home varieties is not completely absent, however (see Table 32). This lexical set is the least frequently occurring of those under examination in this section, so the chances of picking these vowels up in the recordings is smaller than the others, which may explain the absence of home variants in the data until age 4;02. The lack of appearance until this age does not mean that they didn't occur at all.

Example (4) 'I can't'
['aɪ **'kɑnt**]

The back vowel's appearance in the word 'can't' does not appear to be tied to this specific lexical item; there are plenty of examples of Henry producing 'can't' with a local vowel:

Example (5) 'I can't see Daddy's car'
['aɪ **'kaːnt** 'si 'dadiz 'kaː]

Of particular note is that there is no steady progress towards local variants fully taking hold as there was with FOOT and STRUT. In the later recordings of PALM, home variants still appear alongside local variants:

Example (6) 'I can't even find them'
['aɪ **'kɑnt** 'iʏ̯ən 'faɪnd 'ðɛm]

PALM and START vowels behave in the same way in both local and home varieties, so we now look to the START lexical set to compare the progress of its vowel realisations with PALM.

5.2.2 START

Unlike PALM, we see home variants of START from the earliest recordings, probably due to its higher frequency. To some extent, it mirrors the pattern we have seen so far in STRUT – early realisations are home variants or blends, but the local realisation appears earlier and more consistently than in STRUT, from 3;07 (Figure 53).

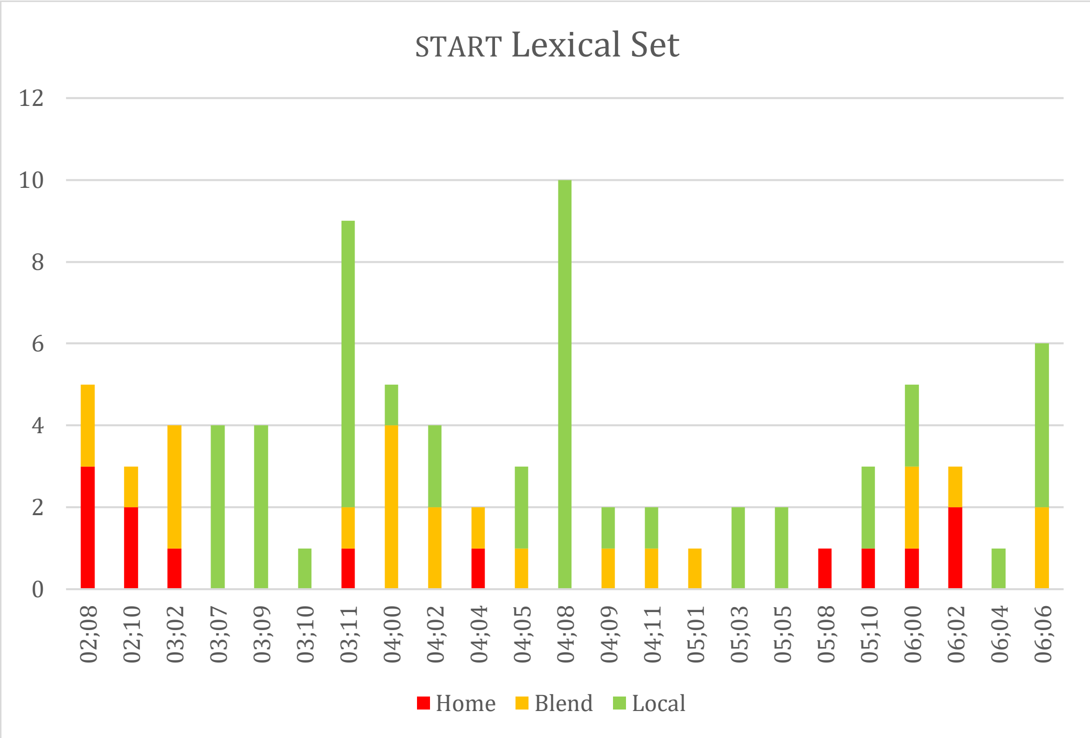


Figure 53: Realisations of the START *lexical set*

It is possible that the START vowel is more salient to Henry than STRUT. Its prevalence in Henry's speech was certainly obvious to us as parents through his articulation of the nursery rhyme 'Twinkle Twinkle Little Star' which had been sung at home since he was a baby but was also sung frequently at nursery, where the staff all spoke with local accents. At age 2;10 he sang:

Example (1) 'twinkle, twinkle, little star'
['wɪŋkʊ 'gɪŋkʊ 'ɪtʊ **'stɐ**]

And by the age of 3;07, we have a recorded example of a fully front realisation of START in this nursery rhyme:

Example (2) 'twinkle, twinkle, Christmas star'
['dwɪŋkʊ 'dwɪŋkʊ 'kwɪsməs **'staː**]

Perhaps the repetitions in song which were so frequent helped to kick-start the process of acquiring the local vowel. Henry had begun attending nursery for four full days each week at age 2;0 when his language development was still at a very early stage featuring only single word utterances, therefore the influence of nursery on his linguistic development was strong.

Home variants are very rare after 3;02, appearing only once at 3;11 and 4;04. Between the ages of 3;07 and 5;05, the local and blended realisations dominate Henry's articulations (see Figure 53). At 5;08, however, the home realisation begins to re-emerge. There was no significant life event which corresponds with this stage, though it was around Christmas time where Henry would have had a few weeks away from school. I remember Henry reading aloud from a book about sharks at that time, which we had already had for a year or so, and many of the frequent appearances of the word 'shark' were realised consistent with the home varieties. Within the overall pattern of the development of the START vowel, the presence of these home variants indicate that the acquisition of the local variant is not yet complete.

There are many examples of lexical items being realised in more than one way within a recording, or even within the same turn. For example, at 3;11, stressed tokens of 'are' are realised as ['ɑ] and as ['ɐ], and at 4;11, within the same turn, 'car' is realised as both ['kaː] and ['kɑ̟]. At 6;0, 'dark grey' is articulated as ['dɑk 'gʊeɪ], ['dɐːk 'gʊeɪ], and ['daːk 'gʊeɪ]. See Table 33 for more examples.

	Home	**Blend**	**Local**
02;08	[jɔ **'ʔɑm**]	[**'kɐ**] [**'kɐ**]	
	'your arm'	*'car, car'*	
02;10	['haɪ 'waɪ 'wɒndə 'wɒt 'ju **'ɑ**]	['wɪŋkʊ 'gɪŋkʊ 'ɪtʊ **'stɐ**]	
	'how I wonder what you are'	*'twinkle, twinkle, little star'*	

	Home	Blend	Local
03;02	[aɪ ˈwɒm maɪ **ˈkɑ**]	[ˈlɪtʊ **ˈstɐː** ˈhaɪ wˈaɪ ˈwɒnda ˈlaɪk ju **ˈɐː**]	
	'I want my car'	*'little star, how I wonder like you are'*	
03;07			[ˈdwɪŋkʊ ˈdwɪŋkʊ ˈkwɪsməs **ˈstaː**]
			'twinkle, twinkle, Christmas star'
03;09			[ˈðɪs ˈðɪs **ˈkaːz** ˈgɒt ə mɪsɪn ˈwiəl]
			'this- this car's got a missing wheel'
03;10			[ˈhu **ˈfaːtɪd**]
			'who farted?'
03;11	[əʊ ˈdɛə deɪ **ˈɑ**]	[ˈdɛə deɪ **ˈɐ**]	[ˈnaʊ ə **ˈkaː**]
	'oh there they are'	*'there they are'*	*'now a car'*
04;00		[ˈtwɪŋkʊ ˈtwɪŋkʊ ˈlɪtʊ **ˈstɐ**]	[ˈbət ɪs **ˈʧaːʤd** ˈnaʊ]
		'twinkle, twinkle, little star'	*'but it's charged now'*
04;02		[ˈɔ ˈmeɪbi ˈpleɪ wɪv də **ˈʃɐˑk** ˈʤɪgθɔ]	[bət **ˈhaːvi** ˈdɒz ˈdat]
		'or maybe play with the shark jigsaw'	*'but Harvey does that'*
04;04	ˈðiz **ˈkɑz** nid tə ˈstɒp]	[ˈstɒp ju **ˈkɐˑz**]	
	'these cars need to stop'	*'stop you cars'*	
04;05	[( ) ˈkɒmbaɪn **ˈhaːvɪstə**]	[**gɪˈtɐː**]	
	'combine harvester'	*'guitar'*	
04;08	[**ˈmaːbʊz**]		
	'marbles'		

	Home	Blend	Local
04;09		[ˈwɒɾ **ˈɐ˞** wi ˈlʊkɪn ˈfɒ ˈfɒ]	[ˈju **ˈaː** ˈwɒt ( ) ˈju ˈseɪɪŋ (ˈlʊk)]
		'what are we looking for?'	*You are what ( ) you saying look'*
04;11		[an də ˈkaː ˈgəʊz ˈhɪə də ˈbɪg **ˈkɑ̟**]	[an də **ˈkaː** ˈgəʊz ˈhɪə də ˈbɪg ˈkɑ̟]
		'and the car goes here, the big car'	*'and the car goes here, the big car'*
05;01		[ˈwɪʧɪz **ˈɐ˞** ˈblak]	
		'witches are black'	
05;03			[ˈɔ ˈðɛə ˈðeɪ **ˈaː**]
			'oh, there they are'
05;05			[ˈlaɪk ˈju ˈsɛd ˈɪt ˈwɒz ˈ**daːk**]
			'like you said it was dark'
05;08	[**ˈɑ** ˈðeɪ]		
	'are they?'		
05;10	[**ˈɑ** ˈbinz]		[ˈiʧ ˈʤɒg ʤʊg ɒv ˈmɪlk ðə ˈkaʊz ˈmɪlk tu **ˈmaːkɪt**]
	'are beans'		*'each jug- jug of milk – the cow's milk to market'*
06;00	[**ˈdɑk** ˈgʊeɪ]	[**ˈdɐːk** ˈgʊeɪ]	[ˈɜ **ˈdaːk** ˈgʊeɪ]
	'dark grey'	*'dark grey'*	*'er, dark grey'*
06;02	[ˈlɛts ˈhav a **ˈpɑti**]	[ˈθaŋk ˈju ˈfɔ ðə **ˈpɑ̟ti**]	
	'let's have a party'	*'thank you for the party'*	
06;04			[ˈbʌɾ ˈɪʔ ˈmaɪʔ **ˈstaːt** əˈgɛn]
			'but it might start again'

	Home	Blend	Local
06;06	[ˈwɛəu̯ **ˈɑː** ˈðeɪ]	[ˈkwaɪt **ˈhɐːd**]	[ˈdəʊnʔ ˈdu ˈɪʔ **ˈhaːd** ˈðəʊ]
	'where are they?'	*'quite hard'*	*'don't do it hard though'*

Table 33: Examples of START *realisations*

In a case similar to 'plus' initially being allocated to the TRAP lexical set by Henry, he made a similar classification error with the word 'scarf'. Though never appearing in a recording, the first time I heard him say this word, he produced it with a BATH vowel [a], which for him overlaps with TRAP rather than START as it does in my variety. In spite of being corrected, [skaf], quite an infrequent word for Henry as I struggled to get him to wear one when he was younger, continued to appear in this form until he was around 8 years old. I cannot be sure that he never said the word as ['ska:f] (local realisation) or ['skɑf], as would be typical of the home variety, as it would not have stood out as unsual in any way. I can only be sure that he produced ['skaf], which belongs in neither dialect, on at least three or four occasions.

In comparison to PALM, START occurs more frequently. In both cases, blended and local variants are present early on, and although home variants still appear sporadically throughout, local variants dominate. As for STRUT above, Figure 54 illustrates the acquisition of local variants of START and PALM as a percentage of the total number of tokens per recording. Here, the data from START and PALM has been merged, as these sets are not differentiated in the home or local accents. Until 3;07, the local variant did not appear, but at that point it appeared in all potential lexical items. The pattern of acquisition of this variant is mixed; in some recordings, only the local variant is found, while in others, only the home variant is present. Viewing the variants as a percentage reveals an interesting pattern, where the home variant appears in waves, with the local variant not being fully established, even at the end of the data collection period.

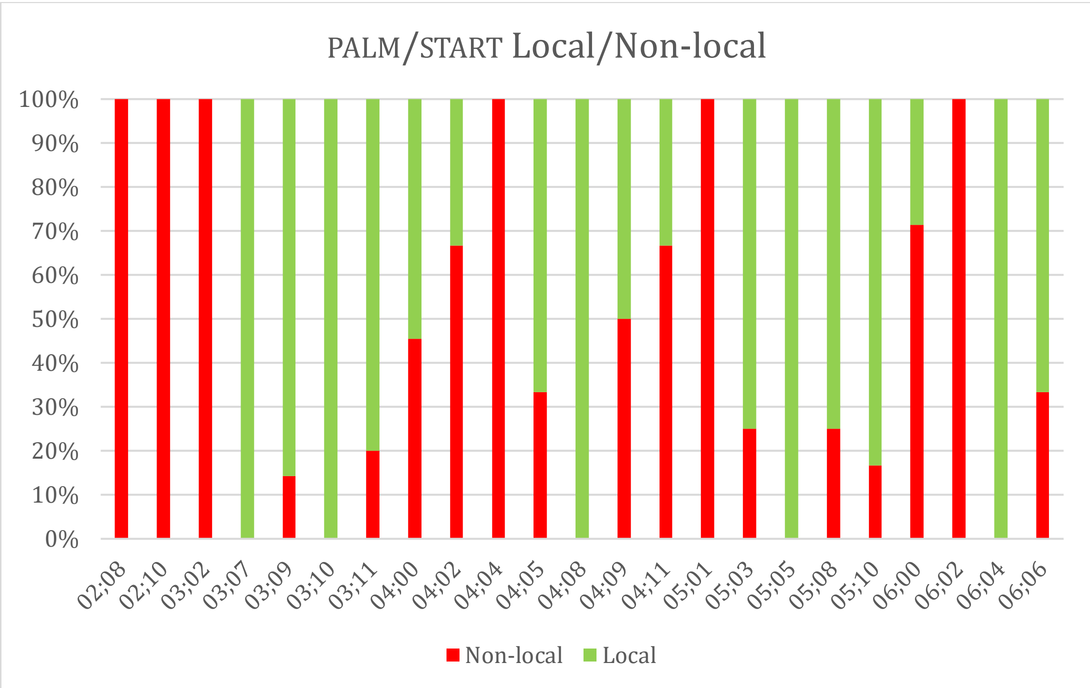


Figure 54: Acquisition of PALM/START

Turning now to an acoustic analysis of the vowels realised in the PALM/START sets, F1 and F2 have been plotted on to a scattergram as seen in the analysis of STRUT, above.

At age 3, there is a single token classed impressionistically as being a home variant. This vowel features a lower F1 and F2 than the local variants (Figure 55).

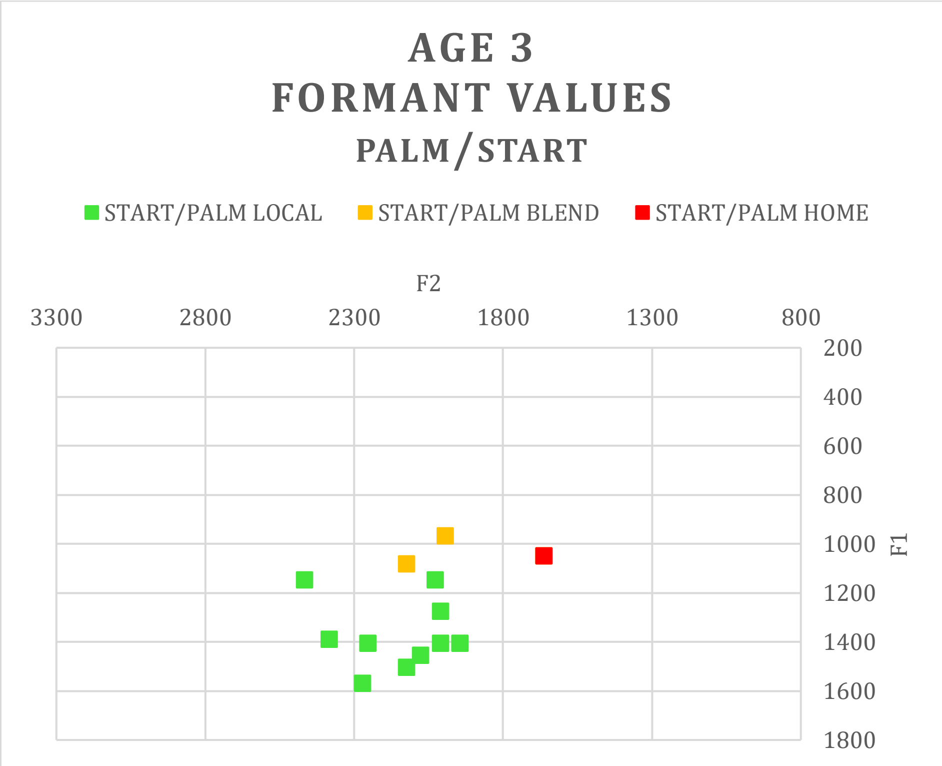


Figure 55: Formant values PALM/START *lexical sets at age 3*

Most tokens are of local variants due to their prevalence in the dataset.

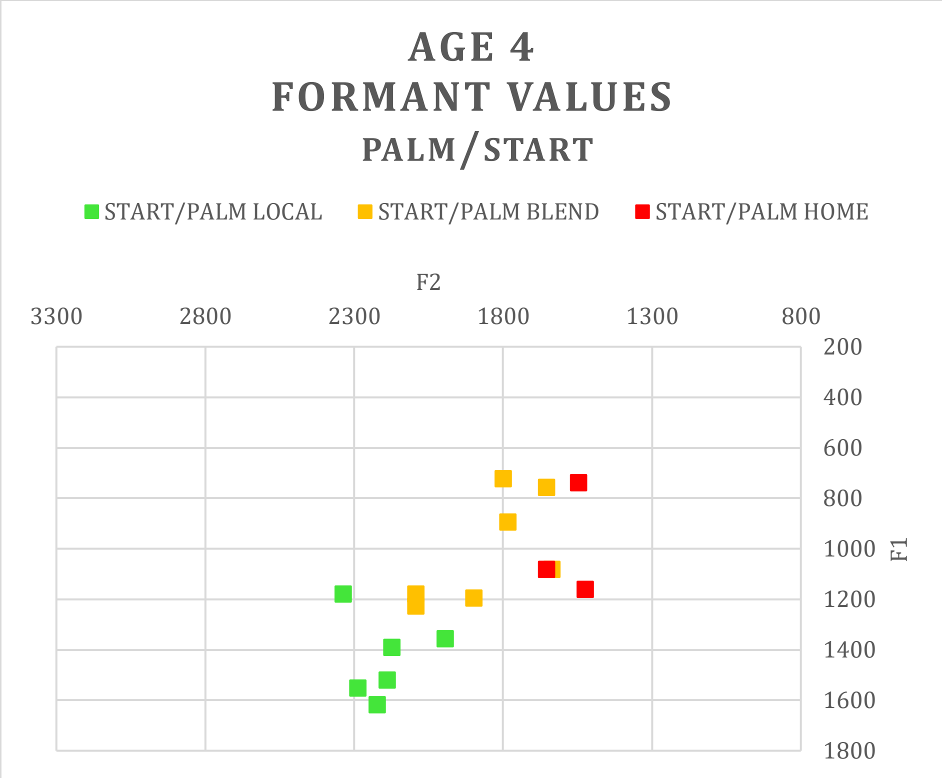


Figure 56: Formant values PALM/START *lexical sets at age 4*

F1 in some blended variants overlaps with tokens which were classified as local (Figure 56). This may partially be due to the gradual lowering of the formants as Henry ages. Having said that, one local example features an even higher F1 than those seen at age 3.

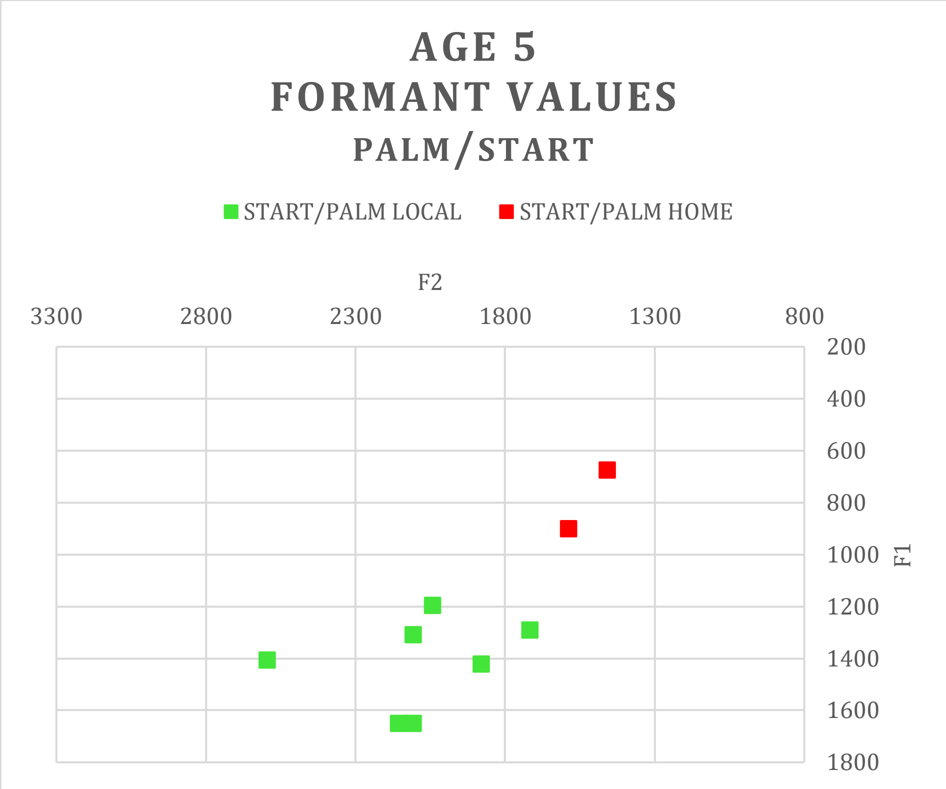


Figure 57: Formant values PALM/START *lexical sets at age 5*

At age 5, fewer tokens are available for acoustic analysis (Figure 57). Very high F1 values, ranging from around 600 Hz to 1650 Hz continue. Tokens were checked to ensure that high F1 values were not due to increased amplitude. This variation appears to be a combination of the dialect, which, as discussed above, is not currently represented in any study providing formant reference values, and the difference in the data collection method here. As expected, this naturalistic data exhibits much wider range of variation than is seen in the laboratory data available for comparison, particularly in jaw movement in these open vowels.

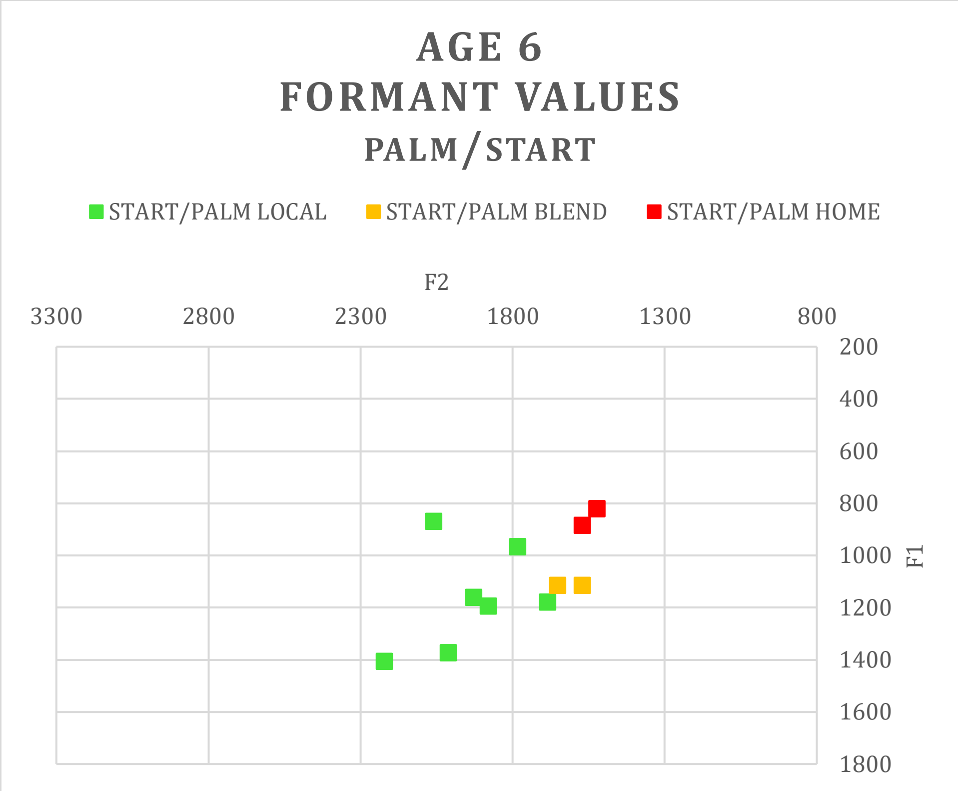


Figure 58: Formant values PALM/START *lexical sets at age 6*

At age 6 (Figure 58), few tokens are available for acoustic analysis. These exhibit slightly less variation than the variants seen at ages 4 and 5 years.

Ranging from 1684 Hz to 2220 Hz, Henry's local variants are generally lower than Vorperian and Kent's (2007) average low front corner vowel F2 value for boys aged 6 of around 2600 Hz, though we did see F2 values of up to 2594 Hz between 3-5 years old in his speech. At age 6, Henry's F1 values suggest that he produces much more open vowels than those seen in the literature averages. Both Vorperian and Kent (2007) and Busby and Plant (1995) place the average F1 value for 6-year-olds' and 7-year olds' open front vowels at around 1000 Hz (Table 34). Even though Henry's F1 variants are more tightly clustered at age 6 than at ages 4 and 5 years old, his highest F1 values for the PALM/START vowel still exceed the literature average by around 400 Hz — however, we must remember that these are averages, and that the data was collected under laboratory conditions.

	Busby and Plant (1995) (Estimated age 5 average)		**Vorperian and Kent (2007) (Estimated Age 6 average)**		**Henry's range at age 6 in PALM/START**
	Lowest avg. in low vowels	**Highest avg. in low vowels**	**Lowest avg. in low vowels**	**Highest avg. in low vowels**	
F1	760 Hz	1000 Hz	990 Hz	1090 Hz	968 Hz -1570 Hz
F2	1400 Hz	2250 Hz	1500 Hz	2600 Hz	1684 Hz to 2220 Hz

Table 34: Comparison of formant value ranges for low vowels[12]

In the case of Vorperian and Kent, their averages are based on multiple studies of many different dialects, therefore they may be smoothing out some very different values in the underlying data. This will undoubtedly be contributing to the differences between Henry's vowels and the reference data.

[12] Data in the literature is presented in graph form, so values have been estimated by eye.

5.2.3 BATH

BATH behaves differently to the PALM/START set. Tokens belonging to this lexical set do not appear in the recordings at all until 3;09 (see Figure 59). The peaks in tokens at 5;10 and 6;06 are attributable to repetitions of the word 'minecraft' and 'castle' respectively.

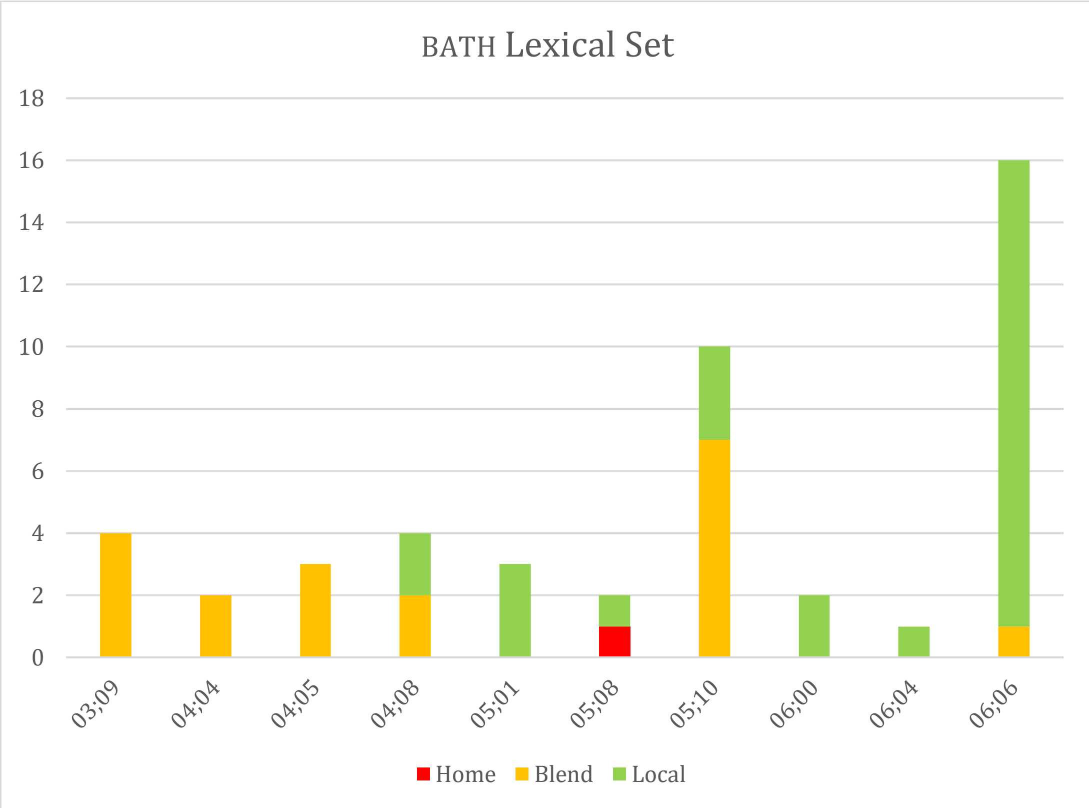


Figure 59: Realisations of the BATH *lexical set*

Early realisations appear to take vowel quality (i.e. jaw height and tongue advancement) from the local variety, [a] but length from the home variety, resulting in [aː], or, as in PALM and START, the vowel is articulated further back than in the local variety: [ɐ].

Example (1) *'fast, fast, fast'*

[**'faˑst 'faˑst 'faˑst**]

Example (2) *'on the top of that, um, grass'*

['ɒn də 'tɒp ɒv 'dat ʌm '**gɹɐːs**]

While the number of tokens for this lexical set is quite low, these blends were a salient feature of Henry's accent outside of recordings as well as during recorded sessions (see Table 35 for examples).

	Home	**Blend**	**Local**
03;09		[**'faˑst 'faˑst 'faˑst**]	
		'fast, fast, fast'	
04;04		[ˈɒn də ˈtɒp ɒv ˈdat ʌm ˈ**gʊɐːs**]	
		'on the top of that, um, grass'	
04;05		[**'gʊɐːs**]	
		'grass'	
04;08		[ˈɪz ˈðaʔ **'paːstə** ˈʃʊgə]	[**'kastə** ˈʃʊgə]
		'is that paster (caster) sugar?'	*'caster sugar'*
05;01			[**'kasəl**]
			'castle'
05;08	[ˈɪt ˈɪz ə **'dɑns**]		[ˈaɪm ˈnɒt ˈduɪn ə ˈ**dans**]
	'it is a dance'		*'I'm not doing a dance'*
05;10		[**maɪŋkʊɐf**]	[ˈhi s **'askt**]
		'Minecraft'	*'he s- asked'*
06;00			[ˈhad ə **'baθ**]
			'had a bath'
06;04			[ˈɪz ˈðɪs **'last** ˈpeɪʤ ˈtɛn ˈmɪnɪts]
			'is this last page ten minutes?'
06;06		[ˈwɛəz nadim **'ɐˑskt** ˈʧɪpt]	[ˈɪt ˈwɒz a ˈspeɪs **'kʊaft**]
		'where's Nadim, asked Chip?'	*'it was a space craft'*

Table 35: Examples of realisations of BATH

The first local variant is not seen until 4;08.

Example (3) *'caster sugar'*

['**kastə** 'ʃʊgə]

Unlike in the other lexical sets however, home variants are almost non-existent. Its only appearance is at 5;08, which is also realised with the local variant in the same recording:

Example (4) *'I'm not doing a dance'*

['aɪm 'nɒt 'duɪn ə '**dans**]

Example (5) *'it is a dance'*

['ɪt 'ɪz ə '**dɑns**]

Once the home and blended variants are combined, the transition from home to local influence is clearer. Non-local variants dominate Henry's early articulations, with local variants appearing more consistently from 5;01. Influences of home diminish to almost nothing by 6;0 (see Figure 60).

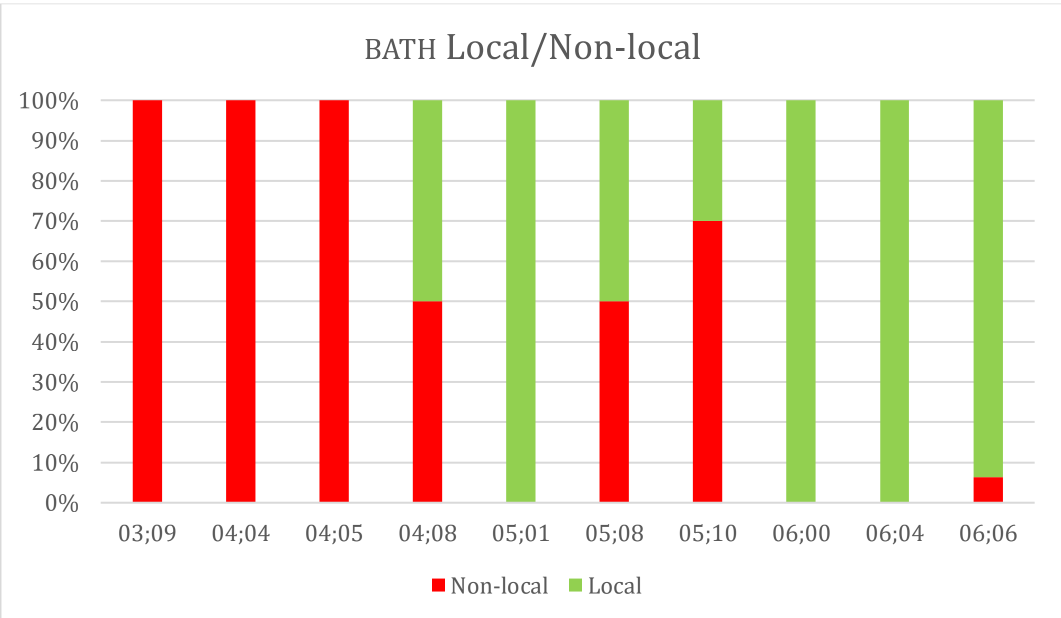


Figure 60: Acquisition of BATH

An acoustic analysis of realisations of the BATH lexical set follows below. No tokens of BATH were suitable for acoustic analysis before the age of 4, as few were recorded due to its low frequency. At the end of this section, BATH will be presented together with START, PALM and TRAP so that their distribution in the acoustic space can be compared. In the meantime, local and blended realisations will be compared.

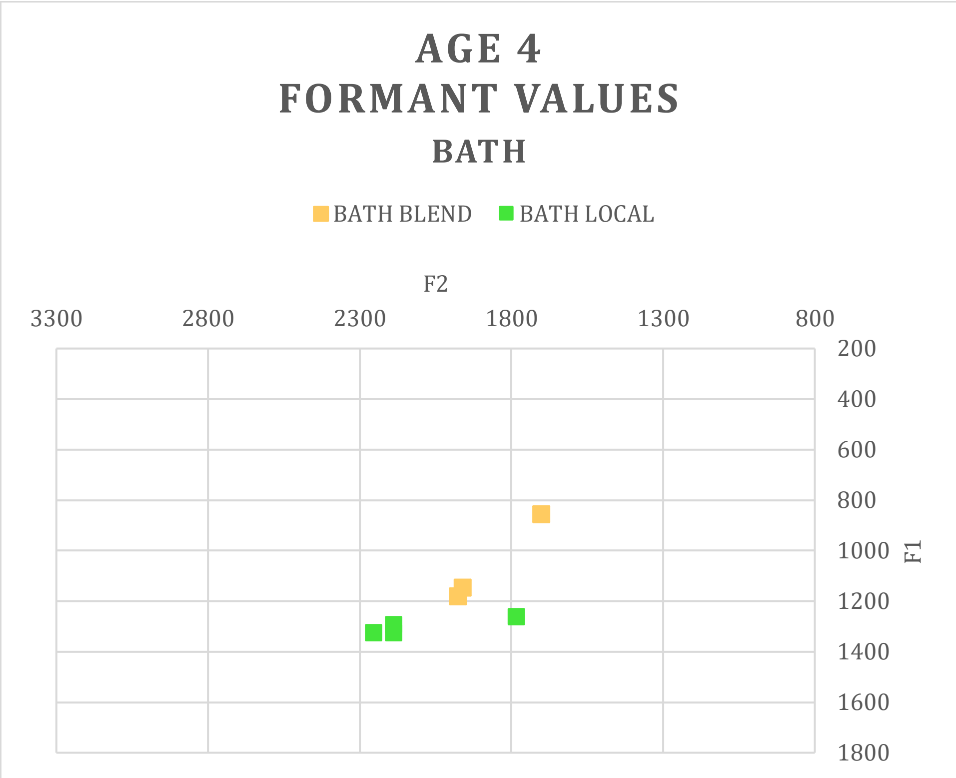


Figure 61: Formant values BATH *lexical set at age 4*

At age 4, there are few tokens available for acoustic analysis (Figure 61). Blended realisations feature lower F1 values compared to local variants, and F2 overlaps with local variants in two cases out of three.

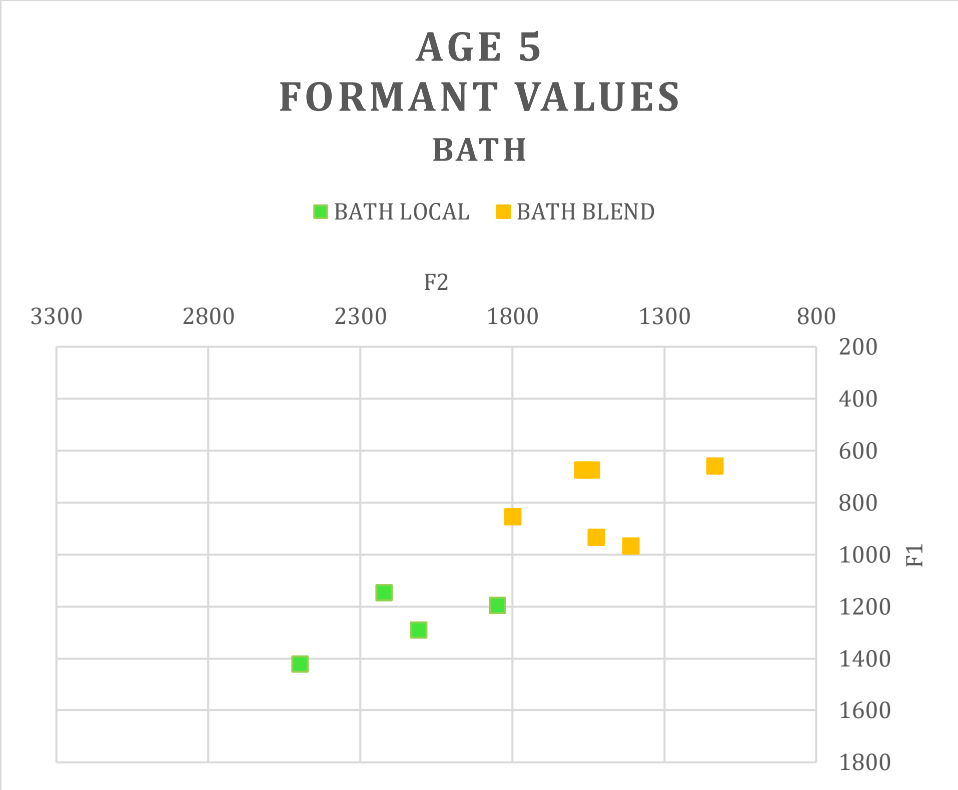


Figure 62: Formant values BATH *lexical set at age 5*

At age 5, there are more examples available. Again, F1 is lower in blended variants, but here we see a clearer distinction between F2 in local and blended variants (Figure 62). The lowest F2 (1131 Hz) has a much lower value than one of the home realisations of PALM above (1456 Hz), yet impressionistically it sounds further forward.

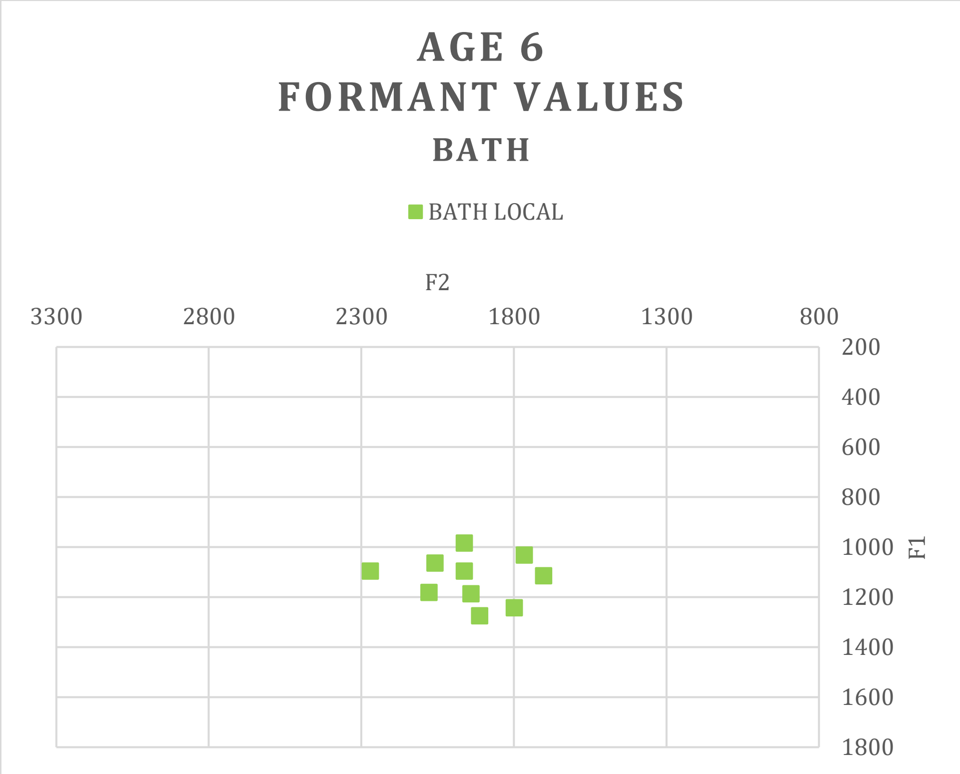


Figure 63: Formant values BATH *lexical set at age 6*

By the age of 6, only local realisations of BATH suitable for acoustic analysis are present (Figure 63). These form a relatively tight cluster in comparison to the wide ranges of F2 seen at ages 4-5 years, though it should be noted that six of these tokens are repetitions of the word 'castle' in read speech, three of which were in consecutive turns. In fact, all but one token in this set are of read speech. We have seen the same word being pronounced with considerable differences even within the same turn at talk previously, however, so in Henry's case, repetitions are not always produced in a similar way. The reading style is likely to have had an impact on the variation in Henry's realisations of this lexical set, here.

5.2.4 TRAP

Members of the TRAP lexical set in the home variety are articulated in broadly the same way as for local varieties – in this case as an open front vowel. These do, however, contain some anomalous articulations (see Figure 64).

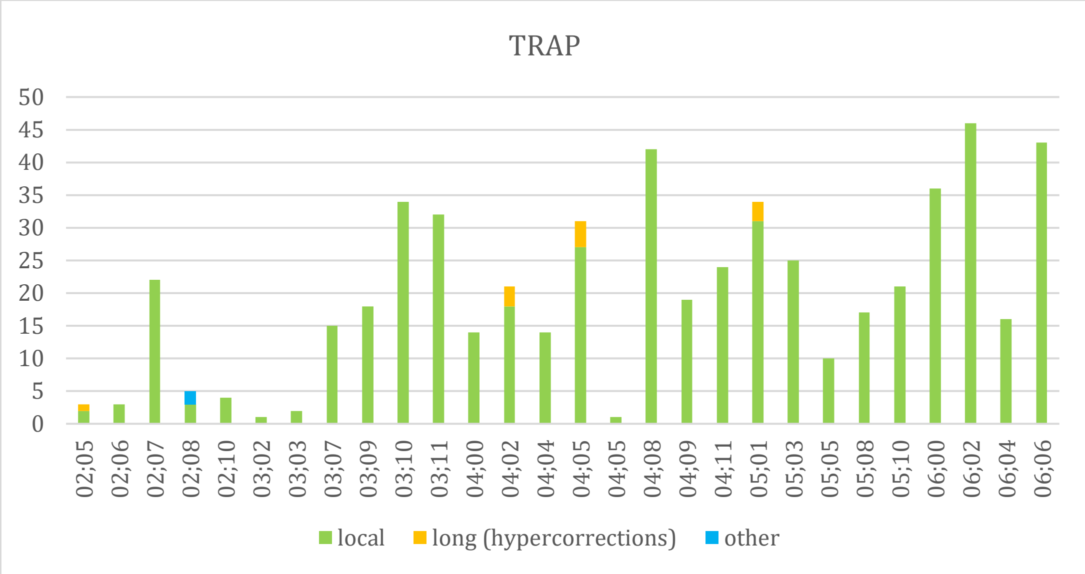


Figure 64: Realisations of TRAP

Articulations classed as ‘other’ are attributable to early vowel articulations not consistent with any variety in Henry’s environment.

Example (1) ‘thank you’
['**e** 'ku]

The tokens classed as long, however, present a more consistent pattern of articulation divergent from the TRAP set in either variety. This appears to be a form of hypercorrection, aligning a small subset of TRAP variants with BATH. Most of these variants share phonetic features with the BATH set, although the TRAP vowel in ‘Grandma’ appears once with lip rounding not seen elsewhere in either TRAP or BATH.

Example (2) ‘Grandma’
['**gʊɒ**ˑmɑ]

The data in Table 36 is consistent with my observations of Henry's speech outside of recorded sessions, and this idiosyncratic articulation persists occasionally at the time of writing. There is only one example of 'Grandma' being articulated with a short TRAP vowel in the whole dataset. This occurs at age 5;01.

Example (3) 'one of Grandma's'
[wən əv ˈ**gʊa**mːəz]

'Dad' doesn't appear here, but I have noted that Henry frequently but inconsistently articulates the vowel in 'Dad' as a long or half long central open vowel [ɐ] outside of recorded sessions. In most of the examples in Table 36, the vowel appears before /d/, though not in 'Grandma'. This vowel lengthening does not occur in front of other /d/ articulations in the data however, such as 'Daddy', 'bad', and 'ladder', which are all short.

Age		
02;05	Grandma	ˈgʊɐˑmə
04;02	bad	ˈbɐˑd
04;02	bad	ˈbaˑd
04;02	bad	ˈbaˑd
04;05	baddies	ˈbaˑdiz
04;05	baddies	ˈbaˑdiz
04;05	baddies	ˈbɐˑd̪iz
04;05	baddies	ˈbɐˑdiz
05;01	Grandma	ˈgʊɒˑmɑ
05;01	Grandma's	ˈgɹaˑməz
05;01	Grandma	ˈgʊɑmɑ

Table 36: Long (hypercorrected) TRAP *tokens*

Hypercorrected tokens of TRAP are compared against home TRAP tokens in Figure 65 below. There is some slight overlap of the tokens judged impressionistically to be hypercorrections, but in general they feature lower F2 values than typical TRAP vowels, confirming tongue retraction.

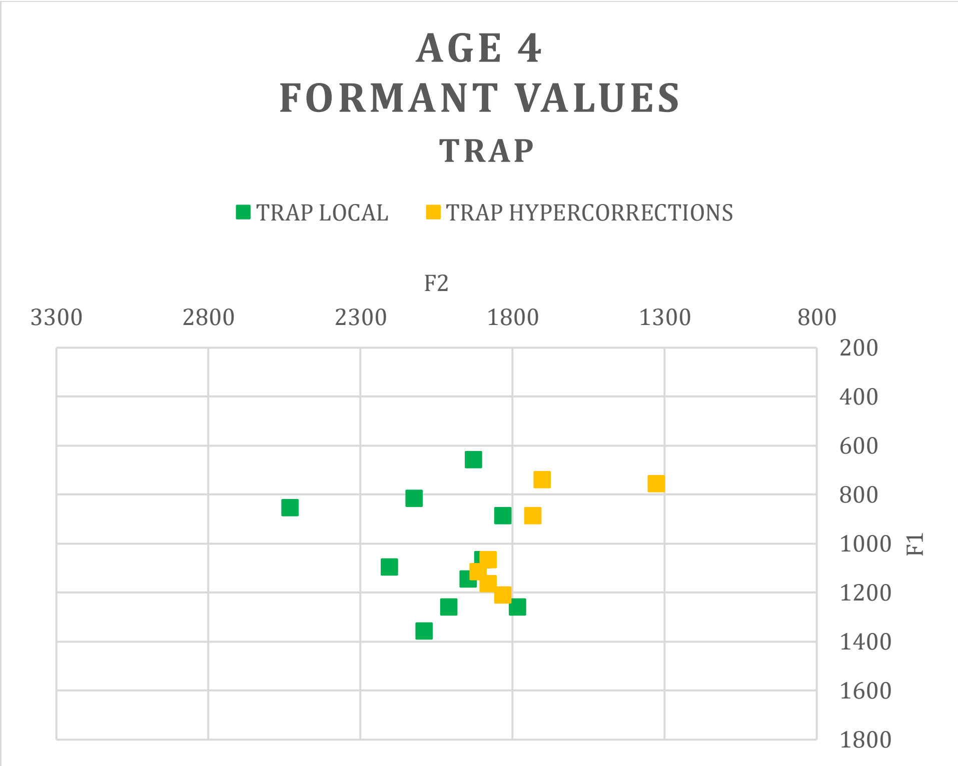


Figure 65: Formant values for TRAP *at age 4*

Fewer tokens are present at age 5 (Figure 66), but again, these occupy the same space as BATH blends and home realisations.

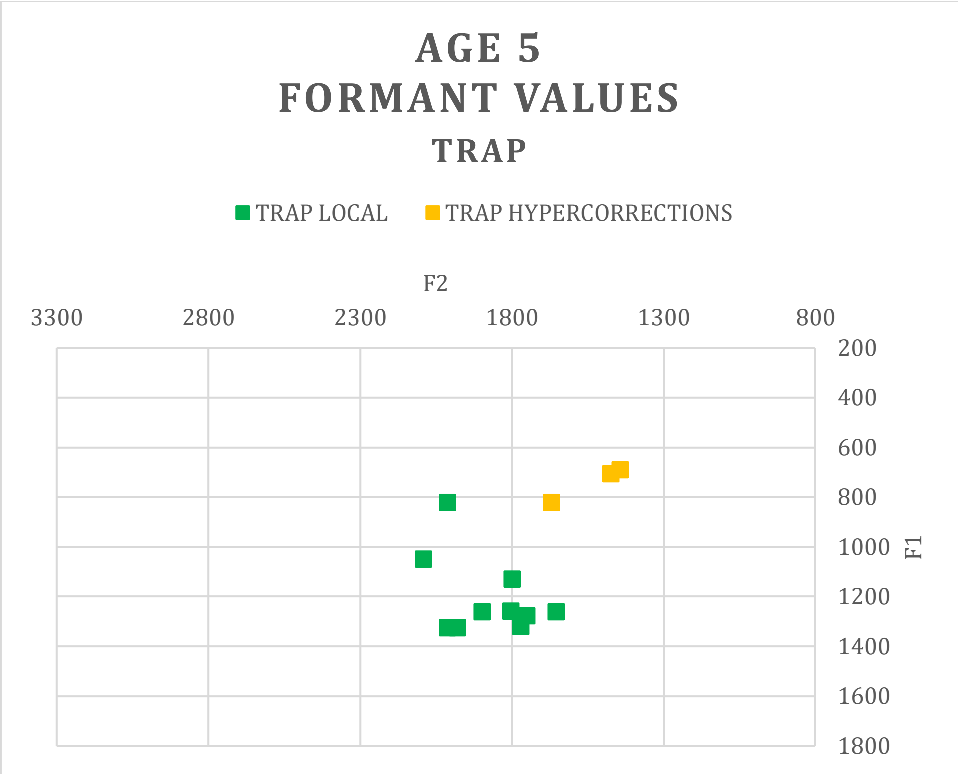


Figure 66: Formant values for TRAP *at age 5*

Figure 67 shows the formant values for TRAP between the ages of 3 and 6 plotted on a single graph. While there is a general trend of a lowering of frequencies as would be expected as Henry's vocal tract grows (shifting up and right) at age 3 and 4, and to a lesser extent at age 5. At age 6 there are both very front and very back tokens – these exhibit the highest and lowest frequencies for TRAP found at any age. This suggests that as Henry ages, he begins to use more of his available acoustic space (as seen in Veatch, 1991).

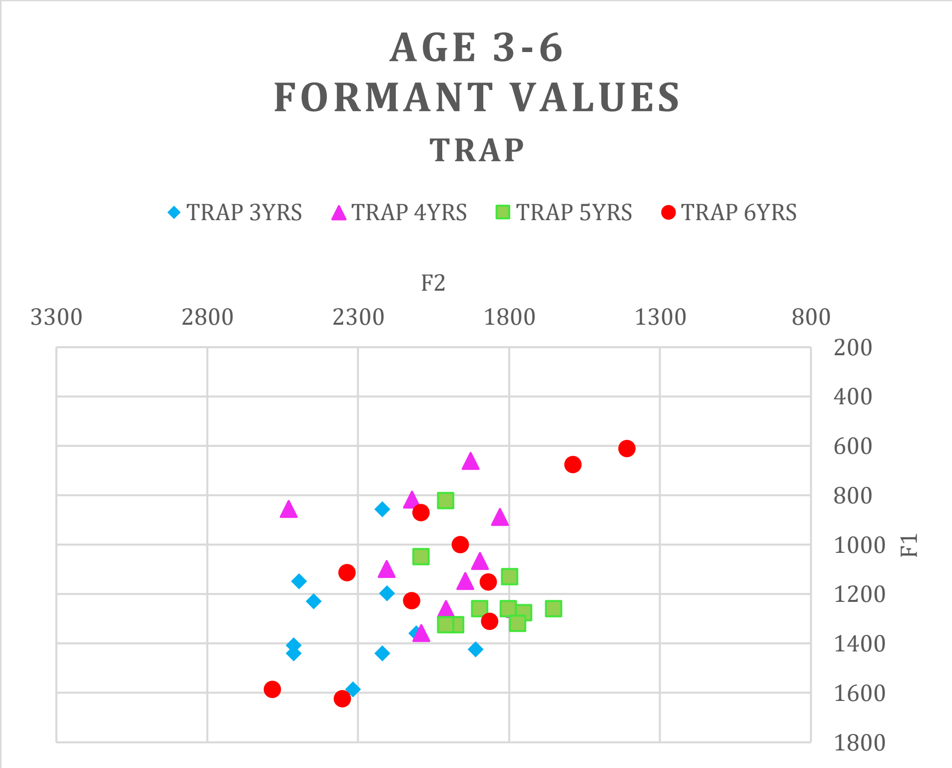


Figure 67: Age 3-6 Formant values TRAP

Further acoustics of TRAP are presented below in comparison with PALM/START and BATH.

5.2.5 Acoustic analysis: PALM, START, BATH, TRAP comparison

Here we consider the distribution of START, PALM, BATH and TRAP in the vowel space from age 3 to 6 years. TRAP tokens are included here to offer a comparison of the tongue position between the longer vowels of START and PALM against the short vowel, TRAP.

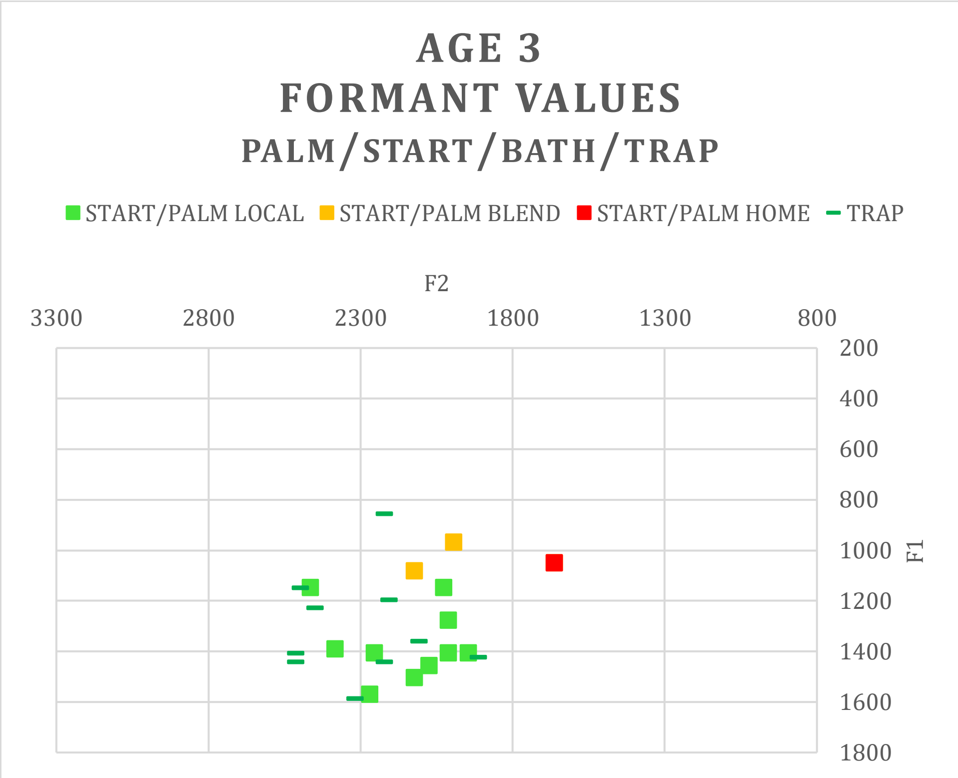


Figure 68: Formant values PALM/START/BATH/TRAP *at age 3*

At age 3, TRAP consistently occupies the same space as the vowels impressionistically identified as local variants of START and PALM (Figure 68). As mentioned above, BATH does not appear in the impressionistic analysis at age 3.

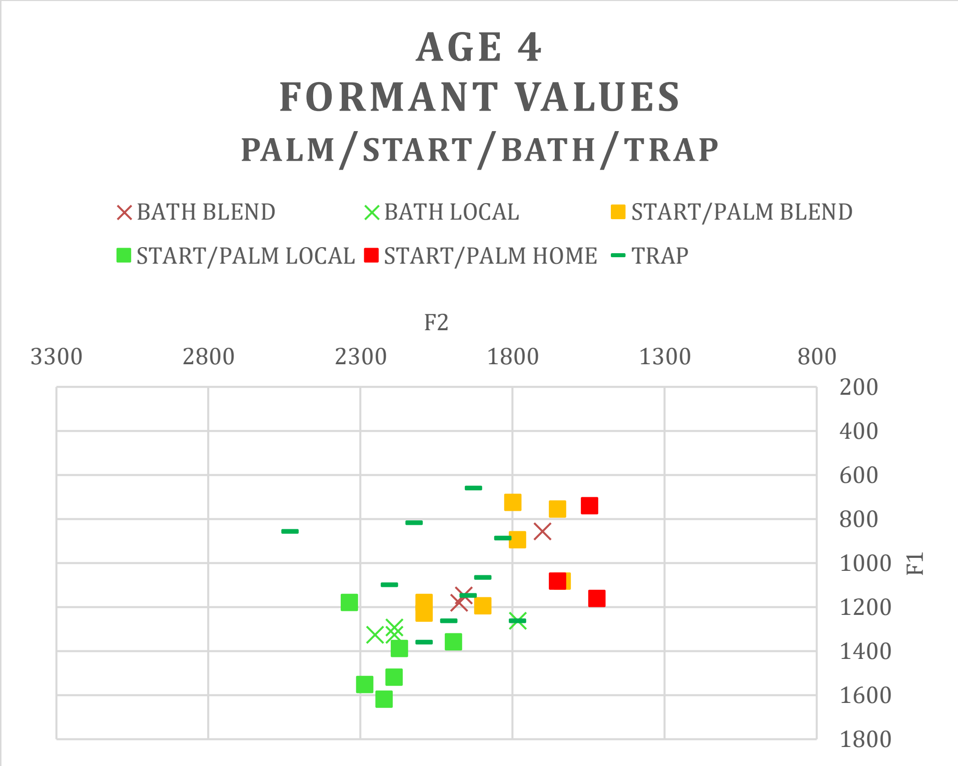


Figure 69: Formant values PALM/START/BATH/TRAP *at age 4*

By age 4, it is notable that PALM and START local variants are articulated more open and front than TRAP, which seems to be overlapping with the blended variants to an extent (Figure 69). Both TRAP and PALM/START have lower F2 values than at age 3. Local BATH variants overlap with TRAP, as expected, while BATH blends have similar formant values to the PALM/START blends.

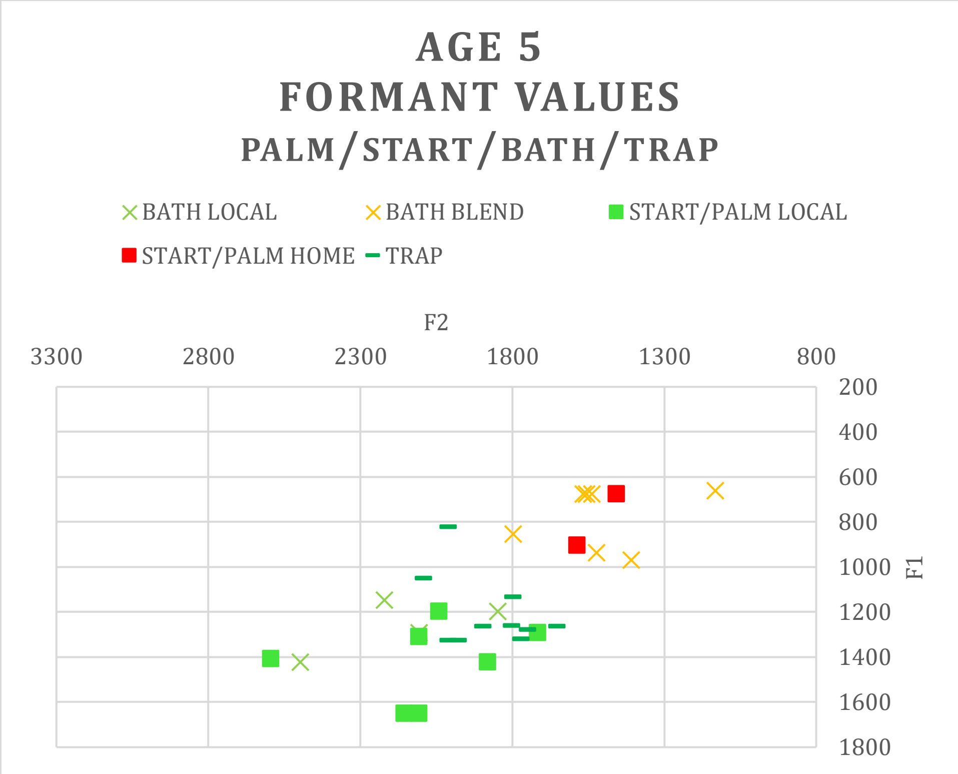


Figure 70: Formant values PALM/START/BATH/TRAP *at age 5*

Distribution at age 5 appears slightly more dispersed, and the blended variants of PALM/START and BATH are more clearly distinct from local variants, though the variant assigned to the home category appears in the same space as the blended variants (see Figure 70). Again, these examples have been carefully scrutinised and I remain confident in the impressionistic classification. While these articulations appear to be becoming more widely distributed, Yang and Fox (2013) claimed that the range of articulations produced by children becomes more tightly distributed as they age. In contrast to this study however, their data was collected in a laboratory environment. Alternatively, Local (1983), whose data was naturalistic, claimed that children increase their range of articulations within the constraints of the adult dialect as they learn the range of permissible variation. As at age 4, PALM and START may be more front and open than TRAP. BATH tokens also begin to appear in this very open front space.

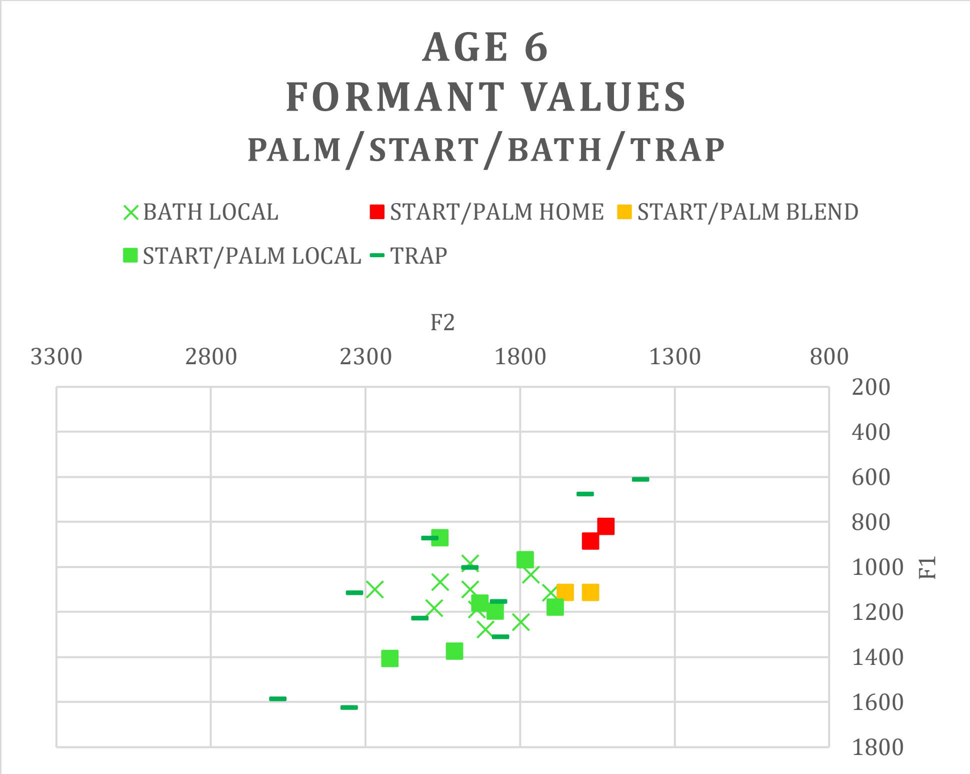


Figure 71: Formant values PALM/START/BATH/TRAP *at age 6*

At age 6, local tokens of PALM/START and local realisations of BATH occupy the same space as TRAP. BATH blends are slightly longer than the local BATH/TRAP vowel, but share the same acoustic space, in this case realised as [aˑ]. The PALM/START blends are, as before, occupying the space between the home and local realisations, impressionistically [ɐː] (see Figure 71).

More variation in tongue retraction and height in TRAP vowels is evident in the sample above than at previous ages (see Figure 72 and Figure 73). This is borne out by the impressionistic analysis. These TRAP tokens overlap impressionistically with FOOT or STRUT tokens. For example, 'Jack Black' is realised as ['ʤək 'blak] and 'carry on' as ['kəwi 'ɒn]. As discussed above, this increase in the range of articulations is at odds with Yang and Fox's (2013) observations but corresponds with Local's (1983) and Veatch's (1991) claims which were based on naturalistic data.

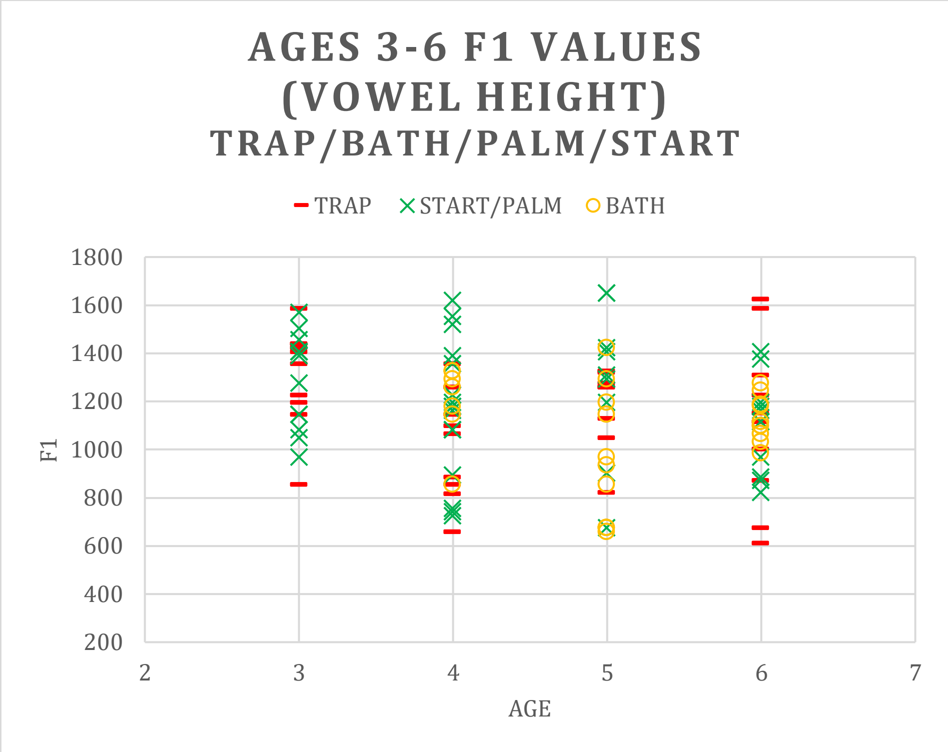


Figure 72: Age 3-6 F1 values TRAP/BATH/PALM/START (*does not include* TRAP *hypercorrections*)

The patterns observed above are evident in a separate analysis of F1 and F2 values. In both Figure 72 and Figure 73, frequencies decrease at the lower end as would be expected for a growing vocal tract, and PALM/START variants seem to be shifting in correspondence with that. However, at age 6 there are some examples in the higher frequencies for TRAP possibly reflecting the child's increasing variation, as he figures out what articulations are an acceptable 'hit' (Local, 1983). Unlike at ages 4 and 5, when PALM/START were the most open front vowels, at age 6, TRAP is more open.

We see much more variation (around 1000 Hz) in vowel height at age 6 in these open vowels than we saw in the close STRUT/FOOT vowels (300 Hz between the lowest and highest frequencies except for one outlier). This difference will be revisited in Chapter 7.

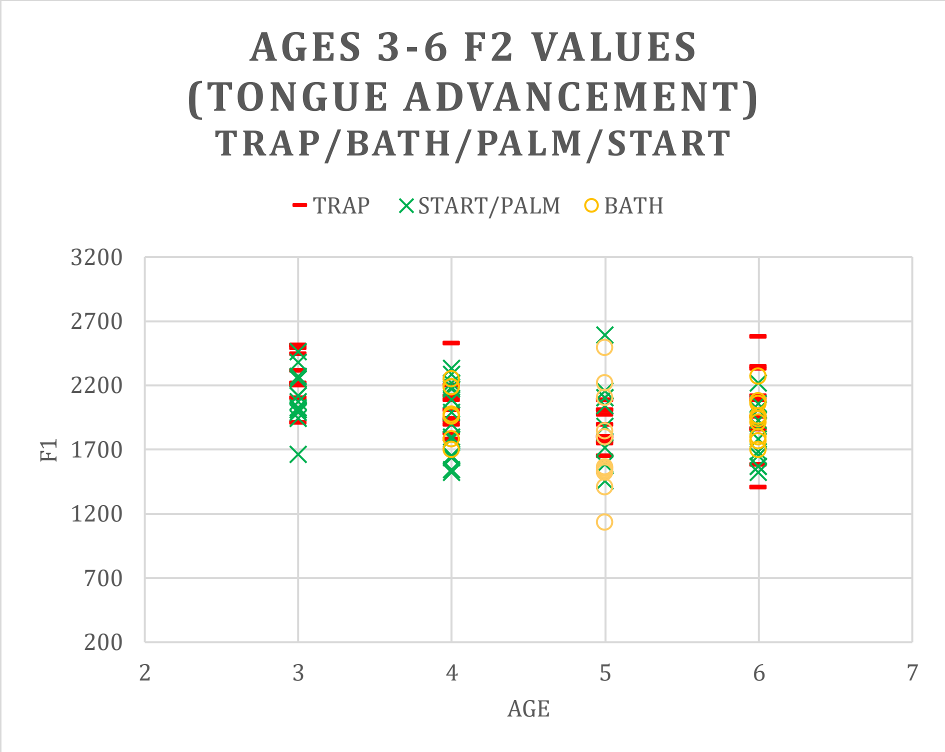


Figure 73: TRAP/BATH/PALM/START *Age 3-6 F2 values (does not include* TRAP *hypercorrections)*

As mentioned above, there are no tokens of BATH at age 3 in the acoustic analysis. TRAP tokens feature the most advanced tongue position at ages 3, 4 and 6, but at age 5, some realisations of PALM/START and BATH have higher F2 values. At age 5 in particular, Henry's realisations of PALM/START and BATH exhibit more tongue movement than at any other age.

5.3 Summary

In the lexical sets STRUT, START and PALM, Henry starts by realising members of these sets with vowels heard in the home while vowels found in the local area appear increasingly over time. In all these sets, Henry also realises what have been classified here as 'blends', that is, realisations which share features of both home and local vowels. In the BATH lexical set, blended variants appear first in the recordings, though this could be due to its low frequency – members of this lexical set didn't appear at all until 3;09. The way that blended variants manifested themselves varied according to lexical set. In the case of STRUT, Henry's vowels exhibited different heights and levels of lip rounding. For START and PALM, blends appeared as central vowels rather than the back home vowels and front local vowels. In the case of BATH, blends shared length with the vowels heard in the home, with the front articulation of the local vowels [a:]. This articulation aligned with those in the PALM/START sets, representing a compromise between front and back articulations of the vowels in the input.

Figure 74 shows the data presented in Figure 24, Figure 43, Figure 54 and Figure 60 combined. As in those figures, home and blended variants are combined, to divide variants into those exhibiting parental influence (non-local) and those showing the influence of the local speech community (local). In the case of FOOT, these were not divided into local and non-local in the FOOT analysis as Henry's realisations did not fit neatly into those categories for either the home or local varieties. Instead they were divided into front vs. central and back. However, this data is presented in Figure 74 as '% local', as his variants do appear to show progress towards a 'typical' northern realisation of FOOT.

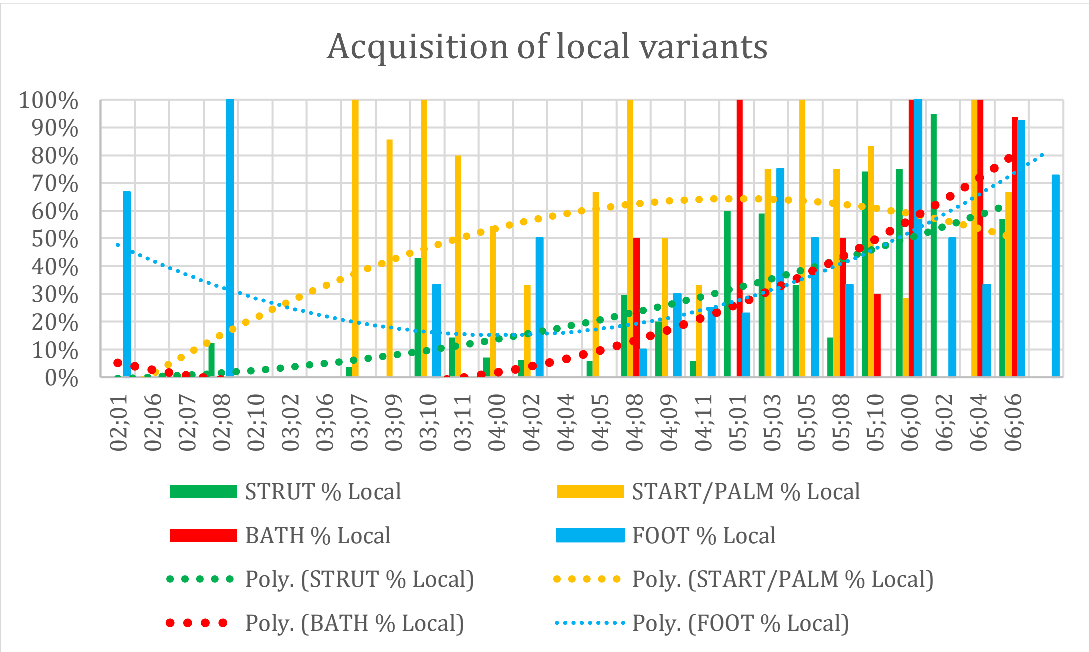


Figure 74: Comparison of acquisition of local variants showing polynomial trendlines

Local variants of PALM and START established themselves in Henry's accent earlier than STRUT, which has manifested itself more slowly but steadily. However, while local forms of PALM and START were evident in Henry's accent earlier than STRUT, home variants are still apparent, and the acquisition process seems more sporadic than for STRUT. BATH variants were later to establish but appear to have been acquired more consistently than the other variants by 6;06. A small number of local FOOT variants were present early on, but these were soon replaced by a preference for very front variants. Central and back variants appeared sporadically until they begin to establish at around age 5.

In Chapter 4, the shift from recording in .MP3 format to .WAV at age 3;08 was discussed. F1 and F2 were given special attention between the ages of 3 and 4 years to see if any effects of the change in file compression were evident, but variation between tokens was so great, given the naturalistic data collection method, that any effect was impossible to detect.

Chapter 6 Accommodation Analysis

As seen in section 2.3.2, there is evidence of children accommodating to the speech of others from an early age (see for example, Dyer, 2007 and Khattab, 2013). This chapter explores the possibility that Henry may shift his accent when speaking to his friends in comparison to the accent he uses at home.

As outlined in the method (4.3.4), at age 6;11, Henry was recorded playing with Lego with his friend James and his younger brother. This recording was made to determine whether there were any signs of him accommodating to his friend's accent. Some key features we saw in Chapter 5 above were a shift towards a rounded vowel [ʊ] from an open unrounded vowel [ʌ] in STRUT, a disappearing split between FOOT and STRUT, with FOOT featuring a front rounded articulation [ʏ] that appeared to be moving to a back vowel [ʊ] over time, and a lumpier shift from home realisations of PALM and START [ɑ] to the local front variant [aː]. Henry had acquired the local variant of BATH, [a] the most quickly and completely. The lexical sets STRUT and FOOT will be analysed in both children, followed by BATH, and PALM/START. In addition to the lexical sets analysed in Chapter 5, the GOAT and FACE lexical sets will also receive attention, as James produces these vowels as monophthongs, as is typical for some speakers of the local varieties. Finally, a brief analysis of features susceptible to style shifting will be considered.

6.1 STRUT/FOOT

6.1.1 STRUT

Impressionistically, Henry's friend James, who was born in North Yorkshire and has parents from the same area, produces all STRUT vowels consistently as a close back rounded vowel. As expected, and consistent with the analysis of his speech up to 6;06, Henry's STRUT vowel shows more variation. As seen previously, home variants are realised as an open back unrounded vowel [ʌ], or a schwa.

In conversation with his friend, examples of Henry's variants consistent with the home variety include:

Example (1) 'Can you make an - can you make an enderportal[13] big enough for me?'
['kən ju 'meɪk ən 'kən ju 'meɪk ən 'ɛndəˌpɔtəl 'bɪg ə'**nə̹f** fɔ 'mi]

Example (2) 'Another one'
[ə'**nʌðə** wɒ̹n]

Example (3) 'I'm doing a multicoloured one of those including fire'
[aɪm 'duɪŋ ə '**məlti**ˌkʊləd wə̹n ə 'ðəʊz ɪŋ'kludɪŋ 'faɪə]

Henry also realises a proportion of STRUT vowels as sounds sharing characteristics of both the home and local variants. As in the preceding analysis in Chapter 5, some of these take the form of a rounded open back vowel [ʌ̹], while others are realised as an open back rounded vowel [ɒ].

Examples of Henry's variants of STRUT showing elements of both the local and home varieties are as follows:

Example (4) 'I'm doing a melt- multicoloured one of those'
[aɪm 'duɪŋ ə 'mɛlt **'mɒlti**ˌkʊləd 'wʊn ə 'ðeʊz]

Example (5) 'I'm doing a multicoloured fire'
[aɪm 'duɪŋ ə **'mʌ̹lti**ˌkʊləd 'faɪə]

Example (6) 'You're going to build a person- a dumb dumb person a stupid person'
['jɔ gənə 'bɪld ə 'pɜsən ə **'dʌ̹m 'dʌ̹m** 'pɜsən ə 'stupɪd 'pɜsən]

It is of note that 'multi' is realised in three different ways in the examples above – with both a home vowel and two different blended realisations. Even when these articulations are found within the same turn at talk, they are realised with slightly different vowels.

[13] A feature in the computer game 'Minecraft'.

As with the interactions between Henry and me at age 6;06, the local variant dominates, however. The following examples illustrate the range of variation found in STRUT during the period of play with his friend.

Below are examples of Henry's variants of STRUT consistent with local realisations:

Example (7) 'I- I built this without instructions'
['aɪ 'aɪ 'bɪlʔ 'ðɪs wɪð'aʊʔ ɪn'**stɹʊk**ʃənz]

Example (8) 'I found another head found another head'
['aɪ 'faʊnd ə'nʊðə 'hɛd 'faʊnd ə'**nʊð**ə 'hɛd]

Example (9) 'I have- I have got the Scuttler and the Jokermobile'
[aɪ 'av aɪ 'av gɒʔ ðə **'skʊt**lə 'and ðə 'dʒəʊkəməˌbil]

The distribution of these variants is weighted heavily in favour of local pronunciations, though a small number of realisations were consistent with the home variety. Figure 75 presents the data from the main analysis of STRUT in Chapter 5 (Figure 23) alongside Henry's STRUT realisations in conversation with James at age 6;11 (rightmost column).

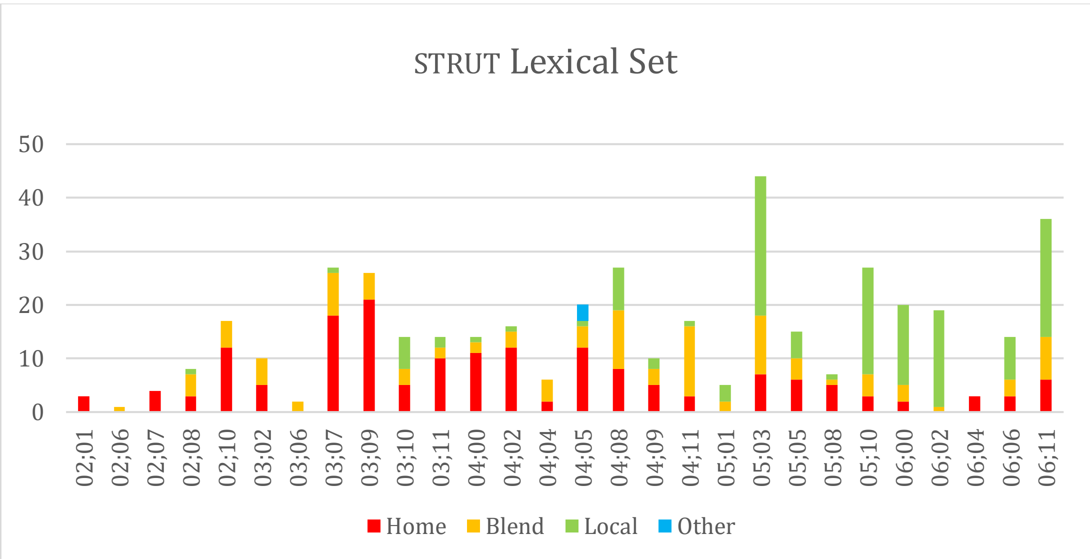


Figure 75: Realisation of the STRUT lexical set including data from 6;11

In the final recordings of Henry interacting with me in the main analysis at age 6;06, variants with any influence of the home variety have shrunk, although there is some degree of inconsistency between the different recordings. After his sixth birthday, realisations of STRUT without rounding (home variants) are *usually* down to 30% or less in each recording. This can be seen in Figure 76 below, where data from Henry's conversation with James is shown alongside the data presented in Chapter 5. There is one recording where all realisations of STRUT are consistent with the home variety (6;04), but this is unusual, and there are only three tokens in the whole recording.

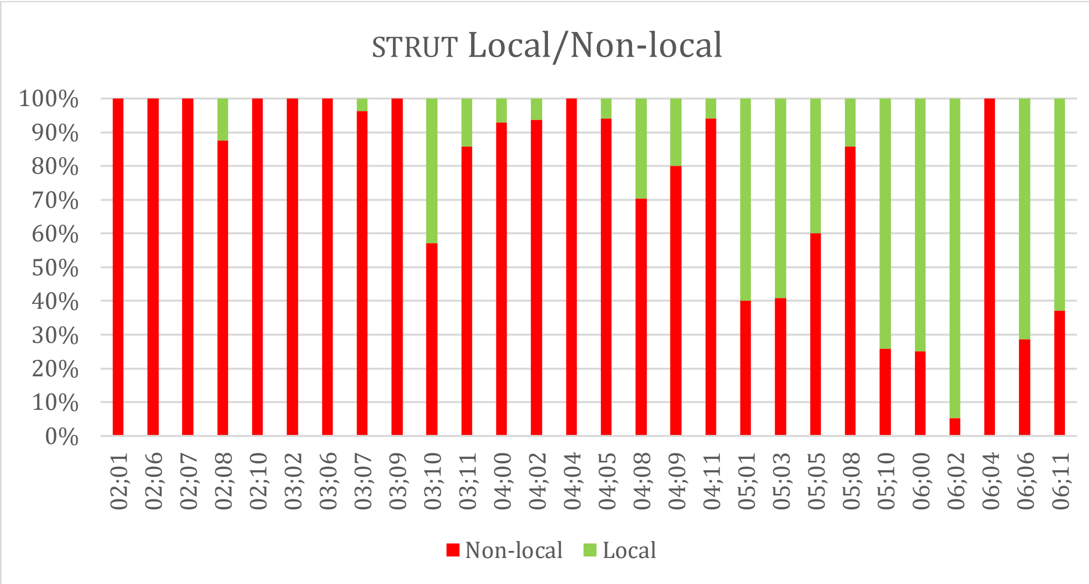


Figure 76: Realisations of STRUT *including data from Henry's conversation with James (age 6;11)*

When home and blended variants are combined together to indicate realisations which show some influence of the home variety, local variants still dominate, but it is clear that the influence of home dialects is still present, even when in conversation with a friend with a local accent who consistently shows no differentiation between STRUT and FOOT. It is clear that if any accommodation towards James's accent is present at all, it is not affecting Henry's STRUT vowel realisations.

As in Chapter 5, an acoustic analysis was performed to support the impressionistic analysis. Ten tokens from each child were selected for acoustic analysis, where available (see Table 37). In some cases, duplicated words were included for analysis, as some lexical items appeared multiple times. These duplications appeared for multiple reasons. Some words happened to appear more than once in a single utterance, and in some cases, these formed part of a self-initiated self-repair. In other cases these duplicates appeared in different turns.

Orthography	Henry Home	Henry Blend	Henry Local	James
enough	ə'nə̝f			
another	ə'nʌðə		ə'nʊðə	ə'nʊvə
Scuttler			'skʊʔlə	skʊtlə
stuff			'stʊf	
multi-	'məlti	'mɒlti ~ 'mʌ̹lti		
coloured			'kʊləd	
dumb		'dʌ̹m	'dʊm	
dumber			'dʊmə	
something	'sə̝mθɪŋ			
hundred				'hʊndɹəd
just				'ʤʊst
unstoppable			ʊn'stɒpəbl̩	

Table 37: Examples of Henry and James' STRUT *realisations selected for acoustic analysis*

As before, not all tokens analysed impressionistically were suitable for acoustic analysis due to background noise from the toys or overlapping talk. Realisations of 'one' were excluded for the same reason that they were kept separate in the initial analysis: some northern accents produce 'one' with a LOT vowel, and we saw in Chapter 5 that Henry inconsistently produces 'one' with a LOT vowel. As in the method used for the main analysis, some examples were excluded due to being unstressed even though they were unrounded, such as the STRUT vowel in 'some' being realised as a schwa.

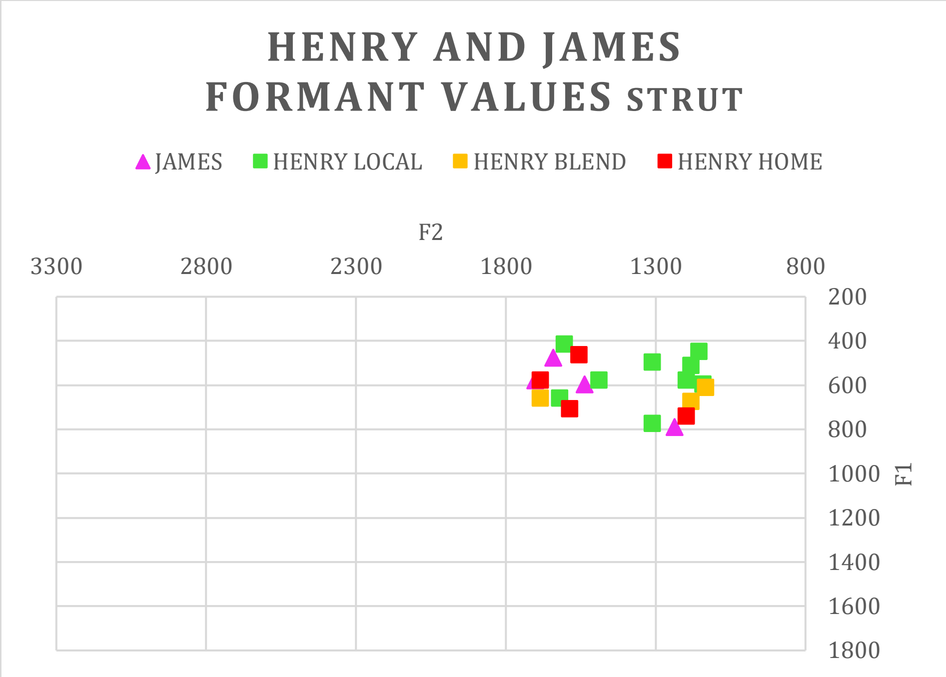


Figure 77: Henry and James formant values STRUT *lexical set*

As in Chapter 5, frequency ranges for the scatterplots have been kept consistent across all vowels so the reader may compare realisations across the whole vowel space.

Henry's speech contains more examples of STRUT than James', but the overall distribution of F1/F2 overlaps with James' realisations of this lexical set (see Figure 77). Henry's realisations are notably quite tightly clustered in comparison with the much larger acoustic space used in his conversations with me. This could be attributed to him being excited by the presence of his friend, and a degree of showing off. Speech was rapid, which led to many vowels being very short. Indeed, on close listening, many individual words were not easily recognisable in isolation. As in the main analysis in Chapter 5, there was not a clear acoustic division between vowels that were impressionistically judged to be closer to the home variant (open) than the local variant (close). Though Henry's local variants typically featured a first formant in the region of 400-600 Hz, there were a few with a more open position, 'scuttler' (a vehicle in the Lego Batman movie) realised as ['skʊʔlə], and 'coloured', realised as ['kʊləd]. Impressionistically, these were both clearly rounded. Equally, some home variants appear to have a relatively close jaw position – in Chapter 5, we saw that home variants typically featured a first formant in the region 800-1000 Hz when Henry was younger, but as he has grown older, F1 has lowered. In this recording F1 ranged from 465-741

Hz, overlapping with the close back rounded vowels. Impressionistically, these were transcribed as raised schwa:

Example (10) ‘fine I'll throw something at him’
['faɪn aɪl 'fɹəʊ '**sə̝m**θɪŋ 'aɾ ɪm]

Although there were not many tokens suitable for acoustic analysis, the blended variants, those which shared elements of both home and local variants, occupied the same acoustic space as the home and local vowels. We saw in section 3.2.4.4 how even in reading style, realisations of one vowel can overlap with another (Keating & Huffman, 1984), so in these rapidly articulated examples of naturalistic speech, this is not unexpected.

6.1.2 FOOT

Although Henry’s early realisations of FOOT are mostly front, by 6;11, most variants are realised as back rounded [ʊ], consistent with the local variety. Figure 78 presents the data from Henry’s conversation with James alongside the data from the main analysis seen in Chapter 5.

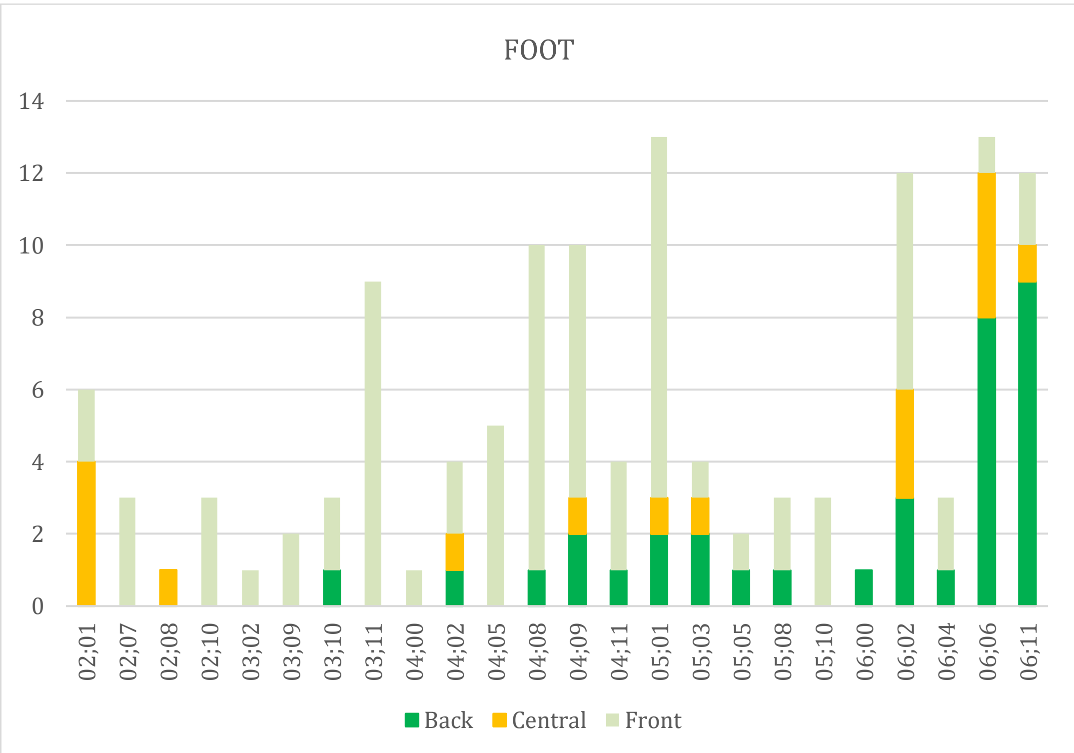


Figure 78: Realisations of the FOOT *lexical set including data from 6;11*

Front realisations are still present at 6;11, for example:

Example (1) ‘Could you build a ender dragon’
['kʏd ju 'bɪld ə 'endə 'dɹagən]

Example (2) ‘could you- can you join to me?’
['**kʏd** ju 'kan ju 'ʤɔɪn tə 'mi]

Back variants of FOOT consistent with local realisations are more frequent, however:

Example (3) ‘what does obsidian look like anyway?’
['wɒdəz ɒb'sɪdiən **'lʊk** 'laɪk 'ɛniweɪ]

Example (4) ‘I found a foot’
[aɪ 'faʊnd ə **'fʊt**]

Example (5) ‘Shall we put this in the nether?’
[ʃəʊ wi **'pʊt** ðɪs ɪn ðə 'nɛðə]

Henry also continues to produce one variant of FOOT with a central realisation:

Example (6) Could you- could you put that on top of it?’
[**'kʊ̟d** jə 'kʊ̟d jə 'pʊʔ 'ðaʔ ɒn 'tɒp əv ɪʔ]

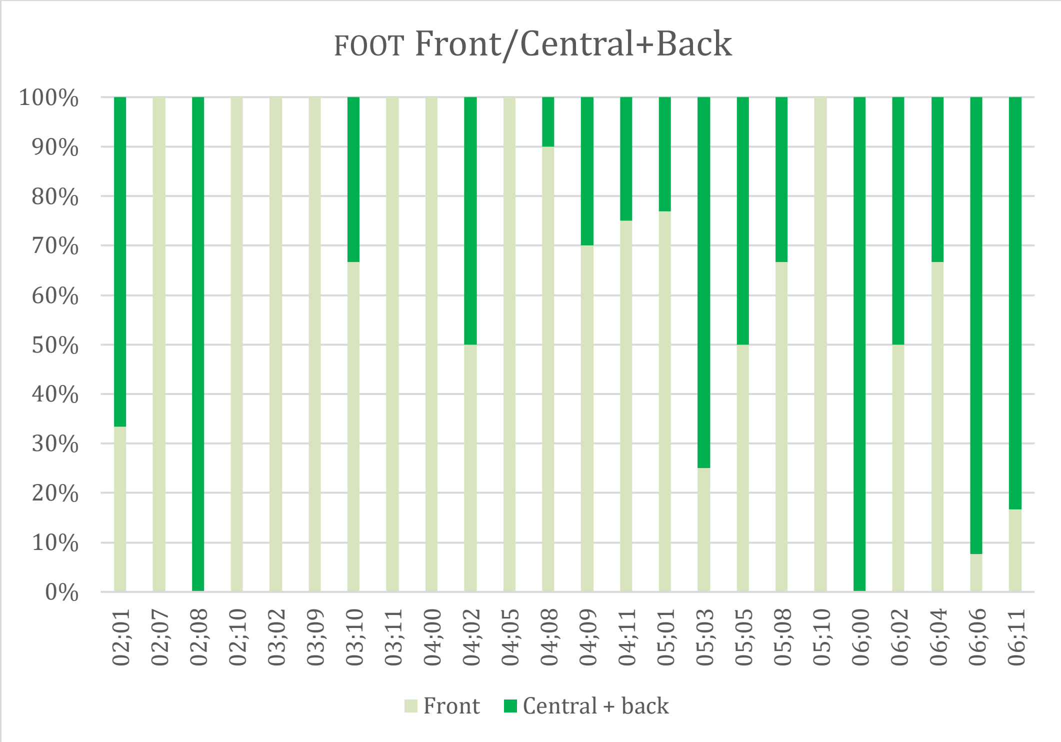


Figure 79: FOOT *front/central+back including data from Henry's conversation with James (age 6;11)*

Figure 79 re-presents the progression of FOOT seen in Chapter 5 when separated into front articulations and the central and back articulations more typical of the local area. At 6;11, Henry's FOOT realisations feature slightly fewer (local) central and back articulations, but the overall pattern is one of progression away from front realisations.

The following examples were selected for acoustic analysis.

Orthography	**Henry IPA Front**	**Henry IPA Central (blend)**	**Henry IPA Local (back)**	**James IPA**
good				ˈgʊd
look				ˈlʊk
could	ˈkʏd	ˈkʊ̟d		ˈkʊd
foot			ˈfʊt	
would			ˈwʊd	ˈwʊd
put			ˈpʊʔ	ˈpʊt

Table 38: Henry and James FOOT *realisations selected for acoustic analysis*

Only one token of Henry's six realisations of the word 'foot' was selected for acoustic analysis as most of these were articulated in a peculiar manner, out of line with the rest of his FOOT realisations (see Table 38).

Example (7) 'I found a foot'
[aɪ ˈfaʊnd ə **ˈfʊət**]

These diphthongal realisations are unusual and seem to be only found in a few turns at talk in this recording; I have neither heard them in his speech at home nor observed them in any other data analysed. As observed by Local (1983), children exhibit a wide range of speaking styles which lead to vowel variation, such as singing or whingeing. Here, these variants appear to be a form of playfulness or experimentation — possibly an attempt at accommodation. This will be discussed below in section 6.5.

In Chapter 5, we saw that Henry's FOOT realisations had a tendency to be a front vowel, [ʏ], around the F2 of Busby and Plant's (1995) Australian [u] in 7 year olds (approximately 2450 Hz – 2500 Hz), and in some cases around the F2 of [ɪ] (approximately 2700 Hz – 2900 Hz). In comparison, his local realisations of STRUT [ʊ] shared similar F2 values to Busby and Plant's [ʊ] (around 1050 Hz to 1150 Hz). See Figure 80 for a reproduction of Figure 48 from the main analysis.

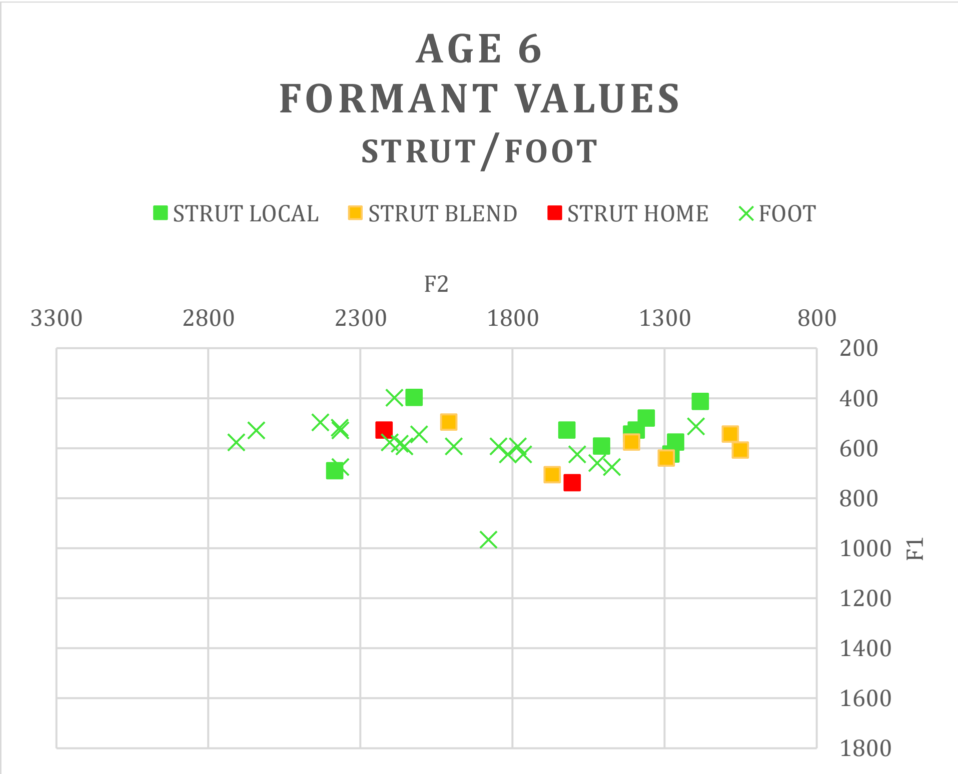


Figure 80 (copy of Figure 48): Formant values STRUT/FOOT *lexical sets at age 6*

At 6;11, this variation continues. Here, his articulations of FOOT range from very front to back, with the back tokens overlapping the same acoustic space as James' FOOT vowels occupy (see Figure 81). The most back token, 'foot' shares a similar F2 to James' most retracted token, 'look'.

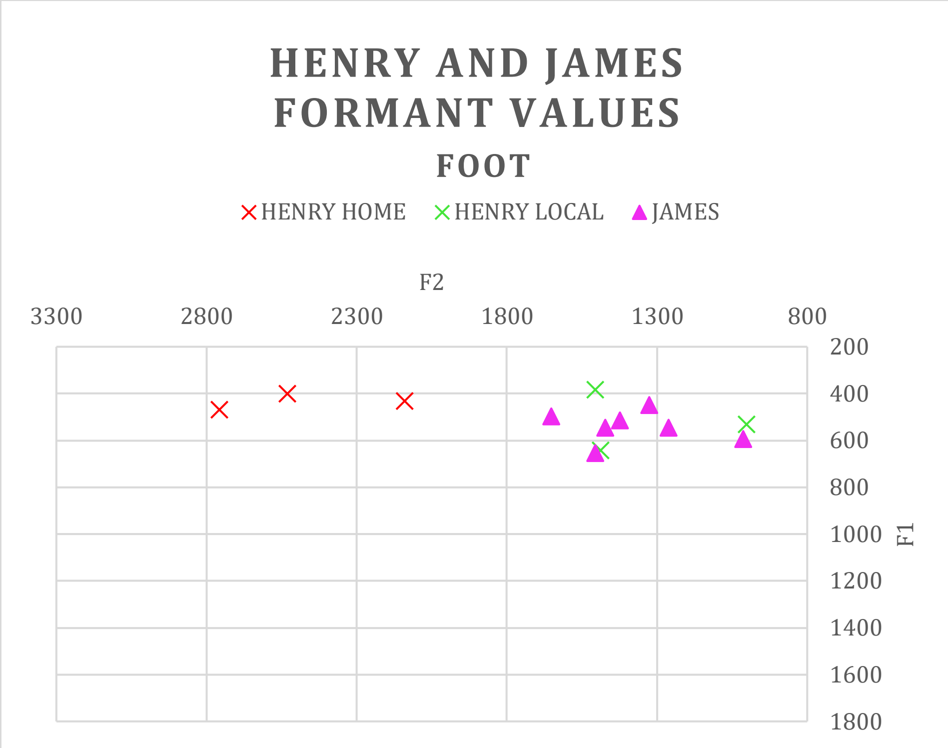


Figure 81: Henry and James formant values for the FOOT *lexical set*

By the age of 6, F2 in FOOT increasingly overlapped with the F2 of Henry's STRUT articulations, but very front realisations were still present (Figure 82).

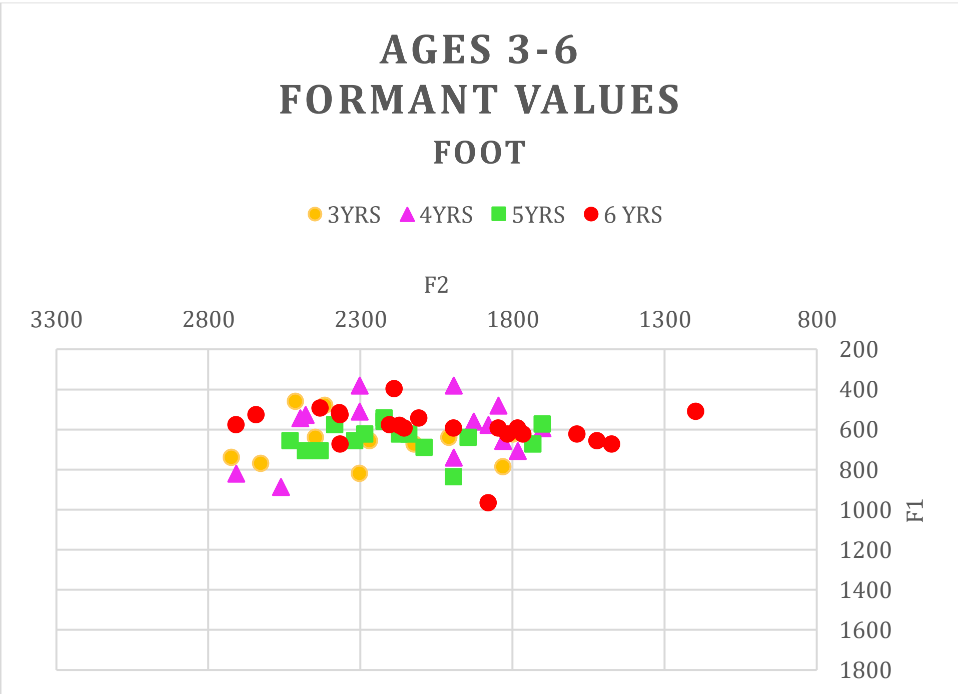


Figure 82 (copy of Figure 49) FOOT *formant values age 3-6*

The very front articulations (F2 ranging from 2138 Hz -2756 Hz) heard in Henry's conversation with James (Figure 81) all appear in the word 'could' and correspond with the impressionistic analysis of a close front rounded vowel [ʏ], consistent with many of the realisations in the main analysis in Chapter 5. At the other extreme, the vowel with the lowest F2 (1001 Hz) is in the word 'foot' which has an unusually back quality (for Henry). This corresponds with the impressionistic analysis, ['fʊt]. This example was the only monophthongal realisation of 'foot' out of six repetitions in this recording all in the space of a few turns, where the other five were articulated with some velarisation and a shift in vowel quality, as seen in example 7. This vowel overlaps with one of James' FOOT articulations. Henry's F2 values at age 6;11 (Figure 81) generally correspond with his FOOT vowels in Chapter 5. In comparison, James' FOOT vowels are consistent with both Henry and James' STRUT realisations (see Figure 77).

Before 6;11, there is only one example of a back realisation with an F2 of lower than 1472 Hz. Henry's back variants typically cluster around an F2 of 1500 Hz at age 6. However, there is one example of FOOT at 1192 Hz at that age in conversation with me, which is closer to Henry's most back FOOT realisation in conversation with James (F2=1013 Hz). So, while the back realisation Henry produced while playing with James could be evidence of an attempt to accommodate to his friends' FOOT vowels, there is limited evidence to support this hypothesis.

6.2 PALM/START/BATH/TRAP

We saw in Chapter 5 that the acquisition of the PALM/START vowel followed a different path from the acquisition of STRUT and FOOT. While STRUT followed a fairly stable trajectory towards the local variant becoming dominant, PALM/START seemed to go through peaks and troughs, and by the end of the data collection period, a clear dominant variant was not yet established.

6.2.1 PALM

Figure 83 shows Henry's production of PALM while playing with James alongside the rest of his PALM productions seen in Chapter 5. Here we see no home variants, but the number of local variants is smaller than blended variants. As in the main analysis in Chapter 5, the PALM and START (analysed below) variants will be combined as they are not differentiated in either the local or home accents, apart from for Henry's grandmother, who speaks a variety of Scottish English. Here, they will be considered separately initially, before being combined.

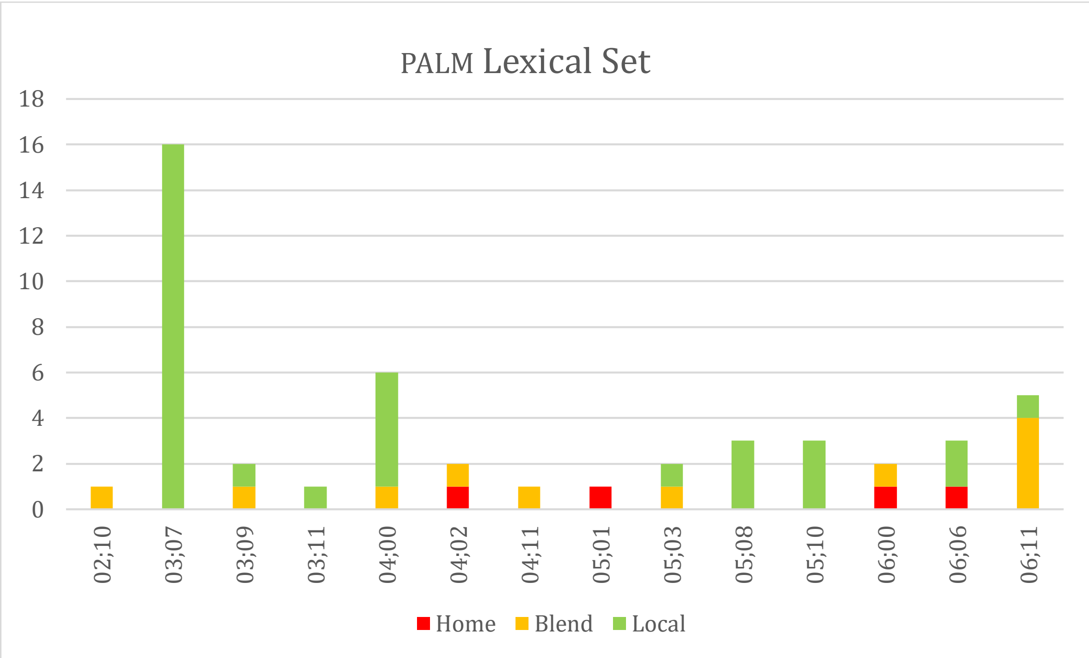


Figure 83: Realisations of the PALM *lexical set including data from 6;11*

Henry's blended variants of PALM, showing elements of both the local and home varieties include the following:

Example (1) 'I'm looking for some pieces but I can't find any'
['aɪm 'lʊkɪŋ 'fɔ səm 'pisɪz bəɾ aɪ **'kɐːnt** 'faɪnd ɛni]

Example (2) 'I'm trying to find some stuff but I can't'
[əm 'tɹaɪn tə 'faɪnd səm 'stʊf bəɾ aɪ **'kɐːnʔ**]

Example (3) ‘A lava ball, we don’t need it’

[ə **ˈlɐːvə** ˈbɔl wi ˈdəʊnʔ ˈnid ɪt]

Local variants of PALM include the same words as are produced with a blended articulation, confirming that Henry’s pronunciations are not connected to specific lexemes. Here, ‘can’t’ and ‘lava’ are produced as both blends and local variants.

Example (4) ‘He can’t even balance’

[hi **ˈkaːn** ˈivən ˈbaləns]

Example (5) ‘I found a lava thingy’

[ˈaɪ ˈfaʊnd ə **ˈlaːvə** ˈθɪŋi]

There are only five realisations of this lexical set in the entire recording, but they were all suitable for acoustic analysis (Table 39).

Orthography	**Henry Home**	**Henry blend**	**Henry local**	**James**
Lava		ˈlɐːvə	ˈlaːvə	ˈlaːvə
Can’t		ˈkɐːnt	ˈkaːn	

Table 39: Henry and James’ PALM *realisations selected for acoustic analysis*

Although there are few tokens, most have a higher and more retracted articulation than James' PALM vowel, featuring F2 between 1685 and 1960 Hz, while James' only example of this vowel is at the front edge of that space (Figure 84).

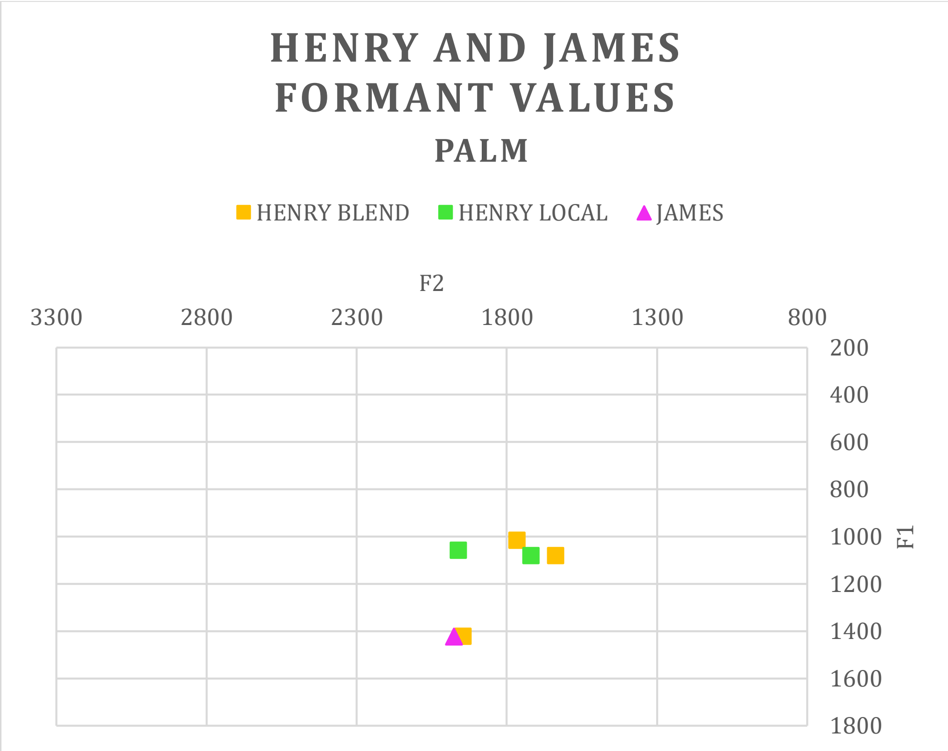


Figure 84: Henry and James' formant values for the PALM *lexical set*

The combination of these PALM vowels together with START, below, will provide a more substantial dataset.

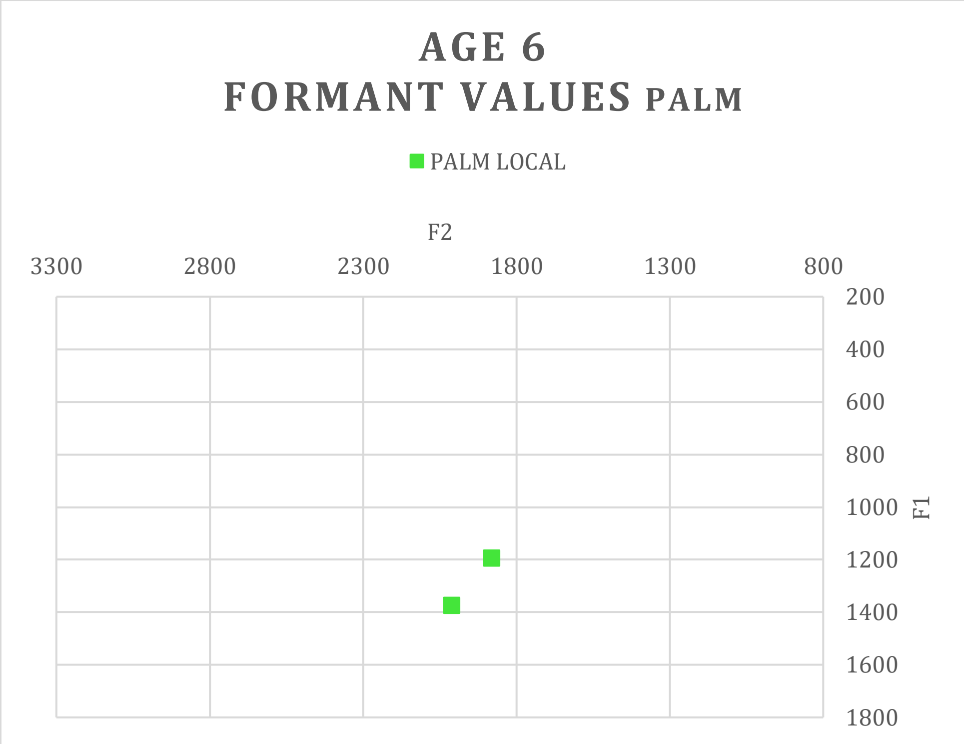


Figure 85: Henry's formant values for the PALM *lexical set at age 6 in conversation with me*

In the original analysis of Henry in conversation with me at age 6, there were very few examples of PALM, and even fewer were suitable for acoustic analysis, so only two tokens are available as a comparison, and both were judged to be local variants (see Figure 85). The acoustic space Henry used for this vowel in conversation with James overlaps with the acoustic space used in conversation with me.

6.2.2 START

Figure 86 below presents the data from the main analysis in Chapter 5 alongside Henry's realisations of START at age 6;11 in his play with James.

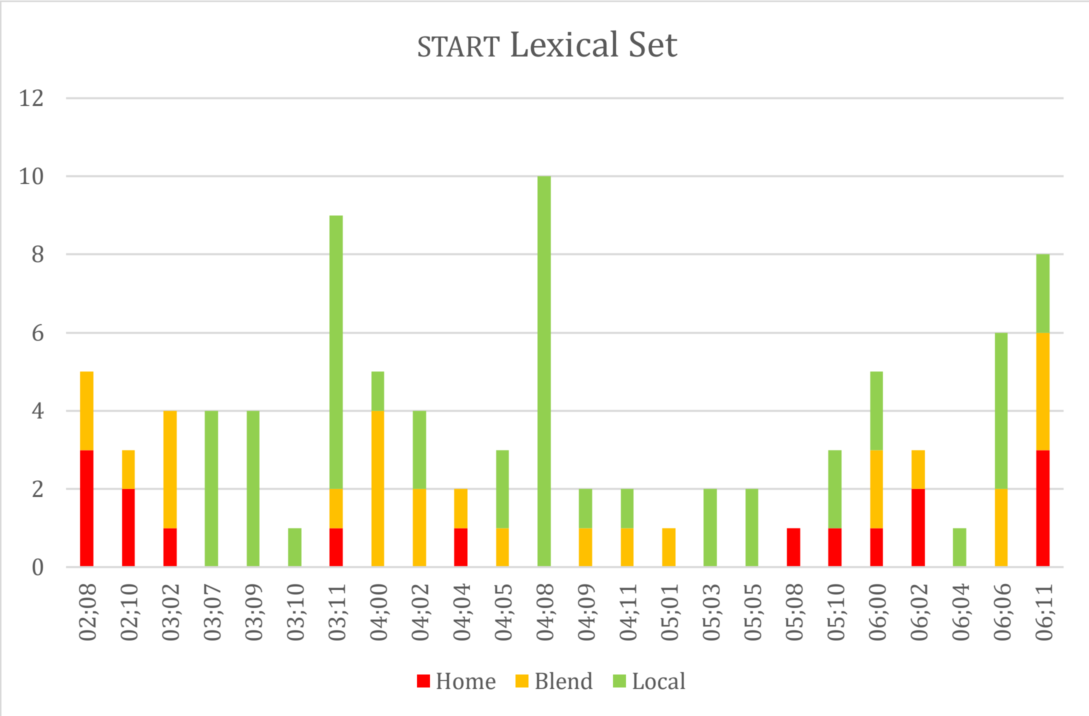


Figure 86: Realisations of the START *lexical set including data from 6;11*

Home and blended realisations of START are evenly balanced at three examples of each, with slightly fewer (two) local realisations of this vowel being present. No home realisations appear in the previous two recordings, yet these vowels re-emerge in conversation with James, a speaker who only uses local variants.

Examples of Henry's variants of START are looked at below, to check for patterns in variation. Examples consistent with home realisations include:

Example (1) 'I found Darth Vader'
[aɪ ˈfaʊnd **ˈdɑθ** ˈveɪdə]

Example (2) 'You have this card'
[ju ˈhav ˈðɪs **ˈkɑd**]

Example (3) 'Does anyone have this card?'
[dəz ˈɛniwən ˈhav ðɪs ˈ**kɑ̞d**]

‘Darth Vader’, a character in the Star Wars films, is a name that Henry will have heard his father say many times, using the home variant [ɑ], as they often watch these films together. This may have influenced Henry’s tacit choice of pronunciation of this word; however, James is also a fan, and bought Henry a Star Wars book for his birthday, so the boys are likely to have discussed the characters in the films at school as well.

Examples of Henry’s variants of START showing elements of both the local and home varieties:

Example (4) ‘I found some armour’
[ə ˈfaʊnd səm **ˈɐmə**]

Example (5) ‘There’s another arm’
[ðɛəz əˈnʌðə **ˈɐˑm**]

Example (6) ‘Can you make Minecraft cars?’
[ˈkan ˈju ˈmeɪk ˈmaɪnkɹaft ˈ**kɐːz]**

None of these variants have the same front quality as the local variety, but neither do they sound far back enough to be considered canonical home variants as seen in examples 1-3. The two variants of START consistent with local realisations are of the word ‘arm’ – however, we saw above that ‘arm’ also features a blended realisation, so there doesn’t appear to be straightforward connection between the lexical items and their production.

Example (7) ‘He's got one leg and one arm’
[hiz ˈgɒt wɒn ˈlɛg ən wən ˈ**aːm**]

Example (8) ‘Here's my - here's the arm’
[ˈhɪəz ˈmaɪ ˈhɪəz ði **ˈaːm**]

Both the home and local variant phonemes are long, but there is some variation in length seen in the blended realisations here due to the naturalistic data collection methods. This was also seen in the home variants.

As before, only a small number of tokens were suitable for acoustic analysis, due to the naturalistic data collection method (Table 40), but there are more than were available for PALM, above. Overlapping talk and noisy toys rendered some tokens uninterpretable.

Orthography	**Henry Home**	**Henry blend**	**Henry local**	**James**
arm			ˈaːm	
cars		ˈkɐːz		ˈka̰ːz
armour		ˈɐmə		
Darth Vader	ˈdɑθ ˈveɪdə			
card	ˈkɑd ~ ˈkɑ̝d			
Aren't				ˈaːnt
are				ˈaː
far				ˈfaː

Table 40: Henry and James START *realisations selected for acoustic analysis*

The sample of James' START vowels analysed acoustically are noticeably more front than Henry's START vowels in the same recording (see Figure 87).

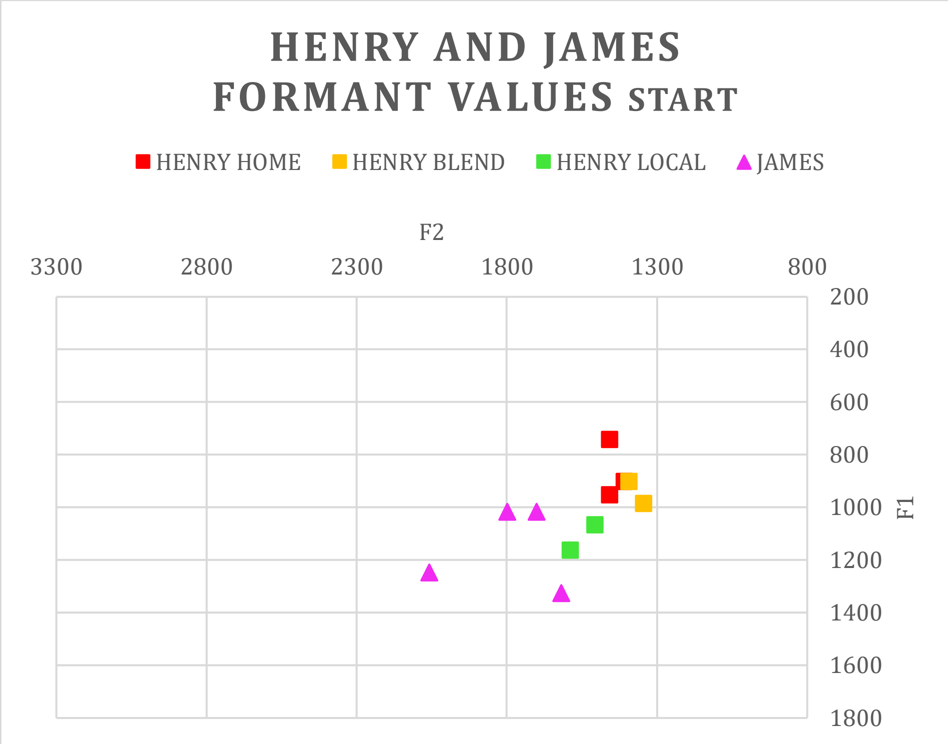


Figure 87: Henry and James' formant values for the START *lexical set*

Variants judged to be local in conversation with James (Figure 87) have a lower F2 than those produced at age 6 with me (Figure 88), ranging from 1505 Hz to 1586 Hz while speaking to James. This F2 shift reflects variants being articulated closer to the acoustic space used for the home variant, a rather different outcome than might be expected if Henry were accommodating to James' START vowels.

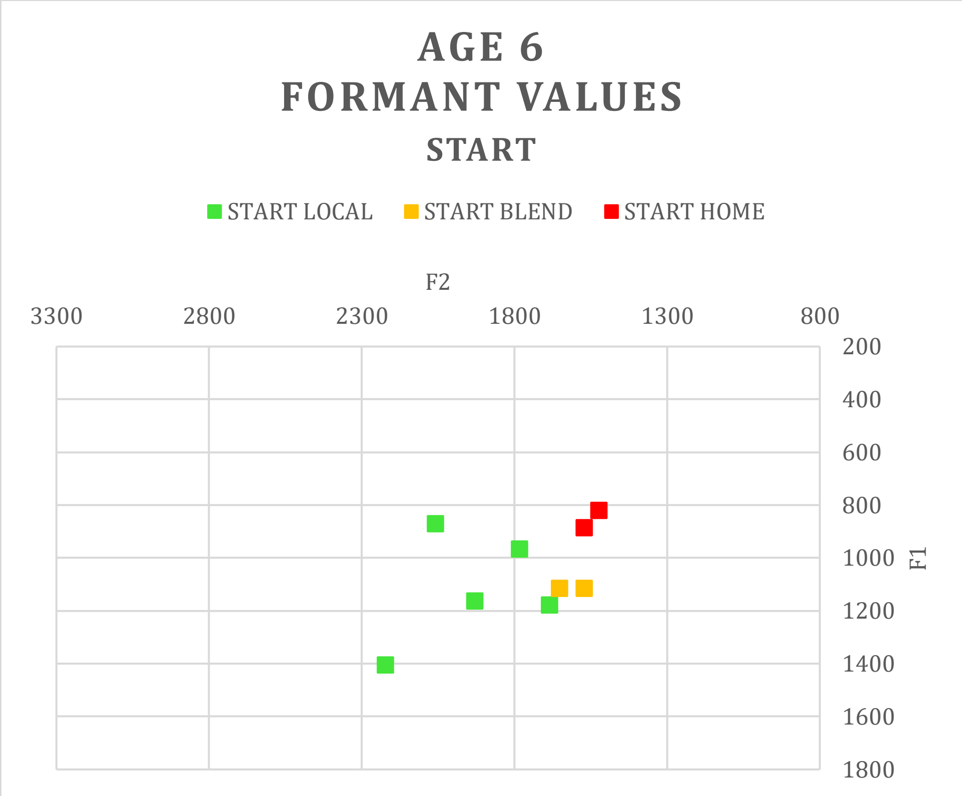


Figure 88: Henry's formant values for the START *lexical set at age 6 in conversation with me*

6.2.3 BATH

As in the main analysis (Chapter 5), and in PALM above, there are not many tokens of BATH for analysis, as this is a relatively infrequently occurring vowel. Again, the data from the recording of Henry playing with James has been combined with the data from the main analysis of Henry talking to me presented in Chapter 5. There are only four tokens of the lexical set articulated by Henry in this recording with James (Figure 89).

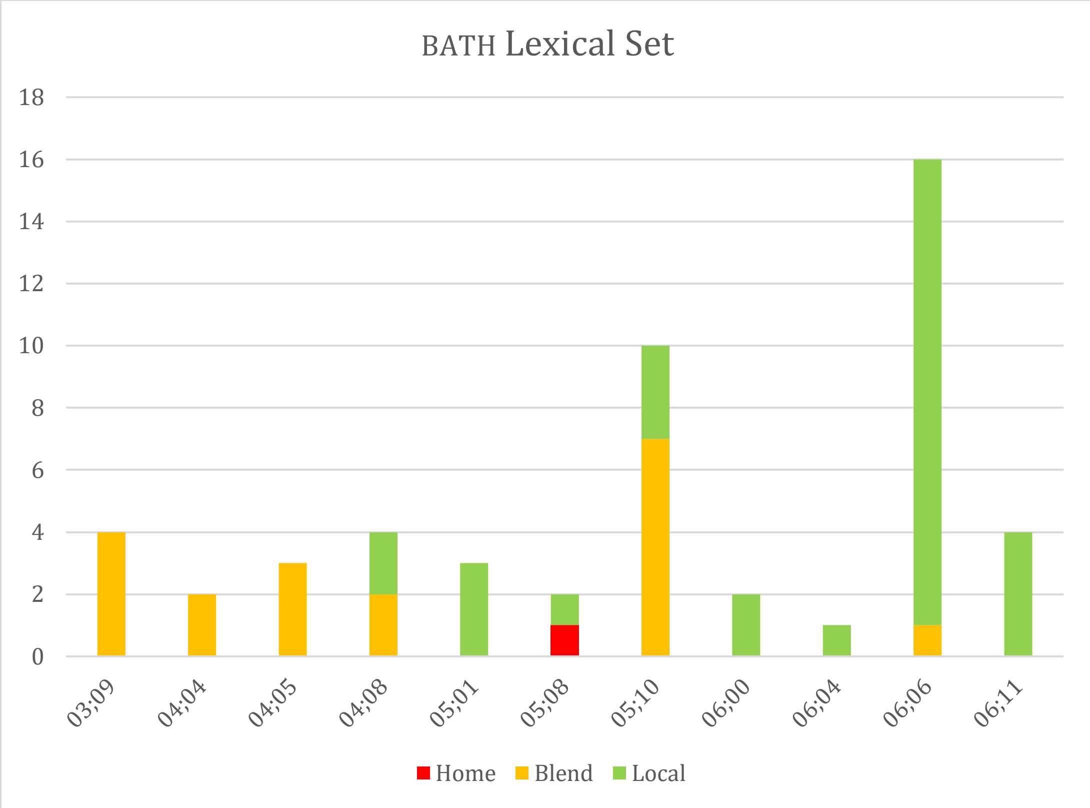


Figure 89: Acquisition of BATH *including data from 6;11*

All of Henry's variants of BATH are consistent with local realisations, a short front vowel [a], rather than the back vowel found in the home variety, [ɑ]. For example:

Example (1) ‘I built my Minecraft figure’
[aɪ ˈbɪlt ˈmaɪ **ˈmaɪŋkɹaf** ˈfɪgə]

Example (2) ‘Can you make Minecraft cars?’
[kən ju ˈmeɪk **ˈmaɪnkɹaft** ˈkɐːz

Example (3) ‘I don't play Minecraft anymore’
[aɪ dəʊn ˈpleɪ **ˈmaɪnkɹaft** ˌɛniˈmɔ]

Orthography	Henry Home	Henry Blend	Henry Local	James
Minecraft			ˈmaɪnkɹaft ˈmaɪnkɹaf	ˈmaɪnkɹaft
crafting				ˈkɹaftɪŋ

Table 41: Henry and James BATH *realisations selected for acoustic analysis*

Even fewer examples are suitable for acoustic analysis due to the children's noisy play (see Table 41). Combining the two boys, there were a total of 15 tokens of this lexical set and all but one was an articulation of the word 'Minecraft'. The only clear tokens are in the words 'crafting', and 'Minecraft', which appear multiple times.

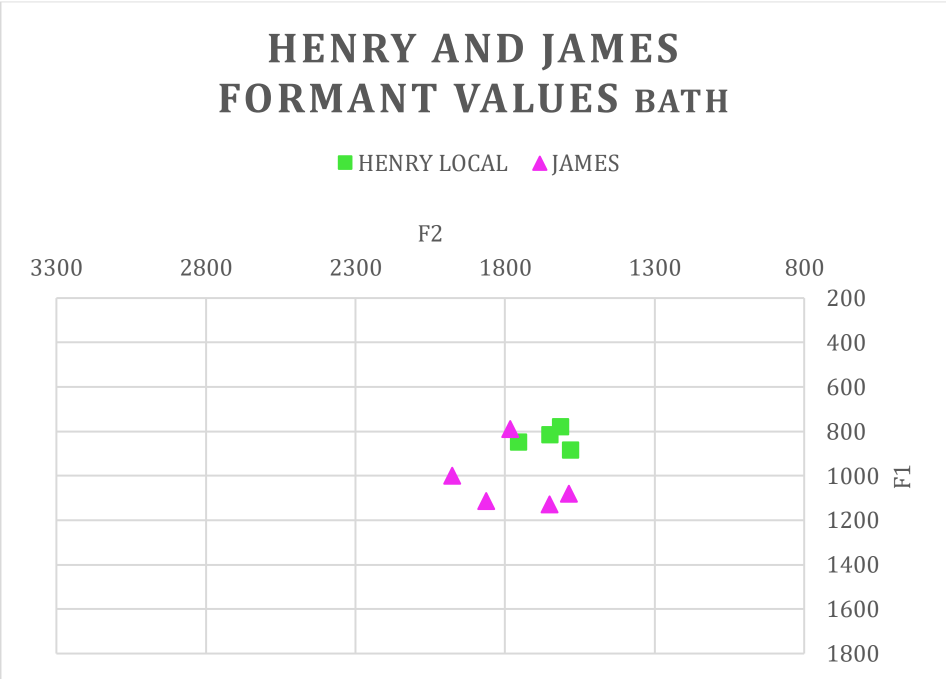


Figure 90: Henry and James' formant values for the BATH *lexical set*

Once again, even though these tokens were judged as local, the plot of F1 and F2 reveals a slightly higher, backer realisation than seen in James' speech, and are tightly clustered together (Figure 90).

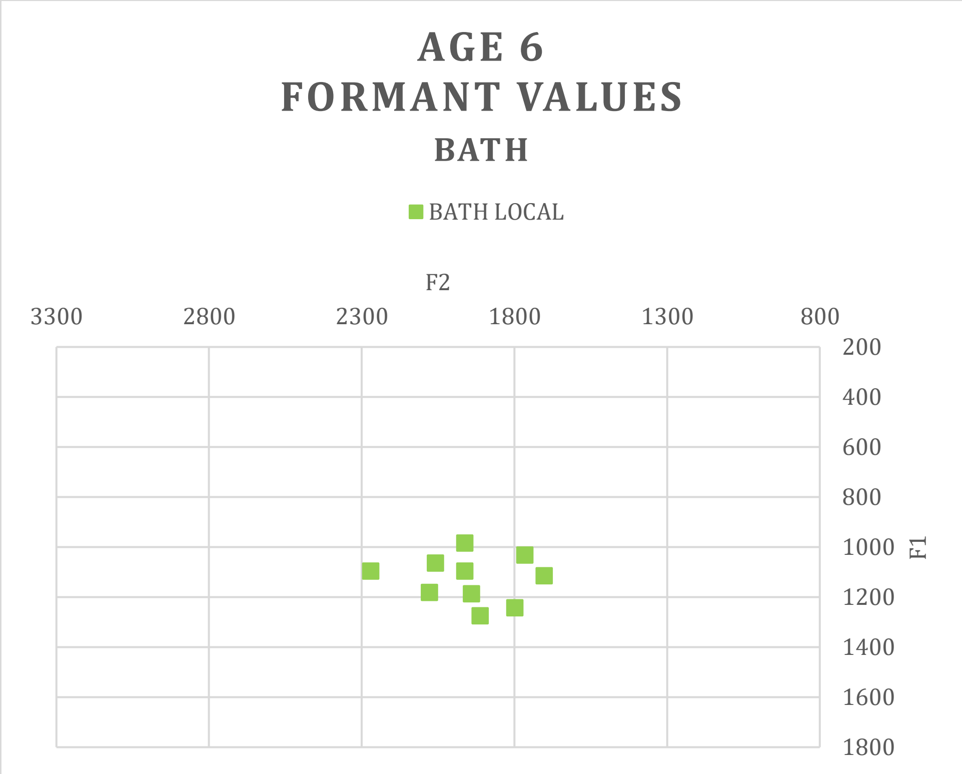


Figure 91: Henry's formant values for the BATH *lexical set at age 6 in conversation with me (copy of Figure 63)*

Looking back to Henry's articulation of the vowel in BATH words in conversation with me at age 6 (Figure 91), while he did still produce the occasional BATH word as a blend, only local variants were suitable for acoustic analysis. When we consider the distribution of these BATH variants, they occupy a larger acoustic space, even though they represent six repetitions of the same word and are examples of read speech. Surprisingly, these read tokens are more variable than Henry's rapid, excited realisations when he is talking to his friend.

6.2.4 PALM/START/BATH comparison

Above, PALM, START and BATH were presented by lexical set, drawing comparisons between the two children's realisations of the vowels in these sets, and comparing Henry's articulations to those when he was in conversation with me at age 6. Here, we look at these lexical sets together to see whether any patterns in their articulations are evident.

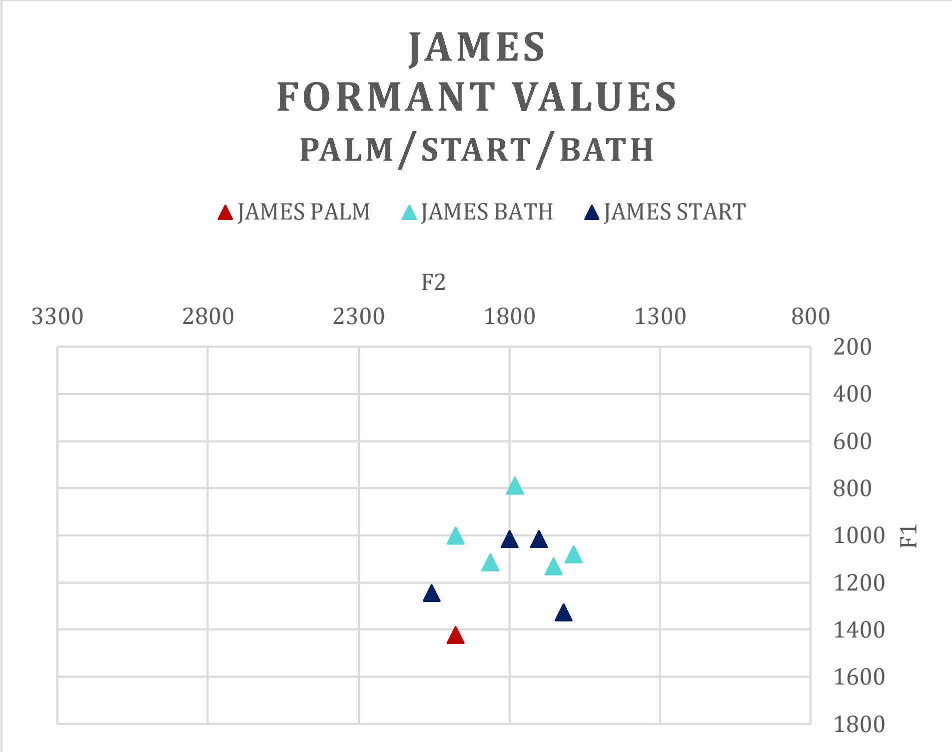


Figure 92: James' formant values for the PALM/START/BATH *lexical sets*

When PALM, START and BATH are combined into a single representation of the vowel space (Figure 92), we can see that James' realisations of BATH and START overlap. There is only one example of PALM, so it is difficult to say whether this lexical set overlaps acoustically, but it would be expected that he would produce these in the same acoustic space as these two lexical sets are not differentiated in the local accent. BATH and PALM/START are differentiated by length in James' accent as is usual for the local accent, but not in quality.

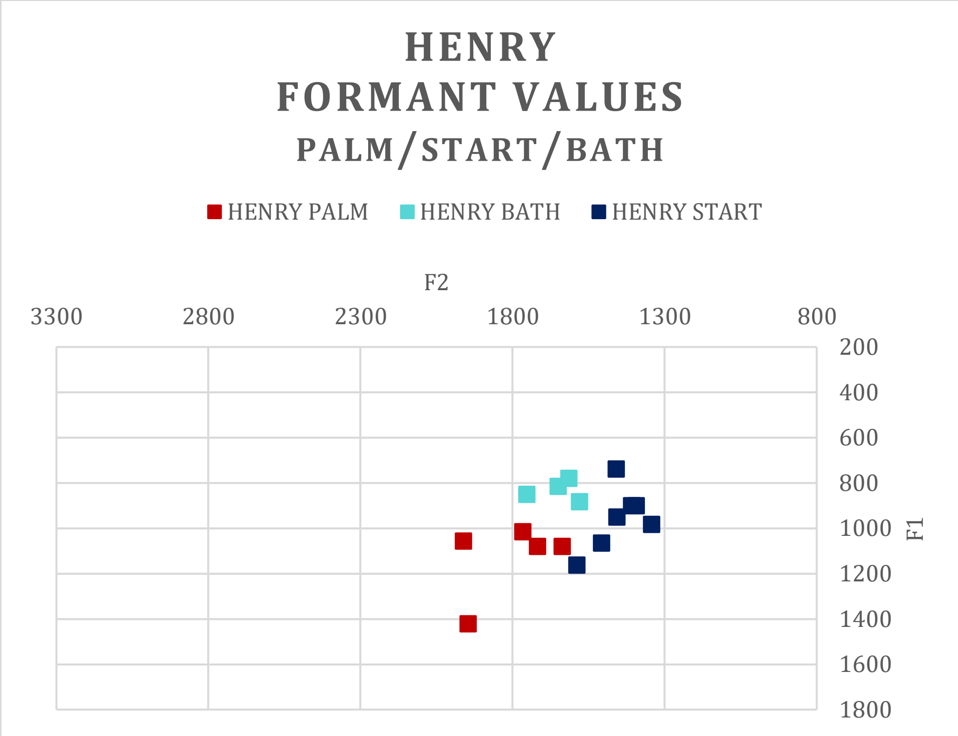


Figure 93: Henry's formant values for the PALM/START/BATH *lexical sets in conversation with James*

When we consider Henry's productions of these vowels, however, there is a noticeable distinction between the acoustic space each of these lexical sets occupies (Figure 93). Realisations of each lexical set are clustered together.

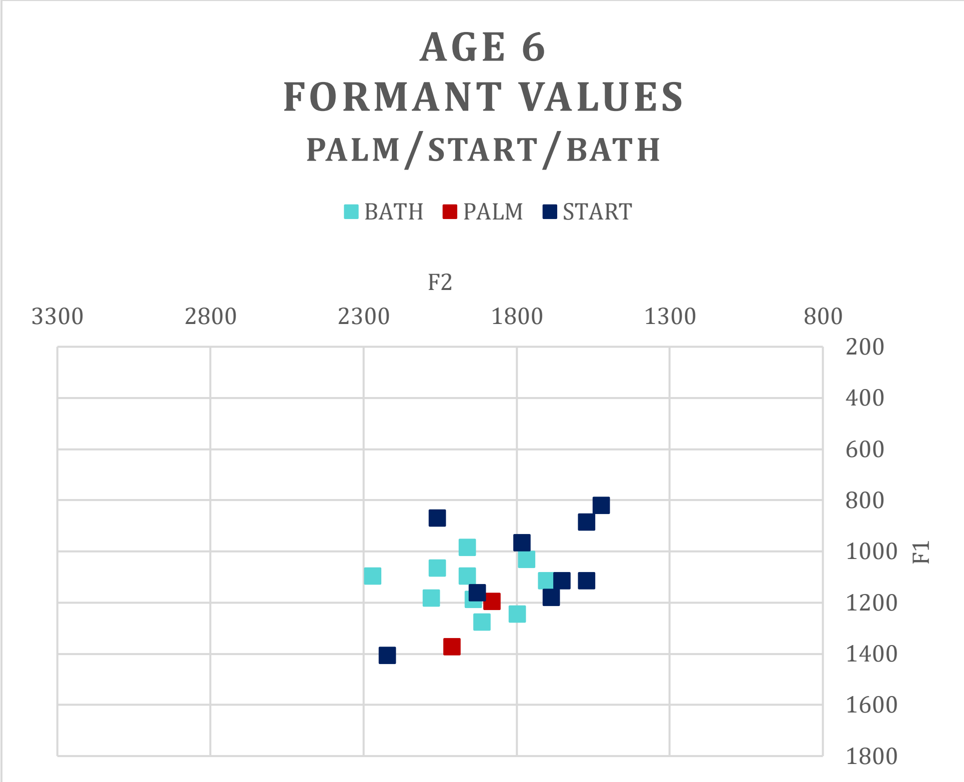


Figure 94: Formant values for PALM/START/BATH *at age 6*

Looking back at Henry's realisations of the vowels in these lexical sets at age 6 in conversation with me (Figure 94), the picture is less clear. Although BATH vowels appear to form a visible cluster, PALM and START overlap with the BATH set in both F1 and F2 values, though these do represent tokens which were judged as home, blended and local variants. Figure 95 shows these variants split as they were seen in Chapter 5.

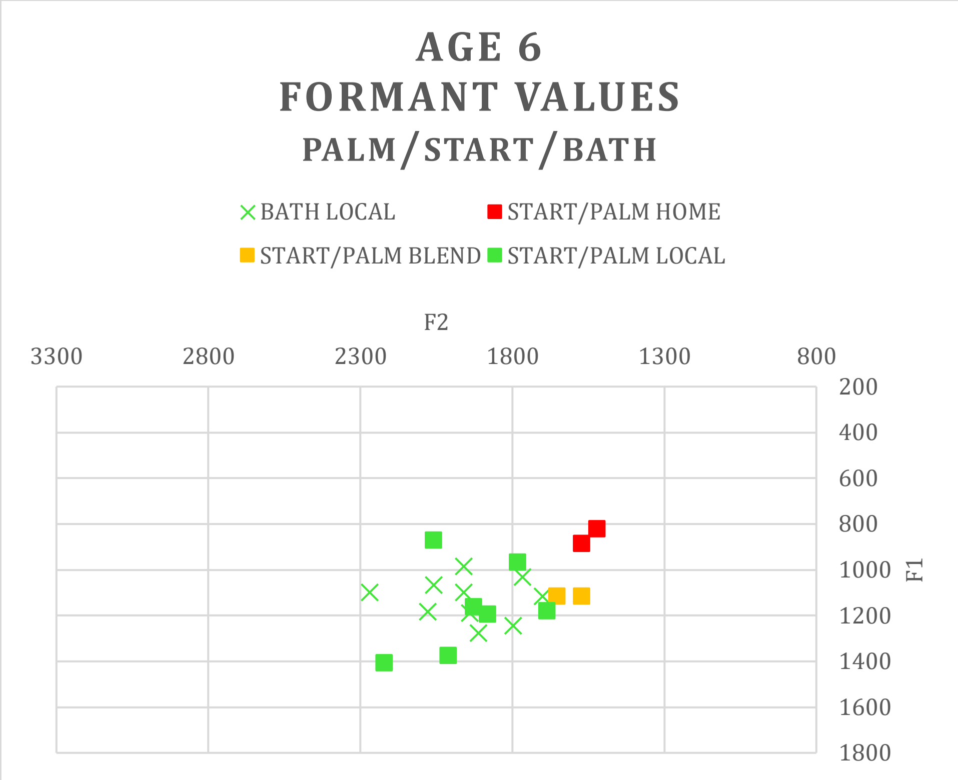


Figure 95 (copy of Figure 71 with TRAP *values removed): Formant values* PALM/START/BATH *at age 6*

There are two possible interpretations of this difference between the distribution of the PALM/START/BATH tokens in conversation with me and in conversation with Henry. Perhaps the tightly clustered realisations are as a consequence of analysing data from a single recording, in the same style — rapid and excited — featuring repetitions of a small number of lexical items rather than being indicative of a more significant phonological distinction between the lexical sets PALM and START (Table 42). On the other hand, we have noted that Henry can produce multiple different realisations of a vowel within the same turn, including in his play with James.

PALM	START
can’t	arm
lava	armour
	car
	card
	far
	are
	aren’t

Table 42: PALM *and* START *words at age 6;11 in conversation with James*

Alternatively, it is possible that Henry’s PALM and START vowels are still moving. What about the split in distribution between PALM and START (Figure 94), as there is no overlap in acoustic space in the recording with James? It seems implausible that Henry would differentiate these lexical sets when they are not differentiated in the home and local varieties, when contact with the only person in his network who differentiates these vowels, his grandmother, is so limited. Although the division between these two sets is connected to their spelling, with START vowels featuring a non-prevocalic <r> after the vowel while PALM vowels do not, Henry’s awareness of spelling was quite limited at this age, so it is very unlikely that he has made this connection and manifested it in his speech. It seems much more likely that the first explanation is responsible; that this apparent complementary distribution is a consequence of a small number of tokens of the same word being spoken in the same recording in the same style.

In Figure 96 below, the data from Henry's meeting with James has been added to the data from Chapter 5. Here, home and blended variants have been combined to create a binary opposition between variants which show some influence of the home variety and those which do not.

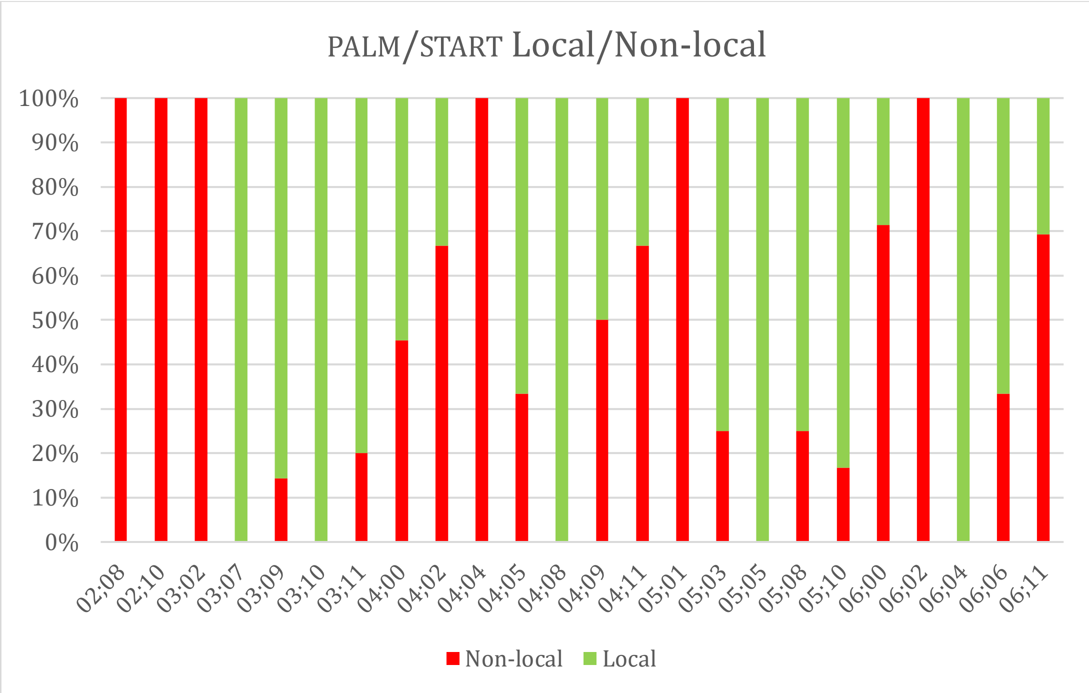


Figure 96: Acquisition of PALM/START *including data from 6;11*

When START and PALM are considered together, the data exhibits a striking pattern where the local variant looks to be gaining a toehold, only to see the home variant re-emerging. There appear to be no signs of accommodation to James' accent affecting Henry's realisations of these vowels; the influence of the home variants appears to be stronger than his last recording where he was speaking to me alone, continuing the pattern of re-emergence seen in previous months.

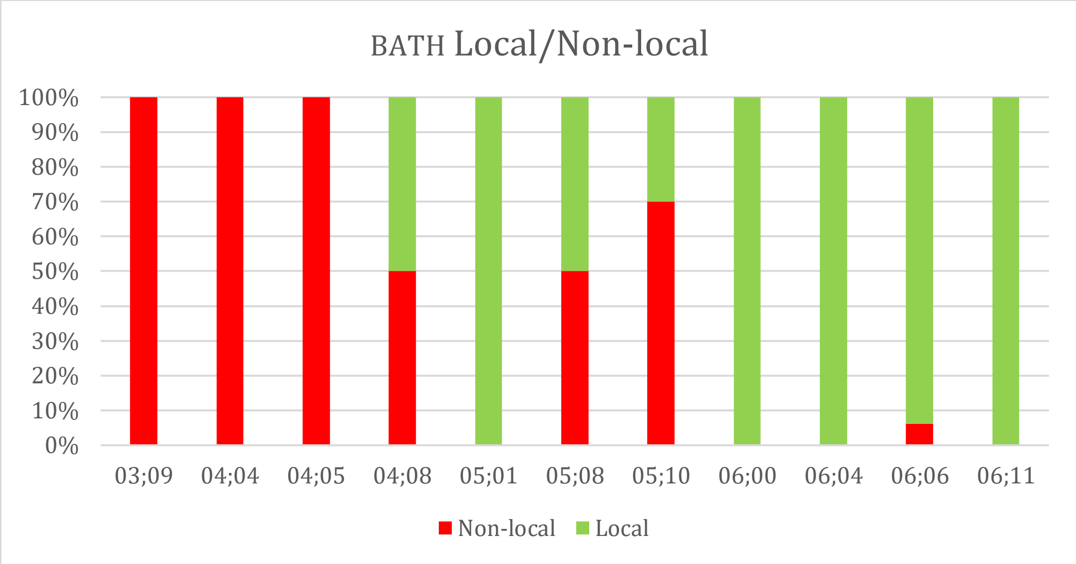


Figure 97: Acquisition of BATH *including data from 6;11*

Conversely, although BATH was already well on its way to being acquired in the recordings at age 6, Henry's BATH variants are all local in his conversation with James. Any influence of the home variants has disappeared altogether (Figure 97).

6.2.5 TRAP

In the main analysis in Chapter 5, TRAP realisations were found to be much more straightforward than the realisations of the other lexical sets considered in this thesis. TRAP words are identical in both the home and local varieties, and the vast majority of realisations of words in this frequently occurring lexical set were produced consistently with those varieties. However, the influence of the overlap between TRAP and BATH did cause Henry some issues. In a small number of tokens there was evidence of hypercorrection as he produced some TRAP vowels as a long vowel, particularly in the word 'Grandma', which was produced variously as ['gɹaˑməz], ['gʊɑmɑ] and ['gʊɒˑmɑ]. In conversation with James, no TRAP variants were produced with a long or back vowel, so an acoustic analysis has not been performed.

6.3 GOAT/FACE

While the GOAT and FACE lexical sets have not been analysed as part of the main analysis in Chapter 5, I offer an impressionistic comparison here of Henry and James's realisations of these vowels, as these are a site of difference between the home and local accents in the case of many speakers, including James. Wells (1982) describes the vowel system of the nearby city of Leeds as realising GOAT and FACE as monophthongs [oː] and [eː] respectively. These monophthongal variants are widely described as being 'pan-northern' (Beal, 2004; Haddican et al., 2013; Watt & Milroy, 1999), though there is evidence that GOAT may be fronting over time in some parts of Yorkshire (Watt & Tillotson, 2001). Speakers in 'some of the urban middle north' may use the diphthongs [oʊ] or [əʊ] for GOAT and [ɛɪ] for FACE (p. 364).

Henry consistently produced both vowels as diphthongs, as is normal for him. No signs of monophthongal realisations of these lexical sets have ever been observed in his speech, and there was no alteration to this in conversation with James. No acoustic analysis has been performed as the phonetic differences are very clear impressionistically.

6.3.1 GOAT

GOAT was typically realised as [o] or [oː] by James, but also [o̞] and on a couple of occasions as the diphthongs [əʊ] and [oə]. Henry's realisations of GOAT vary between a diphthong starting with a front vowel, [eʊ] and one starting from a central position, [əʊ], but never a monopthong (Table 43).

Orthography	**Henry IPA**	**James IPA**
closer		ˈklosə
no	ˈneʊ ~ ˈnəʊ	ˈno̰ːː
Joker(s)	ˈʤe̞ʊkəz	ˈʤokə
going		ˈgoɪŋ ~ ˈgəʊɪŋ
Jokermobile	ˈdʒəʊkəməˌbil	
totally	ˈtəʊʔəli	
robot	ˈɹəʊbɒʔ	
bones		ˈbonz
Don't	ˈdəʊ̃ʔ ~ ˈdəʊnʔ	ˈdəʊnʔ ~ ˈdoənʔ
spoken		ˈspokən
those	ˈðeʊz (2)	
Ghost(ie)		ˈgo̰̞ːsti ~ ˈgo̰̞ːst

Table 43: Henry and James GOAT *realisations*

There are no signs of Henry accommodating to James' accent in this vowel. James' occasional diphthongs may be evidence of him accommodating to Henry's accent, but as I have not recorded James in any other context, this is speculation.

6.3.2 FACE

Henry's productions of FACE were consistently realised as [eɪ], while James realised members of this lexical set as [e] or [eː] (see Table 44).

Orthography	Henry IPA	James IPA
blaze		ˈbleːz (2)
make	ˈmeɪk (2)	
anyway	ˈɛniweɪ	
today		təˈdḛ
decorate		ˈdɛkəɹeːt
faces	ˈfeɪsɪz	
grey	ˈɡɹeɪ (3)	
take		ˈtek
stay	ˈsteɪ	
Jacob		ˈʤekəb (2)
lightsaber	ˈlaɪʔseɪbə	
make/making		ˈmek (2) ~ ˈmekɪŋ
say	ˈseɪ	
played	ˈpleɪd	
laser	ˈleɪzə	
plane	ˈpleɪn	

Table 44: Henry and James FACE *realisations*

Again, there is no sign of Henry accommodating to James' FACE vowel in this recording.

6.4 Other aspects of accommodation

While Henry shows little to no sign of accommodating to James' accent in his vowels, I did observe high rates of glottal replacement of /t/ in Henry's speech in comparison to James. His speech was rapid, and he appeared excited to have his friend at home. The high levels of glottal replacement led me to believe that Henry might be accommodating to James' speech in other ways. Glottal replacement, h-dropping, and the 'non-standard' realisation of words ending in verbal -*ing* as [ɪn] are all features which speakers may have some degree of control over and can shift according to their interlocutor. An initial investigation was therefore conducted in order to establish whether Henry's speech exhibited higher levels of these features than James', and whether there is any difference between his use of these features when talking to his friend in comparison to his conversations with me.

Labov (1973) differentiates between features which he labels as 'indicators' and 'markers'. Indicators, he claims, are linguistic units not subject to style shifting. Markers on the other hand, shift according to style. Indicators and markers differ according to the individual. For example, for Henry's father, the realisation of -*ing* as [ɪn] is consistent – it is an indicator which does not change according to style. For me, however, the same feature is a marker, as it is socially indexical (Rácz, 2012) and varies according to formality. We have seen above that Henry's vowel usage appears to fall into the category of indicator; he does not change his vowel usage according to his interlocutor. Below, I will compare the usage of linguistic features which are variable in Henry's speech to James' usage to attempt to establish whether he is accommodating to James.

6.4.1 Glottal replacement of /t/

The closure of the vocal folds as a realisation of /t/ is commonly found in many speakers. Wells (1982) writes that glottal replacement spread quickly during the 20th century, as it had only been noted in a few locations (London, East Anglia and parts of Scotland) during the early part of the century, and yet by 1982 it was reported in some phonological environments in R.P.. In 1999, Trudgill described /t/ glottalisation as 'one of the most dramatic, widespread and rapid changes to have occurred in British English in recent times (Trudgill 1999: 136). Once a stigmatised feature, glottal stops are now 'entirely standard' in pre-consonantal positions (e.g. 'football'), including in word final position (e.g. 'that one') (Lindsey, 2019), though it remains 'non-standard' in some environments. The literature suggests that realisation of intervocalic /t/ as [ʔ] is generally much less frequently found than word final glottal replacement (Smith & Holmes-Elliott, 2018) and probably remains the most socially stigmatised environment for /t/ glottaling, for example in words such as 'crater'. /t/ glottaling is not possible in all environments, for example, in word initial positions such as 'top', and it is not usually possible before a stressed vowel, for example, 'attack'. However, this is possible in some dialects such as Buckie in North-West Scotland (Smith & Holmes Elliott, 2017). Its frequency tends to be conditioned by phonological environment.

Candidates for glottal replacement were identified, and those examples of /t/ which were realised as a glottal stop are presented as a percentage in Table 45, below. /t/ elisions (e.g. 'definitely' ['dɛfnli]) and tapped realisations (e.g. 'what do you need' ['wɒɾəjə 'nid]) were excluded from this data.

	[t]		**[ʔ]**		
	N	**%**	**N**	**%**	**N**
Henry	20	**21**	75	**79**	95
James	28	**43**	37	**57**	65

Table 45: Overall glottal replacement of /t/ in conversation with James

Henry's overall realisation of eligible /t/ as a glottal stop is high (79%) in this recording with James. Henry realises /t/ as a glottal stop much more than his friend, whose percentage of total realisations is 57%. However, when Henry's total level of glottal replacement was cross-checked against his total level of glottal replacement in conversation with me, he exhibits similar levels of this feature (Table 46). For purposes of comparison, a recording from age 6;0 was selected. In this recording, we were playing with Lego. This recording was selected as in some of the later sessions, Henry was reading for part of the recording, which would most likely have led to a more careful speech style, and therefore be likely to elicit more standard variants.

	[t]		**[ʔ]**		
	N	**%**	**N**	**%**	**N**
Henry	12	**23**	40	**77**	52
Mother	28	**21**	105	**79**	133

Table 46: Overall glottal replacement of /t/ in conversation with Henry's mother

As Henry's total level of glottal replacement is in line with my own, and he shows very similar levels in conversation with his friend, no further analysis of glottal replacement was performed.

6.4.2 H-dropping

'H-dropping' refers to the phonological process of eliding realisations of /h/in stressed syllables. /h/ is phonologically restricted in English, only appearing at the start of syllables. Its elision is considered non-standard and is socially stigmatised (Wells, 1982), though the elision of /h/ in unstressed function words such as 'him' and 'her' is typically found in standard speech (Lindsey, 2019). While it is assumed that all English dialects used to pronounce /h/, it has been lost by most local dialects over the past few hundred years (Trudgill, 1999), though there is some evidence that this trend could be reversing (Cheshire et al., 2008; Jansen, 2021). Wells claims that most working-class dialects of English (in England) feature h-dropping, and that it remains 'the single most powerful pronunciation shibboleth' (1982: 254). Upton and Widdowson (2006) provide a map of /h/ retaining and /h/ dropping areas in the British Isles (Figure 98). Of course, not all speakers elide /h/ one hundred percent of the time, as this feature is subject to stylistic variation (Trudgill, 1999). Henry's home in North Yorkshire falls clearly within the largest h-dropping area in England.

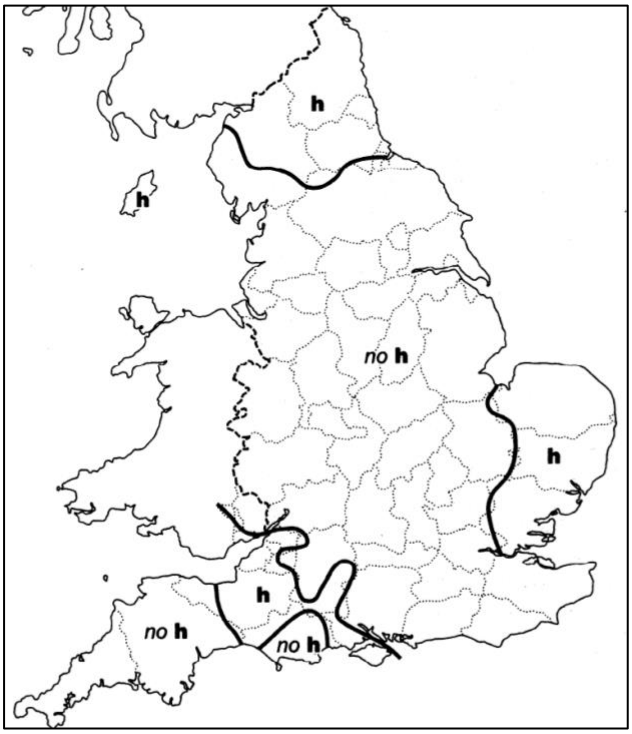


Figure 98: Presence/absence of /h/ in the British Isles (Upton & Widdowson, 2006: 58)

Given Henry's heavy use of glottal replacement, and the widespread status of /h/ as a feature subject to style shifting, the boys' use of /h/ dropping was considered.

However, Henry only elided stressed /h/ on two occasions during his conversation with James (both in the word 'have', which commonly features no /h/ in unstressed auxilliary positions in standard accents), and James did not delete /h/ at all, even in unstressed positions.

6.4.3 -ing

Many words ending in -*ing* may be pronounced as [ɪŋ] or [ɪn]. The velar nasal ending is typically associated with higher social classes or formal speech styles, while [ɪn] is more commonly found in lower class speech or more informal speech styles (Wells, 1982: 262). All verbs with an -*ing* suffix are candidates for this pronunciation variation, but nouns (e.g. 'something, morning, ceiling') and adjectives (e.g. 'cunning') can also feature this variant. 'Thing' is not eligible for this variant (though I have personally heard /θɪŋ/ consistently articulated as [θɪn] in one speaker as an idiolectal feature). Again, this feature has been considered here due to Henry's use of glottal replacement, and the associated status of this feature as one which can alter according to formality. As with /h/ dropping, Henry only produces the non-standard variable on a very small number of occasions (2), as does James (1). There is not enough evidence of this feature being produced to warrant further analysis or discussion.

6.5 Accommodation summary

The lexical sets analysed in Chapter 5 were analysed in Henry's speech while he played with his friend, James, to look for evidence of further local influence or reduction in influence of the home accents. In addition to these lexical sets already analysed, the GOAT and FACE sets were analysed impressionistically in order to check for local influences. Henry shows no signs of realising these lexical sets as monophthongs at home, as is the case for some local speakers, including many of Henry's friends, but as he spends approximately six or seven hours at school each day, mostly interacting with children with a local accent, it is important to establish whether his accent changes when he is speaking to friends.

STRUT showed slightly more evidence of the home influence in conversation with James than it did at age 6 when Henry was talking to me. The acoustic analysis revealed that the whole range of local, blended and home variants took up a much smaller acoustic space than in previous recordings. Perhaps this can be attributed to Henry's articulators being subject to a smaller range of movement due to the rapid speech in this recording with James. This corresponds with Lee et al.'s (1997) claim (raised in section 3.2.4.4) that fast speech rates in teenagers led to less variation in vowel production.

FOOT also showed a small step back from the overall move to local variants in the conversation with James. While there were fewer blended examples in this conversation than at age 6 in the main analysis, some front realisations were still present. An unusual diphthongal pronunciation of 'foot' was highlighted:

Example (7) 'I found a foot'
[aɪ ˈfaʊnd ə **ˈfʊət**]

This was repeated several times before being articulated as a back vowel:

Example (4) 'I found a foot'
[aɪ ˈfaʊnd ə **ˈfʊt**]

Is this an example of Henry attempting to accommodate to James' accent? Previously he has produced some similarly back realisations of FOOT in conversation with me, so it is not only in conversation with James that back variants are found. As there is only one example of such a back articulation in conversation with James, evidence to support this hypothesis is weak.

There were very few tokens of PALM in Henry's speech in the recording with James (five tokens in total), all realisations of two words, 'can't' and 'lava'. In conversation with James, Henry's PALM vowels featured lower F1 and F2 than James' realisations, indicating a closer jaw position and a more retracted tongue, i.e. closer to the home realisation than James'. This is the same acoustic space as Henry uses in the recordings with me, and may suggest home influence, as these lexical sets are realised with a more retracted tongue position in the home variety. START vowels also show the influence of home, as home and blended variants outnumber local variants.

As in the main analysis, there were not many tokens of BATH. In this recording, all realisations of this lexical set were local, continuing (and completing) the trend of acquisition of the local vowel. As was the case for PALM and START, the acoustic analysis places this vowel as closer and more retracted than James' BATH vowels. However, in conversation with me, his BATH vowels overlap with James' realisations, so this appears to be an issue with this recording in particular, perhaps for the same reasons as we saw STRUT realisations forming a tighter cloud of realisations in comparison to the earlier recordings; the effect of rapid speech and data all coming from a single recording.

The analysis of GOAT and FACE did not show any differences in Henry's vowels in comparison to the earlier recordings. He continued to produce these vowels consistently as diphthongs and showed no signs of accommodating to James' monophthongal realisations of these lexical sets.

This chapter has presented a comparison of Henry's vowels alongside those of his friend James, and has compared these to Henry's articulations presented in Chapter 5. In the following chapter I draw together the main results from Chapter 5 and this chapter, and compare these to the literature presented in Chapter 2 and Chapter 3.

Chapter 7 Discussion

7.1 Introduction

This thesis has followed the phonological development of a child living in a multi dialectal environment from the age of 2 to nearly 7 years old. The lexical sets STRUT, FOOT, PALM, START, BATH and TRAP were analysed, as these represent the site of the most significant phonological differences between the accents spoken in the child's home by his parents, and those spoken in the local area (Chapter 5). The vowels produced in words belonging to these lexical sets all showed evidence of phonetic compromises between the home and local variants, representing a blend of articulatory features of the vowels found in both the home and local accents. As the child was recorded in conversation with me (his mother) and may have been accommodating his speech to mine at home while using a different variety outside of the home, he was also recorded in conversation with a school friend in order to look for any accent differences between these two settings (Chapter 6).

As set out in Chapter 1, the linguistic development of children who live in multidialectal environments is rarely represented in the language acquisition literature. In particular, there is a lack of systematic *longitudinal* research in this area. We know, for example, that children's accents are usually heavily influenced by their peer group (e.g. Hazen, 2002; Kerswill, 1996), and that they may not fully acquire complex features of a new dialect if they move to a new area with a different linguistic system (e.g. Payne, 1980; Roberts, 1997). However, little has been reported on how children negotiate this process over time. The child's job of acquiring a phonological system is made even more complex by the huge range of variation in their environment — not just the range of speakers and their differing accents, but *intra*-speaker variation and idiolectal features (Local, 1983). Smoothing out these complexities is tempting, as it conveniently simplifies the dataset (cf. Local, 1983), but it does not reflect the reality for most children. Even within a supposedly homogenous speech community, variation will undoubtedly be found.

Here, a case study of a single child with parents from different dialect areas presents an opportunity to examine the development of his phonological system over an extended period, as he navigates acquisition in the face of multiple phonological models in his environment.

The results in Chapter 5 and Chapter 6 expose a good deal of variation in Henry's articulations of the vowels examined: variation occurs over the four-year period of data collection, but also within individual recorded sessions and even within a single turn-at-talk. Does this variability fall within normal ranges? We saw in Local's (1983) research in section 2.3 a description of extensive variation in realisations of a single vowel in a child's speech, some of which were neither stylistically nor lexically determined. This suggests that variability in realisations is to be expected, even in children with local parents whose input is less diverse than Henry's, as they have not yet developed consistent motor control. How much of Henry's variability can be attributed to these expected ranges of variation, and what can be specifically attributed to Henry's exposure to more than one dialect? As Local points out, there is no stable phonetic model, even within a single adult variety, but in Henry's case, he has been exposed to *multiple* systematically different *and* potentially unstable varieties. So how does he resolve this mixed, probably unstable input, and how was this manifested in his phonetic realisations? We began by considering what aspects of speech sounds children are capable of interpreting as they develop.

7.2 Accent perception in children

The literature on children's perception of accents mostly reports an uncontroversial trend: Children get better at recognising/differentiating between unfamiliar accents as they get older. In most cases, this is attributed to increased exposure to different accents.

Some research focusses on accents which are 'unfamiliar' or even 'foreign' or asks children to identify accents as 'local'. For children in multidialectal environments, these can be problematic labels. Are not all accents in a child's environment local, as they form part of the child's input, which is necessarily local to them? Beck (2014) asked children whether a range of speakers' accents were local or non-local, and discovered that children with a parent from outside the local area didn't know what a 'local' accent *was*. Whether Henry knows that the different accents in his environment are local or not is unimportant. Here, the aim is only to track how he processes and resolves these multiple, and sometimes contradictory, phonetic and phonological inputs by examining his speech output.

I have drawn a distinction throughout this thesis between home and local accents; these two categories are broadly different in Henry's case, especially in the lexical sets examined here. I have taken care to refer to local 'varieties' rather than a single monolithic local 'variety', but still, for convenience and by necessity (because in practice it would be unrealistic to fully account for *all* of the variation in an area), I have assumed these groupings of home and local varieties to be broadly homogenous.

In the face of exposure to new accents (local or non-local), children are able to quickly update their phonological systems (White & Aslin, 2011) while adults can be more constrained by their own phonology (Maye et al., 2002). Exposure to variation builds a child's tacit understanding of what level of variation leads to phonological contrast, and what does not; something that van der Feest and Johnson (2016), term 'phonetic noise'. Of course, non-contrastive phonetic details are not necessarily just 'noise'. Much of this detail might be relevant to sociolinguistic meaning, and children need to work out what these details are and how they work. Children in multidialectal environments have the extra-complex task of doing this at the same time as tracking phonological variants in more than one variety, which can lead to them simplifying or 'collapsing' information (van der Feest & Johnson, 2016). This perceptual process is also evident in children's speech production, as reported in the literature in section 2.3.1, and as we have seen in Henry's blended articulations. Henry started nursery at 24 months old, so it was at this point when substantial and sustained exposure to varieties outside the home began. At 25 months old, van Heugten and colleagues (2015) claim that prior exposure to an accent offers no additional benefit in understanding that accent compared to children who have had no previous exposure. So even at this very young age, children are capable of comprehending unfamiliar accents, though it seems likely that in real-world contexts their understanding will not be perfect, and their expertise will grow gradually with increased exposure. In Henry's case, he was spending 4 days each week in nursery, so these accents are likely to have become familiar quickly. We will see in section 7.10 below, how soon after starting nursery he started to exhibit local variants.

7.3 Parent and peer influence

There is no agreement in the literature about what it is that causes a child to fully acquire a local variety (or not) when they have parents from out of the area. The outcomes for the children investigated vary in each case. All arguments offered previously seem plausible explanations for the contexts described. Complexity of rule to be acquired (Payne, 1980; Roberts, 1997), age on arrival (Payne, 1980), social integration (Labov, 2001) and the child's personality (Deser, 1989) have all been implicated as factors in the acquisition (or lack thereof) of local variants. Thomas and Scobbie (2015) showed that even within the same family, children could acquire different variants, some more influenced by parental varieties, others by the accents of their peers. Where children showed mixed influence, this varied in its presentation. Kerswill (1996) described what he called a 'compromise' in the vowels of one of the children in his study, where some element of a sound was shared with the parents' variety while another was found in the local accent. Thomas and Scobbie's (2015) analysis of a child of parents with two different accents observed something similar, but found that each vowel showed a different level of influence of each parent. Examples of similar compromises, called 'blends' here, were evident in Henry's speech throughout. The specific effects on each lexical set will be discussed in the following sections.

7.4 STRUT/FOOT

A number of researchers cited complexity as a key factor in the acquisition of local variables. The more complex the rule to be acquired, the less likely a child is to acquire it fully if they lack a consistent input model, i.e. because they have a parent with a non-local accent (e.g. Hewlett et al., 1999; Payne, 1980). This complexity can be hard to define, quantify, or compare between varieties. Complexity might be cited as where the rules of one variety do not directly correspond with the rules in another, but these may be compounded by layers of multiple phonological environments interacting, and also lexical exceptions. Let us consider the complexity of relationship between the two varieties in Henry's input, following Wells' (1982) system for describing accent differences.

The relationship between the STRUT and FOOT vowel oppositions in the home accent and realisation by single vowel in local varieties is systemic – that is, Henry needs to learn that there are two vowels in the home variety which correspond to only one in the local varieties. If we believe that Henry acquired the home variety first (discussed below, Table 50) and is then acquiring the local variety second, this should be a simple matter in that all STRUT vowels can be replaced with a FOOT vowel. However, Henry's acquisition of local STRUT variants has been slow and steady rather than rapid.

The STRUT/FOOT division in the home variety had an unexpected impact on Henry's acquisition of FOOT. Henry's early FOOT vowels were very front. My own realisations of FOOT are rather front, [ʊ̟], in comparison to a typical local vowel [ʊ], so Henry has been exposed to these at home. But his early articulations were *very front indeed*, impressionistically much more advanced than my own, with some being realised as [ʏ] and sharing the F2 of the close front vowel [ɪ] in the closest reference data (Busby & Plant, 1995). Henry's realisations of STRUT, on the other hand, generally featured a lower F2, indicating a more retracted tongue, [ʊ], as would be expected for the local variant. As mentioned above, FOOT and STRUT are not differentiated in the local accent, and yet Henry's FOOT vowels were of a totally different quality from most of his early local STRUT realisations. Some STRUT tokens did have a high F2, indicating tongue advancement, but these were home variants – they were not rounded and they had higher F1 values indicating an open jaw position; they were impressionistically more open and TRAP-like. Henry did not start to merge the FOOT and local STRUT variants until he was around 6 years old. Until then, he maintained the phonological contrast between FOOT and STRUT by keeping FOOT realisations front. Henry's realisations of these lexical sets have been typically around [ʏ ~ ʊ̟] in FOOT and [ʊ ~ ɒ ~ ʌ̹ ~ ə ~ ʌ] in STRUT. This differentiation persisted to some extent up to the end of the data collection period, though over time his FOOT and local STRUT realisations began to overlap. The front realisations were, however, still evident impressionistically and acoustically. It would appear that the split between FOOT and STRUT in the home varieties is having a lasting impression on Henry's articulations of this vowel. Even though an increasing number of back tokens of FOOT appeared in his dialect towards the end of the data collection period, the front realisations persisted, and even occasional local STRUT variants were produced in a front position, suggesting a continuing phonetic influence from the accents at home.

7.5 PALM/START/BATH

The PALM and START lexical sets are not differentiated from each other in either the home or local varieties, apart from in Henry's grandmother's speech. This pair is generally realised around [ɑ] at home and as [aː] in local accents. However, some complexity is added by BATH, which is realised in the same way as TRAP in local varieties, [a], but as PALM, [ɑ], in the home varieties. Wells classifies the difference in how BATH patterns as a lexical-distributional difference between the accents. Lexical items belonging to the BATH set are the result of what Wells describes as a 'half completed' sound change, as TRAP and BATH 'probably' split in the early 19th century (1982: 232-3). The resulting membership of the set is complex as the sound change halted before completing lexical diffusion. Consequently, Henry must figure out which words belong to the BATH set; this is difficult as at home they sound like PALM words and outside of the home they sound like TRAP words. So how has this contradictory input affected Henry's acquisition of these vowels? PALM/START appeared as both a home and blended vowel early on, while local variants appeared from 3;07. Compared to STRUT, local variants of PALM and START were prevalent much earlier and in greater numbers, but less consistently than STRUT, perhaps because of this complexity. BATH may be interfering with the acquisition of the local vowel /aː/ which would otherwise be straightforward to acquire. While BATH occurred infrequently, Henry's preference for the local vowel is clear as the home variant is so rare – only one home variant appeared in the whole corpus – the others were blends or local vowels. Local variants did not appear until relatively late compared to the acquisition of the other vowels being examined here (4;08). By this time the acquisition of the local vowel appears to be nearly complete, with very few non-local variants appearing by 6;06, and none at all at 6;11.

An unexpected artefact of this BATH/PALM/START relationship appears to be Henry's occasional hypercorrection of TRAP. Although the number of tokens affected is small, some TRAP tokens were realised with what appear to be PALM/START variants on 12 occasions over five recordings, with the earliest example being at 2;05 and the latest 5;01. These realisations have also been noted outside of recorded sessions. Dyer (2007) describes a similar situation with her son growing up in Michigan, USA with an English mother – her son pronounced 'laugh' (BATH) and 'can't' with *r*-colouring by analogy with START vowels. Compare this longer-term categorisation with the anomalous example of 'plus' from the STRUT dataset. Henry appeared to initially allocate this to the TRAP lexical set, but it quickly resolved itself into STRUT/FOOT. Moreover, there were no hypercorrections apparent in the FOOT lexical set; FOOT words such as 'book' were never pronounced with a STRUT vowel. Given the complexity of relationship between START, PALM and BATH, Henry's ongoing blends, his inconsistent pattern of acquisition, and his hypercorrections are not, on the face of it, surprising, given Payne's (1980) conclusions that more complex phonological rules are harder to acquire. However, on closer inspection, the most complex job appears to be well underway – that of using the appropriate local BATH vowel. The remaining task – the one that Henry appears to be struggling with more – seems relatively simple. The difference between the home and local variants of PALM and START is a phonetic one. All instances of /ɑ/ should be replaced with /aː/, and yet this is the vowel which shows the home influence in distinct phases over the whole of the data collection period, as the dominant variant switches between home influence and local influence and back again.

7.6 Variation

How does one differentiate between general unstructured variation and the influences of the multiple dialects in Henry's input? Local (1983) described huge variation in the FLEECE vowel of a single child and highlighted how this kind of variation is usually smoothed out in published research. Veatch's (1991) research shows us how enormously variable real speech can be. His plots of the pronunciations of a single vowel, /ɛ/ in an adult speaker demonstrated that their realisations of this one phoneme filled their entire vowel space. Henry also exhibits this kind of variation, for example, in his realisation of 'Jack Black' as ['ʤək 'blak]. This kind of free variation is at least partly responsible for his overlapping vowel realisations seen in the vowel plots in Chapter 5.

We have seen that Henry's vowels show variation across all of the lexical sets analysed. On careful inspection of the whole dataset, it would appear that there is some systematicity to his blended articulations. The blended articulations are neither home, nor local, but neither are they just random variation. There is systematicity in his articulations. Realisations of STRUT, for example, might take on the feature of rounding from the local variety, but degree of openness from the home vowel, though this happens to a varying extent over the course of the data collection period. While it is not possible to predict whether Henry will produce a home, local or blended variant of each of the lexical sets, their variation is systematic; they will broadly fall into one of the three categories – those which pattern with the home variants, those which share local features, or those which exhibit articulatory features of both (cf. Hazen, 2002). This will be laid out in detail below, in section 7.10.

Some of this variation is attributable to the target sound. A great deal more variation was evident in the F1 of open vowels (PALM/START/BATH variants) than of close vowels (FOOT variants). At age 6, variants of START, PALM, BATH and TRAP spanned a 1000 Hz range, while in comparison, FOOT variants were spread out over only 300 Hz. In Chapter 2 we saw that some researchers have found F1 (jaw height) to be less variable than F2 (Green et al., 2002; Nittrouer, 1993), yet while Henry's F2 is very variable, particularly in FOOT, there is a definite distinction here between F1 variation in close vowels compared to open vowels. This distinction was recognised in Koenig and Fuchs' (2019) laboratory experiments on adults, where they noticed that the F1 of open vowels were more affected by vocal effort than close vowels. The recordings of Henry take in a range of different speech styles, so this may explain the variation. The peak of widest variation differed according to lexical set and by formant number. In the STRUT and FOOT sets, F1 variation peaked at age 5, while F2 was consistently variable across all ages. In all three of the BATH, PALM and START lexical sets, F1 and F2 variation peaked at age 5. Even though Henry's speech has been captured in a naturalistic setting, this pattern corresponds with Yang and Fox's (2013) laboratory experiment looking at variation ranges in children. They observed that children's articulations varied more in 3–5-year-olds than in 5–7-year-olds. However, some of the recordings of Henry at age 6 included read speech, which is likely to have constrained variation in comparison to recordings which were completely naturalistic. We should also bear in mind Veatch (1991), whose naturalistic adult realisations of a single phoneme filled the whole vowel space. Further data would need to be collected to establish the direction of travel for Henry's variability.

In early recordings of Henry, there was a great deal of variability in his blended articulations. Some blends were so far from either the home or local varieties that they overlapped with other lexical sets, for example, some blended realisations of STRUT were realised with a LOT vowel, e.g. 'running' [ɹɒ̝nɪn] and 'monkey' ['mɒ̝ŋki]. There was also additional evidence of how Henry might be classifying the STRUT vowel in the examples of his writing where he represents this vowel with 'o'. This could be interpreted as evidence of some blurred boundaries between LOT and STRUT. As Henry has been exposed to 'one' being realised as part of the LOT, STRUT *and* FOOT lexical sets, could this have contributed to his understanding of how this vowel functions? His realisations of 'one' and 'once' appeared to be as variable as the rest of the STRUT lexical set.

7.7 Accommodation

At age 6;11, Henry was recorded playing with one of his closest friends from school, James, and his younger brother. The lexical sets STRUT, FOOT, PALM, START and BATH were analysed in an extension of the work done in Chapter 5 as well as two additional lexical sets, GOAT and FACE.

The concern was that the variety being spoken at home might be subject to change at school. Kobayashi (1981) and Dyer (2007) both observed this phenomenon in their research. In section 2.3.2 I described how my own brother recalled using two distinct accents, one for home and one for school, each serving the needs of the two audiences. However, the analysis of Henry's speech playing with a friend revealed no substantial difference across any of the features investigated. The only lexical set to show an increase in local realisations in conversation with James was BATH. BATH has been the least problematic lexical set for Henry. In the earliest recordings, he was already articulating BATH vowels as blends (though it was very late to appear at all due to its limited frequency in English), and only one home realisation was ever recorded. In conversation with James, all BATH tokens were local, but this only represented a small increase since his recordings at age 6. In all other lexical sets, Henry showed an increase in the proportion of home realisations in this recording, demonstrating that the influence of home varieties was not lessened when speaking to interlocutors with a local accent. Glottal replacement, h-dropping, and -*ing* realisations as [ɪn] were also investigated as possible candidates for style shifting in conversation with his friend, but the analysis did not reveal any difference between these features in conversation with his friend in comparison to speaking to his mother in the data analysed in Chapter 5.

As Henry's mother, I have witnessed him in conversation with other children on many occasions, and therefore it is not surprising to me that his vowels did not significantly alter in conversation with his friend. The influence of the home accents is below the level of Henry's consciousness; he is not aware that his vowels are any different to those of his friends, and therefore they do not change according to interlocutor or social situation. Although BATH showed an increase in local variants in comparison to the previous recordings at age 6, it had been five months since the last recording with me. The vowel continued on an existing trajectory towards complete acquisition of the local realisations of this lexical set, showing only a modest increase in local realisations. This is most likely attributable to time rather than James' influence. Moreover, rather than an increase in local realisations of STRUT and FOOT, Henry produced more home variants. In START and PALM, home-influenced realisations also increased in this recording, continuing the established overarching pattern of alternating dominance of the home and local influences.

Although I had not expected to see evidence of accommodation in Henry's vowels, I find it surprising that he showed no increase in his levels of glottal replacement, h-dropping or non-standard *-ing* realisations. Even though James used all of these features to a lesser extent than Henry, the use of non-standard variants would have been a way for Henry to signal his social credentials in conversation with his friend. His rapid and excited speech certainly pointed to *maximum* informality, so it was surprising to find that his speech did not exhibit any more non-standard features in conversation with his friend than it did in conversation with me.

7.8 Phonological acquisition

If, as Smith (1973) claims, a child's phonological system is mapped from an adult's, how does this work when the child has multiple competing inputs, as in Henry's case? As discussed in Chapter 2, children can readily update their phonological system in the face of new data (Smith, 1973; White and Aslin, 2011), which Smith argues is evidence for his claim that the adult phonological system is the same as the child's underlying system.

This position has been subject to some criticism, e.g. by Macken (1980), who claims that children may mis-hear a sound in adult speech, and therefore build a phonological system which is different from the adult's. Henry did this at a lexical level in two cases discussed in Chapter 5, where 'plus' was briefly assigned to the TRAP lexical set, and the similar, longer-term assignment of 'scarf' to BATH. If we overlook these cases, as they are anomalous individual lexical items, can Henry's phonological system be mapped to an adult system? In the case of the STRUT/FOOT split (discussed in section 7.4 above) in Henry's phonological system, the division of these lexical sets (as in home varieties) was in evidence in his early recordings, though many of Henry's FOOT realisations were much more front than those heard at home. As time went on however, an increasing number of back tokens overlapping with STRUT appeared. By the end of the recording period, Henry was still articulating some tokens of FOOT with a very front position, but most were either central or back, as are typically seen in the local varieties. The continued presence of these front tokens indicates that Henry was still being influenced by the phonological systems at home. The phonetic realisations were unique to Henry, as they overlap with both home and local varieties, but the increasing number of back realisations suggest that he was shifting towards the local system. Although the overall pattern is one of a trajectory towards a STRUT/FOOT merger, this process was not yet complete at the end of the data collection period. It is impossible to say whether Henry will eventually have a fully local system, or whether there will always be anomalous pronunciations indicative of an underlying split, but we *can* say that the system is unique to him as it is in flux; it is not one that can be directly mapped on to an adult speaker of either the home or local varieties. While it may be possible that the systems of some children can be mapped directly on to an adult system, it seems likely that many will feature anomalies due to the variability of input that most speakers are exposed to. This suggests that a more plausible explanation which could be applied to all children would be that they construct their own underlying system with information taken from multiple sources. In some cases, this may be identical to their adult caregiver, but more likely this will be constructed from input from caregivers at home as well as (for example) nursery care workers, schoolteachers, classmates and siblings. The child will continuously restructure their system as they are presented with new information. In some cases, this will result in alterations to the underlying system, while in others, adjustments may be purely phonetic. Of course, not all interactions or exposures will result in changes to the underlying system. In Henry's case, there were multiple influences, and while exposed to local varieties, he was still exposed to home variants alongside these, creating a tension between at least two models. In section 2.2 I discussed Smith's (1973) claim that children might take a few days to roll out a new piece of phonological learning to analogous words, but in Henry's case, his exposure to multiple phonological systems appears to extend this process to years, and looks likely that parts of his system may remain permanently unstable as a consequence.

We have seen how Henry is well on the path to a complete acquisition of the local BATH variant, representing a split from the BATH/PALM/START group which are all articulated in the same way in the varieties spoken at home. PALM and START, however, are still showing clear influence of the home varieties, although this appears to come in alternating phases. Because BATH has split off from this group, and it appears to have done so successfully, the adjustment needed to achieve local PALM and START is purely a phonetic shift from the low back vowel [ɑ] to a front [aː]. Henry has successfully acquired the phonological shift required to align with the local varieties, but the apparently simple shift to consistent phonetic realisations appears to be lagging behind. In FOOT and STRUT, which requires Henry to shift all STRUT words to align with the FOOT vowel, STRUT has made this transition slowly and steadily, but the underlying contrast between /ʊ/ and /ʌ/ at home has left its trace in Henry's phonetic realisations of this set, as he uses a great deal of variation in tongue advancement and retraction in his realisations of both FOOT and STRUT in a way not seen in either the home or local varieties.

Henry's acquisition of stable phonological representations was still incomplete at the end of the data collection period, where a dwindling number of home and blended realisations are still present across all lexical sets. This continuing instability was demonstrated by the range of articulations of a single word in multiple ways in a single recorded session, sometimes within the same turn, (Table 47 - Table 49).

		STRUT		
		Home	**Blend**	**Local**
4;08	love	[ˈlʌv]	[ˈlɒv]	
4;09	suckers	[ˈsʌkəz]	[ˈsʌ̹kəz]	

Table 47: Multiple realisations of the same STRUT word in the same recorded session

		FOOT		
		Front	**Central**	**Back**
2;01	Woody	[ˈwʏdi]	[ˈwədi]	
3;10	put	[ˈpʏt]	[ˈpʊt]	
4;02	look	[ˈlʏk]		[ˈlʊk]
4;08	sugar	[ˈʃʏgə]		ˈ[ʃʊgə]
4;09	look	[ˈlʏk]	[ˈlʊ̟k]	[ˈlʊk]
4;11	book	[ˈbʏk]	[ˈbʊ̟k]	
5;01	push	[ˈpʏʃ]	[ˈpʊ̟ʃ]	
5;05	look	[ˈlʏk]	[ˈlʊ̟k]	

Table 48: Multiple realisations of the same FOOT word in the same recorded session

		PALM/START		
		Home	**Blend**	**Local**
3;09	can't		[ˈkɑ̟nt]	[ˈkaːnt]
3;11	are	[ˈɑ]	[ˈɐ]	
4;00	can't		[ˈkɐːnt]	[ˈkaːnt]
4;02	can't	[ˈkɑnt]	[ˈkɐːnt]	
4;04	cars	[ˈkɑz]	[ˈkɐˑz]	
4;09	are		[ˈɐˑ]	[ˈaː]
4;11	car		[ˈkɐˑ]	[ˈkaː]
5;08	danse	[ˈdɑns]		[ˈdans]
6;00	dark	[ˈdɑk]	[ˈdɐːk]	[ˈdaːk]
6;06	can't	[ˈkɑnt]		[ˈkaːnt]

Table 49: Multiple realisations of the same PALM/START word in the same recorded session

These examples also clearly evidence that Henry's realisations are generally not tied to particular words. All members of the lexical sets examined in Chapter 5 appear to be eligible for blending. It is clear therefore that Henry's articulations are phonologically rather than lexically driven.

7.9 GOAT and FACE

It would appear that there is no straightforward answer to why Henry has adopted some elements of the local varieties and not others, notably the local GOAT and FACE monophthongs. Possible factors affecting acquisition may be salience (whether Henry implicitly notices a vowel), frequency, status of the interlocutor (whether they are a parent or friend), complexity of the distributional relationship between the local and home varieties and ubiquity (how many people in Henry's environment use the variant).

Local realisations of the BATH lexical set were most easily acquired, and yet this is a rather infrequently occurring lexical set. Is it more salient than GOAT or FACE? Haddican and colleagues (2013: 373) describe monophthongal FACE and GOAT as 'a principal shibboleth of northern English speech', indicating its social salience, but do children of Henry's age notice it, even implicitly? In an experiment with pre-school children, Jeffries (2016) found that the GOAT vowel was the most reliable indicator enabling children to identify a familiar nursery worker with a monophthongal realisation against a range of unfamiliar speakers with different accents. This suggests that children have at least a tacit awareness of these vowels. However, her experiment tested this vowel against LOT, MOUTH, GOOSE and FOOT rather than STRUT, PALM, START and BATH, so this does not answer the question of whether the GOAT vowel is more salient than the other lexical sets investigated here.

According to Cruttenden (2014), the frequency of /eɪ/ (FACE) makes up 1.57% of all vowels and /əʊ/ (GOAT) accounts for 1.55% of vowels. This is more than double the frequency of /ɑ/, which makes up only 0.68% of all vowels. The figure for BATH would, of course, be further reduced as 0.68% includes PALM, START *and* BATH lexical sets. As discussed above, in spite of its relative infrequency, BATH was the first lexical set where Henry fully acquired local realisations, so high frequency does not seem to be a major reason for local variants to be acquired. Could it be that low frequency is the important factor in securing the early success of this lexical set? Perhaps, but I am not convinced that its success can be attributed to low frequency alone. Of all of the lexical sets, FOOT has proven to be maximally variable in its tongue position, and yet, the frequency of FOOT is also low at 0.62% of all vowels (Cruttenden, 2014). While this is higher than BATH, the FOOT set is mainly made up of a very limited set of frequently occurring words, for example, 'look', 'put', and 'could', which would suggest that he has the opportunity to practice a small set of words, and yet FOOT has ended up with a very large range of phonetic variation in Henry's speech.

I established in Chapter 6 that Henry does not alter his vowels (or other features commonly subject to style-shifting) when talking to different audiences, in this case, his mother and his close friend. However, as discussed above in section 2.3, it is widely thought that a child's peers are the dominant influence on their accent. In Henry's case, this is clear – he has adopted several key local features, though this has been with varying rates of success at the end of the data collection period. The literature highlights complexity of the features in determining a child's successful acquisition of a feature (e.g. Payne, 1980; Roberts, 1997; Trudgill, 1986), but of all the features investigated in this thesis, GOAT and FACE would likely be the simplest to acquire. The difference between the local pronunciation of these vowels and those spoken at home are simply realisational; a phonetic difference rather than a phonological one (Wells, 1982). All GOAT and FACE vowels would be realised as a monophthong instead of a diphthong, with no exceptions, lexical or otherwise. Complexity is clearly not a factor in determining which features Henry has acquired; it has played a role in the completeness of the acquisition process, but not in predicting *which* features will be adopted from the local varieties. In Chapter 2 we learned that prestige features may be transferred from parents to children in some cultural settings (Kazazis, 1970; Stanford, 2008; Surek-Clark, 2000), though this has not happened across the board for Henry. In Chapter 4, I explained that some northern speakers may find the pronunciation of STRUT vowels as [ʊ] 'vulgar' (Wells, 1982), and yet Henry has acquired this variant readily. It seems unlikely that the prestige of family variants would only be responsible for affecting GOAT and FACE. The inconsistency of acquisition of prestige variants suggests that covert prestige (Trudgill, 1972) is not responsible on its own.

Perhaps the most compelling factor is a sound's ubiquity among different speakers in Henry's environment. Although Haddican and colleagues (2013) highlight the social salience of monophthongal FACE and GOAT among teenagers, they point to its reduction in the city of York, the nearest city to Henry's home, in favour of diphthongal realisations. The monophthongal realisations are traditionally found in working class speakers, and while several of Henry's closest friends use these vowels, they are not universal in his cohort at school, or in the teaching staff. BATH vowels, on the other hand, are likely to be realised as the short /a/ by most northern speakers (Wells, 1982: 358). This ties in with Floccia and colleagues' (2012) claim that infants may only store the community representation of a sound, but here it is the dominant variant rather than the community variant that surfaces, as there is not a single consistent variant found in all local dialect speakers.

7.10 Implications of this research

What are the implications of the blends seen in Henry's developing phonological system? There is evidence in the literature of children exposed to multiple inputs developing their own idiosyncratic systems, though previous research has been limited to looking at a single point in time, with the exception of Kerswill and Williams (2000), who looked at two points in time 18 months apart. The evidence presented here provides a longer term, more gradual view; how phonetic realisations can change over time, how this ties in with the complexity of the relationships between the phonological systems of the home and local varieties and the lasting influence that the phonetics in the home accents may have, even once a new phonological system is acquired. Henry was 3;07 before local variants began to establish themselves with any regularity, but blended variants appeared much earlier, from 2;05. This data illustrates the way in which Henry's accent changes; not as a sudden shift from one variety to another, or even a slow and steady process affecting each variant equally, but rather, slowly, with articulatory features of local vowels appearing increasingly over time. In the case of STRUT/FOOT, this manifested itself in changes in jaw openness or rounding, and for PALM/START and BATH, changes in tongue advancement or vowel length. However, it is notable that blends and home variants remain across the PALM, START, FOOT and STRUT lexical sets even if this is at a low level. This could be evidence that the home phonology does not disappear completely but is retained at some level, if only latently (cf. Khattab, 2002).

Can Henry's blended articulations tell us about the way that children acquire their phonological systems more broadly? Might children orient to phonetic features in their input rather than phones? Children's first language acquisition often demonstrates *features* of a target phone before the full phone is acquired, as has been recognised in the acquisition of consonants. For example, in developing speech, children often achieve target manner or place of articulation, but not both, as in the case of realising fricatives as plosives, for example (stopping) (see for example, Johnson & Reimers, 2010). Studdert-Kennedy and Goodell (1992: 97) wrote of features "sliding along the timeline" in children's early speech, as in gestural harmony, for example. This tells us that in children's beginning articulations, features are not so tightly bound to a phoneme as in adult speech, where they are more constrained. Fowler and colleagues (1991) found that children were more likely than adults to switch phonological features than phonemes in speech errors, and that this tendency decreased with age. Browman and Goldstein (1992) proposed that after initial words, a child's system functions according to articulatory gestures before higher-level phonological units such as phonemes. This seems to indicate an orientation to sub-segmental features in young children which disappears by adulthood. This ties in with Henry's blended vowel realisations which incorporate sub-segmental aspects of both the home and local vowels, and other examples of phonetic compromises seen in the literature, discussed in section 2.3.1.2.

In their discussion of the evidence for the emergence of phonological features, Menn and Vihman (2011) claim that rather than being innate, they become part of a child's mental grammar as the child discovers them, and that these features become more fully realised over time, after the emergence of first words. In the data presented here, Henry's first words are already well established. There is evidence that he always realises phones in the variety spoken at home, or as a blend, before the local variant is realised (Table 50). When the shift towards the local variant begins, *in the first instance* it appears to happen at the level of feature rather than phoneme.

	First variant	First home variant	First blended variant	First local variant
STRUT	2;01	2;01	2;06	2;08
FOOT	2;01	2;01	2;01	3;10
START	2;08	2;08	2;08	3;07
PALM	2;10	4;02	2;10	3;07
BATH	3;09	5;08	3;09	4;08

Table 50: Appearance of variants by age

Ladefoged's claim that we orient to articulatory parameters rather than phonological features offers a plausible explanation for Henry's linguistic behaviour. He extracts articulatory features from the phonemes in each phonological system and combines them. These new combinations result in phonological contrasts in his own system, so, as Menn and Vihman (2011) explain, if articulatory parameters behave systematically, then we can usefully conclude that the child is functioning as if they have a phonological feature. For example, in this case, Henry orients to the tongue advancement of BATH, adopting this articulatory parameter for use in his own system, creating a phonological opposition with PALM/START, which exists in the local but not the home system. In realisations of STRUT, he also orients to and adopts (albeit inconsistently) the jaw position and rounding of FOOT, which are distinct from the articulatory parameters in the home system. This articulatory adjustment should merge two phonological oppositions, /ʊ/ and /ʌ/ into one. However, unexpectedly, while he adjusts to the close jaw position of STRUT in the local variety, this does not merge with the FOOT category, as he produces realisations of FOOT so far forward, overshooting the position of home FOOT. Therefore, instead of collapsing two categories, he maintains the phonological contrast but with /ʏ/ and /ʊ/ rather than /ʊ/ and /ʌ/ (Figure 99).

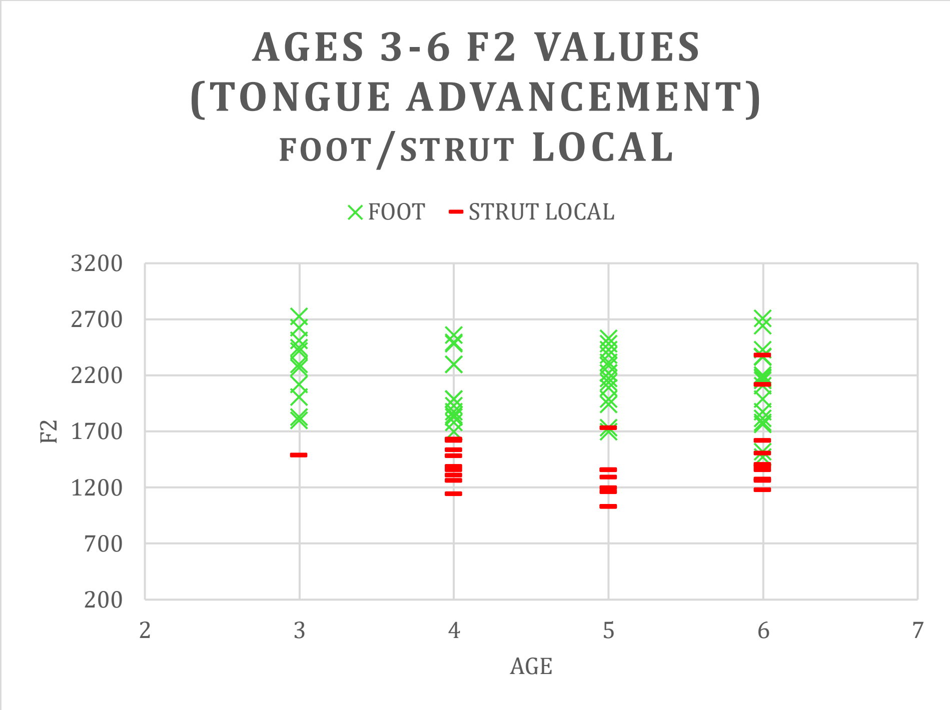


Figure 99: F2 values STRUT/FOOT *local (copy of Figure 50)*

Over time, this phonological contrast begins to neutralise as he produces tokens from each of the categories which overlap, but the phonetic legacy of the home dialect remains as he retains a much wider range of tongue movement than is usual for either dialect.

Henry has shown evidence of adjusting the articulatory parameters shown in Table 51.

Articulatory Parameter	**Phonemic adjustment**	**Lexical Set**
Back → Front	ɑː → a	BATH
Long → Short	ɑː → a	BATH
Back → Front	ɑː → aː	PALM/START
Back → Front	ʊ → ʏ	FOOT
Open → Close	ʌ → ʊ	STRUT
Unrounded → Rounded	ʌ → ʊ	STRUT

Table 51: Articulatory features adjusted by Henry

The existing feature sets/articulatory gestures discussed in Chapter 2 (e.g. SPE, Ladefoged's articulatory parameters) are generally well specified for consonants but less so for the description of vowels. Donegan's (2013) work on vowels presented in section 2.2.4.1 is helpful. She acknowledges variation in both child and adult speech, and accounts for this in her proposal that features are both binary and gradient. These proposals accept that a feature may be present to a greater or lesser extent depending on a range of factors. In the case of the vowels analysed here, those factors are not just related to phonological environment and free variation but are also connected to Henry's variable realisations of the articulatory features he hears in the voices around him. Donegan (2013) describes her feature set for vowels as "basic", thus I find that it is not the perfect set for circumscribing the features that Henry adjusts in his own speech. She proposes three vowel heights, palatal and non-palatal, labial and non-labial, and tense and lax as sufficient to describe her vowel system. Donegan's distinction between lax and tense does not account for Henry borrowing the length of the home BATH vowel in blended variants of this vowel, for example, [ˈɡʊɐːs], as for Donegan, the only distinction between /a/ and /ɑː/ is one of palatality; they are both classed as lax. I therefore propose the features long/short in their place (Table 52).

Donegan's (2013) Features (revised)	**Phonemic adjustment**	**Lexical Set**
Non-palatal → Palatal	ɑː → a	BATH
Long → Short	ɑː → a	BATH
Non-palatal → Palatal	ɑː → aː	PALM/START
Non-palatal → Palatal	ʊ → ʏ	FOOT
Mid → High	ʌ → ʊ	STRUT
Non-labial → Labial	ʌ → ʊ	STRUT

Table 52: Articulatory features (revised) adjusted by Henry according to Donegan (2013)

A featural analysis offers a compelling explanation for the ongoing development of Henry's vowels. I do not claim that he orients to abstract features (cf. Chomsky and Halle, 1968; Jakobson et al., 1963), but I do propose that I have provided strong evidence that he tunes into articulatory features or gestures of the speakers in his environment rather than larger linguistic segments such as phonemes. He then adjusts his own articulatory parameters accordingly (cf. Browman and Goldstein, 1986; Ladefoged, 2005). In a child living with less obvious accent variation in their environment, this might go unnoticed (Hazen, 2002), but here, due to the differences between the home and local phonological systems, Henry's ongoing and variable acquisition of articulatory features from both accents within individual phonemes is clearly evident.

7.11 Summary

In this chapter I have discussed the changes to Henry's realisations of the lexical sets PALM, START, BATH, FOOT and STRUT over time, and described the impact of his exposure to multiple dialects on his developing phonological system. Overall, he has acquired these sets at varying rates and to varying levels of completeness. Henry's vowels are not subject to accommodation in conversation with a friend with a local accent. The recordings made at home with me appear to be wholly representative of his speech in general. During the transition from the home dialect to the local dialect, Henry has demonstrated that he combines articulatory parameters or features from both varieties within each of the phonemes under scrutiny here, which function as phonological features. Though the process is not yet complete at the end of the data collection period, it appears that the varieties spoken at home have left a lasting phonetic impression upon Henry's accent, even when the phonological shift is almost complete.

Chapter 8 Conclusion

8.1 Overview

This thesis has analysed data from a child living in multidialectal environment between the ages of 2;01 and 6;11. The analysis has focussed on the lexical sets STRUT, FOOT, PALM, START, and BATH, which have all shown some evidence of influence from both the home and local accents. The data was analysed impressionistically in the first instance, with a subset of the data being subject to acoustic analysis in order to support the impressionistic analysis.

8.2 Overall findings

The following research questions were stated at the beginning of this thesis:

(RQ1) Does the child eventually acquire all of the vowels of their peers or does any parental influence remain?

(RQ2) Are vowels acquired at word level, phonemic level, phonetic level or is there evidence of acquisition at a more abstract level, for example, distinctive features?

(RQ3) Is each vowel acquired in the same way and at the same time/rate?

(RQ4) Is there evidence of the vowels being subject to accommodation?

(RQ5) Is the acquisition process complete by the age of 6;06?

Henry has made progress in acquiring most of the vowels of his peers. There is some variation among his school cohort, however, representing variation in the local model. Henry's closest friends realise GOAT and FACE as monophthongs, but he shows no signs of acquiring these local vowels. The main areas of difference between the local and home varieties are in the lexical sets STRUT, FOOT, PALM, START and BATH. Henry's realisations of the vowels in these lexical sets all showed evidence of local influence, though this was at the level of articulatory feature in the first instance (RQ2). Realisations made up of articulatory features drawn from both the home and local dialects were called 'blended' articulations. In all lexical sets, Henry articulated a mixture of home variants, local variants and blended articulations. There was no further shift towards local variants in conversation with a friend; the vowels under examination here were free from signs of accommodation (RQ4). The path from home to local vowel varied across the lexical sets, with BATH being closest to completion by the end of the data collection period. The path to acquisition of the local STRUT vowel has been slow and steady, while the acquisition of PALM and START has been the least successful as the influence shifts in an alternating pattern from home influence to local influence and back again (RQ3).

Even after a phonological shift has mostly been achieved, phonetic traces of the multidialectal input remain (RQ1; RQ5). In the case of STRUT and FOOT, Henry has been left with much more variability in tongue advancement than would be expected for either the home or local variants. Although local realisations of BATH dominated Henry's articulations of this lexical set quickly, what should have been a simple phonetic adjustment to PALM and START has proved problematic.

Multiple factors influencing the success of acquisition of a local dialect have been suggested in the literature, but none of these alone can explain the nuances of Henry's phonological acquisition. The most complex phonological relationship was between the home and local distributions of BATH, PALM and START, and yet the local BATH vowel was the most readily adopted. Collapsing STRUT and FOOT lexical sets from two phonemes to one should be relatively simple as there are no lexical influences, yet this merged set continues to feature more home realisations than the lexically populated BATH lexical set. An unexpected effect of the STRUT/FOOT merger was Henry's fronting of the FOOT vowel, which led to a continued phonological contrast between STRUT and FOOT until age 6. This was in spite of the adoption of local articulatory features for STRUT; FOOT articulations were mostly articulated towards [ʏ] while local STRUT realisations were around [ʊ].

High frequency appears not to be a factor in improving speed of acquisition, as BATH is the least frequently occurring lexical set, and yet is the closest to complete acquisition. A variant's ubiquity in the local speech community, as is the case for the local realisation of BATH, may increase its salience and therefore could be responsible for its speed and completeness of acquisition.

Though variable, Henry's realisation of vowels between the ages of 2-6 years suggest that the shift happens at the level of his underlying phonological system. As Henry adjusts his articulatory features to match those in the local dialect, he creates new phonological contrasts, though these are unstable.

8.3 Contribution

While the impact of multiple dialects on a child's phonological acquisition has received some attention in the literature, this has been from an almost exclusively synchronic perspective. Acoustic analysis of naturalistic data in children is also rare. Here it serves to support impressionistic transcriptions, but it also reveals the unexpected distribution of F2 values between FOOT and local realisations of STRUT. This long-term approach has indicated that there can be residual consequences of phonological shifts for a child's phonetic realisations, even when the phonological shift is near completion.

8.4 Limitations

This case study has followed in the tradition of linguists researching the language acquisition of their own children. While there are limitations to case study research, in this case, these are offset by its advantages. The privileged position of a parent enables in-depth and long-term access to a child's developing system in a way that would be impractical for any other researcher. Extending a study of this kind to include more children would of course allow more robust conclusions to be drawn, but it is possible that compromises would have to be made; parents might not all speak the same varieties, so the lexical sets may have to vary between families, age of starting nursery may vary, and each family's unique networks could involve exposure to differing varieties in their environment.

Henry knew he was being recorded. In early recordings this is very unlikely to have had any impact, as he was so young. Even in later recordings, the effects were of showing off and being silly when the recording device was switched on, but this was rare, and where it did occur, the effect quickly wore off once recording was underway. Perhaps the conversation with James was likely to be the most affected by the children's knowledge of being recorded. Henry's high excitement levels were evident in his speech in that it was fast, there was some high-spirited shouting, and their play was very imaginative, which suggests a lack of self-consciousness. Speech collected in the school playground through a concealed mic would have yielded speech even more uninhibited, however. This would of course have brought further challenges of overlapping speech and noise as more children would have been present. As it is, the naturalistic data collection techniques of recording typical activities at home came at a price to the acoustic analysis. Many tokens were overlapping with other talk or background noise such as the television or washing up. Later recordings often drew on Henry reading from his school reader; this means that a reading style was represented in later recordings, but these were not separated out in any way for individual consideration. Where options were available, non-read speech was always selected for acoustic analysis, but in some cases only read speech was available. The impact of reading style tends to make speech more formal, with more standard features being selected, however we see no movement towards the standard in these later recordings; there is no shift towards home variants in Henry's reading style, though a resulting reduction in variability in the BATH vowel was evident at age 6. The naturalistic data collection method led to lower representation of some infrequently occurring lexical sets (e.g. BATH and PALM), which further impacted upon the number of tokens suitable for acoustic analysis. This led to these lexical sets having few tokens available for analysis at some ages.

Reference data for the acoustic analysis of Yorkshire vowels in children of an appropriate age was non-existent, and even British dialect data across a range of vowels was extremely limited, therefore Australian English proved to be the closest available reference for the vowels under examination here. There is also a lack of acoustic analysis of naturalistic data in children represented in the literature. We saw in Chapter 7 that Henry's open vowels varied much more than his close vowels, yet the vast majority of acoustic analysis of children's vowels is of careful, laboratory speech. It was therefore necessary to extrapolate from adult naturalistic data rather than having concrete evidence of articulatory behaviour in children, specifically, the increased use of the vowels space in naturalistic data seen in adults (Veatch, 1991). The development of naturalistic reference data is important to establish credible benchmarks for children's speech.

8.5 Recommendations for further research

The most obvious next step would be to continue to observe Henry's speech into adulthood. Analysis of his speech as a teenager could reveal whether the acquisition period is ever complete, and whether the phonetic legacy of his phonological shifts has become a permanent feature. At the end of the data collection period, Henry's phonological system was still unstable, as revealed by his continuing use of home or blended variants. It seems likely that BATH will eventually be completely acquired, but unclear whether or not STRUT will eventually stabilise. So long as STRUT continues to be realised by variants that FOOT does not, this will indicate continued, perhaps permanent, instability of this lexical set. Whether the phonetic legacy of the phonological shifts left for FOOT and the PALM/START group are a permanent feature would also be of interest.

Larger scale research building on this study would of course be desirable but would take considerable resources and the identification of more families, ideally with similar realisations of the lexical sets examined here at home. There are four important questions to be answered:

1) How variable is children's naturalistic speech in comparison to lab speech and how does this vary across individual children?
2) Do all children in multidialectal environments make use of blended realisations, and how do these vary across lexical sets?
3) Do all children who make phonological shifts exhibit a phonetic legacy of their home variety and if so, is it permanent?
4) Are the phonetic realisations of children from multidialectal environments more variable than children from broadly monodialectal environments?

There is a gap in the literature concerning the acoustic reality of children's phonetic realisations *per se*; Local (1978; 1983) wrote of the enormous variability in children's naturalistic speech, yet 40 years on this has still not been documented acoustically.

Though this study does not present evidence of the shift from whole words to the emergence of segments, the data presented here provides preliminary evidence of the primacy of sub-segmental features (or gestures) over the phoneme, (cf. Studdert-

Kennedy, 1987). As data from children raised in multidialectal communities 'un-masks' the sources of their phonological input (Hazen, 2002), here, data from Henry's continuing vowel acquisition uncovers the significance of articulatory features in a child's journey to phonemic representations. Further longitudinal investigations of children living in multidialectal environments will surely strengthen this claim.

Appendices

Appendix 1: Audio Files

Sample audio files have been submitted, including one example each of home, blended and local variants for every lexical set. The files have been submitted inside a zipped folder with the following structure. The file name is composed as follows:

Age_classification _lexical set_word produced_Phon record number

6_LOCAL_STRUT_BUMPER_190

STRUT

4_HOME_STRUT_DRUM_89
4_BLEND_STRUT_MONKEY_178
6_LOCAL_STRUT_BUMPER_190
4_HOME_STRUT_PLUS_46

FOOT

3_FRONT_6_FOOT_PUTTING_40
6_CENTRAL_2_FOOT_GOOD_37
6_BACK_4_FOOT_LOOK_13

PALM/START

5_HOME_PALM_CANT_27
4_BLEND_1_PALM_CANT_41
3_LOCAL_START_CAR_2

BATH

5_HOME_BATH_DANCE_23
5_BLEND_BATH_MINECRAFT_87
6_LOCAL_BATH_BATH_174

TRAP

4_BLEND_TRAP_BAD_2
4_BLEND_TRAP_BADDIES_81

Appendix 2: Inter-rater reliability

Cohen's Kappa Calculations

STRUT

		Listener 2			
		Home	**Blend**	**Local**	**Total**
Listener 1	**Home**	21	3	11	35
	Blend	9	3	3	15
	Local	4	0	12	16
	Total	34	6	26	66

Table A: STRUT judgements for Kappa calculations

Number of observed agreements: 36 (54.55% of the observations)
Number of agreements expected by chance: 25.7 (38.93% of the observations)
Kappa= 0.256

The strength of agreement is considered to be 'fair' according to Landis and Koch's (1977) strength of agreement scale.

Weighted Kappa= 0.278
Assessed this way, the strength of agreement is considered to be 'fair' according to Landis and Koch's (1977) strength of agreement scale.

FOOT

		Listener 2			
		Front	**Mid**	**Back**	**Total**
Listener 1	**Front**	4	0	2	6
	Mid	1	2	3	6
	Back	1	0	5	6
	Total	6	2	10	18

Table B: FOOT judgements for Kappa calculations

Number of observed agreements: 11 (61.11% of the observations)
Number of agreements expected by chance: 6.0 (33.33% of the observations)
Kappa= 0.417

The strength of agreement is considered to be 'moderate' according to Landis and Koch's (1977) strength of agreement scale.

Weighted Kappa= 0.423

Assessed this way, the strength of agreement is considered to be 'moderate' according to Landis and Koch's (1977) strength of agreement scale.

PALM/START/BATH

		Listener 2			
		Home	**Blend**	**Local**	**Total**
Listener 1	**Home**	13	1	0	14
	Blend	3	4	1	8
	Local	0	1	3	4
	Total	16	6	4	26

Table C: *PALM/START/BATH judgements for Kappa calculations*

Number of observed agreements: 20 (76.92% of the observations)

Number of agreements expected by chance: 11.1 (42.60% of the observations)

Kappa= 0.598

The strength of agreement is considered to be 'moderate' according to Landis and Koch's (1977) strength of agreement scale, but is very close to being 'substantial'.

Weighted Kappa= 0.693

Assessed this way, the strength of agreement is considered to be 'substantial' according to Landis and Koch's (1977) strength of agreement scale.

TRAP

		Listener 2		
		Short	**Long**	**Total**
Listener 1	**Short**	93	0	93
	Long	0	3	3
	Total	93	3	96

Table D: TRAP judgements for Kappa calculations

Number of observed agreements: 96 (100.00% of the observations)

Number of agreements expected by chance: 90.2 (93.95% of the observations)
The strength of agreement is almost perfect according to Landis and Koch's (1977) strength of agreement scale.

Second inter-rater tests

Tokens selected for analysis	77
Tokens analysed	53
Initial agreement	35
Disagreements	18
Resolved	14
Outstanding disagreements	4

Table E: STRUT

	rounded	**blend**	**Unrounded**
Agreed	12	3	21
Initial disagreement	18		
Resolution	1	8	5
No resolution	4		

Table F: STRUT

1st listener shifts	8
2nd listener shifts	5
Compromise	1
No resolution	4
Total	18

Table G: STRUT listener shifts

	Rounded (local) to blend	**UR (home) to blend**	**Blend to UR**	**Rounded (local) to UR(home)**	**Blend to rounded (local)**	**UR(home) to rounded (local)**	**Total**
Listener 1	2	1	3	2	1		9
Listener 2	3	3					6
Total	5	4	3	2	1	0	
Examples	bubble, rubbing	up	up	some	funny		

Table H: STRUT Listener shifts detail

UR= unrounded

Tokens selected for analysis	23
Tokens analysed	19
Initial agreement	11
Disagreements	7
Resolved	6
Outstanding disagreements	1

Table I: FOOT

	front	**mid**	**back**
Agreed	4	2	5
Initial disagreement	7		
Resolution	1	5	0
No resolution	1		

Table J: FOOT

1st listener shifts	3
2nd listener shifts	0
Compromise	3
No resolution	1
Total	7

Table K: FOOT listener shifts

	Back to mid	**Front to mid**	**Mid to front**	**Mid to back**	**Front to back**	**Back to front**	**Total**
Listener 1	5	0	1	0	0	0	6
Listener 2	0	3	0	0	0	0	3
Total	5	3	1	0	0	0	
Examples	foot, could	look, full	look				

Table L: FOOT listener shifts detail

Tokens selected for analysis	38
Tokens analysed	30
Initial agreement	20
Disagreements	6
Resolved	6
Outstanding disagreements	0

Table M: PALM/START/BATH

	front	**mid**	**back**
Agreed	13	3	4
Initial disagreement	6		
Resolution	2	4	
No resolution	0		

Table N: PALM/START/BATH

1st listener shifts	3
2nd listener shifts	3
Compromise	0
No resolution	0
Total	6

Table O: PALM/START/BATH listener shifts

	Back to mid	**Front to mid**	**Mid to front**	**Mid to back**	**Front to back**	**Back to front**	**Total**
Listener 1	1	1	1	0	0	0	3
Listener 2	1	1	1	0	0	0	3
Total	2	2	2	0	0	0	0
Examples	are, can't	are, hard	castle, asked	N/A	N/A	N/A	

Table P: PALM/START/BATH listener shifts detail

Appendix 3
Tokens selected for acoustic analysis

Example number	**Age**			
Home	3	4	5	6
1	Up_5	Mummy_37	But_101	Mum_229
2	Mummy_93	Love_54	Mum_121	Mummy_66
3	Just_19	Other_12	Yummy_1	Does_170
4	Tummy_24		Doesnt_12	
5	Coming_63	Month_115	Sunk_26	
6	Does_77	Up_40	Sunk_28	
7	Lovely_204	Drum_89	Sunk_33	
8	Nothing_38	Trump_103	Love_31	
9	Covered_31	Cup_22	Mummy_61	
10	Us_104			
Blend	3	4	5	6
1	Up_2	Other_10	Coming_118	Love_2
2	Sun_4	Button_29	Up_139	Other_9
3	Muffet_6	Sunny_24	Brothers_3	Mummy_214
4	Tuffet_6	Just_19	Instructions_45	Up_69
5	Coming_63	Up_19	Thunder_53	Up_71
6	Does_106	Plum_67	Funny_43	Us_164
7	Just_140	Coming_154	Does_67	
8	Come_163	Monkey_178	Rubbing_96	
9	Sun_189	Much_14	Just_3	
10	Another_37	Lumpies_16	Mummy_52	
Local	3	4	5	6
1	Mummy_41			Front_15
2	Done_4	Just_31	Done_53	Colours_26
3		Just_30	Cover_139	Just_56
4		Mummy_16	Instructions_18	Come_72
5		Tons_31	Other_49	Other_98
6		Love_16	Thunder_50	Done_186
7		Run_29	Bubble_36	Up_188
8		Some_29	Rub_90	Bumper_190
9		Button_12		But_4
10		Somebody_36	Yummy_76	Some_17

Table Q: STRUT *tokens selected for acoustic analysis*
**Numbering of token reflects record number assigned in PHON.*

Example number	**Age**			
Front rounded	3	4	5	6
1	Look_14	Whoosh_47	Books_91	
2	Put_77	Put_31	Took_7	
3	Put_133	Book_27	Could_30	
4	Look_192	Put_13**	Look_55	Full_38
5	Put_296	Goodie_89	Look_8	Full_85
6	Putting_40	Put_13**	Push_149	
7	Put_1		Good_30	Look_53
8	Book_4	Cooker_9	Look_82	Put_73
9	Would_18	Push_46	Look_98	
10	Woman_38	Book_26	Took_23	Look_155
Mid	3	4	5	6
1		Woody_6	Looking_123	Good_30
2		Put_8	Puss_143	Good_37
3				Good_35
4		Book_28	Push_151	Good_38
5			Puss_144	Good_81
6			Pushed_151	Good_147
7			Push_151	
8			Put_14	
9			Good_30	
Back	3	4	5	6
1	Put_296	Look_4		Look_5
2		Goodie_81	Took_154	Could_14
3		Goodie_81	Look_10	Put_30
4		Couldn't_81		Look_13
5		Putting_21		Goodbye_85
6		Sugar_15	Looks_14	Looked_80
7		Look_43	Look_101	Foot_97
8		Look_47		Put_33
9		Looking_15		Could_120
10				Took_162
Other	3	4	5	6
1		Look_40		

Table R: FOOT *tokens selected for acoustic analysis*
**Numbering of token reflects record number assigned in PHON.*
***2 unique tokens with the same name in the same turn at talk. Both realised as front rounded.*

Example number	**Age**			
Home	3	4	5	6
1	Car_18	Cant_25	Cant_27	Dark_248
2		Cars_22	Cant_8	Are_48
Blend	3	4	5	6
1	Are_18	Cant_41		Car_230
2	Are_18	Cant_64		Dark_235
3		Shark_11		
4		Guitar_85		
5		Are_15		
6		Car_17		
7		Car_17		
8		Cant_41		
Local	3	4	5	6
1	Cant_3	Cant_6	Cant_51	Far_5
2	Cars_1		Are_2	Dark_28
3	Car_2	Harvester_130	Are_20	Start_68
4	Car_2	Marbles_37	Darn_35	Are_1
5	Car_3	Marbles_37	Dark_125	Cant_76
6	Parping_161	Are_16	Cant_7	Cant_76
7	Car_1	Cant_19	Cant_7	Hard_170
8	Parking_31			
9	Parking_31			
10	Parking_31			

Table S: PALM/START *tokens selected for acoustic analysis*
**Numbering of token reflects record number assigned in PHON.*

Example number	**Age**			
Home	3	4	5	6
No tokens				
Blend	3	4	5	6
1		Grass_67	Dance_23	Asked_140
2		Brass_103	Minecraft_49	Asked_140
3		Paster_1	Minecraft_51	
4			Minecraft_52	
5			Minecraft_55	
Local	3	4	5	6
1		Brass_101	Castle_7	Bath_174
2		Caster_15	Castle_10	
3		Pass_57	Castle_57	Castle_22
4		Pass_57		Craft_31
5			Asked_16	Castle_57
6				Castle_79
7				Castle_141
8				Castle_143
9				Castle_144

Table T: BATH *tokens selected for acoustic analysis*
**Numbering of token reflects record number assigned in PHON.*
***No bath tokens appeared in the data before age 4.*

Example number	**Age**			
Home	3	4	5	6
1	Pants_4	Can_12	Batman_21	Flat_7
2	Cat_75	Tag_26	Black_85	Black_29
3	Tack_106	That_26	Gran_136	Carry_57
4	Accident_140	Can_28	That_137	Thats_75
5	That_142	Daddy_40	Pack_14	Band_44
6	Can_163	Back_43	Alphabet_30	And_144
7	Man_12	Stand_1	Plank_27	Jack_148
8	Thank_42	And_3	Apple_127	Man_154
9	Ladder_44	Has_33	Battleship_12	Had_174
10	At_110	Man_123	Have_16	Standing_188
Blend	3	4	5	6
1			Grandma_13	
2		Bad_6	Grandma_136	
3		Bad_13	Grandma_76	
4		Baddies_167		
5		Baddies_167		
6		Baddies_81		
7		Baddies_83		
8		Baddies_83		

Table U: TRAP *tokens selected for acoustic analysis*
**Numbering of token reflects record number assigned in PHON.*

References

Aitken, A. J. (1984). *Scottish accents and dialects*. In P. Trudgill (Ed.), *Language in the British Isles* (pp. 94-98). Cambridge University Press.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511527685

Ball, M. J., & Gibbon, F. (Eds.). (2002). *Vowel disorders*. Butterworth-Heinemann.

Ball, M. J., & Gibbon, F. E. (Eds.). (2013). *Handbook of vowels and vowel disorders*. Psychology Press. https://doi.org/10.4324/9780203103890

Bankson, N., & Bernthal, J. (1998). Analysis and interpretation of assessment data. In J. Bernthal & N. Bankson (Eds.), *Articulation and phonological disorders* (pp. 270–298). Butterworth-Heinemann.

Barbu, S., Martin, N., & Chevrot, J. P. (2014). The maintenance of regional dialects: A matter of gender? Boys, but not girls, use local varieties in relation to their friends' nativeness and local identity. *Frontiers in Psychology*. *5*, 1-11. https://doi.org/10.3389/fpsyg.2014.01251

Beal, J. (2004). English dialects in the North of England: phonology. In B. Kortmann & E. Schneider (Ed.), *A handbook of varieties of English: A multimedia reference tool. Volume 1: Phonology.* (pp. 113-133). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110197181-011

Beck, E. L. (2014). *The role of socio-indexical information in regional accent perception by five to seven year old children.* (Doctoral dissertation, University of Michigan).

Beck, E. (2016). What it means to be an outsider: How exposure to regional variation shapes children's awareness of regional accents in their native language. In A. Babel (Ed.), *Awareness and control in sociolinguistic research* (pp. 104-122). Cambridge University Press. https://doi.org/10.1017/CBO9781139680448.007

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *The Journal of the Acoustical Society of America*, *69*(1), 231-238. https://doi.org/10.1121/1.385343

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native-and Jamaican-accented words. *Psychological Science*. *20*(5). 539-542. https://doi.org/10.1111/j.1467-9280.2009.02327.x

Boersma, P., & Weenink, D. (2018). PRAAT: Doing phonetics by computer [Computer program]. http://www.praat.org

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*. *106*(2), 707-729. https://doi.org/10.1016/j.cognition.2007.04.005

Braine, M. D. (1963). The ontogeny of English phrase structure: The first phase. *Language*. *39*(1), 1-13. https://doi.org/10.2307/410757

Braine, M. D. (1974). On what might constitute learnable phonology. *Language*. *50*(2), 270-299. https://doi.org/10.2307/412438

Braine, M. D. (1976). Review of The acquisition of phonology: A case study, by N. V. Smith. *Language*. *52*(2), 489-498. https://doi.org/10.2307/412573

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*. *3*, 219-252. https://doi.org/10.1017/S0952675700000658

Browman, C. P., & Goldstein, L. (1987). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology. 1*, 341-376. Cambridge University Press. https://doi.org/10.1017/CBO9780511627736.019

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*. *6*(2), 201-251. https://doi.org/10.1017/S0952675700001019

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155-180. https://doi.org/10.1159/000261913

Brown, R. (1973). Development of the first language in the human species. *American Psychologist*. *28*(2), 97-106. https://doi.org/10.1037/h0034209

Brown, R., & Bellugi, U. (1964). Three processes in the child's acquisition of syntax. *Harvard Educational Review*. *34*(2), 133-151. https://doi.org/10.17763/haer.34.2.c5263w0q03222r14

Buhr, R. D. (1980). The emergence of vowels in an infant. *Journal of Speech, Language, and Hearing Research*. *23*(1), 73-94. https://doi.org/10.1044/jshr.2301.73

Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *The Journal of the Acoustical Society of America*. *97*(4), 2603-2606. https://doi.org/10.1121/1.412975

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*. *57*(1), 3-16. https://doi.org/10.1159/000028456

Carle, E. (1969). *The very hungry caterpillar*. World Publishing Company.

Chambers, J. (2002). The Ethan experience: Native phonology in immigrant offspring. *Investigating Change and Variation through Dialect Contact*, special issue of Sociolinguistics. *6*(1), 117-30. https://doi.org/10.1111/1467-9481.00180

Chambers, J. K. (2003). *Sociolinguistic theory: Linguistic variation and its social significance*. (2nd ed.). Blackwell.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*. *87*(2), B69-B77. https://doi.org/10.1016/s0010-0277(02)00233-0

Chen, G., Na, X., Wang, Y., Yan, Z., Zhang, J., Ma, S., & Wang, Y. (2020). Data augmentation for children's speech recognition-The" Ethiopian" system for The SLT 2021 Children Speech Recognition Challenge. *arXiv 2020.* arXiv:2011. 04547.

Cheshire, J., Fox, S., Kerswill, P. E., & Torgersen, E. (2008). Ethnicity, friendship network and social practices as the motor of dialect change: linguistic innovation in London? *Sociolinguistica*. *22*(1), 1-23. https://doi.org/10.1515/9783484605299.1

Chomsky, N. (1957). *Syntactic structures*. Mouton.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper and Row.

Clark, E. V. (2016). *First language acquisition* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781316534175

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*. *116*(6), 3647-3658. https://doi.org/10.1121/1.1815131

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement. 20*(1), 37-46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1968). Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*. *70*(4), 213-220. https://doi.org/10.1037/h0026256

Collins, B., & Mees, I. M. (2013). *Practical phonetics and phonology: A resource book for students*. Routledge. https://doi.org/10.4324/9780203080023

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *The Journal of the Acoustical Society of America. 24*(6), 597-606. https://doi.org/10.1121/1.1906940

Cruttenden, A. (2014). *Gimson's pronunciation of English*. Routledge. https://doi.org/10.4324/9780203784969

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, and Computers*. *28*, 125-127. https://doi.org/10.3758/BF03203646

Davis, B. L., & MacNeilage, P. F. (1990). Acquisition of correct vowel production: A quantitative case study. *Journal of Speech, Language, and Hearing Research*. *33*(1), 16-27. https://doi.org/10.1044/jshr.3301.16

Davis, B. L., & MacNeilage, P. F. (1995). The articulatory basis of babbling. *Journal of Speech, Language, and Hearing Research*. *38*(6), 1199-1211. https://doi.org/10.1044/jshr.3806.1199

Davis, B. L., MacNeilage, P. F., & Matyear, C. L. (2002). Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. *Phonetica*, *59*(2-3), 75-107. https://doi.org/10.1159/000066065

De Decker, P., & Nycz, J. (2011). For the record: Which digital media can be used for sociophonetic analysis? *University of Pennsylvania Working Papers in Linguistics*, *17*(2), 7.

DeJesus, J. M., Hwang, H. G., Dautel, J. B., & Kinzler, K. D. (2017). Bilingual children's social preferences hinge on accent. *Journal of Experimental Child Psychology*. *164*, 178-191. https://doi.org/10.1016/j.jecp.2017.07.005

Demuth, K. (2014). The acquisition of phonology. In J.A. Goldsmith, J. Riggle and C.L. Alan (Eds.), *The handbook of phonological theory* (2nd ed.) (pp. 571-595). John Wiley and Sons. https://doi.org/10.1002/9781444343069.ch17

Department of Education. (2019). Childcare and early years survey of parents in England. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/853358/CEYSP_2019_Report.pdf accessed 27/4/2020

Department of Education. (2021). Childcare and early years survey of parents in England. https://explore-education-statistics.service.gov.uk/find-statistics/childcare-and-early-years-survey-of-parents/2021 accessed 20/2/2023

Deser, T. (1989). Dialect transmission and variation: An acoustic analysis of vowels in six urban Detroit families. *York Papers in Linguistics. 13*, 115–28.

Dinnsen, D. A., O'Connor, K. M., & Gierut, J. A. (2001). The puzzle-puddle-pickle problem and the Duke-of-York gambit in acquisition. *Journal of Linguistics*. *37*(3), 503-525. https://doi.org/10.1017/S0022226701001062

Di Paolo, M., Yaeger-Dror, M., & Beckford Wassink, A. (2011). Analyzing vowels. In M. Di Paolo, & M. Yaeger-Dror (Eds.), *Sociophonetics: A students guide*. Oxford: Routledge.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: a normative study of British English-speaking children. *Clinical Linguistics and Phonetics*. *17*(8), 617-643. https://doi.org/10.1080/0269920031000111348

Donegan, P. (2013). Normal vowel development. In Ball, M. J., & Gibbon, F. E. (Eds.). *Handbook of vowels and vowel disorders* (pp. 42-78). Psychology Press. https://doi.org/10.4324/9780203103890.ch2

Dyer, J. (2007). *Mummy is it [pɒti] or [pari]?: a bilingual child's acquisition of a second English dialect.* In Actas del VI Congreso de Lingüística General: Santiago de Compostela, 3-7 de mayo de 2004. Arco Libros.

Eguchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. *Acta Otolaryngologica*. *257*, 1-51.

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*. *1*, 22-30.

Fant, G. (1971). *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations* (No. 2), Walter de Gruyter. (Original work published 1960).

Fant, G. (1975). Vocal-tract area and length perturbations. *STL-QPSR*. *4*, 1-14.

Ferguson, C. A. (1986). Discovering sound units and constructing sound systems: It's child's play. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 36–57), Lawrence Erlbaum Associates, Inc..

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*. *51*(2), 419-439. https://doi.org/10.2307/412864

Fischer, J. L. (1958). Social influences on the choice of a linguistic variant. *Word*. *14*(1). 47-56. https://doi.org/10.1080/00437956.1958.11659655

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*. *106*(3), 1511-1522. https://doi.org/10.1121/1.427148

Flege, J. E. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. In R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (pp. 157-232). https://doi.org/10.1075/sspcl.1.06fle

Floccia, C., Butler, J., Girard, F., & Goslin, J. (2009). Categorization of regional and foreign accent in 5-to 7-year-old British children. *International Journal of Behavioral Development*. *33*(4), 366-375. https://doi.org/10.1177/0165025409103871

Floccia, C., Delle Luche, C., Durrant, S., Butler, J., & Goslin, J. (2012). Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition*. *124*(1), 95-100. https://doi.org/10.1016/j.cognition.2012.03.011

Foulkes, P., Docherty, G., & Watt, D. (1999). Tracking the emergence of sociophonetic variation. In *Proceedings of the 14th international congress of phonetic sciences* (pp. 1625-1628), The Regents of the University of California.

Foulkes, P., Docherty, G., & Watt, D. (2005). Phonological variation in child-directed speech. *Language*. *81*(1), 177-206. https://doi.org/10.1353/lan.2005.0018

Foulkes, P., Watt, D. J., & Docherty, G. R. (2001). The emergence of structured variation. *University of Pennsylvania Working Papers in Linguistics*. *7*, 1-25.

Genesee, F. (2009). Early childhood bilingualism: Perils and possibilities. *Journal of Applied Research on Learning*. *2*(2), 1-21.

Genesee, F. (2015). Myths about early childhood bilingualism. *Canadian Psychology. 56*(1), 6-15. https://doi.org/10.1037/a0038599

Girard, F., Floccia, C., & Goslin, J. (2008). Perception and awareness of accents in young children. *British Journal of Developmental Psychology. 26*(3), 409-433. https://doi.org/10.1348/026151007X251712

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 159–207). Mouton de Gruyter. https://doi.org/10.1515/9783110895094.159

Green, J. R., Moore, C. A., & Reilly, K. J. (2002). Sequential development of jaw and lip control for speech. *Journal of Speech Language and Hearing Research*. *45*(1), 66–79. https://doi.org/10.1044/1092-4388(2002/005)

Haddican, B., Foulkes, P., Hughes, V., & Richards, H. (2013). Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change*. *25*(3), 371-403. https://doi.org/10.1017/S0954394513000197

Hargreaves, R. (1971). *Mr. Greedy*. World International Publishing Limited.

Hargreaves, R. (1976). *Mr. Strong*. World International Publishing Limited.

Harris, C. M. (1953). A study of the building blocks in speech*. The Journal of the Acoustical Society of America*. *25*(5), 962-969. https://doi.org/10.1121/1.1907227

Harrison, P. (2013). *Making accurate formant measurements: An empirical investigation of the influence of the measurement tool, analysis settings and speaker on formant measurements.* (Doctoral dissertation, University of York).

Hazen, K. (2002). The family. In K. Chambers, P. Trudgill, N. Schilling-Estes, (Eds.), *The handbook of language variation and change* (pp. 500-525). Blackwell. https://doi.org/10.1002/9780470756591.ch20

Hewlett, N., Matthews, B., & Scobbie, J. M. (1999). *Vowel duration in Scottish English speaking children*. Proceedings of the 14th international congress of phonetic sciences, University of California*.*

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*. *97*(5), 3099-3111. https://doi.org/10.1121/1.411872

Hirschfeld, L. A., & Gelman, S. A. (1997). What young children think about the relationship between language variation and social difference. *Cognitive Development*. *12*(2), 213-238. https://doi.org/10.1016/S0885-2014(97)90014-9

Hodson, B., & Paden, E. (1978). Phonological feature competencies of normal 4-year-olds. *Acta Symbolica*. *9*, 37-49.

Honeybone, P. (2009). Distinctive features. In *Key ideas in Linguistics and the Philosophy of Language* (pp. 65-66). Edinburgh University Press. https://doi.org/10.1515/9780748631421-025

Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., & Johnson, K. (1999). Formants of children, women, and men: The effects of vocal intensity variation. *The Journal of the Acoustical Society of America*. *106*(3), 1532-1542. https://doi.org/10.1121/1.427150

Huggins, A. W. F. (1980). Better spectrograms from children's speech. *Journal of Speech, Language, and Hearing Research*. *23*(1), 19-27. https://doi.org/10.1044/jshr.2301.19

Ingram, D. (1989). *First language acquisition: Method, description and explanation*. Cambridge University Press.

Irwin, O. C. (1941). Research on speech sounds for the first six months of life. *Psychological Bulletin*. *38*(5), 277. https://doi.org/10.1037/h0058328

Israel, H. (1968). Continuing growth in the human cranial skeleton. *Archives of Oral Biology. 13*(1), 133-IN41. https://doi.org/10.1016/0003-9969(68)90044-7

Israel, H. (1973). Age factor and the pattern of change in craniofacial structures. *American Journal of Physical Anthropology*. *39*(1), 111-128. https://doi.org/10.1002/ajpa.1330390112

Jakobson, R. (1968). *Child language, aphasia and phonological universals*. (A.R. Keiler, Trans.) De Gruyter Mouton. (original work published 1941). https://doi.org/10.1515/9783111353562

Jakobson, R., Fant, C. G. M., & Halle, M. (1963). *Preliminaries to speech analysis: the distinctive features and their correlates*. MIT Acoustics Laboratory Technical Report 13. MIT Press. (Original work published 1952).

James, G. D. H. (2001). An item analysis of Australian English words for an articulation and phonological test for children aged 2 to 7 years. *Clinical Linguistics & Phonetics*, *15*(6), 457-485. https://doi.org/10.1080/02699200110045344

Jansen, S. (2021). Social meaning and the obsolescence of traditional local structures: Loss of h-dropping in Maryport, West Cumbria. *English World-Wide. 42*(1), 1-28. https://doi.org/10.1075/eww.00058.jan

Jeffries, E. (2016). *Children's developing awareness of regional accents: A socioperceptual investigation of pre-school and primary school children in York.* (Doctoral dissertation, University of York).

Jeffries, E. (2019). Preschool children's categorization of speakers by regional accent. *Language Variation and Change*. *31*(3), 329–352. https://doi.org/10.1017/S0954394519000176

Johnson, W., & Reimers, P. (2010). *Patterns in child phonology*. Edinburgh University Press. https://doi.org/10.1515/9780748642489

Joos, M. (1948). Acoustic phonetics. *Language*. *24*(2), 5-136. https://doi.org/10.2307/522229

Kazazis, K. (1970). The relative importance of parents and peers in first language acquisition: The case of some Constantinopolitan families in Athens. *General Linguistics*. *10*(2), 111–20.

Keating, P. A., & Huffman, M. K. (1984). Vowel variation in Japanese. *Phonetica*. *41*(4), 191-207. https://doi.org/10.1159/000261726

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, *19*(3), 421-447. https://doi.org/10.1044/jshr.1903.421

Kent, R. D., & Bauer, H. R. (1985). Vocalizations of one-year-olds. *Journal of Child Language*. *12*(3), 491-526. https://doi.org/10.1017/S0305000900006620

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *The Journal of the Acoustical Society of America*. *72*(2), 353-365. https://doi.org/10.1121/1.388089

Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech.* 2nd ed. Singular/Thomson Learning.

Kent, R. D., & Rountrey, C. (2020). What acoustic studies tell us about vowels in developing and disordered speech. *American Journal of Speech-Language Pathology*. *29*(3), 1749-1778. https://doi.org/10.1044/2020_AJSLP-19-00178

Kenyon, J. S. (1930). Flat a and Broad a. *American Speech*. 5(4), 323-326. https://doi.org/10.2307/451463

Kerswill, P. (1994). Babel in Buckinghamshire? Pre-school children acquiring accent features in the New Town of Milton Keynes. *Stockholm Studies in English*. *84*, 64-83.

Kerswill, P. (1996). Children, adolescents, and language change. *Language Variation and Change*. *8*(2), 177-202. https://doi.org/10.1017/S0954394500001137

Kerswill, P., & Williams, A. (2000). Creating a new town koine : Children and language change in Milton Keynes. *Language in Society*. *29*(1), 65-115. https://doi.org/10.1017/S0047404500001020

Khattab, G. (2002). *Sociololinguistic competence and the bilingual's adoption of phonetic variants: auditory and instrumental data from English-Arabic bilinguals.* (Doctoral dissertation, University of Leeds).

Khattab, G. (2013). Phonetic convergence and divergence strategies in English-Arabic bilingual children. *Linguistics*. *51*(2), 439-472. https://doi.org/10.1515/ling-2013-0017

Khattab, G., & Roberts, J. (2011). Working with children. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A student's guide* (pp. 163–78), Routledge.

Khodareza, M. R., Shabani, G., & Dogolsara, S. (2015). Investigation of the history and methods of studying first language acquisition. *Indian Journal of Fundamental and Applied Life Sciences*. *5*(1), 4636-4644.

King, E. W. (1952). A roentgenographic study of pharyngeal growth. *The Angle* Orthodontist. *22*(1), 23-37. https://doi.org/10.1043/0003-3219(1952)022%3C0023:ARSOPG%3E2.0.CO;2

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*. *27*(4), 623-634. https://doi.org/10.1521/soco.2009.27.4.623

Kobayashi, C. (1981). Dialectal variation in child language. In P. Dale & D. Ingram (Eds.), *Child language, an international perspective : selected papers from the First International Congress for the Study of Child Language* (pp. 5-27). Baltimore: University Park Press.

Koenig, L. L., & Fuchs, S. (2019). Vowel formants in normal and loud speech. *Journal of Speech, Language, and Hearing Research*. *62*(5), 1278-1295. https://doi.org/10.1044/2018_JSLHR-S-18-0043

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*. *51*(2), 141-178. https://doi.org/10.1016/j.cogpsych.2005.05.001

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*. *100*(4), 2425-2438. https://doi.org/10.1121/1.417951

Labov, W. (1964). Stages in the acquisition of Standard English. In R. Shuy, A. Davis & R. Hogan (Eds.), *Social dialects and language learning* (pp. 77–104). Campaign: National Council of Teachers of English.

Labov, W. (1973). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*. *1*(1), 85-97. https://doi.org/10.1017/S0954394500000120

Labov, W. (2001). *Principles of linguistic change: Social factors*. Wiley Blackwell.

Ladefoged, P. (2005). Features and parameters for different purposes. *UCLA Working Papers in Phonetics*. *104*, 1-13.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, *33*(2), 363-374. https://doi.org/10.2307/2529786

Laver, J. (1994). *Principles of phonetics*. Cambridge University Press. https://doi.org/10.1017/CBO9781139166621

Lee, S., Potamianos, A., & Narayanan, S. (1997). *Analysis of children's speech: duration, pitch and formants*. Proc. 5th European Conference on Speech Communication and Technology (Eurospeech 1997). https://doi.org/10.21437/Eurospeech.1997-161

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*. *105*(3), 1455-1468. https://doi.org/10.1121/1.426686

Leonard, L., Schwartz, R., Folger, M., & Wilcox, M. (1978). Some aspects of child phonology in imitative and spontaneous speech. *Journal of Child Language*. *5*(3), 403-415. https://doi.org/10.1017/S0305000900002075

Leopold, W. (1939). *Speech development of a bilingual child. Vol I: Vocabulary growth in the first two years*. Northwestern University Press.

Leung, A. H. C. (2012). Bad influence?–an investigation into the purported negative influence of foreign domestic helpers on children's second language English acquisition. *Journal of Multilingual and Multicultural Development*. *33*(2), 133-148. https://doi.org/10.1080/01434632.2011.649038

Lewis, M. M. (1999). *Infant speech: A study of the beginnings of language (Vol. 15),* Psychology Press. (Original work published 1936).

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*. *18*(2), 201-212. https://doi.org/10.1016/0022-0965(74)90101-5

Lieberman, D. E., McCarthy, R. C., Hiiemae, K. M., & Palmer, J. B. (2001). Ontogeny of postnatal hyoid and larynx descent in humans. *Archives of Oral Biology*. *46*(2), 117-128. https://doi.org/10.1016/S0003-9969(00)00108-4

Lieberman, P. (1984). *The biology and evolution of language*. Harvard University Press.

Liénard, J. S., & Di Benedetto, M. G. (1999). Effect of vocal effort on spectral properties of vowels. *The Journal of the Acoustical Society of America*. *106*(1), 411-422. https://doi.org/10.1121/1.428140

Lindblom, B. (1962). *Accuracy and limitations of sonagraph measurements*. Proceedings of the fourth international congress of phonetic sciences. Mouton.

Lindblom, B. E., & Sundberg, J. E. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *The Journal of the Acoustical Society of America*. *50*(4B), 1166-1179. https://doi.org/10.1121/1.1912750

Lindsey, G. (2019). *English after RP: Standard British pronunciation today*. Palgrave MacMillan. https://doi.org/10.1007/978-3-030-04357-5

Linville, S. E., & Rens, J. (2001). Vocal tract resonance analysis of aging voice using long-term average spectra. *Journal of Voice*. *15*(3), 323-330. https://doi.org/10.1016/S0892-1997(01)00034-0

Local, J. K. (1978). *Studies towards a description of the development and functioning of children's awareness of linguistic variability.* (Doctoral dissertation, Newcastle University).

Local, J. (1983). How many vowels in a vowel? *Journal of Child Language*. *10*, 449-453. https://doi.org/10.1017/S0305000900007881

Macken, M. A. (1979). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua*. *49*(1), 11-49. https://doi.org/10.1016/0024-3841(79)90073-1

Macken, M. A. (1980). The child's lexical representation: the 'puzzle-puddle-pickle' evidence. *Journal of Linguistics*. *16*(1), 1-17. https://doi.org/10.1017/S0022226700006307

Mæhlum, B. (1992). Dialect socialization in Longyearbyen, Svalbard (Spitsbergen): A fruitful chaos. In E. H. Jahr, (Ed.), *Language contact and language change (*pp. 117-30). Mouton de Gruyter. https://doi.org/10.1515/9783110851847.117

Mann, V. A. (1986). Phonological awareness: The role of reading experience. *Cognition.* 24(1-2), 65-92. https://doi.org/10.1016/0010-0277(86)90005-3

Martland, P., Whiteside, S. P., Beet, S. W., & Baghai-Ravary, L. (1996). *Estimating child and adolescent formant frequency values from adult data.* Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96. https://doi.org/10.1109/ICSLP.1996.607439

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *The Journal of the Acoustical Society of America*. *67*(3), 996-1013. https://doi.org/10.1121/1.383941

Mattingly, I. G. (1966). Speaker Variation and Vocal-Tract Size. *The Journal of the Acoustical Society of America*. *39*(6), 1219-1219. https://doi.org/10.1121/1.1942689

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*. *82*(3), B101-B111. https://doi.org/10.1016/S0010-0277(01)00157-3

McCarthy, D. (1930). *The language development of the preschool child*. Reprint (1975). Greenwood Press.

McCarthy, D. (1946). Language development in children. In *Manual of child psychology (*pp. 476-581). John Wiley and Sons Inc. https://doi.org/10.1037/10756-010

McCawley, J. D. (1974). Review of Chomsky & Halle (1968). In IJAL *40*(1), 50-88. https://doi.org/10.1086/465290

McGowan, R. W., McGowan, R. S., Denny, M., & Nittrouer, S. (2014). A longitudinal study of very young children's vowel production. *Journal of Speech, Language, and Hearing Research*. *57*(1), 1-15. https://doi.org/10.1044/1092-4388(2013/12-0112)

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*. *29*(2), 143-178. https://doi.org/10.1016/0010-0277(88)90035-2

Menn, L. (1978). Phonological units in beginning speech. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments* (pp. 157-172), Elsevier.

Menn, L. (1980). Phonological theory and child phonology. In G. H. Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson (Eds.), *Child phonology* (pp. 23-41). Academic Press. https://doi.org/10.1016/B978-0-12-770601-6.50008-5

Menn, L. (1983). Development of articulatory, phonetic, and phonological capabilities. In B. Butterworth (Ed.), *Language production: Vol. 2. Development, writing, and other language processes (*pp. 3-50). Academic Press.

Menn, L., & Stoel-Gammon, C. (1995). Phonological development. In P. Fletcher & B. MacWhinney (eds.), *The handbook of child language (*pp. 335-58). Blackwell.

Menn, L., & Vihman, M. M. (2011). Features in child phonology: Inherent, emergent, or artefacts of analysis? In N. Ridouane & R. Clements (Eds.), *Where Do Phonological Features Come From? Cognitive, physical and developmental bases of distinctive speech categories* (pp. 261-301). John Benjamins. https://doi.org/10.1075/lfab.6.10men

Menyuk, P. (1968). The role of distinctive features in children's acquisition of phonology. *Journal of Speech and Hearing Research*. *11*(1), 138-146. https://doi.org/10.1044/jshr.1101.138

Mielke, J. (2011). Distinctive features. In M. van Oostendorp, C. Ewen, E. Hume & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 391-415). Wiley-Blackwell. https://doi.org/10.1002/9781444335262.wbctp0017

Miller, W., & Ervin, S. (1964). The development of grammar in child language. *Monographs of the Society for Research in Child Development*. *29*(1), 9-34. https://doi.org/10.2307/1165752

Mol, H. (1963). *Fundamentals of phonetics*. Mouton

Montgomery, A., & Mingis, K. (2022). The evolution of Apple's iPhone. https://www.computerworld.com/article/3692531/evolution-of-apple-iphone.html

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). *Literacy training and speech segmentation. Cognition*. *24*(1-2), 45-64. https://doi.org/10.1016/0010-0277(86)90004-1

Mulak, K. E., & Best, C. T. (2013). Development of word recognition across speakers and accents. In *Theoretical and computational models of word learning: Trends in psychology and artificial intelligence* (pp. 242-269). IGI Global. https://doi.org/10.4018/978-1-4666-2973-8.ch011

Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., & Irwin, J. R. (2013). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent. *Child Development*, *84*(6), 2064-2078. https://doi.org/10.1111/cdev.12087

Nardy, A., Chevrot, J. P., & Barbu, S. (2013). The acquisition of sociolinguistic variation: Looking back and thinking ahead. *Linguistics*. *51*(2), 255-284. https://doi.org/10.1515/ling-2013-0011

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: a preliminary investigation. *Journal of Child Language*. *25*(2), 343-365. https://doi.org/10.1017/S0305000998003444

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language*. *43*(1), 1-19. https://doi.org/10.1006/jmla.2000.2698

Nittrouer, S. (1993). The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech, Language, and Hearing Research*. *36*(5), 959-972. https://doi.org/10.1044/jshr.3605.959

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech, Language, and Hearing Research*. *32*(1), 120-132. https://doi.org/10.1044/jshr.3201.120

Norris, M., Harden, J. R., & Bell, D. M. (1980). Listener agreement on articulation errors of four-and five-year-old children. *Journal of Speech and Hearing Disorders*, *45*(3), 378-389. https://doi.org/10.1044/jshd.4503.378

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*. *85*(3), 172-191. https://doi.org/10.1037/0033-295X.85.3.172

Oller, D. K., & Ramsdell, H. L. (2006). A weighted reliability measure for phonetic transcription. *Journal of Speech Language and Hearing Research*. *49*(6), 1391-1411. https://doi.org/10.1044/1092-4388(2006/100)

Paschall, L. (1983). Development at 18 months. In J.V. Irwin & S.P. Wong (Eds.), *Phonological development in children: 18- 72 months* (pp. 27-74). Southern Illinois University Press.

Payne, A. (1980). Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In W. Labov (Ed.), *Locating language in time and space* (pp. 143-178). Academic.

Penney, G., Fee, E. J., & Dowdle, C. (1994). Vowel assessment and remediation: A case study. *Child Language Teaching and Therapy*. *10*(1), 47-66. https://doi.org/10.1177/026565909401000103

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *The Journal of the Acoustical Society of America*. *109*(6), 2988-2998. https://doi.org/10.1121/1.1370525

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America. 24*(2), 175-184. https://doi.org/10.1121/1.1906875

Pisanski, K., Mora, E. C., Pisanski, A., Reby, D., Sorokowski, P., Frackowiak, T., & Feinberg, D. R. (2016). Volitional exaggeration of body size through fundamental and formant frequency modulation in humans. *Scientific Reports*. *6*(1), 34389. https://doi.org/10.1038/srep34389

Pollock, K. E. (1991). The identification of vowel errors using traditional articulation or phonological process test stimuli. *Language, Speech, and Hearing Services in Schools*. *22*(2), 39-50. https://doi.org/10.1044/0161-1461.2202.39

Rácz, P. (2012). Operationalising salience: definite article reduction in the North of England. *English Language & Linguistics*. *16*(1), 57-79. https://doi.org/10.1017/S1360674311000281

Rastatter, M. P., McGuire, R. A., Kalinowski, J., & Stuart, A. (1997). Formant frequency characteristics of elderly speakers in contextual speech. *Folia Phoniatrica et Logopaedica*. *49*(1), 1-8. https://doi.org/10.1159/000266431

Remijsen, B. (2011). amended by Al Tamimi, J. (2011). http://www.lel.ed.ac.uk/~bert/praatscripts.html

Remijsen, B. (2019). http://www.lel.ed.ac.uk/~bert/praatscripts.html

Roberts, J. (1997). Hitting a moving target: Acquisition of sound change in progress by Philadelphia children. *Language Variation and Change. 9*(2), 249-266. https://doi.org/10.1017/S0954394500001897

Roberts, J., & Labov, W. (1995). Learning to talk Philadelphian: Acquisition of short a by preschool children. Language Variation and Change, *7*(1), 101-112. https://doi.org/10.1017/S0954394500000910

Rose, Y., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K., O'Brien, P., & Wareham, T. (2006). Introducing Phon: A software solution for the study of phonological acquisition. In D. Bamman, T. Magnitskaia & C. Zaller (Eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 489-500). Cascadilla Press.

Sachs, J., Lieberman, P., & Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. In R. W. Shuy & R. W. Fasold (Eds.), *Language Attitudes* (pp.74-83) Georgetown University Press.

Samuel, A. G. (1977). The effect of discrimination training on speech perception: Noncategorical perception. *Perception and Psychophysics*. *22*(4), 321-330. https://doi.org/10.3758/BF03199697

Sander, E. K. (1972). When are speech sounds learned? *Journal of Speech and Hearing disorders*. *37*(1), 55-63. https://doi.org/10.1044/jshd.3701.55

Schmale, R., Cristia, A., Seidl, A., & Johnson, E. K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy*. *15*(6), 650-662. https://doi.org/10.1111/j.1532-7078.2010.00032.x

Schulman, R. (1989). Articulatory dynamics of loud and normal speech. *The Journal of the Acoustical Society of America*. *85*(1), 295-312. https://doi.org/10.1121/1.397737

Scobbie, J. M. (2006). Flexibility in the face of incompatible English VOT systems In L.M. Goldstein, D.H. Whalen & C.T. Best (Eds.), *Laboratory Phonology 8 Varieties of Phonological Competence* (pp. 367-392), Mouton de Gruyter. https://doi.org/10.1515/9783110197211.2.367

Scukanec, G. P., Petrosino, L., & Squibb, K. (1991). Formant frequency characteristics of children, young adult, and aged female speakers. *Perceptual and Motor skills*. *73*(1), 203-208. https://doi.org/10.2466/pms.1991.73.1.203

Shriberg, L. D., Kwiatkowski, J., & Hoffmann, K. (1984). A procedure for phonetic transcription by consensus. *Journal of Speech, Language, and Hearing Research*. *27*(3), 456-465. https://doi.org/10.1044/jshr.2703.456

Shriberg, L. D., & Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics. 5*(3), 225-279. https://doi.org/10.3109/02699209108986113

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*. *55*(4), 779-798. https://doi.org/10.1044/jshd.5504.779

Smith, J., Durham, M., & Fortune, L. (2007). ' Mam, ma troosers is fa'in doon!' Community, caregiver and child in the acquisition of variation in Scottish dialect. *Language Variation and Change*. *19*(1), 63-99. https://doi.org/10.1017/S0954394507070044

Smith, J., & Holmes-Elliott, S. (2018). The unstoppable glottal: Tracking rapid change in an iconic British variable. *English Language & Linguistics*. *22*(3), 323-355. https://doi.org/10.1017/S1360674316000459

Smith, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge University Press.

Smith, N. (2010). *Acquiring phonology: a cross-generational case-study* (No. 124). Cambridge University Press. https://doi.org/10.1017/CBO9780511770692

Sosa, A. V., & Stoel-Gammon, C. (2006). Patterns of intra-word phonological variability during the second year of life. *Journal of Child Language*. *33*(1), 31-50. https://doi.org/10.1017/S0305000905007166

Sóskuthy, M., Foulkes, P., Haddican, B., Hay, J., & Hughes, V. (2015). *Word-level distributions and structural factors codetermine GOOSE fronting.* Proceedings of the 18th International Congress of Phonetic Sciences.

Stampe, D. (1969). The acquisition of phonetic representation. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*. *5*(1), 443-454. Chicago Linguistic Society.

Stanford, J. N. (2008). Child dialect acquisition: New perspectives on parent/peer influence*. Journal of Sociolinguistics*. *12*(5), 567-596. https://doi.org/10.1111/j.1467-9841.2008.00383.x

Stoel-Gammon, C., & Pollock, K. (2008). Vowel development and disorders. In M.J. Ball M.R. Perkins, N. Müller & S. Howard (Eds.), *The handbook of clinical linguistics*, (pp. 525-548). https://doi.org/10.1002/9781444301007.ch33

Stoel-Gammon, C., & Sosa, A. V. (2007). Phonological development. In E. Hoff & M. Schatz (Eds.), *Handbook of child language* (pp. 238-56). Blackwell. https://doi.org/10.1002/9780470757833.ch12

Story, B. H., & Bunton, K. (2015). Formant measurement in children's speech based on spectral filtering. *Speech Communication*. *76*, 93-111. https://doi.org/10.1016/j.specom.2015.11.001

Stuart-Smith, J. (2006). The Influence of the Media. In C. Llamas, L. Mullaney, P. Stockwell (Eds.), *The Routledge Companion to Sociolinguistics* (pp. 140-148). Routledge.

Stuart-Smith, J., Pryce, G., Timmins, C., & Gunter, B. (2013). Television can also be a factor in language change: Evidence from an urban dialect. *Language*. *89*(3), 501-536. https://doi.org/10.1353/lan.2013.0041

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. *Haskins Laboratories: Status Report on Speech Research*. *91*, 45-57.

Studdert-Kennedy, M., & Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. *Studies in the Evolution of Language*. *3*, 235-254. https://doi.org/10.1093/acprof:oso/9780199244843.003.0013

Studdert-Kennedy, M., & Goodell, E. W. (1992). Gestures, features and segments in early child speech. *Haskins Laboratories Status Report on Speech Perception*. *SR-111*(112), 89-102.

Sundara, M., & Scutellaro, A. (2011). Rhythmic distance between languages affects the development of speech perception in bilingual infants. *Journal of Phonetics*. *39*(4), 505-513. https://doi.org/10.1016/j.wocn.2010.08.006

Surek-Clark, C. (2000). Dialect acquisition and prestige. *University of Pennsylvania Working Papers in Linguistics*. *6*(3), 16.

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech*. *46*(2-3), 265-294. https://doi.org/10.1177/00238309030460021001

Templin, M. C. (1957). *Certain language skills in children; their development and interrelationships*. University of Minnesota Press. https://doi.org/10.5749/j.ctttv2st

Thomas, S., & Scobbie, J. M. (2015). Mixed accents: Scottish children with English parents. In J. A. Mompean & J. Fouz-González (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 83–104). Palgrave Macmillan.

Traunmüller, H., & Eriksson, A. (2000). Acoustic effects of variation in vocal effort by men, women, and children. *The Journal of the Acoustical Society of America*. *107*(6), 3438-3451. https://doi.org/10.1121/1.429414

Treiman, R., & Zukowski, A. (1996). Children's sensitivity to syllables, onsets, rimes, and phonemes. *Journal of Experimental Child Psychology*. *61*(3), 193-215. https://doi.org/10.1006/jecp.1996.0014

Trubetzkoy, N. S. (1969). *Principles of phonology.* (C.A.M Baltaxe, Trans.) University of California Press. (Original work published 1939).

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*. *1*(2), 179-195. https://doi.org/10.1017/S0047404500000488

Trudgill, P. (1986). *Dialects in Contact*. Blackwell.

Trudgill, P. (1999). *The Dialects of England* (2nd ed.). Blackwell.

Upton, C. (2004). Received pronunciation. In B. Kortmann & E.W. Schneider (Eds.), *A handbook of varieties of English* (pp. 217-230). Mouton de Gruyter.

Upton, C., & Widdowson, J. D. A. (2006). *An atlas of English dialects: region and dialect*. Routledge. https://doi.org/10.4324/9781315887807

Ureta, J., Brito, C. I., Dy, J. B., Santos, K. A., Villaluna, W., & Ong, E. (2020). *At home with Alexa: a tale of two conversational agents*. Text, Speech, and Dialogue: 23rd International Conference, TSD 2020, Brno, Czech Republic, September 8–11, 2020, Proceedings 23. Springer International Publishing. https://doi.org/10.1007/978-3-030-58323-1_53 van der Feest, S. V., & Johnson, E. K. (2016). Input-driven differences in toddlers' perception of a disappearing phonological contrast. *Language Acquisition*. *23*(2), 89-111. https://doi.org/10.1080/10489223.2015.1047096

van Heugten, M., & Johnson, E. K. (2014). Learning to contend with accents in infancy: benefits of brief speaker exposure. *Journal of Experimental Psychology. 143*(1), 340-350. https://doi.org/10.1037/a0032192

van Heugten, M., Krieger, D. R., & Johnson, E. K. (2015). The developmental trajectory of toddlers' comprehension of unfamiliar regional accents. *Language Learning and Development*. *11*(1), 41-65. https://doi.org/10.1080/15475441.2013.879636

Veatch, T. C. (1991). *English vowels: Their surFACE phonology and phonetic implementation in vernacular dialects*. (Doctoral dissertation University of Pennsylvania).

Velleman, S., & Vihman, M. (2002). Whole-word phonology and templates : trap, bootstrap, or some of each? *Language and Speech. 33*(1), 149-70. https://doi.org/10.1044/0161-1461(2002/002)

Vihman, M. (1982). A note on children's lexical representations. *Journal of Child Language*. *9*(1), 249-253. https://doi.org/10.1017/S0305000900003767

Vihman, M. M. (2014). *Phonological development: The first two years*. Wiley.

Vihman, M., & Croft, W. (2007). Phonological development: toward a 'radical' templatic phonology. *Linguistics*. *45*(4), 683–725. https://doi.org/10.1515/LING.2007.021

Vihman, M. M., Ferguson, C. A., & Elbert, M. (1986). Phonological development from babbling to speech: Common tendencies and individual differences. *Applied Psycholinguistics*. *7*(1), 3-40. https://doi.org/10.1017/S0142716400007165

Vihman, M. M., & Keren-Portnoy, T. (Eds.). (2013). *The emergence of phonology: Whole-word approaches and cross-linguistic evidence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511980503

Vihman, M. M., & Velleman, S. L. (2000). The construction of a first phonology. *Phonetica*. *57*(2-4), 255-266. https://doi.org/10.1159/000028478

Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*. *50*(6), 1510-1545. https://doi.org/10.1044/1092-4388(2007/104)

Vorperian, H. K., Kent, R. D., Gentry, L. R., & Yandell, B. S. (1999). Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: Preliminary results. *International Journal of Pediatric Otorhinolaryngology*. *49*(3), 197-206. https://doi.org/10.1016/S0165-5876(99)00208-6

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., & Yandell, B. S. (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *Journal of the Acoustical Society of America. 117*(1), 338–350. https://doi.org/10.1121/1.1835958

Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Ziegert, A. J., & Gentry, L. R. (2009). Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. *The Journal of the Acoustical Society of America*. *125*(3), 1666-1678. https://doi.org/10.1121/1.3075589

Vorperian, H. K., Wang, S., Schimek, E. M., Durtschi, R. B., Kent, R. D., Gentry, L. R., & Chung, M. K. (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*(4), 995–1010. https://doi.org/10.1044/1092-4388(2010/10-0097)

Wagner, L., Clopper, C. G., & Pate, J. K. (2013). Children's perception of dialect variation. *Journal of Child Language*. *41*(5), 1062-1084. https://doi.org/10.1017/S0305000913000330

Walsh, B., & Smith, A. (2002). Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech,* *Language, and Hearing Research*. *45*(6), 1119–1133. https://doi.org/10.1044/1092-4388(2002/090)

Waterson, N. (1971). Child phonology: A prosodic view. *Journal of Linguistics*. *7*(2), 179-211. https://doi.org/10.1017/S0022226700002917

Watt, D. J. L. (1998). *Variation and change in the vowel system of Tyneside English*. (Doctoral dissertation, Newcastle University).

Watt, D., & Milroy, L. (1999). Patterns of variation and change in three Newcastle vowels: Is this dialect levelling? In P. Foulkes & G. J. Docherty (eds.), *Urban Voices: Accent studies in the British Isles* (pp. 25-46). Arnold.

Watt, D., & Tillotson, J. (2001). A spectrographic analysis of vowel fronting in Bradford English. *English World-Wide*. *22*(2), 269-303. https://doi.org/10.1075/eww.22.2.05wat

Watt, D., Fabricius, A. H., & Kendall, T. (2011). More on vowels: Plotting and normalization. In *Sociophonetics: A student's guide* (pp. 107-118). Routledge.

Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann & Y. Malkiel, (Eds.), *Directions for Historical Linguistics: A Symposium.*

Wells, J. C. (1982). *Accents of English*. Cambridge University Press. https://doi.org/10.1017/CBO9780511611759

Wells, J. C. (2001). IPA transcription systems for English. https://www.phon.ucl.ac.uk/home/wells/ipa-english-uni.htm

Wells, J. C., & Colson, G. (1971). *Practical phonetics.* Pitman.

Werker, J. F., & Fennell, C. T. (2004). Listening to sounds versus listening to words: early steps in word learning. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 79-109). MIT Press.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*. *52*(1), 349-355.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*. *7*(1), 49-63. https://doi.org/10.1016/S0163-6383(84)80022-3

White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*. *14*(2), 372-384. https://doi.org/10.1111/j.1467-7687.2010.00986.x

White, P. (1999). Formant frequency analysis of children's spoken and sung vowels using sweeping fundamental frequency production. *Journal of Voice*. *13*(4), 570-582. https://doi.org/10.1016/S0892-1997(99)80011-3

Whiteside, S. P., & Hodgson, C. (2000). Some acoustic characteristics in the voices of 6- to 10-year-old children and adults: A comparative sex and developmental perspective. *Logopedics Phoniatrics Vocology*. *25*(3), 122-132. https://doi.org/10.1080/14015430050175851

Wong, S. P., & Irwin, J. V. (1983). Methodology. In J. V. Irwin & S. P. Wong (Eds.), *Phonological development in children: 18- 72 months*. Southern Illinois University Press.

Xue, S. A., & Hao, G. J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production. *Journal of Speech, Language, and Hearing Research*. *46*(3), 689-701. https://doi.org/10.1044/1092-4388(2003/054)

Yang, J., & Fox, R. A. (2013). Acoustic development of vowel production in American English children. *Proceedings of the 14th annual conference of the international speech communication association* (pp. 1263-1267). Lyon, France. https://doi.org/10.21437/Interspeech.2013-339