Journal of Phonetics 109 (2025) 101388

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

The acquisition of Multicultural London English: Child and adolescent diphthong variation in West London

Rosamund Oxbury [a], Matthew Hunt [a,b], Kathleen M. McCarthy [a,*]

[a] Queen Mary University of London, United Kingdom
[b] University of Southampton, United Kingdom

A R T I C L E I N F O

Article history:
Received 18 April 2023
Received in revised form 20 December 2024
Accepted 23 December 2024
Available online 1 March 2025

Keywords:
Sociophonetics
Diphthongs
Child acquisition
Multiethnolect
Multicultural London English

A B S T R A C T

This study investigated Multicultural London English (MLE) diphthongs as produced by children and adolescents in the London borough of Ealing, UK. We conducted an acoustic analysis of the diphthongs FACE, PRICE and GOAT in the speech of 24 young people aged 16–24 years and, 14 children aged 5–7 years. The results revealed different production patterns between the children and adolescents for some but not all the diphthong variables. We found that the children's and adolescents' diphthongs were similar in the quality of the onset, and similar to the MLE system described in East London, in the London borough of Hackney. However, the children had not acquired monophthongization of the diphthongs, with adolescents producing significantly more monophthongal tokens of PRICE, GOAT and, to a lesser extent, FACE. These findings have implications both for the study of multiethnolects and MLE, and for research on children's acquisition of sociophonetic variation.

Crown Copyright © 2025 Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Urban centres such as London, U.K, are made up of rich multilingual and multidialectal communities. Such environment provides a unique opportunity to study the emergence of new language varieties. Indeed, much of the variationist sociolinguistic work of the last 30–40 years has focused on the mobility of multilinguals and immigrants as a factor in language change (Horvath & Sankoff, 1987; Kotsinas, 1988; Rampton, 1995). Of interest to the current paper are language varieties that arise in diverse multicultural and multiethnic communities in urban centres. Such varieties are thought to arise in linguistically diverse communities where speakers of the local variety (and the standard variety) are outnumbered by new arrivals. The demographic composition of the area means there is no clear target variety in the dominant societal language, resulting in second-generation children acquiring 'combinations of language features from a rich "feature pool" of linguistic forms influenced by a wide variety of languages, dialects, and learner varieties' (Cheshire and Fox, 2016, 268). The term *multiethnolect* was coined to capture the finding of non-standard urban varieties that are not restricted to a particular ethnic group, rather adopted by individuals from various cultural backgrounds within a community, including those from local families with no recent immigrant background, if their friendship network is culturally and linguistically diverse (Cheshire et al., 2011). In particular, the innovative speech practices of adolescents in major urban cities, mostly boys, in multiethnic and multilingual friendship groups have been central to this work (Clyne, 2000).

In London, the urban variety known as Multicultural London English (MLE) - the focus of this paper - has been subject to numerous variationist studies (Fox, 2007, 2015; Cheshire et al., 2011, 2013; Gates, 2018; Ilbury, 2020). The feature pool of MLE includes several phonetic changes in progress, including variation in the onset and trajectory of several diphthongs. These studies have focused on adolescents, with the exception of Cheshire et al. (2011), and East London. Indeed, the majority of studies on multiethnolects have been on the speech of adolescents, a group seen as the principal innovators of sociolinguistic change (Tagliamonte, 2016). Youth languages of this kind have been attested across Europe: including but not limited to, London, the topic of this paper (Cheshire et al., 2011); Manchester (Drummond, 2018a); Berlin (Wiese, 2009, 2013); Amsterdam (Appel & Schoonen, 2005); Flanders (Marzo & Ceuleers, 2011); Oslo (Opsahl, 2009); Copenhagen and Køge in Denmark (Quist, 2005); Stockholm (Young,

* Corresponding author.
E-mail address: k.mccarthy@qmul.ac.uk (K.M. McCarthy).

https://doi.org/10.1016/j.wocn.2024.101388
0095-4470/Crown Copyright © 2025 Published by Elsevier Ltd.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2021). Similar youth languages exist in Kenya, DR Congo, Sudan, Ethiopia, Rwanda, Central African Republic, Zimbabwe and Uganda (see contributions in Nassenstein & Hollington, 2015).

Less well understood is how and when urban varieties are acquired, and the degree to which the features of MLE make their way into the speech of young children from the same communities as the adolescents. Although significant research has been conducted on children's acquisition of sociolinguistic variation (see Smith, 2021 for a review of this work), the majority has been on children growing up in monolingual communities (see Nardy et al., 2013). In the current study we focus on the realization of three diphthongs that have been shown to be innovative phonetic features of MLE, namely FACE, PRICE and GOAT in the speech of children and adolescents. We focus on West London, a different area to the previously investigated speech patterns in East London but similarly diverse, and we compare the speech of children and adolescents from the same community. These two factors allow us to better understand the development of urban varieties and sociophonetic acquisition in two ways. Firstly, investigating a different area of London that is equally diverse, but with different 'input' varieties will help us understand how urban varieties are formed – if the phonetic features already documented in East London are now typical of young Londoners generally (either as a youth variety of a dialect that's beginning to stabilize and found in children), we would expect them to be present outside of East London, in linguistically diverse West London. Alternatively, given the 'input feature pool' might be different in West London, the characteristics of the variety spoken in West London might be different to East London. Secondly, including children in the study of MLE enables us to better understand if it is a transient youth style (as discussed in the sections to follow) or whether new phonetic variants are stabilizing as a new London variety, and therefore being acquired in early childhood. The sections to follow outline the previous research on MLE diphthong realizations and the acquisition of sociolinguistic variation.

1.1. Multicultural London English

Multicultural London English (MLE) has been described by Cheshire et al. (2013: 65) as "an ethnically neutral variable repertoire that contains a core of innovative phonetic, grammatical and discourse- pragmatic features"; it has been claimed to now be the dominant vernacular variety of English in London for many young people (Fox, 2015; Fox et al., 2011). The earliest known examples of what would now be described as MLE come from Fox (2007, 2015). In Fox's (2007, 2015) speech communities, different phonetic variants were primarily associated with specific ethnic groups, but diverse friendship networks simultaneously led to the diffusion of these variants to other groups, and to convergence on levelled variants. Later work by Cheshire et al. (2011) then identified MLE as a new linguistic variety that was developing in inner London, labelling this variety as a *multiethnolect*; this work involved two research projects in Greater London, namely the Linguistic Innovators project (Kerswill et al., 2008) and the Multicultural London English (MLE) project (Cheshire et al., 2011). These projects identified features of MLE from all levels of grammar, including unshifted, monophthongization of diphthongs (Kerswill et al., 2008); TH- and DH-stopping and -fronting; a reduction in H-dropping; k-backing before non-high back vowels; lexis, largely from Jamaican Creole; *man* used as a pronoun (Cheshire, 2013); simplification of definite and indefinite article allomorphy (Cheshire et al., 2011; Fox, 2015).

In the current study we focus on diphthongs FACE, PRICE and GOAT. Table 1 summarizes the specific tendencies of these vowels in terms of onset and monophthongization that is now thought to be typical of young Londoners; based on the findings of Kerswill et al. (2008), Fox (2015) and the more recent work of Gates (2018). MLE is said to show reversal of Diphthong Shift (DS) (Kerswill et al., 2008). Diphthong Shift is a change that is thought to have derived Cockney-like vowels from RP-like vowels, the assumption being that RP is more conservative and being a prestige accent, resists change to an extent. The argument of Kerswill et al. (2008) is that East End London speech shows a diachronic shift away from Diphthong Shifted (Cockney) vowels to more RP-like vowels and diphthong onsets. They show this through comparison between older Londoners and the 16–19-year-olds in the Linguistic Innovators project. They note that this is a reversal of Diphthong Shift only in the sense that the changes go in the opposite direction to that described by Wells (1982; see also Labov, 1994), and that they do not believe that before Cockney, popular London speech resembled modern-day RP (Kerswill et al., 2008). Similar moves away from diphthong-shifted vowels have been identified in Reading and Milton Keynes (Kerswill & Williams, 2005).

More recent research has suggested gendered and ethnically diverse patterning of diphthong usage. Gates (2018) reported "stark gender differentiation" in FACE and PRICE in a cohort of secondary school students in Newham, East London: the boys in her study showed a more front PRICE and raised FACE compared to the girls. Gates (2018) suggests that gender, ethnic, and peer group identities intersect, such that the White girls used distinctly more conservative diphthong variants in comparison to the majority ethnic girl peer groups. Similarly, Fox (2015) found that, in an East London youth centre, Bangladeshi boys were leading in the innovation of MLE-like FACE and PRICE vowels, and that the innovative variants were diffusing through friendship networks to White British boys; but the White British girls did not seem to be adopting the innovative variants to the same extent and favored more conservative variants. Cheshire et al. (2011) also found complex ethnicity

Table 1
Summary of MLE tendencies in terms of onset and monophthongization in the diphthongs FACE, PRICE, and GOAT.

Diphthong	MLE tendency	Possible realisations found in previous MLE research
FACE - /eɪ/	• Closer realisation than /eɪ/ • Monopthongization	[e] ~ [eɪ]
PRICE - /aɪ/	• More front realisation than /aɪ/ • Monophthongization	[a(ɪ)] ~ [ɐ(ɪ)] ~ [ɐ(e)] ~ [a]
GOAT - /əʊ/	• More back realisation than /əʊ/ • Monophthongization	[oː] ~ [ɔʊ]

and gender-based differences in the adoption of MLE vowels among Hackney 16–19-year-olds. In their study, they distinguish between “Anglo” i.e. speakers whose families were White British and had been based in London for several generations, and “non-Anglo” speakers whose parents or grandparents were from an immigrant background, a broad ethnically diverse group. They found that Anglo girls led in fronting of FOOT; Anglo boys and girls tended to have a more open FACE than non-Anglos; and non-Anglo boys had a markedly further back GOAT onset than the other groups.

While MLE was initially described in East London, recent research has shown that these diphthong tendencies are present in the speech of adolescents in other similarly diverse urban areas in England – namely Manchester (Drummond, 2018a, 2018b) and Birmingham (Khan, 2006). There are trends in the minority ethnic adolescents in all locations to favour: [aɪ] for PRICE, rather than the Birmingham [ɔɪ] variant or the East London [ɐ(ɪ)] variant; [o:] or [ɔʊ] for GOAT, rather than the Birmingham or Cockney [ʌʊ] pronunciation; and [e:] or [eɪ̯] for FACE, rather than Diphthong Shifted [æɪ]. Meanwhile, White British adolescents were more likely to favour traditional local dialect variants for these diphthongs (i.e. Cockney variants in London). Adolescents whose friendship networks included White British and majority ethnic members, were converging on diphthong variants that were intermediate between the most innovative and most conservative variants. Fox et al. (2011) name this emerging variety “Multicultural English”, and similarly Drummond (2018b) argues for a more general ‘Multicultural Urban British English” (MUBE) described as “some kind of overarching variety or repertoire of shared features, with each urban centre having its own local version or sub-variety”.

The presence of MLE features in other urban centres in England raises the question of whether the shared features between these areas arise from similar histories of cultural, linguistic, and ethnic diversity or from geographical diffusion of an adolescent youth style, perhaps related to engagement with youth culture and music, as discussed by Drummond (2018a, 2018b). In the current study, to separate out features that are part of an adolescent style, vs. those that appear to be endogenous to the local community, the speech of children will be compared with that of adolescents. The following section discusses the role of children in our understanding of the development of urban language varieties.

1.2. The role of children and the acquisition of MLE

Children are thought to play a key role in sociolinguistic change, with the traditional view being that they acquire sociolinguistic variation from their primary caregiver (Labov, 2001b). In the case of changes in progress, it has been traditionally thought that children acquire a variable’s rate of use from the caregiver and then increase this rate as they grow up, a process known as incrementation (Labov, 2001b: 427–429); the process is thought to continue up until the child’s phonology stabilizes at around age 17 (Labov, 2001b: 455), giving rise to the eventual adolescent peak. Children also learn from their caregivers which variants are associated with formal vs. informal contexts, matching their caregiver’s situation-dependent style-shifting (Smith & Durham, 2019).

In other cases, changes get transmitted and incremented despite caregiver input not providing a source of community variation i.e., when the child’s caregiver is not from the local community; in these cases, there is a mismatch between the language heard in the home and the language heard outside in the community. A seminal example of this is the work of Kerswill and Williams (2000) in Milton Keynes, in the South of England, UK. Though a largely monolingual community, the majority of children in their data had caregivers from other parts of the UK. Notably, the children aged 8 and 12 showed different rates of use for several variables compared to their caregivers, for example favoring the RP variant, [aʊ], for MOUTH (vs their caregivers’ preference for [æʊ]) and favoring [əʏ] for GOAT (vs a range of variants used by caregivers). In contrast, the younger children in Kerswill and Williams (2000), aged around 4, appeared to have acquired the dialect of their principal caregiver. The idea is that, as they grow up, progressing through school and moving towards adolescence, their peer group takes on greater importance in their lives (Kerswill & Williams, 2000). Similar findings have been made in communities that are radically different from the one in Milton Keynes, including Stanford’s (2008) work on three Sui clans of Guizhou province, China, and Habib’s (2014) work on children in the Oyoun Al Wadi village of Northern Syria. In each of these communities, there appears to be a key shift in children’s productions from matching their caregivers to matching their peers at around 8 years old.

Studies in other communities have shown that the acquisition of sociolinguistic variables starts earlier, as young as 2 years old (Chevrot & Foulkes, 2013; Díaz-Campos, 2005; Foulkes et al., 1999; Kushartanti, 2015; Roberts, 1997; Smith et al., 2007, 2009, 2013). It has also been suggested that acquisition patterns may differ depending on the feature being acquired – with the use of specific speech patterns being driven by social status and linguistic complexity (Chevrot et al., 2000). For example, Roberts (1997) in her Philadelphia study identifies a number of children (aged 3;4–4;11), each with at least one non-local caregiver, who had acquired some of the phonetic features used by their peers, such as [aʊ] fronting (e.g., onset closer to [ɛ]). Smith et al. (2007, 2009, 2013) in their extensive research on the speech of children and their caregivers in Buckie, Scotland, found age-graded differences in the use of the lexical phonological *hoose* variable (the alternation between the diphthong [ʌʉ] and the monophthong [uː], with [ʌʉ] being the local form) by 2–4-year-olds. Specifically, younger children (and their caregivers, when speaking to their children) used the local form less than the older children.

The situation in London, which includes comparatively multilingual and diverse communities, appears to show a different acquisition pattern in young children. Cheshire et al. (2011) found that even children aged 4 had converged on the vowels used by their peers, rather than using their caregivers’ non-local vowel variants. Instead, these children oriented to their peers as their target in the acquisition of GOOSE, rather than their caregivers. The critical difference between the findings of Kerswill and Williams (2000) and those of Cheshire et al. (2011) is that the apparent-time results of the latter did not show a change between the ages of 4 and 8: even the 4-year-olds in inner-London were more similar, linguistically, to their peers than to their caregivers. This suggests that children's participation in language change begins at an earlier age than might be expected in less diverse communities. Additionally, Cheshire et al. also found increasing gender differentiation with age (Cheshire et al., 2011: 172). The 4–5-year-old group showed no significant effect of gender on the vowels, while the 8-year-old group showed gender and ethnicity differences in FOOT but not for FACE or GOAT. Furthermore, in the 12–13- year-old group, they found that Anglo girls were more conservative than their peers with respect to FACE and PRICE; while GOAT and FOOT appeared to show an interaction of broad ethnicity groupings and gender: Non-Anglo boys showed a further back realization of these two vowels compared to girls and Anglo boys (Cheshire et al., 2011, p. 170). Taken together, these findings suggest the acquisition of a multiethnolect may show a different developmental trajectory to what is observed in relatively less diverse communities. That is, in linguistically diverse communities where there is limited local variety, children may navigate to peer speech earlier, with gender differences emerging later in adolescence.

Finally, while not directly investigated in the current study, given the multilingual nature of large urban cities such as London, the role of multilingualism on children's developing repertoires should be mentioned. In such settings, the societal language input from caregivers may be L2-accented (i.e. their second language, L2, is influenced by their heritage language), which potentially influences the children's initial production patterns. For example, Khattab's (2007, 2013) study of three English-dominant bilingual children from a Lebanese family living in Yorkshire, aged 5, 7 and 10, found variation in the use of English in different contexts. In the monolingual English recording sessions, the bilingual children were shown to favor the same set of variants used by their monolingual English friends, including the fronted GOAT diphthong undergoing a change in progress in the community, and to avoid the L2-accented variants that their caregivers sometimes produced (Khattab, 2007). In the Arabic-language recordings with their caregiver, the children switched to English occasionally. In these switches to English, the children frequently used Arabic phonetic features, potentially as part of a "compromise strategy", complying with the mother's encouragement to use Arabic but also using the child's preferred language, English (Khattab, 2013). The data showed that the Arabic-accented variants present in the caregiver input must have been part of the children's repertoires, and that these were used in an addressee- and context-appropriate way.

1.3. The present study

The focus of the current study is on the realization of the FACE, PRICE and GOAT vowels, identified in 2011 by Cheshire et al. (2011) in children and adolescents. These vowels have been reported widely in previous research on MLE in East London and are considered as one of the most salient features of MLE (Cheshire et al., 2011; Kerswill et al., 2008, 2013). The current study is the first to collect data from child and adolescent speakers of MLE from a diverse community in a different area of London, contributing to understanding how urban varieties are formed and acquired in childhood beyond East London. Specifically, this paper aims to address the following research questions:

1. Do children in West London differ from adolescents in their realization of the FACE, PRICE and GOAT? If the MLE is stabilizing as a London variety, rather than solely a transient youth style, we might expect to see similar phonetic realizations between young children and adolescents from this community.
2. Do adolescents and children in West London show similar age and gender patterns in their production of FACE, PRICE and GOAT, to what has been found in East London? We assume that similar production patterns will emerge in West London, where there is a similar extent of linguistic diversity and interactions between friendship groups. Alternatively, the specific realizations might differ from East London, given the different ethnic demographics in West London.

To address these questions, and for the sake of comparability with the existing Cheshire et al. (2011) data, the current study employs the same variationist methodology. In the current study, recordings were taken of 24 young people aged 16–24 and 14 children aged 5–7. Focusing on FACE, PRICE and GOAT, we present acoustic analyses of the speech of both groups, comparing diphthong onsets and trajectories. In doing so, we are significantly extending the study of MLE and multiethnolects more generally, in terms of both their geography and their acquisition.

2. Methodology

2.1. Fieldsite: The London borough of Ealing

Ealing was chosen as a fieldsite for its comparability to East London boroughs (Newham, Hackney, Tower Hamlets) where previous research on MLE had been conducted (see Cheshire et al., 2011; Fox, 2015; Gates, 2019 for details). The key criteria were linguistic and ethnic diversity, and socioeconomic deprivation – circumstances which are argued to lead to the emergence of multiethnolects (Cheshire et al., 2011: 152–153). In the 2011 Census, Ealing was the 3rd most diverse borough in England and Wales. In 2016, "[t]he most common ethnic groups in Ealing's school population are white British (15%), Indian (14%), Eastern European (10%), Somali (8%), Pakistani (7%), Asian Other (7%), Afghan (4%), Arab Other (4%) and black Caribbean (4%)" (Mangara, 2017: 13). The ethnic diversity of Ealing is comparable to Hackney, the field site of Cheshire et al. (2011), however, Ealing has a significantly larger proportion of people with Somali or Arab heritage, reflected both our adolescent and child samples. While we are not directly investigating the influence of heritage languages on English (nor did previous MLE research), to understand the ambient language environment in Ealing, "[a]t the time of the 2011 Census, around a third (35%) of pupils in Ealing's primary schools spoke English as a first language, while in the secondary schools the figure was 45%. Pupils in Ealing schools speak over 100 different languages and the 10 most common languages spoken are: English, Polish, Punjabi, Somali, Arabic, Urdu, Tamil, Persian/Farsi, Gujarati, Pashto/Pakhto (in order of the numbers of speakers)" (Mangara, 2017: 14). In terms of socioeconomic deprivation, the 2011 Census found that Ealing was the 18th most deprived borough in the country. The ward where the youth club was located was one of the poorer areas of the borough and has significantly lower life expectancy than the national average (Mangara, 2017).

2.2. Participants

The child participants were aged 5;5–7;3 years (*n* = 14, 7 girls, 7 boys). They were at school together in an area of West London and lived within the London Borough of Ealing. All children passed a speech production screen (Dodd et al., 2003) and were reported by their teachers to have no known speech, language, or hearing issues. All child participants were born in the UK except for one girl who arrived aged 3. The adolescent participants were aged 16;0–24;0 years (*n* = 21, 7 girls, 14 boys). While this age range is slightly higher than that used by Cheshire et al. (2011), 16–19 years old, it should be noted that all but one of the participants were 20 or under. The single 24-year-old participant was included due to his centrality to this friendship group. All lived in the London borough of Ealing or neighboring West London boroughs and attended a secondary school in Ealing. They were recruited at a youth club near the child participants' primary school. All adolescent participants were born in the UK except for one boy who arrived aged 3 years. See Table 2 for a summary of the participants' ethnicity. It should also be noted that in comparison to the original MLE research, our sample included only 3 white British adolescents ("Anglo" in Cheshire et al., 2011), which reflected the ethnic diversity in the youth club and primary school at the time of the study. This means that we were unable to conduct the broad ethnic group comparisons as conducted in previous MLE research (i.e., Anglo, non-Anglo, as in Cheshire et al., 2011).

Recruitment and recordings were carried out with the approval of the Queen Mary University of London Research Ethics Committee (Ref: QMERC2016/71). The data reported here are part of a larger dataset collected by the first author as part of her PhD.

2.3. Recording methods

All recordings were conducted in peer pairs using a Zoom H4 recorder with Audio-Technica lavalier microphone (sampling rate 44,100 Hz, 16-bit resolution). Age-appropriate methods were used to elicit spontaneous speech from the children and adolescents. While traditional Labovian sociolinguistic interviews were possible with the adolescents, this method cannot be used with children without significant adaptation (Roberts, 1997). To create an environment in which a) children were comfortable and confident speaking at length and b) the target vowels could be elicited, a co-operative game with pairs of children was used in place of a sociolinguistic interview.

2.4. Child recordings

The children were recorded in a quiet room in their school. For the child pairs, we used an adapted version of the spot-the-difference Diapix task (Baker & Hazan, 2011). The children took part in two recording sessions: a training session, using Baker & Hazan's pilot Diapix scene; and a recording session using the modified version of the Diapix task. Although the Diapix task has successfully been used on children as young as 9 (Granlund, 2015), our pilot showed that our child participants found this version too difficult to complete.

Our adapted version of the Diapix task was simpler than the original picture sets in a number of ways. The number of target words was reduced to 16 instead of 36 and spread across 4 picture sets instead of three. Participants in Baker and Hazan's (2011) Diapix task were instructed to describe the scene one quadrant at a time. To facilitate this among 5–7-year-olds, the task was modified by making gridlines visible on the scenes (see Fig. 1 for an example). Following an initial training session, the two children were sat either side of a barrier, obscuring their view of the other child's picture. The children then took it in turns to describe the images in the different quadrants of their picture, working cooperatively to spot the differences between each of their pictures. The adult interviewer was present throughout the recording.

The images for the target words were selected from standardized databases (Duñabeitia et al., 2018; Snodgrass & Vanderwart, 1980). The target words had an average age of acquisition of 4.45 years old (Kuperman et al., 2012). The target vowels and words used to elicit them can be found in Table 3. Target words were coded for age of acquisition (Kuperman et al., 2012), imageability and familiarity (Stadthagen-Gonzalez & Davis, 2006), and raw word frequency in the British National Corpus (BNC; Davies, 2004). In addition to FACE and PRICE, tokens of the point vowels FLEECE, TRAP, LOT, and FOOT were elicited to map the boundaries of the children's vowel spaces.

2.5. Adolescent recordings

The adolescents were recorded in pairs in a quiet space in their youth club. To elicit spontaneous speech, we used a sociolinguistic interview protocol adapted from Labov (1972), conducted by the first author. Sociolinguistic interviews took place in a room in their youth club, with one interviewer and two interviewees, following the method used by Cheshire et al. (2011). The interviews covered topics including race and ethnicity, including discrimination; fights, childhood, the local area and growing up in London, music, religion and superstition, future plans, and language.

2.6. Acoustic analysis

The recordings were transcribed at the utterance level in ELAN, 2020) and force-aligned using FAVE (Rosenfelder et al., 2014). For the child recordings, all productions of the tar-

Table 2
Summary of ethnicity for children and adolescents.

		Ethnicity							
		Arab	Black African	Black Caribbean	Brazilian	Somali	Sri Lankan	White	Total
Gender	Adolescents	4	6	2	0	3	0	3	21
	Children	3	1	0	1	7	1	0	14
	Total	7	7	2	1	10	1	3	35

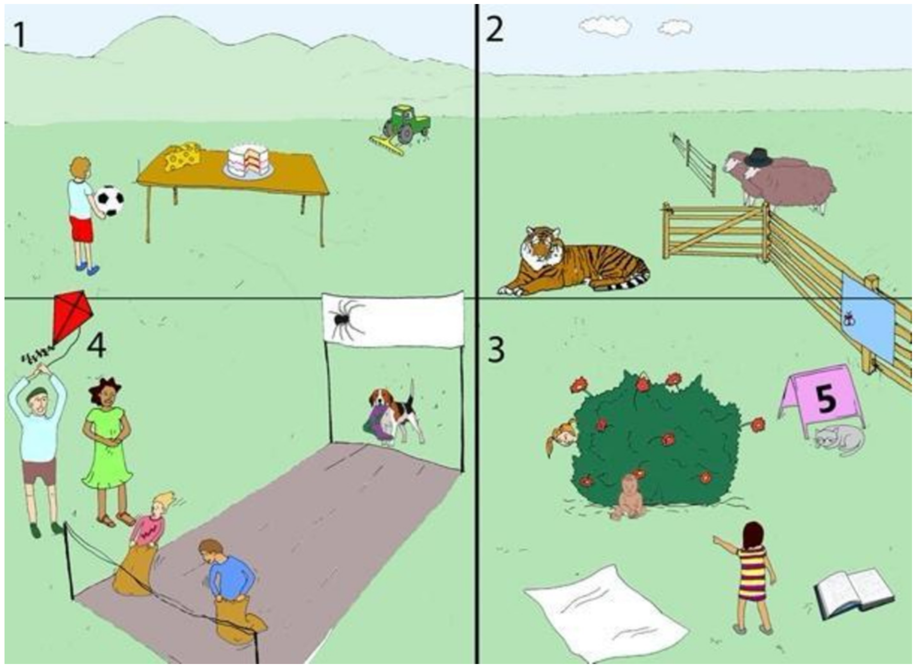


Fig. 1. Example adapted Diapix scene.

Table 3
Target words. No. of phonemes, no. syllables, age of acquisition, imageability, familiarity (both mean transformed) and BNC frequency for the keywords selected for the modified Diapix task. Empty cells indicate that the keyword had no rating for that factor.

Word	Vowel	Phonemes	Syllables	Age of acquisition (years and months)	Imageability	Familiarity	Word frequency
cake	FACE	3	1	4.4	624	594	2700
gate	FACE	3	1	4.4	632	532	3398
baby	FACE	4	2	3.84			8480
table	FACE	4	2	4.39			19,128
kite	PRICE	3	1	4.1			701
five	PRICE	3	1	4.15			39,453
spider	PRICE	5	2	3.43			648
tiger	PRICE	4	2	4			870
ghost	GOAT	4	1	5	622	482	1300
goat	GOAT	3	1	3.9	636	443	593
rope	GOAT	3	1	5			1469
clothes	GOAT	5	1	3.11			6858
sheep	FLEECE	3	1	4.18	641	484	2942
cheese	FLEECE	3	1	4.41	592	588	2504
cat	TRAP	3	1	3.68			3788
hat	TRAP	3	1	3.33			2872
dog	LOT	3	1	3.62	636	598	7780
sock	LOT	3	1	3.94			938
football	FOOT	6	2	4.82	597	565	6536
book	FOOT	3	1	4.47	591		24,142

get vowels were used in the analysis i.e., productions of the Diapix target words and those produced during the task. Tokens were included where the duration of the vowel segment was no less than 50 ms, and where the diphthong was not reduced based on the same 50 ms limit and an auditory impression of a more schwa-like quality. The 50 ms cut-off follows precedent in sociophonetic studies in which the analysis of naturally occurring (rather than read) speech is more common (Tanner et al., 2019). Given that monophthongization is thought to be a change-in-progress in MLE, removing tokens that are more likely to be monophthongized for reasons that do not pertain to sociolinguistics is a more conservative approach aimed at increasing the robustness of the data.

Tokens were also only included where there was not substantial speaker overlap or background noise. Tokens with a coda lateral or approximant, including /ɹ/, /w/, /j/ and /l/, were excluded because of the tendency for these types of segments to produce allophonic changes in preceding vowels (see e.g., Lee-Kim et al., 2013). The final number of tokens analyzed per vowel are included in Table 4 (see Appendix A1 for a more detailed summary). PRICE in *like* was statistically analyzed separately from other PRICE words because the lemma *like* accounted for almost half of the PRICE tokens among the adolescents, and previous research, both in urban and non-urban settings, has shown that the various syntactic functions of *like* systematically influence its phonetic realizations (Drager, 2011, 2016), potentially being more monophthongized and open than other PRICE tokens (Drummond, 2018a; Schleef & Turton, 2018). There were 866 *like* tokens from the adolescents, compared to only 117 from the children.

Table 4
Token numbers by vowel and age.

	FACE	PRICE	PRICE in *like*	GOAT
Children	841	842	117	942
Adolescents	1590	1076	866	1546
Total	2431	1918	983	2488

Due to the nature of child speech data in which the harmonics are typically more spaced out, thus obscuring formants in a spectrogram (Thomas, 2011: 160), different acoustic parameters for formant extraction were used for the children and the adolescents (see Appendix A2 for the acoustic parameters, and Appendix A4 for example spectrograms).

The target diphthongs were manually segmented to ensure consistent placing of the boundaries. Measurements of F1 and F2 frequencies at 20%, 35%, 50%, 65% and 80% of vowel duration were extracted using hand-corrected LPC analyses in Praat (Boersma, 2019) – these measurements were visually inspected and manually corrected, if necessary, before being extracted using a script. The parameters used in the annotation script are given in the Appendix (Table A2). To enable comparison between adolescent boys, adolescent girls, and children, the formant frequencies were normalized using the modified Watt-Fabricius method (Fabricius, Watt, & Johnson, 2009). This method reduces the differences in formant frequencies that arise due to speaker physiology, whilst preserving other variation. This version of the Watt-Fabricius (Fabricius et al., 2009) method has been used in many recent studies in sociophonetics (e.g., Wormald, 2015; Podesva et al.,et al., 2015). See Appendix A3 for more detail.

From these measurements we calculated two metrics for our diphthong analysis: (1) Trajectory length and (2) Onset F1/F2 frequency. Fox & Jacewicz's (2009) Trajectory Length measure was adopted for the current study as a measure of monophthongization. It was desirable to have a measure of diphthong dynamics that took account of change in both F1 and F2 together, and Fox and Jacewicz (2009) found that Trajectory Length outperformed other metrics in measuring the monophthongization of [ɑɪ]. The calculation of trajectory length involves calculating the Euclidean distance in F1xF2 space within four sections of the vowel – 20–35%, 35–50%, 50–65% and 65–80% – and then taking the summing of these four vectors. As such, measurements of the first and second formant were taken at 20%, 35%, 50%, 65% and 80% time points in each vowel token. As Trajectory Length can only take on positive values, the natural logarithm was taken and used as the dependent variable in the statistical models. Trajectory Length is hereafter abbreviated to TL.

The onset qualities of FACE, PRICE and GOAT are operationalized as the F1 frequency of FACE, and the F2 frequency of PRICE and GOAT, at the 20% duration point. Hereafter, "onset F1" or "onset F2" may be used to mean the F1/F2 frequency at the 20% duration point. The choice of selecting F1 or F2 as the dependent variable for each diphthong is based on the MLE tendencies described above in Table 1 (Kerswill et al., 2008). That is, an MLE realization of FACE is indicated by a lower F1 at onset, i.e., more close realization. For PRICE, MLE shows a more front onset diphthong, i.e., higher F2. MLE shows monophthongization of both these diphthongs, i.e., the change in F1 and F2 from onset to offset is less.

2.7. Statistical analysis

Statistical analysis was carried out in the Bayesian paradigm, using the brms package in the open-source software R (Buerkner, 2017; R Core Team, 2020). A total of 6 Bayesian mixed-effects linear regression models were fit: one for the diphthong onset and one for the trajectory of each of the 3 diphthongs. The independent predictors included in the models were: log(duration); age (adolescent or child); speaker gender (girl or boy); an interaction of age and gender. By-speaker and by-word random intercepts were included, and by-speaker random slopes for preceding environment and following environment. Preceding environment included approximant, coronal, labial, nasal, velar and other. Following environment included coda nasal, coda voice obstruent, coda voiceless obstruent, word-final open syllable, and word-medial open syllable. Note, as outlined above, the broad analysis of ethnicity investigated in previous MLE research (i.e., Anglo, non-Anglo) was not included in our modelling because our sample, while ethnically diverse, did not include such broad categorization (see Participants section above).

Duration was log-transformed and included in the models to control for this factor – we would expect that tokens with a longer duration would also be more diphthongal. Continuous variables were z-scored before being entered into the models, and categorical variables were sum-coded.

The following are reported in the presentation of results:

1. The estimated regression coefficient ($\widehat{\beta}$), i.e., the median of the posterior distribution
2. 95% Highest Density Interval (HDI) about this estimate
3. Probability of direction (PD): the probability that the effect is positive or negative – the proportion of the posterior distribution that is above or below 0. Close to 50% = low probability, close to 100% = high probability.

If 95% of the posterior is > or <0, and the 95% HDI does not include 0, we will conclude that there is strong evidence for a positive or negative effect (Tanner et al., 2019; Franke & Roettger, 2019; Nicenboim & Vasishth, 2016). If 95% of the posterior is > or <0, but the 95% HDI includes 0, we will say that there is marginal evidence for a positive or negative effect.

3. Results

3.1. Descriptive analysis of vowel plots

Fig. 2 shows the normalized mean F1 and F2 and 68% confidence ellipses for the point vowels FLEECE, TRAP, LOT, and the diphthongs FACE, PRICE, and GOAT. The trajectories the 35%,

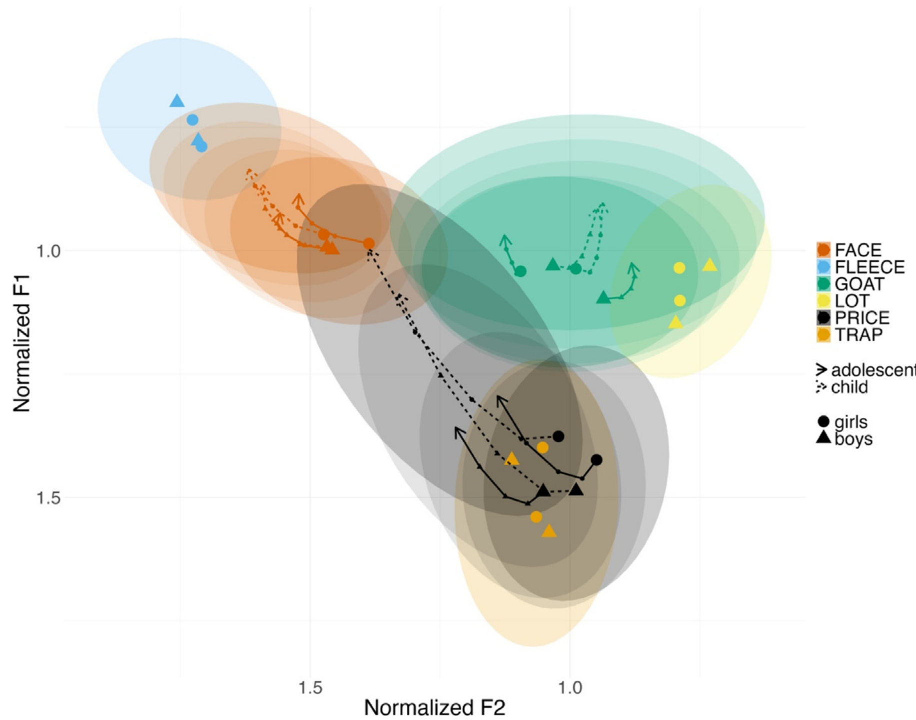


Fig. 2. Vowel plot by age and gender showing 68% confidence ellipses and means for the target vowels FACE, GOAT, PRICE, and point vowels FLEECE, LOT and TRAP from normalized F1 and F2 measurements taken at the 20% time point. Further ellipses and trajectories at the 35%, 50%, 65% and 80% timepoints are included for the vowels FACE, PRICE and GOAT.

50%, 65% and 80% timepoints are included for the vowels FACE, PRICE and GOAT. A few observations can be drawn from this plot, with regard to the target vowel trajectories. Firstly, the trajectory of FACE is very consistent between the two age and gender groups, and within these groups. The plot for PRICE is especially revealing. The trajectory shapes for adolescents and children are similar, i.e., F1 increases initially and then rapidly decreases, while F2 shows a constant increase over time. But for the children this pattern is exaggerated, leading to a crossover pattern: between the 50% and 65% time points, the values of the two normalized formant values crossover and F2 finishes at a high value (i.e., front target), while F1 finishes low (i.e. close target). For the adolescents, the overall amount of change in each formant is smaller. For GOAT, for all groups, the F1 starting point is similar, between 1.00 and 1.10. However, while for children the F2 starting point is around 1.00, for adolescent girls it is 1.09, and for adolescent boys, it is 0.93. For adolescent boys, and for children of both genders, F2 tends on average to decrease over the duration of the vowel. However, for adolescent girls, the F2 tends to increase slightly over time. For both ages and genders, F1 decreases over time, indicating that the end point of the vowel is closer than the onset.

3.2. Bayesian analysis: Diphthong onsets

3.2.1. FACE onset F1

There was marginal evidence for a main effect of age on FACE F1 onset, with adolescents predicted to have a higher onset F1 than children ($\widehat{\beta}$ = 0.16; 95% HDI [−0.03, 0.36]; 95% PD). But there did not appear to be a main effect of gender on onset F1 ($\widehat{\beta}$ = 0.02; 95% HDI [−0.15, 0.20]; 58% PD), nor was there strong evidence for an interaction of age and gender ($\widehat{\beta}$ = 0.04; 95% HDI [−0.14, 0.20]; 66% PD). This can be seen in Fig. 3, where a gender difference is not evident for either age group, but the children are predicted to have a lower onset F1 than the adolescents.

3.2.2. PRICE onset F2

There was no evidence for an age difference with regard to PRICE F2 onset, as the posterior was centered at zero ($\widehat{\beta}$ = 0.00; 95% HDI [−0.20, 0.19]; 51% PD). However, there was strong evidence for a main effect of gender ($\widehat{\beta}$ = −0.27; 95% HDI [−0.45, −0.10]; 99.85% PD), and also strong evidence for an interaction of age and gender ($\widehat{\beta}$ = −0.25; 95% HDI [−0.40, −0.08]; 99.80% PD). This is because there is a stark gender difference in the F2 onset of PRICE among adolescents that is not reflected in the children, who instead tend to have a PRICE F2 onset that is intermediate between that of the adolescent boys and girls. This is shown in Fig. 3 which shows the posterior predicted median onset F2 for the age and gender groups.

3.2.3. PRICE in like onset F2

The model did not find strong evidence for an age difference in PRICE onset F2 in *like*; however, it does allow us to be 90% confident that the children have a higher onset F2 than the adolescents in the PRICE vowel in *like* ($\widehat{\beta}$ = −0.18; 95% HDI [−0.46, 0.08]; 90.75% PD). The evidence for a main effect of gender also did not reach the criterion for strong evidence ($\widehat{\beta}$ = −0.21; 95% HDI [−0.49, 0.06]; 93% PD), though this means that we can be 93% confident that boys have a higher

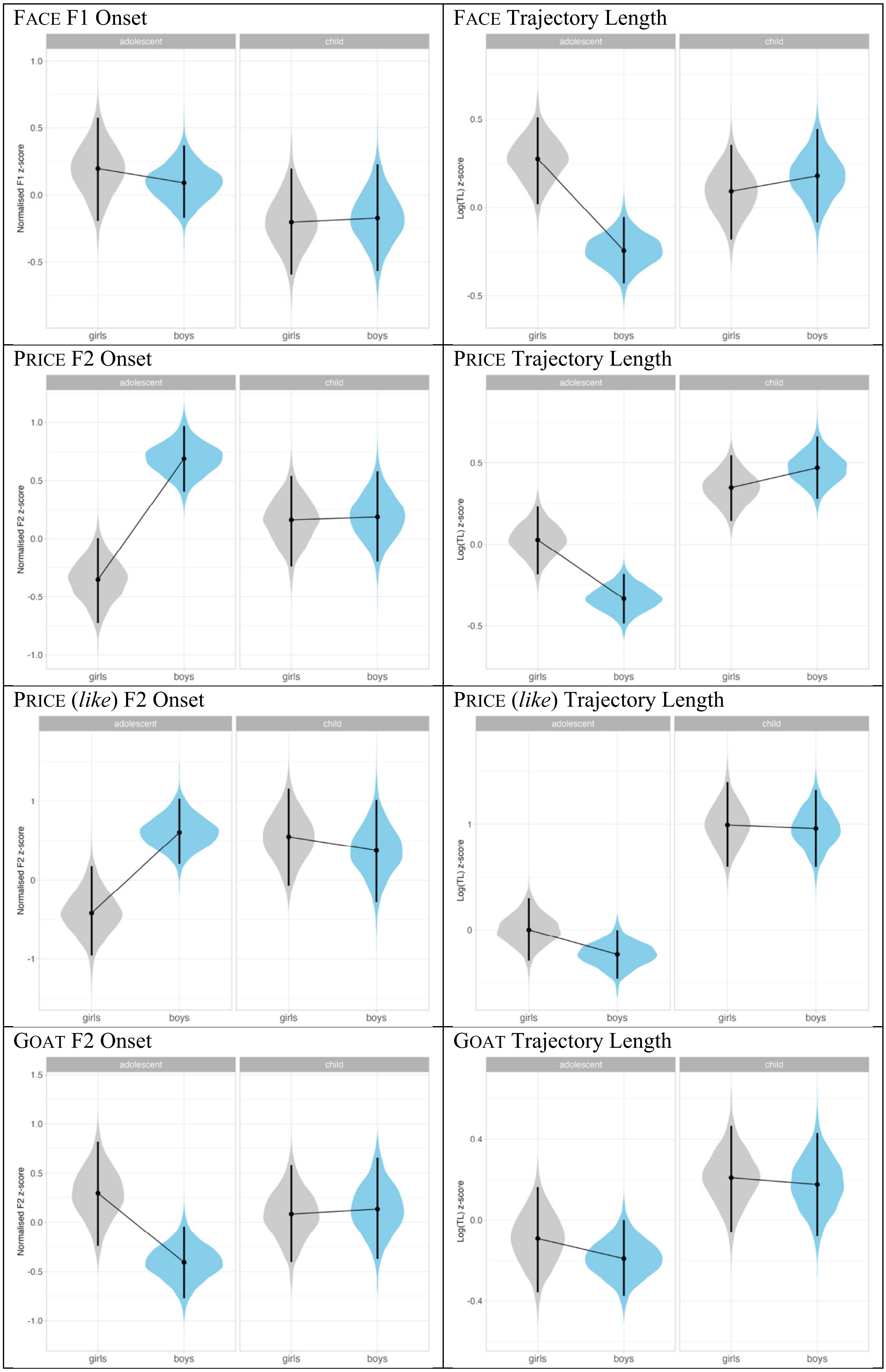


Fig. 3. F1/F2 onset and Trajectory Length (TL) for FACE, PRICE, PRICE (*like*) and GOAT by age group (adolescent, child) and gender (girls = grey, boys = blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

onset F2 than do the girls. There was also strong evidence for an interaction of age and gender, similar to that found for the other PRICE words ($\widehat{\beta}$ = −0.30; 95% HDI [−0.58, −0.02]; 98% PD): among the children, girls are predicted to have a slightly higher onset F2 than the boys, but among the adolescents, the girls are predicted to have a much lower onset F2 than the boys. These findings are represented graphically in Fig. 3; the adolescent girls are predicted to have a lower onset F2 than the adolescent boys; the children are predicted to show an onset F2 that is more similar to the adolescent boys than to the adolescent girls.

3.2.4. GOAT onset F2

Adolescents are estimated to have a slightly lower F2 onset compared to the children, but the evidence for this effect was not strong ($\widehat{\beta}$ = −0.08; 95% HDI [−0.34, 0.17]; 75% posterior < 0). There was some evidence for a main effect of gender on GOAT F2 onset, though this did not quite reach the criterion for strong evidence ($\widehat{\beta}$ = 0.16, 95% HDI [−0.09, 0.39]; 91% PD). There was also marginal evidence for an interaction of age and gender ($\widehat{\beta}$ = 0.19; 95% HDI [−0.03, 0.39]; 96% PD). This is because the evidence for a gender difference is stronger among the adolescents than among the children: among adolescents, girls are estimated to have an F2 onset that is 0.70 standard deviations greater than the boys (95% HDI [0.09, 1.32]), while among the children, girls are estimated to have an F2 onset that is −0.05 standard deviations lower than that of the boys (95% HDI [−0.72, 0.60]). It can be seen in Fig. 3 that adolescent girls are predicted to have a much higher F2 onset than the boys, while the children are predicted to have an F2 onset that is intermediate between the adolescent girls and adolescent boys.

3.2.5. Summary

The only diphthong onset to show even marginal evidence for a main effect of age was FACE F1. The children were found to have a lower F1 onset than the adolescents; this means that the children show a more MLE-like realization of the FACE onset than the adolescents. There were also several age-gender interactions, predominantly among the adolescents. For both PRICE onset and PRICE in *like*, adolescent girls were predicted to have a lower F2 onset than adolescent boys, while the reverse was true for GOAT. For all three, the children, regardless of gender, were predicted to have an onset F2 somewhere in-between the adolescent boys and girls.

3.3. Diphthong trajectories

3.3.1. FACE TL

There was some evidence that the children show a greater TL than the adolescents, though this did not reach the criterion for strong evidence ($\widehat{\beta}$ = −0.06; 95% HDI [−0.20, 0.07]; 83% PD). At the same time, there was marginal evidence for a main effect of gender ($\widehat{\beta}$ = 0.11; 95% HDI [−0.01, 0.22]; 96% PD), and strong evidence for an interaction of age and gender ($\widehat{\beta}$ = 0.15; 95% HDI [0.04, 0.26]; 99.48% PD). This is because among the adolescents, there is a clear gender difference, with girls showing a greater TL than boys, but this difference is not found among the children. This can be seen in Fig. 3, which shows that adolescent boys are predicted to have a smaller TL i.e., monophthongal realization of FACE, while adolescent girls are predicted to have a more diphthongal realization, and children are predicted to favor a FACE variant that is intermediately diphthongal. This strongly resembles the pattern found for GOAT onset F2 and PRICE onset F2.

3.3.2. PRICE TL

There was strong evidence for an effect of age ($\widehat{\beta}$ = −0.28; 95% HDI [−0.38, −0.19]; 100% PD), meaning that the adolescents tend to have a smaller TL i.e., more monophthongal realization of PRICE than the children. Although it did not reach the criterion for strong evidence, there was some evidence for a main effect of gender ($\widehat{\beta}$ = 0.06; 95% HDI [−0.03, 0.14]; 91% PD), and also strong evidence for an interaction of age and gender ($\widehat{\beta}$ = 0.12; 95% HDI [0.04, 0.21]; 99.67% PD). This means that while the children generally have a more diphthongal realization of PRICE than do the adolescents, within the adolescents, girls are estimated to have a greater TL than the boys, while among the children, the evidence for a gender difference is not so strong.

3.3.3. PRICE in like TL

There was strong evidence for an age difference ($\widehat{\beta}$ = −0.55; 95% HDI [−0.72, −0.39]; 100% PD). The children have a greater TL i.e., more diphthongal realization of PRICE in like than do the adolescents, by around 1.1 standard deviations (95% HDI [−1.44, −0.79]). Meanwhile, neither the main effect of gender, nor the interaction of age and gender, reached the criterion for strong evidence.

3.3.4. GOAT TL

There was strong evidence that the adolescents tend to have a smaller TL than the children, i.e., a more monophthongal realization of GOAT ($\widehat{\beta}$ = −0.17; 95% HDI [−0.29, −0.04]; 99.5% posterior < 0). This can be seen in Fig. 3. There was no evidence for a main effect of gender on GOAT TL ($\widehat{\beta}$ = 0.03, 95% HDI [−0.09, 0.15]; 69% PD), and similarly, no evidence for an interaction effect of age and gender on GOAT TL ($\widehat{\beta}$ = 0.02, 95% HDI [−0.10, 0.13]; 62% PD).

3.3.5. Summary

Main effects of age were found for both the PRICE (including *like*) and GOAT trajectories, with adolescents predicted to show a more monophthongal realization. The adolescents were also predicted to show a more monophthongal realization of FACE than the children, though this did not reach the criterion for strong or marginal evidence. At the same time, an age-gender interaction emerged for FACE TL, similar to those found for PRICE and GOAT onset F2: the adolescent girls were predicted to have a more diphthongal realization of FACE than the boys; while the children were predicted to have a FACE TL that was intermediate between that of the adolescent girls and boys. A gender difference was not found among the children.

4. Discussion

The current study investigated the production of three diphthongs, FACE, PRICE and GOAT, by children and adolescents in

West London. These vowels have been extensively documented in East London as key features of Multicultural London English (MLE, Cheshire et al., 2011; Fox, 2007; Gates, 2018, 2019; Kerswill et al., 2008, Watt & Fabricius, 2002). To further our understanding of the development and acquisition of urban varieties, this paper expands on the original MLE research to a new area of London and compares children and adolescent from this community. Two key findings arose from this study. Firstly, in the adolescents' speech, we found very similar realizations to what has been shown in East London for FACE, PRICE and GOAT, including a clear gender differentiation between adolescent boys and girls (also found in Gates, 2018). Secondly, similar to Cheshire et al. (2011) we found no age difference for diphthong onsets; however, our analysis of diphthong trajectories showed a more diphthongal realizations by children than adolescents. Taken together, our findings suggest that some features (diphthong onsets) are beginning to stabilize as a London vernacular, whereas other features (monophthongization, gender differentiation,) are possibly age-graded, and acquired later in childhood or adolescence.

A major finding of this paper is that the children in West London showed a similar PRICE and GOAT onsets to the adolescents, partly replicating the findings in East London (Cheshire et al. 2011; Gates, 2018 Kerswill et al., 2013). For PRICE and GOAT, there was not compelling evidence to suggest a significant difference between the two age groups, while for FACE, there was marginal evidence that the children have a more closed FACE onset than the adolescents. For all three vowel onsets, the children, regardless of gender, were predicted to have an F2 onset somewhere in-between the adolescent boys and girls. What we might be capturing is a point in the children's acquisition of sociolinguistic variables, an age-graded change – they have partly acquired the same system as the adolescents but might become like the adolescents and display a gender split as they grow older or become teenagers (i.e., the adolescent peak, Kirkham & Moore, 2013; Labov, 2001b).

Although Cheshire et al. (2011) and Kerswill et al. (2013) claim that 4–5-year-olds in East London had acquired the same vowel system as the adolescents, this claim was based on comparison of diphthong onsets, and not on diphthong dynamics. The current study adds to this by exploring the children's diphthong dynamics (trajectory length, TL). We found similar age patterns to Cheshire et al. (2011) for the vowel onsets, but our findings for diphthong dynamics showed age differences. For PRICE and GOAT TL, we found strong evidence that the adolescents have a more monophthongal realization of these vowels than the children i.e., an MLE realization. For FACE, the findings were complicated by an age-gender interaction: the adolescent girls were more diphthongal than the adolescent boys, and the children were intermediately diphthongal. In fact, similar age-gender interactions were found for PRICE and GOAT onset. For FACE TL, and PRICE and GOAT onset, an age-gender interaction pattern was found: the adolescent boys showed the most MLE-like realization, as described in East London i.e., the most monophthongal FACE, the most front PRICE onset and the most back GOAT onset; the adolescent girls showed a relatively diphthongal FACE, back PRICE onset and central/front GOAT onset. Compared to the adolescents, the children showed no gender differentiation among themselves, and also showed phonetic realizations of these variables that were intermediate between the adolescent boys and girls. These findings align remarkably closely with those of Cheshire et al. (2011), who found that gender differences were not in evidence before age 8. Cheshire et al.'s apparent-time findings suggested "increasing gender differentiation with age" in the MLE vowels in East London (Cheshire et al., 2011, p. 172). In the 4–5-year-old age group, they found no significant effect of gender on the vowels, while the 8-year-old group showed no gender effects for FACE or GOAT, but did show gender and ethnicity differences in FOOT. The lack of gender differentiation in our data aligns with a body of studies that have found no gender differences in sociolinguistic usage in young children; while other studies have found that gender differences are in evidence even in children this young (see Foulkes & Docherty, 2006). This suggests that while children increasingly produce gender-specific speech with age, this may possibly vary depending on the specific variable being explored. Taken together, these findings suggest a more complex developmental pattern than has been shown in East London: children and adolescents are similar for diphthong onsets, but they show age differences in terms of their PRICE and GOAT vowel trajectories, but not FACE. Unlike the adolescents, they do not produce monophthongal realizations PRICE and GOAT.

Why are the children more diphthongal than the adolescents in PRICE and GOAT? The first possibility is that in these children's ongoing acquisition of sociolinguistic norms in their community, the target is not MLE, but potentially the levelled variety of English typical of the southeast of England. While we cannot directly test this without speech data from the community and their caregivers, this would explain the central and open onset to PRICE, the close-mid onset to FACE and the close-mid central onset of GOAT. To some extent, these two possible targets – levelling changes in the southeast, and MLE – overlap (cf. Kerswill et al., 2008), but this possibility would certainly explain why the children have more diphthongal diphthongs than the adolescents. Impressionistically, there is certainly a huge degree of interspeaker variation within the children in our sample. While variability is typical in child speech (McLeod, 2003), some of our children variably used features linked to the southeast rather than to London urban language – for example, among the girls, there were occasional tokens of GOAT fronted in the way described by Kerswill and Williams (2000), i.e. [əʏ]. While detailed individual case studies is beyond the scope of this paper, it might be the case that our recordings are capturing a point in the children's acquisition, where they have acquired some but not all the MLE phonetic characteristic for these vowels. Likewise, the children's phonetic realizations might partly reflect the speech patterns in the children's home environment, something that we did not record in this project, but are currently capturing in child language projects in East London (https://generationsoflondonenglish.org). It might be suggested that the 'feature pool' or ambient languages in the children's environment in West London compared to East London, give rise to an urban English variety that contains more diphthongal realizations of these vowel. We know for example, in studies of British Asian 3–6-year-old children in East London, that they initially exhibit British Asian features in their speech at the start of school – with time, some of these features are no longer present (Kirkham & McCarthy, 2021; McCarthy et al., 2014). Without a systematic study of MLE and ethnicity, we don't know is if some ethnic groups continue to maintain features in their London speech that index their ethnicity. More studies are needed of children's acquisition of changes in progress in urban centres, and especially studies that account for the specific language environment of children acquiring those changes. Without more data on children's acquisition of this kind of change, it is impossible to know whether, in the current case, it is the local 'feature pool' that leads to the children having more diphthongal variants of PRICE and GOAT than the adolescents, or whether it is to do with how diphthongs and monophthongization is acquired.

At the same time, we might expect the children's speech to be closer to the ambient home environment, than that of the adolescents, because adolescents are known to be especially innovative in their language use (Tagliamonte, 2016; Holmes-Elliott, 2021). Even though transmission and incrementation presumably do not occur in the same way in this community as they do in monolingual communities, because of the diverse and multilingual nature of the ambient speech environment as suggested by Cheshire et al. (2011; see also Labov, 2001a, 2007), this does not necessarily mean that we should not expect an adolescent peak: the adolescent peak could occur because of age-grading. However, if this is the correct explanation – i.e., the adolescents are simply more innovative in their use of sociophonetic variation than the children – this does not explain why an age difference has emerged for monophthongization of the diphthongs, and not for the diphthong onset qualities. What we might be seeing in our data is age-graded developments for monopthongization of PRICE and GOAT and gender differentiation for the onsets. In contrast, onset qualities for these vowels, being acquired early, and present in children's speech in East London (Cheshire et al., 2011), suggests that some aspects of MLE are stabilizing, here the diphthong onsets, as part of the new London vernacular. In our study, what we might be capturing is a specific point in the children's acquisition trajectory. To fully understand the developmental process, we need longitudinal research that includes detailed analysis of the children's ambient home and school environment.

Finally, it is worth considering the influence of (a) the nature of the Diapix task and (b) the recording setting. We know that children are sensitive to communicative context and to interlocutor (e.g., Smith & Durham, 2019; Khattab, 2013). The children's recording took place during school hours, and even though it was stressed to the children that the Diapix task was a game, their interpretation of the communicative context may have been influenced by the fact that the activity was carried out in a room at school, and overseen by an adult, the researcher. It is worth noting, however, that the Diapix game was played in peer pairs rather than with the researcher. Recent research using the same Diapix task with various interlocutors in a school setting (peer, researcher, caregiver) has shown an interlocutor effect (Jones & McCarthy, *in prep*). Specifically, similar to Khattab (2007), children show different speech patterns when playing the Diapix game with peers (peer:peer context Diapix game), vs researchers or caregivers. It is therefore unlikely that the researcher simply being present in the room influenced the children's speech. However, the difference in the speech elicitation methods used with adolescents was different to the children - sociolinguistic interviews conducted in the adolescent's youth club – and this difference may have contributed to some of the age differences found in our study. However, without a systematic study of speech in different environments it is hard to fully know how these differing methods drove the age differences found in our study.

5. Conclusion

Multicultural London English began to emerge during the 1980s, and up until now has been exclusively studied in East London, with very few studies exploring how and when MLE is acquired in childhood. In the current study, we have shown that MLE is emerging and possibly stabilizing across the city, in a similarly diverse yet different area of inner London. Our adolescents in West London produced MLE realizations of the FACE, PRICE and GOAT vowels. We also showed that children in our study had acquired some but not all MLE features for these vowels. Taken together, this suggests that MLE is stabilizing as a new London English variety, with some features possibly remaining as an adolescent youth style, acquired later in childhood or adolescents. To date, monolingual homogeneous communities have dominated previous studies of children's sociolinguistic acquisition (Nardy et al., 2013), this study shines a light on children's acquisition of majority language variation in the context of a multilingual and multidialectal urban community. As mobility increases, sociolinguistic studies of people who have lived in one place all their lives, and/or have grown up acquiring one language or dialect, will become increasingly unrepresentative of the general population (Britain, 2016). Future research is needed to explore the role of input in children's changing linguistic environment, from caregiver to school, in their developing speech repertoires. We also need closer systematic studies of ethnic-specific differences and the role of other factors such as the children's perception of the variability in their environment (e.g., Kaiser, 2022) that drive language change in urban settings. Building a better understanding of the linguistic reality of children and young people growing up in urban communities, and their role in language change in that community, will only improve our models of children's acquisition of sociophonetic variation.

Conflict of interest

The authors have declared no conflict of interest.

CRediT authorship contribution statement

Rosamund Oxbury: Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Matthew Hunt:** Writing – review & editing, Visualization. **Kathleen M. McCarthy:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization.

Acknowledgements

We would like to thank the children and adolescents who took part in this project, and their schoolteachers, youth club leads, and caregivers for their support throughout. Thank you to the four anonymous reviewers for their helpful feedback on earlier versions of this paper. Thanks also goes to Jenny Cheshire and Paul Kerswill for their invaluable input and support throughout the project and write-up of this work. This research was funded by an ESRC LISS-DTP Scholarship, awarded to the first author at Queen Mary University of London.

Appendix A1

Table A1

Average number of tokens per vowel per age group.

		FACE	PRICE	PRICE in like	GOAT
Child	Average	59.28	51.23	7.36	73.57
	Max	109	104	24	129
	Min	29	28	1	29
Adolescent	Average	75.52	60.14	40.9	67.21
	Max	126	122	108	37
	Min	45	28	4	119

Appendix A2

Table A2
Formant extraction parameters.

Parameter	Adolescent boys	Adolescent girls	Children
Timestep	0.0025	0.0025	0.0025
No. of formants	5	4	5.5
Max. formant frequency (Hz)	5000	5500	8000
Window length (s)	0.025	0.015	0.01
Pre-emphasis (dB)	50	50	50

Appendix A3

Further details on the Watt-Fabricius normalisation procedure

The method selected for normalisation in this paper was the modified Watt- Fabricius method (Fabricius, Watt, & Johnson, 2009). The Watt-Fabricius method (or the S-centroid procedure) set out in Watt and Fabricius (2002) is intended to be a speaker- intrinsic method of normalising vowel formant data. It requires: the F1 and F2 of the FLEECE vowel, on the assumption that this vowel represents the top left hand corner of a British English speaker's vowel space; the F1 and F2 of the TRAP vowel, on the assumption that this is the most open vowel in the speaker's system; and a hypothetical close back vowel *u'*, where F1 = the F1 of FLEECE, and F2 also = the F1 of FLEECE. The difference between the Watt-Fabricius method as laid out in Watt and Fabricius (2002) and the modified Watt-Fabricius method as described in Fabricius et al. (2009) is that the latter does not use real measurements for the F2 of TRAP, but rather calculates the F2 of TRAP as being the midpoint between the F2 of FLEECE and the F2 of *u'*. A centroid point for F_n for each speaker is calculated using these three corner vowels:

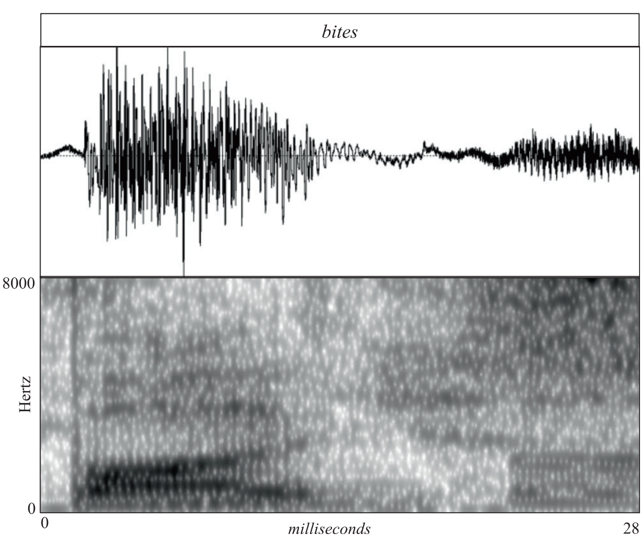
$$S(F_n) = \frac{[\text{fleece}]F_n + [\text{trap}]F_n + [u\prime]F_n}{3}$$

The observed measurements in Hertz of F_n are then divided by $S(F_n)$. The $S(F_n)$ is calculated by speaker, so each speaker's formant measurements are divided by that speaker's centroid measure. Each speaker's centroid value is different, depending on how widely spaced the corners of his/her vowel space are. Dividing the formant measurements by this speaker-specific centroid measure reduces the differences in formant frequencies that arise because of speaker physiology (goal 1), whilst preserving other variation (goal 2).

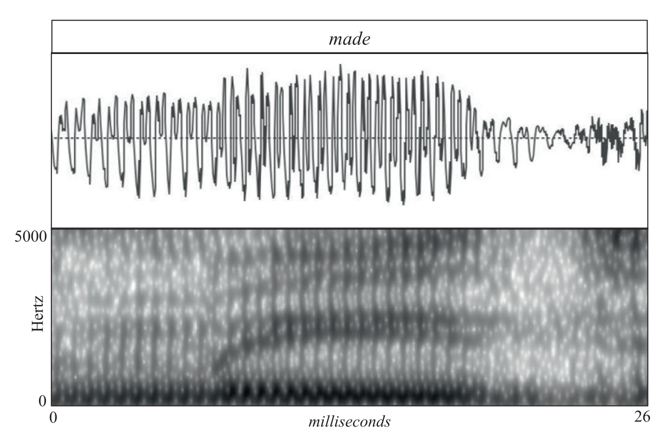
Appendix A4

Example waveform and spectrograms for a child boy (1) and adolescent boy (2) speaker, producing the PRICE (*bites*) and FACE (*made*) vowels.

(1)



(2)



References

Appel, R., & Schoonen, R. (2005). Street language: a multilingual youth register in the Netherlands. *Journal of Multilingual and Multicultural Development, 26*(2), 85–117.

Baker, R., & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods, 43*(3), 761–770.

Boersma, P. (2019). *Praat: doing phonetics by computer*. http://www.praat.org/.

Britain, D. (2016). Sedentarism and nomadism in the sociolinguistics of dialect. In N. Coupland (Ed.), *Sociolinguistics: theoretical debates* (pp. 217–241). Cambridge: Cambridge University Press.

Buerkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28.

Cheshire, J. (2013). Grammaticalisation in social context: the emergence of a new English pronoun. *Journal of Sociolinguistics, 17*(5), 608–633.

Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2011). Contact, the feature pool and the speech community: the emergence of Multicultural London English. *Journal of Sociolinguistics, 15*(2), 151–196.

Cheshire, J., Fox, S., Kerswill, P., & Torgersen, E. (2013). Language contact and language change in the multicultural metropolis. *Revue française de linguistique appliquée, 18*(2), 63–76.

Cheshire, J., & Fox, S. (2016). From sociolinguistic research to English language teaching. *Creating and digitizing language corpora: Volume 3: Databases for public engagement* (pp. 265–290).

Chevrot, J. P., Beaud, L., & Varga, R. (2000). Developmental data on a French sociolinguistic variable: post-consonantal word-final/R. *Language Variation and Change, 12*(3), 295–319.

Chevrot, J. P., & Foulkes, P. (2013). Introduction: language acquisition and sociolinguistic variation. *Linguistics, 51*(2), 251–254.

Clyne, M. (2000). Lingua Franca and ethnolects in Europe and beyond. *Sociolinguistica, 14*, 83–89.

Davies, M. (2004). *BYU-BNC*. Retrieved from https://corpus.byu.edu/bnc/ (Based on the British National Corpus from Oxford University Press).

Díaz-Campos, M. (2005). The emergence of adult-like command of sociolinguistic variables: a study of consonant weakening in Spanish-speaking children. In D. Eddington (Ed.), *Selected proceedings of the 6th conference on the acquisition of Spanish and Portuguese as first and second languages* (pp. 56–65). Somerville, MA: Cascadilla.

Dodd, B. J., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: a normative study of British English-speaking children. *Clinical Linguistics & Phonetics, 17*(8), 617–643.

Drager, K. (2011). Sociophonetic variation and the lemma. *Journal of Phonetics, 39*(4), 694–707.

Drager, K. (2016). Constructing style: phonetic variation in quotative and discourse particle *like*. In H. Pichler (Ed.), *Discourse-pragmatic variation and change in English: new methods and insights* (pp. 232–251). Cambridge: Cambridge University Press.

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). Multipic: a standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71*(4), 808–816.

Drummond, R. (2018a). Maybe it's a grime [t]ing: TH-stopping among urban British youth. *Language in Society, 47*(2), 171–196.

Drummond, R. (2018b). *Researching urban youth language and identity*. Basingstoke: Palgrave Macmillan.

ELAN (version 5.9) [Computer software manual]. (2020). Nijmegen. Retrieved from https://archive.mpi.nl/tla/elan.

Fabricius, A. H., Watt, D., & Johnson, D. E. (2009). A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. *Language Variation and Change, 21*(3), 413–435.

Foulkes, P., & Docherty, G. J. (2006). The social life of phonetics and phonology. *Journal of Phonetics, 34*, 409–438.

Foulkes, P., Docherty, G., & Watt, D. (1999). Tracking the emergence of sociophonetic variation. *Proceedings of the 14th international congress of phonetic sciences*.

Fox, R., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America, 126*(5), 2603–2618.

Fox, S. (2007). *The demise of Cockneys? Language change in London's "traditional" East End* (Unpublished doctoral dissertation). University of Esgender.

Fox, S. (2015). *The new Cockney: new ethnicities and adolescent speech in the traditional East End of London*. New York: Palgrave Macmillan.

Fox, S., Khan, A., & Torgersen, E. (2011). The emergence and diffusion of Multicultural London English. In F. Kern & M. Selting (Eds.), *Ethnic styles of speaking in European metropolitan areas* (pp. 19–44). Amsterdam: John.

Franke, M., & Roettger, T. B. (2019). Bayesian regression modelling (for factorial designs): a tutorial. *PsyArXiv*. https://doi.org/10.31234/osf.io/cdxv3.

Gates, S. M. (2018). Why the long FACE?: Ethnic stratification and variation in the London diphthong system. University of Pennsylvania. *Working Papers Linguistics, 24*(2), 39.

Gates, S. M. (2019). *Language variation and ethnicity in a multicultural East Lon- don secondary school* (Unpublished doctoral dissertation). London: Queen Mary University of London.

Granlund, S. (2015). *Speech communication strategies in older children: acoustic-phonetic and linguistic adaptations to a hearing-impaired peer* (Unpublished doctoral dissertation). London: University College London.

Habib, R. (2014). Vowel variation and reverse acquisition in rural Syrian child and adolescent language. *Language Variation and Change, 26*(1), 45–75.

Holmes-Elliott, S. (2021). Calibrate to innovate: community age vectors and the real time incrementation of language change. *Language in Society, 50*(3), 441–474.

Horvath, B., & Sankoff, D. (1987). Delimiting the Sydney speech community. *Language in Society, 16*(2), 179–204.

Ilbury, C. (2020). *Beyond the offline* (Unpublished doctoral dissertation). London: Queen Mary University of London.

Kaiser, I. (2022). Children's linguistic repertoires across dialect and standard speech: mirroring input or co-constructing sociolinguistic identities? *Language Learning and Development, 18*(1), 41–61.

Kerswill, P., Torgersen, E. N., & Fox, S. (2008). Reversing "drift": innovation and diffusion in the London diphthong system. *Language Variation and Change, 20*(3), 451–491.

Kerswill, P., & Williams, A. (2000). Creating a New Town koine: children and language change in Milton Keynes. *Language in Society, 29*(1), 65–115.

Kerswill, P., & Williams, A. (2005). New towns and koineization: linguistic and social correlates. *Linguistics: An Interdisciplinary Journal of the Language Sciences, 43*(5), 1023–1048.

Kerswill, P., Cheshire, J., Fox, S., & Torgersen, E. (2013). English as a contact language: the role of children and adolescents. In *English as a contact language* (pp. 258–282). Cambridge: Cambridge University Press.

Khan, A. (2006). *A sociolinguistic study of Birmingham English: language variation and change in a multi-ethnic British community* (Unpublished doctoral dissertation). University of Lancaster.

Khattab, G. (2007). Variation in vowel production by English-Arabic bilinguals. In J. Cole & J. L. Hualde (Eds.), *Papers in laboratory phonology IX* (pp. 383–410). Berlin: Mouton de Gruyter.

Khattab, G. (2013). Phonetic convergence and divergence strategies in English-Arabic bilingual children. *Linguistics: An Interdisciplinary Journal of the Language Sciences, 51*(2), 439–472.

Kirkham, S., & McCarthy, K. M. (2021). Acquiring allophonic structure and phonetic detail in a bilingual community: the production of laterals by Sylheti-English bilingual children. *International Journal of Bilingualism, 25*(3), 531–547.

Kirkham, S., & Moore, E. (2013). Adolescence. In J. Chambers & N. Schilling (Eds.), *The handbook of language variation and change* (2nd ed., pp. 277–296). Malden, MA: Wiley-Blackwell.

Kotsinas, U.-B. (1988). Immigrant children's Swedish: a new variety? *Journal of Multilingual and Multicultural Development, 9*(1–2), 129–140.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978–990.

Kushartanti, B. (2015). The acquisition of stylistic variation by Jakarta Indonesian children. *Wacana, Journal of the Humanities of Indonesia, 16*(2), 470–478.

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society, 1*(1), 97–120.

Labov, W. (1994). *Principles of linguistic change (Vol. 1: internal factors)*. Oxford: Blackwell.

Labov, W. (2001a). The anatomy of style-shifting. In P. Eckert & J. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 85–108). Cambridge: Cambridge University Press.

Labov, W. (2001b). *Principles of linguistic change (Vol. 2: social factors)*. Oxford: Blackwell.

Labov, W. (2007). Transmission and diffusion. *Language, 83*(2), 344–387.

Lee-Kim, S.-I., Davidson, L., & Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology, 4*, 475–511.

Mangara, M. (2017). *Ealing JSNA 2017: population characteristics*. Retrieved from https://www.ealing.gov.uk/download/downloads/id/13887/.

Marzo, S., & Ceuleers, E. (2011). The use of Cite ´taal among adolescents in Limburg: the role of space appropriation in language variation and change. *Journal of Multilingual and Multicultural Development, 32*(5), 451–464.

McCarthy, K., Mahon, M., Rosen, S., & Evans, B. G. (2014). Speech perception and production by sequential bilingual children: a longitudinal study of Voice Onset Time acquisition. *Child Development, 85*(5), 1965–1980.

McLeod, S. (2003). General trends and individual differences: perspectives on normal speech development. In S. P. Shovov (Ed.). *Advances in psychology research* (Vol. 22, pp. 189–202). New York, NY: Nova Science.

Nardy, A., Chevrot, J.-P., & Barbu, S. (2013). The acquisition of sociolinguistic variation: Looking back and thinking ahead. *Linguistics, 51*(2), 255–284.

Nassenstein, N., & Hollington, A. (Eds.). (2015). *Youth language practices in Africa and beyond*. Berlin: de Gruyter.

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: foundational ideas – Part II. *Language and Linguistics Compass, 10*(11), 591–613.

Opsahl, T. (2009). "Wolla I swear" this is typical for the conversational style of adolescents in multiethnic areas in Oslo. *Nordic Journal of Linguistics, 32*(2), 221–244.

Podesva, R. J., D'Onofrio, A., Van Hofwegen, J., & Kim, S. K. (2015). Country ideology and the California Vowel Shift. *Language Variation and Change, 27*(2), 157–186.

Quist, P. (2005). *Stilistike praksisser i storbyens heterogene skoleen etnografisk og so-ciolingvistik undersøgelse af sproglig variation [Stylistic practices in the heterogeneous school in the big city – an ethnographic and sociolinguistic study of linguistic variation]* (Unpublished doctoral dissertation). Copenhagen: University of Copenhagen.

R Core Team (2020). *R: A language and environment for statistical computing [Computer software manual]*. Vienna, Austria. Retrieved from https://www.R-project.org/.

Rampton, B. (1995). *Crossing: language and ethnicity among adolescents*. Routledge.

Roberts, J. (1997). Hitting a moving target: acquisition of sound change in progress by Philadelphia children. *Language Variation and Change, 9*(2), 249–266.

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., & Yuan, J. (2014). *Fave (forced alignment and vowel extraction) program suite v1.2.2.* Retrieved from https://github.com/JoFrhwld/FAVE https://doi.org/10.5281/zenodo.22281.

Schleef, E., & Turton, D. (2018). Sociophonetic variation of *like* in British dialects: effects of function, context and predictability. *English Language and Linguistics, 22*(1), 35–757.

Smith, J., & Durham, M. (2019). *Sociolinguistic variation in children's language: acquiring community norms*. Cambridge: Cambridge University Press.

Smith, J. (2021). Child language acquisition and sociolinguistic variation. *Sociolinguistic Variation and Language Acquisition across the Lifespan*, 11–20.

Smith, J., Durham, M., & Fortune, L. (2007). "Mam, my trousers is fa'in doon!": community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change, 19*(1), 63–99.

Smith, J., Durham, M., & Fortune, L. (2009). Universal and dialect-specific pathways of acquisition: Caregivers, children, and t/d deletion. *Language Variation and Change, 21*(1), 69–95.

Smith, J., Durham, M., & Richards, H. (2013). The social and linguistic in the acquisition of sociolinguistic norms: caregivers, children, and variation. *Linguistics: An Interdisciplinary Journal of the Language Sciences, 51*(2), 285–324.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology. Human Learning and Memory, 6*(2), 174–215.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods, 38*(4), 598–605.

Stanford, J. N. (2008). Child dialect acquisition: new perspectives on parent/peer influence. *Journal of Sociolinguistics, 12*(5), 567–596.

Tagliamonte, S. (2016). *Teen talk: the language of adolescents*. Cambridge University Press.

Tanner, J., Sonderegger, M., & Stuart-Smith, J. (2019). Vowel duration and the voicing effect across dialects of English. *Toronto Working Papers Linguistics, 41*(1), 1–13.

Thomas, E. R. (2011). *Sociophonetics: an introduction*. Basingstoke, Hampshire; New York: Palgrave Macmillan.

Watt, D., & Fabricius, A. (2002). Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1–F2 plane. *Leeds Working Papers in Linguistics and Phonetics, 9*, 159–173.

Wells, J. C. (1982). *Accents of English (Vol. 2: The British Isles)*. Cambridge: Cambridge University Press.

Wiese, H. (2009). Grammatical innovation in multiethnic urban Europe: new linguistic practices among adolescents. *Lingua, 119*(5), 782–806.

Wiese, H. (2013). What can new urban dialects tell us about internal language dynamics? The power of language diversity. *Linguistische Berichte, 19*, 207–245.

Wormald, J. (2015). Dynamic variation in 'Panjabi-English': analysis of F1 and F2 trajectories for FACE and GOAT. *Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK*.

Young, N. J. (2021). The sociolectal and stylistic variability of rhythm in Stockholm. *Language & Speech (ETAP4 Special Issue on Sociolinguistic Variation in Prosody)*, 1–37.