Cross-generational phonetic alignment between mothers and their children

Thomas St. Pierre[1], Angela Cooper[1] and Elizabeth K. Johnson[1,2]

[1]*Department of Psychology, University of Toronto Mississauga*

[2] *Department of Psychology, University of Toronto*

Our work was supported by grants awarded to the third author from SSHRC (Social Sciences and Humanities Research Council of Canada) and NSERC (Natural Sciences and Engineering Research Council of Canada). Correspondence concerning this article should be addressed to Thomas St. Pierre, Department of Psychology, University of Toronto, Mississauga, ON, L5L 1C6, Canada. Email: thomas.stpierre@utoronto.ca

Abstract

Over time, people who spend a lot of time together (e.g., roommates) begin sounding alike (Pardo et al., 2012). Even over the course of short conversations, interlocutors often become more acoustically similar to one another (Pardo et al., 2017). This phenomenon—known as phonetic alignment—has been well studied in adult interactions, but much less is known about alignment patterns in intergenerational, adult-child dyads. In the current study, we investigated alignment between mothers and their children in a picture-naming task, as assessed using a perceptual similarity test and acoustic measures. Experiments 1 and 2 examined alignment in 2.5- and 4-year-old children and their mothers, both when mothers shadowed their children (Experiment 1), and when children shadowed their mothers (Experiment 2). Experiments 3 and 4 investigated long-term similarity between mothers and children when they were recorded separately. Results show that children and mothers aligned to one another in the shadowing task, regardless of who shadowed whom, and while there was no evidence for long-term alignment in younger children, there *was* some evidence of long-term alignment with 8-year-old children and their moms, but only for male children.

Keywords: speech production, speech perception, toddlers, phonetic alignment, developmental sociolinguistics

Cross-generational phonetic alignment between mothers and their children

During communication, interlocutors frequently align their speech to one another on a variety of linguistic dimensions (Pickering & Garrod, 2004). For example, speakers often repeat recently heard syntactic constructions (Bock, 1986), converge upon the same labels for objects in discourse (Brennan & Clark, 1996; Garrod & Anderson, 1987), and will even adjust their own productions to more phonetically match their interlocutor's speech patterns (e.g., Pardo, 2006). Many of these forms of linguistic alignment have additionally been found in children, who, like adults, are also more likely to produce recently heard labels (Branigan et al., 2016) and syntactic constructions (Shimpi et al., 2007) over equally plausible, unmentioned alternatives.

One type of alignment, phonetic alignment, has remained largely unexplored in the developmental literature beyond infancy, despite being heavily implicated in theories/debates about how speech sounds and adult-like productions are successfully acquired by children (Messum & Howard, 2015). In the present paper, we investigated the extent to which young children phonetically align to their mothers and vice versa, shedding light on possible social factors influencing gender differences in alignment across child development.

Phonetic alignment in adults

In adults, phonetic alignment—also referred to as phonetic convergence, imitation, entrainment, or accommodation—has been well-documented both during dyadic social interactions (Kim et al., 2011; Lewandowski & Jilka, 2019; Pardo, 2006) as well as within non-interactive shadowing tasks (e.g., Babel & Bulatov, 2012; Clopper & Dossey, 2020; Goldinger, 1998; Lewandowski & Nygaard, 2018; Namy et al., 2002; Pardo et al., 2013; Pardo et al., 2017). During phonetic alignment, a talker's speech pattern becomes temporarily more similar to another speaker's along such dimensions as duration (e.g., Pardo et al., 2017), fundamental frequency (f0; e.g., Babel & Bulatov, 2012; Garnier et al., 2013), speech rate (Street, 1983), spectral characteristics (e.g., Babel, 2010, 2012; Dufour & Nguyen, 2013; Pardo et al., 2017), and voice-onset time (e.g., Nielsen, 2014; Olmstead et al., 2013), as measured by acoustic analysis and listener similarity judgments (Schertz et al., 2019).

By some theoretical accounts, phonetic alignment arises from low-level links between the perceptual and production systems (e.g., Pickering & Garrod, 2004, 2013; Shockley et al., 2004). During the process of speech perception, for example, listeners may recruit resources involved in production (e.g., activation of motor sequences, perception of vocal tract activity, etc.), causing listeners' own productions to resemble those of their interlocutor. In this way, phonetic alignment occurs naturally and automatically in conversation, with newly-formed perceptual representations (created to understand incoming speech) guiding subsequent productions.

Although there is support for low-level accounts, with alignment operating as a largely automatic process, there are some findings that suggest that alignment is not necessarily an automatic, inevitable process. For example, numerous social factors have been found to modulate the degree of phonetic alignment, including the gender (Namy et al., 2002), accent (Babel, 2010; Walker & Campbell-Kibler, 2015), social status (Gregory & Webster, 1996), and attractiveness (Babel, 2012) of the talker. Overall, positive attitudes towards an interlocutor have been found to increase phonetic alignment; in one study, for example, New Zealand participants were more likely to align to an Australian speaker's productions if they had pre-existing positive views towards Australia, as measured by an Implicit Association Task (Babel, 2010). In some cases, listeners may even diverge from interlocutors, making their speech more dissimilar from other talkers (e.g., Bourhis & Giles, 1977; Kim et al., 2011). Little work has addressed the influence of age as a social factor in alignment, particularly in inter-generational (e.g., caregiver-child) contexts.

Research has additionally looked at accommodation over much larger time scales, when accents over time shift to resemble community norms. This has been observed in adults adjusting to pronunciation changes within their local communities (Harrington, 2007; Sankoff & Blondeau, 2007) and more commonly when individuals move to another geographic area exhibiting different dialect features (e.g., Evans & Iverson, 2007; Harrington, 2007; Sancier & Fowler, 1997; Shockey, 1984; see Nycz, 2015 for a review of second dialect acquisition). For example, in one study, Canadians living in Alabama were rated as sounding "more American" relative to Canadians living in Canada (Munro et al., 1999), suggesting that the Canadians in Alabama had become more aligned with the locally dominant speech patterns. Other studies have even found some evidence of long-term alignment to individual speakers, for example, in randomly assigned college roommates (Pardo et al., 2012) or reality TV stars confined to a house for months at a time (in this case, one pair of individuals who developed an intimate relationship across the show; Sonderregger et al., 2017).

Taken together, there is robust evidence that adults dynamically vary their speech patterns in response to their language input; that is, there is a tendency for individuals to phonetically align their speech patterns to more closely approximate other talkers' speech patterns, both ephemerally in moment-to-moment conversations as well as over longer stretches of time in response to individuals or community norms. Much less is understood about alignment in cross-generational interactions, particularly those between caregivers and their children, and to our knowledge, no study has investigated social factors that might influence the degree of alignment between caregivers and children (e.g., gender), as has been observed in some studies with adults (e.g., Namy et al., 2002; Pardo, 2006).

Alignment in Children

It is taken for granted that children exhibit long-term alignment; in the process of acquiring a first language, children must converge on the properties of the language in their ambient environment to become native speakers. Even in cases where children move to a new community *after* acquiring a particular dialect, they are still able to quickly learn the new regional accent, arguably to a greater extent than adults (Chambers, 1992; Payne, 1976; Tagliomonte & Molfenter, 2007). In addition, there is evidence that children—like adults—may also exhibit long-term alignment to individual speakers, observed most commonly in situations where children receive only limited or idiosyncratic exposure to a language from a small number of speakers (Khattab, 2003; Klinger, 1962). For example, in one study, bilingual Dutch-German children living in the Netherlands exhibited VOT lengths more like their L1 German mothers—whose German VOTs had changed through increased exposure to Dutch—compared to monolingual children living in Germany, suggesting that children's German productions had aligned with those of their mothers (Stoehr et al, 2019).

Studies looking at children's short-term alignment in real time (e.g., in dyadic interactions or in a lab setting) have produced more mixed results. While some studies have found evidence of children aligning to adult speech along a variety of dimensions such as pitch (Ko et al., 2016), speaking rate (Eaton & Ratner, 2013), response latency (Welkowitz et al., 1976), utterance duration (Street, 1983), amplitude (Oviatt et al., 2004), and VOT (Nielsen, 2014; Stoehr et al., 2019), other studies have found no evidence of alignment (Garvey & BenDebba, 1974; Paquette-Smith et al., 2021; Seidl et al., 2018; Wynn et al., 2018; Wynn et al., 2019), making it difficult to draw general conclusions about when, how, and under what circumstances children and adults align their speech to one another.

Most of these studies examining short-term alignment in adult-child dyads have tended to examine just a subset of attributes of interest (e.g., speech rate), which may have left unnoticed other dimensions where alignment *was* occurring. Indeed, there is often a large degree of variability in how interlocutors align with one another, with talkers exhibiting “a unique profile of convergence and divergence on multiple dimensions” (p. 653, Pardo et al., 2017). In addition, in many of the studies, which employ more naturalistic, interactional paradigms, it is often difficult to determine which interlocutor is aligning to whom, as highlighted by several researchers (Kuhl & Meltzoff, 1996; Legerstee, 1990; Papoušek & Papoušek, 1989). For the most part, in studies of prelinguistic children, parents appear to be performing the bulk of the alignment work, with little evidence of infants vocally imitating their parents (Kokkinaki & Kugiumutzakis, 2000; McRoberts & Best, 1997; Pawlby, 1977; Siegel et al., 1990). In slightly older, post-verbal children, however, it is much less clear what the contributions of adults and children in the alignment process actually are, and whether/how that changes across development.

In the current study, we investigated cross-generational, phonetic alignment in child-mother dyads, examining the extent of alignment of children and their mothers in a picture-naming task. Unlike previous literature, which only measured alignment of child-adult dyads along specific dimensions, we additionally assessed phonetic alignment more holistically using an AXB perceptual similarity task, a method frequently used in the adult literature to measure similarity across multiple features simultaneously, but never—to our knowledge—in studies with children (e.g., Babel et al., 2014; Goldinger, 1998; Pardo et al., 2017; Shockley et al., 2004).

This allowed us to investigate the presence or absence of intergenerational alignment without a priori ideas of which acoustic dimensions dyads would align to.

Because social factors have been found to play a role in phonetic alignment in adults (e.g., Babel, 2010), we additionally investigated possible influences of gender, recording equal numbers of mother-daughter and mother-son pairs in each experiment. In the adult literature, results are mixed with respect to how gender influences alignment, with some studies showing greater alignment in women than men (Namy et al., 2002), other studies showing the opposite pattern (i.e., more alignment in men; Pardo, 2006; Pardo et al., 2010), and still other studies finding no evidence for gender-based differences in alignment (e.g.,Pardo et al., 2017). Few studies have investigated social factors influencing alignment in child-adult communication beyond infancy (Paquette-Smith et al., 2021), and since gender is one of the earliest social categories to develop in children (Hines, 2015), we speculated that it might influence the extent to which children and caregivers align. Starting from an early age, research suggests that the way parents speak with their children depends on the assigned gender of their children (Foulkes et al., 2005; Karrass et al., 2002; Kitamura et al., 2002), and from a relatively early age, children may start picking up on gender-based patterns of speech (see Nardy et al., 2013 and Smith & Durham, 2019 for reviews). In fact, recent work has found that children can be perceptibly categorized into their assigned gender as early as 2.5-years-old (Fung et al., 2021). Given that children as young as 2.5-years-old display gender differences in speech, we might expect gender to play a role in mother-child alignment, with boys perhaps aligning less with their mothers than girls.

In Experiments 1 and 2, we examined inter-generational alignment in 2.5- to 4-year-old children and their mothers, including the directionality of alignment, both when mother's shadowed their children in the picture-naming task (Experiment 1), and when the same children

(15 months later) shadowed their mothers (Experiment 2). Experiment 3 investigated whether the alignment observed in Experiments 1 and 2 was due primarily to short-term alignment occurring as a result of the shadowing task or long-term alignment as a result of overall similarities between children and their mothers via prolonged exposure. Finally, in Experiment 4, we looked at alignment in more linguistically developed 8-year-old children and their mothers.

Experiment 1

There is some evidence that parents imitate the vocalizations of their infants, adjusting the pitch of their voices to match their children (McRoberts & Best, 1997; Papoušek & Papoušek, 1989). Most work looking at alignment beyond infancy tends to focus on pre-school aged children or older, with much less work looking at alignment to linguistic utterances of children under the age of three (but see Ko et al., 2016). In Experiment 1, we focus on alignment of mothers to their 2.5-year-old toddler's productions of individual words elicited in a picture-naming task, as measured primarily by an AXB perceptual similarity task, in which adult participants rated whether children's speech productions sounded more similar to their mother's compared to other mothers' productions.

Method

Participants

Twenty-four Canadian English adult listeners (18 women, M_{age} = 27.5 years) participated in this experiment. All listeners learned English before the age of 6 and were students at the University of Toronto Mississauga. These listeners reported currently using English at least 70% of the time. They all self-reported no hearing or vision impairments at the time of testing.

Materials and Procedure

Based on piloting data, we compiled a list of 32 familiar mono- and poly-syllabic words (e.g., *house*, *strawberry, butterfly*) which we could reliably get young children to produce spontaneously (see Appendix for full list of words). We recorded 20 monolingual Canadian-English learning 2.5- to 3-year-old toddlers (10 boys, M_{age} = 33 months, range = 30-35 months) and their Canadian-English speaking mothers saying these words in a sound-attenuated booth. These recordings were elicited using an experimenter-controlled computer game, asking toddlers to teach a Martian labels for common objects. At the beginning of the game, a cartoon alien appeared on the screen, and asked children to help them learn the names of objects on Earth. During elicitation, a clipart image of a referent of a target word (e.g., the picture of a house) was displayed on the screen, and the child was prompted to name the picture (*This is a …*), followed by the child's mother, who was instructed to produce the same word immediately after their child. Children producing the target word with a determiner (e.g., *a ball*) or some other modifier (e.g., *beach ball* instead of *ball*) were prompted to repeat the word without the accompanying morphemes (e.g., *Can you say it without "beach"*). After each trial, a spaceship moved progressively closer to a planet, indicating the child's progress, with checkpoints along the way where children received stickers.

From these recordings—which were normalized for root mean square amplitude—we constructed AXB trials to be used for the experimental task, with each trial consisting of three identical word tokens placed in succession and separated by 500 milliseconds of silence. Each middle X token contained a child's production, and the surrounding A and B tokens came from two mothers: One was from the child's own mother, who had shadowed the child during the recording session, and the other was from another mother, who had shadowed her own (another) child (e.g., *strawberry*$_{\text{MotherX}}$ + *strawberry*$_{\text{ChildX}}$ + *strawberry*$_{\text{MotherY}}$). Although A and B tokens often consist of productions from the same talker, with one production recorded before the shadowing task (the baseline token) and the other during the shadowing task (e.g., Goldinger, 1998), we opted to use two separate speakers instead, given the difficulties of trying to separate 2-year-olds from their mothers to obtain baseline recordings (see Miller et al., 2013, who obtained similar results using this kind of paradigm).

Four well-formed productions from each dyad (i.e., productions that were not shouted or whispered, or contained external noises like tapping, etc.) were randomly selected for testing. The mother of each child (Own Mother) was randomly paired with four different mothers (Different Mother) across these four words. Order of presentation was counterbalanced, such that Own Mother and Different Mother occurred in both first (A) and last (B) position. Each listener heard 8 trials per child (Own Mother paired with 4 Different Mothers x 2 orders), for a total of 160 trials (8 trials x 20 children). Due to a programming error, responses for two trials from one child were not properly recorded in the output, resulting in 158 trials for each participant.

In the AXB perceptual similarity task, participants were instructed to assess whether the first or last item (A or B) sounded most like the middle item (X). They made their selection from two buttons labelled A and B presented on the screen. Listeners completed a total of 160 randomized trials.

Results and Discussion

Regardless of the gender of the child, participants selected children's own mothers as perceptually more similar at above chance rates ($M_{female} = 0.55$, $M_{male} = 0.55$), suggesting that mothers' productions were indeed aligned to those of their children. Using the *glmer* function in the lme4 package in R (Bates et al., 2015), this was tested in a logistic mixed-effects regression model predicting the log odds of selecting children's Own Mother, with Child Gender (Male, coded -0.5, and Female, 0.5) included as a fixed effect. The model—and all subsequent models reported in the study—included the maximum random-effects structure that would converge (Barr et al., 2013), in this case random intercepts for items (each child/word pair) and participants.[1] While no significant effect of Child Gender was found ($\beta = -0.005$, $SE = 0.17$, $z = -0.03$, $p = 0.96$), there was a significant intercept term ($\beta = 0.20$, $SE = 0.09$, $z = 2.32$, $p = 0.02$), indicating that overall the perceived similarity between mothers and their children was significantly above chance.

We additionally examined whether perceived similarity between children and mothers correlated with measures of acoustic similarity. We selected three commonly used dimensions that can reliably be measured in 2.5-year-olds: word duration, mean f0, and f0 range (maximum f0 value minus the minimum f0 value), all of which were extracted/ for each mother and child token using Praat. Given the difficulties of estimating f0 in children's speech using the same static parameters across tokens, we manually inspected each token individually and chose optimal pitch parameters for obtaining accurate f0 values, with 100-500 Hz used the default starting range; duration was obtained by extracting the length of each segmented token. Periods of creak within a token, which were determined auditorily and through visual inspection of spectrographs, were not considered in any of the f0 measures, and any tokens in which either children, their mothers, or other mothers exhibited creak across the entire duration of a syllable or could not be accurately measured (e.g., pitch contour was highly irregular or discontinuous) were excluded. Since the study was primarily interested in perceptual measures of alignment, we did not choose an exhaustive list of acoustic measures, and opted to exclude measurements

[1] glmer(OwnMother ~ ChildGender + (1 | Participant) + (1 | Item))

which are particularly difficult to extract in young children's speech (e.g., vowel formants; Kent, 1976; Story & Bunton, 2016) or are not measurable in all words (e.g., VOT; Schertz et al., 2019).

To test for the presence of alignment/similarity, we first examined whether mothers' mean f0 and f0 range values correlated with those of their children. If mothers were aligning to their children along these dimensions, we might expect higher values in children's productions to lead to higher values in mothers' shadowed productions. To test this, we conducted two linear mixed-effects models (using the *lmer* function from the lme4 package), one for each prosodic trait, in which mothers' f0 and f0 Range values were predicted from their children's values (mean-centered).[2] Child Gender (coded Male, -0.5., Female, 0.5), and its interaction with children's values were also included as fixed effects, along with random by-dyad and by-item intercepts. Five items were excluded due to the presence of creak (4) or an inability to accurately measure f0 (1), and the significance of each term was evaluated using the lmerTest package in R (Kuznetsova et al., 2017). Unsurprisingly—given the lack of a Gender effect in the perceptual similarity task—there were no interactions with Child Gender, both in the f0 model (β = -0.10, SE = 0.15, t = -0.68, p = 0.50) and in the f0 Range model (β = 0.13, SE = 0.21, t = 0.64, p = 0.53). There were, however, significant main effects of Child f0 (β = 0.29, SE = 0.07, t = 3.97, p < .001) and Child f0 Range (β = 0.36, SE = 0.10, t = 3.42, p < .01), such that larger values in children's productions led to higher values in mothers' shadowed responses (see Figure 1), suggesting that mothers' productions were aligned to their children's along these prosodic dimensions. Note that no such effects were found in similar models examining the relationship between children's tokens and those of other mothers (i.e., the other mother tokens that were

[2] lmer(Motherf0/Motherf0Range ~ Childf0/Childf0Range * ChildGender + (1 | Dyad) + (1 | Item))

paired with the child tokens in the AXB task; $\beta_{Childf0}$ = -0.06, *SE* = 0.11, *t* = -0.57, *p* = .57 and $\beta_{Childf0Range}$ = 0.15, *SE* = 0.13, *t* = 1.19, *p* = .24).

[FIGURE 1 NEAR HERE]

We next tested whether listeners relied (at least partly) on similarities in f0, f0 Range, and Duration when choosing which of the mother tokens in the AXB task most resembled the child token. For each AXB trial, differences were calculated for each acoustic measure between Own Mother and Child and between Different Mother and Child. These differences were then used to compute difference-in-distance (DID) scores for each trial (Pardo et al., 2017), subtracting the absolute values of the Own Mother/Child differences from the absolute values of Different Mother/Child differences (DID = Different Mother distance – Own Mother distance). This yielded a range of difference scores above and below zero, with positive values denoting greater acoustic similarity between children and their own mothers in a particular trial, and negative values denoting greater similarity to different mothers. If listeners were sensitive to similarities along these dimensions, then more positive DID scores were expected to lead to increased selection of Own Mothers as more similar than Other Mothers (and vice versa for negative DID scores).

To determine whether acoustic similarity—as measured by these difference scores—correlated with perceived similarity in the AXB task, a logistic mixed-effects model was created, predicting the selection of Own Mother versus Different Mother in the AXB task from *z*-score transformed DID scores for word Duration, f0, and f0 Range. Child Gender was additionally included as a fixed factor, along with random intercepts for subjects and items (child/word pairs), as well as uncorrelated by-subject slopes for f0, f0 Range, and Duration.[3] Twelve items were excluded due to creak (10) and an inability to get accurate measurements (2) (288/3792 trials). The model revealed significant main effects of DID duration ($\beta = 0.36$, $SE = 0.07$, $z = 5.23.$, $p < .001$), DID f0 ($\beta = 0.33$, $SE = 0.08$, $z = 4.23$, $p < .001$), and DID f0 Range ($\beta = 0.20$, $SE = 0.07$, $z = 2.77$, $p < .01$), with greater acoustic similarity between children and their mothers (relative to other mothers) in all three dimensions leading to increases in the how likely participants were to select Own Mothers as more similar to children's productions than Different Mothers (see Figure 2). In other words, listeners appeared to make use of at least these three acoustic properties (or other properties that covary with these dimensions) when determining which mothers' productions were more similar to the children's.

[FIGURE 2 NEAR HERE]

Altogether, Experiment 1 provides evidence that mothers were aligned to the productions of their toddlers in a shadowing task, as assessed by holistic pronunciation judgments on an AXB similarity task as well as acoustic measures. This alignment was not found to be modulated by Child Gender; mothers of both female and male children were considered to sound more similar to their child than different mothers. Moreover, these judgments appeared to be driven in part by child-mother similarity on multiple acoustic attributes, with greater similarity in duration, f0 and f0 Range leading to greater accuracy in perceiving mothers as sounding similar to their children.

These results are consistent with research investigating alignment with prelinguistic infants, which has found that caregivers adjust their vocalizations to match those of their children (McRoberts & Best, 1997; Papoušek & Papoušek, 1989). In Experiment 1, we showed that

[3] glmer(OwnMother ~ zDIDf0 + zDIDf0Range + zDIDDuration + ChildGender + (zDIDf0 + zDIDf0Range + zDIDDuration || Participant) + (1 | Item))

caregivers also appear to align with their children's productions at a later age, after they have begun producing linguistic utterances. In Experiment 2, we investigated alignment in the same child-mother dyads approximately 15 months after the recordings in Experiment 1, additionally testing whether we also find evidence for alignment in situations where toddlers shadow their mothers.

Experiment 2

In Experiment 2, we sought to replicate the findings from Experiment 1 in older children, testing whether mothers align to their children's productions while shadowing them in a picture-naming task. In addition, we also tested whether toddlers too show evidence of phonetic alignment while shadowing their mothers in the same task, and whether we find the same degree of alignment in each shadowing condition.

Method

Participants

Fifty Canadian English listeners (34 women, M_{age} = 19.6 years, range = 17-35 years), who satisfied the same criteria as Experiment 1, participated in this study. Participants were randomly assigned to either the Mother 1st condition (n=25) or the Child 1st condition (n=25). One participant was removed from the Mother 1st condition for selecting the B response 99% of the time, yielding 24 participants in the Mother 1st condition.

Materials and Procedure

As in Experiment 1, recordings of the same 32 highly familiar mono- and poly-syllabic words (e.g., *house*, *strawberry, butterfly*) were used as stimuli. These recordings were obtained from 18 of the 20 child-mother dyads from Experiment 1, who returned to the lab approximately 15 months after their original recording session (two additional child-mother dyads were recruited to replace the missing dyads, resulting in 20 child-mother dyads; 10 boys, M_{age}= 48 months, range = 46-50 months). Using the same picture-naming paradigm as in Experiment 1, each session consisted of 2 recordings, one with the child naming each picture first followed by the mother, as in Experiment 1 (Child 1^{st}), and another with the mother naming each picture first followed by the child (Mother 1^{st}). The Child 1^{st} recordings always occurred first, where children were asked to label a series of pictures for a Martian, followed by the Mother 1^{st} recordings, in which children were then instructed to repeat the words produced by their mothers, who were responsible for teaching the same words to another Martian.

The recordings were spliced into individual files, which were then used to construct AXB trials with the same structure as Experiment 1 (e.g., *strawberry*$_{\text{MotherX}}$ + *strawberry*$_{\text{ChildX}}$ + *strawberry*$_{\text{MotherY}}$). The same 4 words that were used from each dyad in Experiment 1 were also used to construct AXB trials for both the Child 1^{st} and Mother 1^{st} condition. The procedure of Experiment 2 was identical to Experiment 1, with participants completing either the Child 1^{st} condition (where all tokens were taken from the Child 1^{st} recording format) or the Mother 1^{st} condition (where all tokens were taken from the Mother 1^{st} recording format). Since we wanted to keep the nature of the task consistent for listeners across both the Child 1^{st} and Mother 1^{st} conditions, we opted to present children's productions in the X position in both conditions, having no reason to believe our results would differ based on whether shadowed or model tokens occurred in the X position (see Miller et al., 2013).

Results and Discussion

Observationally, mothers and daughters appeared to phonetically align to one another to the same degree no matter who shadowed whom ($M_{Child1st}$ = 0.59, $M_{Mother1st}$ = 0.60). In contrast, in mother-son dyads, alignment seems to be stronger in cases where male toddlers shadowed their mothers and not vice versa ($M_{Child1st}$ = 0.53, , $M_{Mother1st}$ = 0.57).This was tested in a logistic linear mixed-effects regression model, with selection of Own Mother versus Different Mother as the dependent variable and Child Gender (Male, -0.5, Female, 0.5) and Recording Condition (Child 1^{st}, -0.5, Mother 1^{st}, 0.5)—along with their interaction—as the independent variables. Intercepts were allowed to vary by-subject and -item (child/word pairs), as well as the slope for Recording Condition by item.[4] The model revealed a significant intercept term (β = 0.32, *SE* = 0.07, *z* = 4.49, $p < .001$), indicating that overall, across all conditions, participants were more likely to rate children as perceptually more similar to their mothers than other mothers. This overall alignment did not differ significantly by any of the conditions, however, as suggested by the non-significant coefficients of Child Gender (β = 0.21, *SE* = 0.14, *z* = 1.46, *p* = .14), Recording Condition (β = 0.10, *SE* = 0.10, *z* = 1.04, *p* = .30), and their interaction (β = -0.21, *SE* = 0.19, *z* = 1.10, *p* = .27). Thus, while mothers were numerically less likely to align to their sons compared to mothers of daughters, this difference was not significant.

As in Experiment 1, we additionally investigated the extent to which a model talker's mean f0 and f0 Range values predicted their shadower's corresponding values. We ran four mixed-effects linear regression models, two for the Mother 1^{st} condition and two for the Child 1^{st} condition, in which the mean f0 values and f0 Range values from the model talkers (centered) were used to predict the corresponding values from shadowers. Child Gender (Male, -0.5,

[4] glmer(OwnMother ~ ChildGender * RecordingCondition + (1 | Participant) + (RecordingCondition | Item))

Female, 0.5) and its interaction with the acoustic predictor (mean f0 or f0 Range) were also included, as well as random by-dyad and by-item intercepts.[5] Six of the 80 items were excluded from the Mother 1st analysis due to creak, and seven from the Child 1st analysis. As shown in Figure 3, while both f0 and f0 Range values from model speakers positively correlated with corresponding values from shadowers, unlike in Experiment 1, these trends did not significantly differ from zero, both when children shadowed their mothers ($\beta_{f0} = 0.15$, $SE = 0.16$, $t = 0.91$, $p = .37$ and $\beta_{f0Range} = 0.16$, $SE = 0.18$, $t = 0.93$, $p = .36$), and when mothers shadowed their children ($\beta_{f0} = 0.15$, $SE = 0.08$, $t = 1.99$, $p = .052$ and $\beta_{f0Range} = 0.06$, $SE = 0.07$, $t = 0.83$, $p = .41$). Thus, while caregivers have been shown to align to their children prosodically in infancy (e.g., McRoberts & Best, 1997) and throughout toddlerhood (Experiment 1), it appears that over time, as children develop more and more language proficiency, the degree to which children and parents prosodically align decreases (perhaps because there is more alignment across a wider array of dimensions, e.g., F1, F2, VOT, etc.).

[FIGURE 3 NEAR HERE]

Although children and their mothers did not significantly align to one another along the two prosodic variables of interest, there is evidence that listeners' selections were nevertheless guided (in part) by similarities along these dimensions (or unmeasured, covarying dimensions), as revealed by analyses investigating the influence of similarity in word Duration, f0, and f0 Range on listeners' performance in the AXB task. Two mixed-effects logistic regressions (one for each shadowing condition) were used to predict the selection of Own Mother versus Different

[5] lmer(Motherf0/Motherf0Range ~ Childf0/Childf0Range * ChildGender + (1 | Dyad) + (1 | Item)) for the Child 1st condition and lmer(Childf0/Childf0Range ~ Motherf0/Motherf0Range * ChildGender + (1 | Dyad) + (1 | Item)) for the Mother 1st condition

Mother from *z*-score transformed DID scores for word Duration, f0, f0 Range, and Child Gender.[6] Due to the presence of creak, 17 out of the 80 items were excluded in the Mother 1st analysis and 18 in the Child 1st analysis. In the Mother 1st model—which included random intercepts for items along with random by-subject intercepts and slopes for f0 and f0 Range (all uncorrelated)—there were significant main effects of f0 ($\beta = 0.17$, $SE = 0.07$, $z = 2.33$, $p = .02$) and f0 Range ($\beta = 0.21$, $SE = 0.07$, $z = 2.82$, $p < .01$), but no effect of Duration ($\beta = 0.12$, $SE = 0.14$, $z = 0.53$, $p = .60$). The Child 1st model, which included random by-item and by-subject intercepts, as well as random by-subject slopes for f0 and f0 Range (all uncorrelated), similarly revealed a significant main effect of f0 ($\beta = 0.31$, $SE = 0.11$, $z = 2.91$, $p < .01$), but unlike the Mother 1st data, the effect of Duration ($\beta = 0.23$, $SE = 0.10$, $z = 2.23$, $p < .05$) was significant while the effect of f0 Range was not ($\beta = 0.18$, $SE = 0.10$, $z = 1.71$, $p = .09$).

Age-related changes in perceived similarity.

Because the same recording format and most of the same child-mother dyads were used in Experiments 1 and 2 (Child 1st condition), we were additionally interested in whether the magnitude of perceived similarity changed over time from the initial recording (~2.5- to 3-years-old) to the second recording approximately 15 months later. Using only trials including mothers and children from the 18-dyads that completed both recordings (128/180 trials per participant), a mixed-effects logistic regression model was constructed, with selection of Own Mother versus Different Mother as the dependent variable and Child Gender (Male, -0.5, Female, 0.5) and Recording Time Point (First, -0.5, Second, 0.5) included as fixed effects, along with their

[6] glmer(OwnMother ~ zDIDf0 + zDIDf0Range + zDIDDuration + ChildGender + (zDIDf0 + zDIDf0Range || Participant) + (1 | Item))

interaction. Random by-subject and by-item intercepts as well as a random by-item slope for Recording Time were also included.[7]

As expected, the model revealed a significant intercept term (β = 0.25, *SE* = 0.09, *z* = 2.88, $p < .01$), indicating that overall mothers and their children were aligned in both recording sessions ($M_{1stRecordingFemale}$ = 0.57, $M_{1stRecordingMale}$ = 0.53, $M_{2ndRecordingFemale}$ = 0.60, $M_{2ndRecordingMale}$ = 0.54 **)**. While there appears to be an increase in perceived similarity from the 1st to 2nd recording session in mother-daughter but not mother-son dyads, the interaction between Gender and Recording Condition was not significant (β = 0.12, *SE* = 0.25, *z* = 0.47, *p* = .64), suggesting that the level of alignment between the two recording sessions did not differ by Child Gender. And while there was numerically more alignment in Recording Session 2 compared to Session 1, as well as more alignment in mother-daughter compared to mother-son dyads, these main effects were also not significant (β_{Gender} = 0.18, *SE* = 0.18, *z* = 1.04, *p* = .30 and $\beta_{Session}$ = 0.09, *SE* = 0.12, *z* = 0.74, *p* = .46). This consistent level of perceived alignment between Recording 1 and 2 is particularly interesting, given that there was little evidence that mothers and their children were aligned with respect to f0 and f0 Range in Session 2 compared to Session 1.

All in all, the results of Experiments 1 and 2 provide evidence for alignment between mothers and their children. Regardless of whether mothers shadowed their children's productions (Experiments 1 and 2) or children shadowed their mothers' productions (Experiment 2), listeners were more likely to perceive children's productions as sounding more similar to children's own mothers than to other mothers, presumably based in part on perceived similarities in duration, f0, and f0 range. While we think children were indeed aligning to their mothers when shadowing them in Experiment 2, we should reiterate that the recordings for these trials were obtained

---

[7] glmer(OwnMother ~ ChildGender * RecordingTime + (1 | Participant) + (RecordingTime | Item))

immediately after recording mothers shadowing their children. Because of this, it is possible that the alignment which occurred during the previous, Child 1st recordings facilitated alignment during the subsequent Mother 1st recordings, though we find this unlikely, given previous work suggesting that short-term convergence in children is quite short-lived (Paquette-Smith et al., 2021).

In addition, the strength of alignment in Experiments 1 and 2 did not differ significantly by the gender of the child, unlike what has been found with some studies on adults (e.g., Namy et al., 2002). This might be related to the fact that at this age, children's linguistic systems tend be modelled after their primary caregiver's (typically their mother), so it might be natural for children and their mothers—regardless of a child's gender—to phonetically resemble one another, especially in a shadowing task (Labov, 2001). Finally, alignment did not differ by age group, with both 2-year-olds and 4-year-olds and their mothers exhibiting similar patterns of alignment.

We assume that the alignment we observed in Experiments 1 and 2 was short-term, whereby shadowers were aligning to the tokens produced immediately before their own productions rather than to exemplars stored in long-term memory. However, given that we did not gather any baseline recordings before the shadowing task—which would have been difficult to do with 2.5- to 4-year-olds, on top of the shadowing task—it is difficult to claim with certainty that the alignment we observed was due primarily to short-term alignment, or to overall similarities between mothers and their children as a result of long-term alignment. To help to better adjudicate between these two possibilities using the recordings we had, we ran an additional experiment to more directly test the presence of long-term alignment, testing the perceived similarity of children and their mothers between the two different recording sessions, creating a new set of AXB trials using children's productions in the first recording session (Experiment 1) and mothers' productions from the second recording session (Experiment 2). If the observed alignment in Experiments 1 and 2 was primarily driven by overall similarities between children and their mothers, we might expect to find a similar magnitude of perceptual similarity across different recording sessions, as assessed by an AXB perceptual task. If, however, there is no evidence of overall similarity between children and their mothers (compared to other mothers), that would lend support to the idea that the alignment observed in Experiments 1 and 2 was primarily due to short-term rather than long-term alignment.

Experiment 3

To assess the extent of overall similarity between child-mother dyads, Experiment 3 examined whether tokens of the same words produced by mothers and their children on separate occasions would still be evaluated as perceptually similar to one another compared to productions from other mothers. To test this, the Mother 1st recordings obtained in Experiment 2 were paired with the original child recordings in Experiment 1 (in which the child produced the word stimuli first), resulting in the child and mother tokens from two separate recording sessions. In this way, any perceived similarity found in Experiment 3 between children and their own mothers (compared to other mothers) would stem from overall, long-term similarities in mother-child productions.

Method

Participants

Twenty-five Canadian English listeners (22 women, M_{age} = 20.3 years, range = 18-26 years), with no self-reported speech or hearing deficits, participated in this experiment. All listeners learned English before the age of 6, and reported using English at least 70% of the time at the time of testing.

Materials and Procedure

Recordings of the mother-child dyads who participated in both the recording session for Experiments 1 and 2 (18 pairs) were used to construct stimuli for Experiment 3. Specifically, the tokens produced by the mothers in the Mother 1st recording session in Experiment 2 were spliced into the original trials of Experiment 1, yielding trials in which the mother tokens were produced in a different recording session than the child tokens (e.g., *strawberry*$_{\text{MotherX-Experiment2}}$ + *strawberry*$_{\text{ChildX-Experiment1}}$ + *strawberry*$_{\text{MotherY-Experiment2}}$). Trials which included tokens produced by mothers who did not come in for the second recording session were removed, resulting in 126 trials per participant. Apart from the reduced number of trials, the procedure for participants was identical to Experiments 1 and 2.

Results and Discussion

The same analyses were conducted as in Experiment 1, using a logistic mixed-effects regression model to predict the log odds of selecting children's Own Mother compared to the Different Mother. Child Gender (Male, -0.5, and Female, 0.5) was included as a fixed effect, as well as random intercepts for items (each child/word pair) and participants.[8] Unlike Experiment 1, overall perceptions of similarity between children and their own mothers were not significantly above chance ($\beta = 0.06$, $SE = 0.13$, $z = 0.48$, $p = 0.63$). While there was numerically greater perceived similarity between mothers and their daughters (M = 0.53) compared to mothers and

8 glmer(OwnMother ~ ChildGender + (1 | Participant) + (1 | Item))

their sons (M = 0.49), this difference was not statistically significant ($\beta = 0.17$, $SE = 0.25$, $z = 0.70$, $p = 0.49$), suggesting that there was no evidence of long-term similarity in either mother-daughter or mother-son pairs, as measured by an AXB perceptual similarity task.

Thus, while mothers and their children appear to align with one another when directly shadowing one another, as demonstrated in Experiments 1 and 2, there is no evidence that in the long-term, children's productions are perceptually similar to their mother's (compared to other mothers' productions), despite the fact that children are thought to begin their language development by reproducing the linguistic patterns exhibited by their female caretakers (Foulkes & Docherty, 2006; Labov, 2001). We think that this lack of evidence for long-term alignment is related to toddlers' underdeveloped precision in articulating and producing speech. Given that that toddlers' spontaneous productions are highly variable (Levy & Hanulíková, 2019) and not adult-like in all their acoustic detail, any two tokens from mothers and children are unlikely to match along a variety of phonetic dimensions, making the detection of long-term alignment via an AXB perceptual similarity task improbable. With the presence of recent exemplars, however, as was the case in the shadowing task in Experiments 1 and 2, where partners could imitate an immediately preceding token, toddlers and mothers may have been better able to match one another's productions. For this reason, we believe that the alignment we observed in Experiments 1 and 2—but not in Experiment 3—was due to short-term rather than long-term alignment; specifically, mothers and their children imitated certain aspects of each other's immediately preceding productions rather than exhibit overall more similarity to one another (in relation to other speakers) as a result of prolonged contact.

Another possibility is that the mothers' speech may have changed in the 15-month gap between when the child and mother tokens were recorded, making the mother recordings sound dissimilar from the child recordings 15 months earlier. However, we find this possibility unlikely; while the children's speech certainly changed between the two recordings, we have no reason to believe the mothers underwent significant language change in the same time period, given that speech patterns are generally presumed to remain relatively stable across adulthood, all things being equal (Sankoff, 2006). Thus, the alignment we observed in Experiment 1 (and by extension Experiment 2), rather than due to long-term similarity, is instead likely due to short-term convergence, with mothers adjusting their productions in real-time in response to their child's immediately preceding productions.

If the failure to detect long-term convergence in Experiment 3 is indeed due to young children being highly variable in their productions, as suggested earlier, then we might expect to find long-term alignment in older children and their mothers when children are able to produce language with more adult-like precision, in contrast to younger children whose speech motor control is still developing. To explore this possibility, we tested for the presence of long-term alignment between older children (8-year-olds) and their mothers. Because children were older and could be easily separated them from their mothers, we were able to record both mothers and children during the same visit, removing potential confounds that were present in Experiment 3.

Experiment 4

In Experiment 4, we tested whether children with more adult-like proficiency in English would exhibit some degree of long-term alignment with their mothers, unlike the younger children assessed in Experiment 3. We selected 8-year-old children, who, unlike pre-school-aged children, have largely mastered the sound patterns of their native language; and although they have begun schooling, associating increasingly more with their peers and other adult figures (e.g., teachers, coaches, etc.), they have not yet shown the same degree of deviation from adult/parental norms as pre-adolescents and adolescents, who tend to increasingly focus on peer relationships and carving out an identity distinct from adult figures (e.g., Deser, 1989; Eckert, 2000; Kerswill & Williams, 2000). At the same time, evidence suggests that children at this age may also exhibit gender-based variation in their own speech (Fischer, 1958; Ladegaard & Bleses, 2003), so if long-term alignment varies as a function of gender, we might expect mothers to more closely align with daughters than their sons.

Method

Participants:

Twenty-six Canadian English listeners (17 women, M_{age} = 19.6 years, range = 18-31), with no self-reported speech or hearing difficulties participated in this experiment. All listeners learned English before the age of 6.

Materials and Procedure

The stimuli consisted of recordings of the same 32 words used in the previous experiments; this time obtained from 20 monolingual, Canadian-English learning 8-year-olds (11 boys, M*age* = 8;6, range = 8;0 – 9;5) and their Canadian-English speaking mothers. Unlike in the previous experiments, productions were elicited from mothers and their children separately during the same visit, with each speaker taking a turn in the sound booth. Because these children were older, we removed the cover story involving the Martians, and simply asked them to name pictures that appeared on the screen when prompted by the computer, as in the previous experiments (*This is a …*). As before, these recordings were used to construct AXB trials, with each of the 20 mother-child pairs appearing in 8 trials (4 words x 2 orders, OwnMother in A position and OwnMother in B position) for a total of 160 trials per participant. The procedure was also identical to the previous experiments.

Results and Discussion

As in the previous experiments, we ran a mixed-effects logistic regression predicting the log odds of selecting a child's own mother from Child Gender (Male, -0.5, Female, 0.5), with random by-subject and -item intercepts included in the model.[9] Although overall mothers and children did not exhibit long-term alignment (M = 0.52), as suggested by a non-significant intercept term ($\beta = 0.08$, $SE = 0.08$, $z = 1.03$, $p = 0.30$), there was a marginally significant effect of Child Gender ($\beta = -0.29$, , $SE = 0.16$, $z = -1.81$, $p = 0.07$). Specifically, as follow-up tests revealed (re-running the test two times with each gender coded as 0), while participants were equally likely to select a female child's own mother as compared to another mother (M = 0.49; $\beta = -0.06$, , $SE = 0.12$, $z = -0.52$, $p = 0.60$), they selected male children's own mother at rates significantly above chance (M = 0.55; $\beta = 0.23$, , $SE = 0.11$, $z = 2.12$, $p = 0.03$), providing some evidence that mothers and their sons—and not mothers and daughters—exhibited some degree of long-term alignment. While it is not clear why male children were perceived as more acoustically similar to their mothers but not female children, it may be the case that girls at this age have already started orienting more towards their peers than boys. Indeed, studies have shown that girls—especially early in life—are on average more advanced in language performance than their similarly-aged, male counterparts (Bornstein et al., 2004). If higher language competence extended also to sociolinguistic competence, we might predict that female children might start orienting more towards peer speech earlier than male children, who might

[9] glmer(OwnMother ~ ChildGender + (1 | Participant) + (1 | Item))

retain features of their caregiver's speech for longer.

We next assessed the extent to which mothers and their children were similar with respect to mean f0 values and f0 Range. As in Experiments 1 and 2, we created two linear mixed-effects models predicting one speaker's values from the other, with Child Gender (Male, -0.5, Female, 0.5), and its interaction with the independent variable (mean-centered), included as additional fixed effects, and by-dyad and by-item intercepts included as random effects.[2] Eight items were excluded due to creak, and given that there were no model speakers in this experiment, we arbitrarily chose mothers' values as the dependent variable. As in Experiment 2, there were no significant correlations between children's and mothers' productions, both in the f0 model ($\beta_{Childf0}$ = -0.09, *SE* = 0.12, t = -0.74, p = .46) and in the f0 Range model ($\beta_{Childf0Range}$ = 0.01, *SE* = 0.12, t = 0.07, p = .94). In addition, there were no significant interactions with or main effect of Child Gender.

As in previous experiments, we additionally examined whether greater similarity between children and mothers (as measured by DID scores) predicted listeners' choices in the AXB task. A mixed-effects logistic regression was conducted, predicting listeners' choice of Own Mother (versus Different Mother) from Child Gender (Male = -0.5, Female = 0.5) and z-transformed DID scores for f0, f0 Range, and word Duration. Uncorrelated by-subject random intercepts and slopes for f0, f0 Range, Duration, and Child Gender were included in the model as well as random by-item intercepts.[10] Ten items were excluded from the analysis (260/4160 trials) due to creak. Results revealed significant main effects of Duration (β = 0.26, *SE* = 0.08, z = 3.27, p < 0.01) and f0 (β = 0.20, *SE* = 0.08, z = 2.33, p = 0.02), with greater similarities between children

[10] glmer(OwnMother ~ zDIDf0 + zDIDf0Range + zDIDDuration + ChildGender + (zDIDf0 + zDIDf0Range + zDIDDuration + ChildGender || Participant) + (1 | Item))

and their mothers (relative to other mothers) along those dimensions leading to increased selection of children's mothers over other mothers.

General Discussion

Previous research has found that vocal imitation appears quite early in development, and has been posited to play a pivotal mechanistic role in the process of acquiring a language, helping to develop turn-taking skills (Masur & Olson, 2008), word knowledge (Masur & Eichorst, 2002), and pronunciation (Gros-Louis et al., 2006). Some theories have argued that imitation on the part of infants is central for learning the sounds of their language, which is due in large measure to their ability to imitate stored representations of words and sounds in memory (Kuhl et al., 2008; Sundqvist et al., 2016). Other theories have instead emphasized the crucial role played by caregivers, whose imitations of infants are said to facilitate language acquisition (Rasilo & Räsänen, 2017). Mothers, for example, may differentially imitate only certain types of structures (e.g., CV sequences) to scaffold vocal development (Gros-Louis et al., 2006), or mirror infant vocalizations that sound most speech-like in order to encourage more speech-like productions in the future (Messum, 2007).

Despite some evidence of immediate, vocal imitation in infants (Imafuku et al., 2019; Kuhl & Meltzoff, 1996; Legerstee, 1990), for the most part, caregivers have been found to perform the bulk of imitative work when engaging with preverbal infants (Kokkinaki & Kugiumutzakis, 2000; Pawlby, 1977; Siegel et al., 1990; see Jones, 2009 for a review), with relatively less known about alignment between caregivers and children after the second year of life. In the current study, we investigated a form of vocal imitation—phonetic/prosodic alignment—in older intergenerational (mother-child) dyads, looking at the directionality of alignment and doing so using perceptual measures in addition to acoustic ones.

In Experiments 1 and 2, we overall found alignment between mothers and their children during a shadowing task, with greater perceived similarity between children and their own mothers compared to other mothers. In contrast to studies with infants, which have found caregivers contributing most to parent-child alignment, in our study alignment appeared to be bidirectional in nature, with *both* children and mothers equally likely to align to one another, suggesting that as children develop linguistic abilities, they begin to engage increasingly more in imitative verbal behavior (Masur, 1993; Masur & Rodemaker, 1999).

This interpretation assumes that the shadowers in Experiments 1 and 2 were imitating features of the immediately preceding token (short-term alignment), but because no baseline recordings were obtained before the shadowing task (due to difficulties involved in separating young children from their parents), it was alternatively possible that shadowers simply exhibited overall similarities with the model speaker as a result of long-term exposure to one another (long-term alignment). To help adjudicate between the two possibilities, we ran an additional experiment (Experiment 3) testing for the presence of overall similarities between the same children and their mothers (i.e., long-term alignment) using recordings from two separate recording sessions (children in Experiment 1 and mothers in Experiment 2). In this experiment, we found no evidence of long-term alignment, lending support to the idea that the alignment in Experiments 1 and 2 was due to short-term rather than long-term alignment.

The presence of no long-term alignment in Experiment 3 could be seen as somewhat surprising, given previous work showing long-term similarities between children's productions and their caregivers (e.g., Stoehr et al, 2019) and the fact that children's early grammars are believed to be modelled after their primary caregivers (typically their mother; Labov, 2001). This could suggest that, given how variable young children's productions are in the absence of adult exemplars (MacDonald et al., 2012), the speech of toddlers and their mothers do not perceptibly resemble one another without temporally close exemplars in a shadowing task.

Interestingly, in older children (8-years-old) and their mothers, who *were* recorded separately in the same visit, we did find some evidence of long-term alignment, but only between boys and their mothers. This finding ran counter to what we expected, since by 8-years-old, children have already been shown to exhibit gender-based variation in their speech (Fischer, 1958; Munson et al., 2015; Perry et al., 2001), making mothers and daughters arguably more likely to align than mothers and sons. Instead, we speculate what might be happening is that the female children in our sample have already begun orienting more towards the speech of their peer groups, with male children continuing to exhibit characteristics of the caregiver model. Given that our effect of Gender was only marginally significant, however, more research is needed to confirm whether primary-school aged boys indeed resemble their caregivers more than girls, particularly research looking at multiple time points across the school-aged years.

In future work, it may be additionally useful to explore other factors that may influence the degree to which children are phonetically aligned with interlocutors. While a factor like gender is a highly salient category for children which leads to early linguistic differentiation between boys and girls, it may be too coarse of a predictor in the absence of other variables which also influence the degree of phonetic alignment. In Experiment 4, for example, while male 8-year-olds were overall more likely to exhibit long-term alignment to their mothers than girls were, there was nevertheless a fair amount of unexplained variability in the degree to which caregivers and their children phonetically aligned (see Figure 4). Gathering more detailed information about the nature children's interpersonal relationships with peers, caregivers, siblings, etc., might help to shed light on the kinds of social processes that lead to more or less alignment between children and their interlocutors across development (see Barbu et al., 2014; Milroy, 1980; Nardy et al., 2014).

[FIGURE 4 NEAR HERE]

Studies examining vocal imitation in children emphasize its importance in helping children acquire the sounds of their language, but there is also a sense in which imitation additionally plays a social role. Adults use imitation to increase liking (Chartrand & Bargh, 1999) and closeness to others (Ashton-James et al., 2007), with phonetic alignment serving as one form of imitative behavior through which to achieve these outcomes (e.g., Babel, 2012). By 5 months of age, infants too expect characters who imitate each other to socially affiliate with one another (Powell & Spelke, 2018), and by 14 months, will selectively imitate in-group over outgroup members (Buttelmann et al., 2013; see Kinzler et al., 2011 for similar results with older children). Very little is known, however, about when and how children use begin using imitative linguistic behaviors (e.g., phonetic alignment) to convey group affiliation and liking. In the same way they imitate in-group behaviors more than out-group ones, would children be more likely to phonetically align with in-group members more than out-group members? Would they be more likely to phonetically align to socially desirable children who are perceived as "cool" compared to children who are not? Are there gender differences in the extent to which children align with their peers? Would female children be as likely to align with their fathers as male children? Answering these types of questions would help to shed light on when and how children begin using language to achieve social goals.

Overall, as children develop, they must learn how language is used to convey social—in addition to propositional—meaning. They must learn how speakers of different social groups (e.g., age groups, gender, SES, school cliques, etc.) tend to vary with respect to how they use language, and begin exploiting these features of language to mark their group affiliation(s) and carve out their own sociolinguistic identity. Children may begin the language acquisition process by speaking like their primary caregivers, but eventually they will come to speak more like their peers (Eckert, 2000) and the broader community more generally (Tagliomonte & Molfenter, 2007), a phenomenon most apparent in children whose parents speak with a non-local accent (Kerswill & Williams, 2000). For the most part, the process of how this occurs and its trajectory across development remains largely unexplored. Phonetic alignment could prove to be a useful tool for investigating these types of questions, by exploring ways in which different types of dyads (e.g., child-child, child-father in addition to child-mother, between same-sex and different-sex children, etc.) resemble one another across different stages in development in different interactive contexts. In other words, phonetic alignment could provide a way of probing how children's sociolinguistic knowledge develops across time. The current study has laid the ground work for exploring these types of questions in the future.

Acknowledgements

We would like to thank Lisa Hotson, Arshnoor Khaira, Jasleen Grewal, Jenna Laking, and Tamim Fattah, as well as the other members of the Child Language and Speech Studies Lab for their support. This work was supported by grants from the Social Sciences and Humanities Research Council and Natural Sciences and Engineering Research Council. Portions of this work were presented at the 176th Meeting of the Acoustical Society of America.

Declaration of Interest Statement

We have no conflicts of interest to disclose.

References

Ashton-James, C., van Baaren, R. B., Chartrand, T. L., Decety, J., & Karremans, J. (2007). Mimicry and me: The impact of mimicry on self-construal. *Social Cognition*, *25*(4), 518-535. https://doi.org/10.1521/soco.2007.25.4.518

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437–456. https://doi.org/10.1017/S0047404510000400

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189. https://doi.org/10.1016/j.wocn.2011.09.001

Babel, M., & Bulatov, D. (2012). The Role of Fundamental Frequency in Phonetic Accommodation. *Language and Speech*, *55*(2), 231–248. https://doi.org/10.1177/0023830911417695

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, *5*(1), 123–150. https://doi.org/10.1515/lp-2014-0006

Barbu, S., Martin, N., & Chevrot, J.-P. (2014). The maintainence 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*, Article 1251. https://doi.org/10.3389/fpsyg.2014.01251

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Bock, J. K. (1986). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning Memory and Cognition*, *12*(4), 575-586. https://doi.org/10.1037/0278-7393.12.4.575

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., Pêcheux, M.-G., Ruel, J., Venuti, P., & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, *75*(4), 1115-1139. https://doi.org/10.1111/j.1467-8624.2004.00729.x

Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, Ethnicity, and Intergroup Relations* (pp. 119-135). Academic Press.

Branigan, H. P., Tosi, A., & Gillespie-Smith, K. (2016). Spontaneous lexical alignment in children with an autistic spectrum disorder and their typically developing peers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(11), 1821-1831. https://doi.org/10.1037/xlm0000272

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development*, *84*(2), 422-428. https://doi.org/10.1111/j.1467-8624.2012.01860.x

Chambers, J. K. (1992). Dialect Acquisition. *Language*, *68*(4), 673–705. https://doi.org/10.2307/416850

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893-910. https://doi.org/10.1037/0022-3514.76.6.893

Clopper, C. G., & Dossey, E. (2020). Phonetic convergence to Southern American English: Acoustics and perception. *The Journal of the Acoustical Society of America*, *147*(1), 671-683. https://doi.org/10.1121/10.0000555

Deser, T. (1989). Dialect transmission and variation: An acoustic analysis of vowels in six urban Detroit families. *York Papers in Linguistics*, *13*, 115-128.

Dufour, S., & Nguyen, N. (2013). How much imitation is there in a shadowing task? *Frontiers in Psychology*, *4*, Article 346. https://doi.org/10.3389/fpsyg.2013.00346

Eaton, C. T., & Ratner, N. B. (2013). Rate and phonological variation in preschool children: Effects of modeling and directed influence. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1751-1763. https://doi.org/10.1044/1092-4388(2013/12-0171)

Eckert, P. (2000). *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Blackwell Publishers.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, *121*(6), 3814-3826. https://doi.org/10.1121/1.2722209

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

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*(4), 409-438. https://doi.org/10.1016/j.wocn.2005.08.002

Foulkes, P., Docherty, G. J., & Watt, D. (2005). Phonological variation in child-directed speech. *Language*, *81*(1), 177-206. https://doi.org/10.1353/lan.2005.0018

Fung, P., Schertz, J., & Johnson, E. K. (2021). The development of gendered speech in children: Insights from adult L1 and L2 perceptions. *The Journal of the Acoustical Society of America Express Letters*, *1*(1), 014407. https://doi.org/10.1121/10.0003322

Garnier, M., Lamalle, L., & Sato, M. (2013). Neural correlates of phonetic convergence and speech imitation. *Frontiers in Psychology*, *4*, Article 600. https://doi.org/10.3389/fpsyg.2013.00600

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218. https://doi.org/10.1016/0010-0277(87)90018-7

Garvey, C., & BenDebba. (1974). Effects of age, sex, and partner on children's dyadic speech. *Child Development*, *45*(4), 1159-1161. https://doi.org/10.2307/1128114

Goldinger, S. D. (1998). Echoes of echoes? An Episodic Theory of Lexical Access. *Psychological Review*, *105*(2), 251-279. https://doi.org/10.1037/0033-295X.105.2.251

Gregory, S. W., Jr., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, *70*(6), 1231–1240. https://doi.org/10.1037/0022-3514.70.6.1231

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*(6), 509-516. https://doi.org/10.1177/0165025406071914

Harrington, J. (2007). Evidence for a relationship between synchronic variability and diachronic change in the Queen's annual Christmas broadcasts. In J. Cole & J. Hualde (Eds.), *Laboratory Phonology 9: Phonetics and Phonology* (pp. 125–144). Walter de Gruyter.

Hines, M. (2015). Gendered development. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of Child Psychology and Developmental Science: Socioemotional Processes* (pp. 842–887). John Wiley & Sons Inc. https://doi.org/10.1002/9781118963418.childpsy320

Imafuku, M., Kanakogi, Y., Butler, D., & Myowa, M. (2019). Demystifying infant vocal imitation: The roles of mouth looking and speaker's gaze. *Developmental Science*, *22*(6), Article e12825. https://doi.org/10.1111/desc.12825

Jones, S. S. (2009). The development of imitation in infancy. *Philosophical Transactions of the Royal Society B*, *364*(1528), 2325-2335. https://doi.org/10.1098/rstb.2009.0045

Karrass, J., Braungart-Rieker, J. M., Mullins, J., & Lefever, J. B. (2002). Processes in language acquisition: The roles of gender, attention, and maternal encouragement of attention over time. *Journal of Child Language*, *29*(3), 519-543. https://doi.org/10.1017/S0305000902005196

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech, Language, and Hearing Research*, *19*(3), 421-447. https://doi.org/10.1044/jshr.1903.421

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. (2003). Age, input, and language mode factors in the acquisition of VOT by English-Arabic bilingual children. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain* (pp. 3213-3216).

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125–156. https://doi.org/10.1515/labphon.2011.004

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science*, *14*(1), 106-111. https://doi.org/10.1111/j.1467-7687.2010.00965.x

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2002). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and a non-tonal language. *Infant Behavior and Development*, *24*(4), 372-392. https://doi.org/10.1016/S0163-6383(02)00086-3

Klinger, H. (1962). Imitated English cleft palate speech in a normal Spanish speaking child. *Journal of Speech and Hearing Disorders*, *27*(4), 379-381. https://doi.org/10.1044/jshd.2704.397

Ko, E.-S., Seidl, A., Cristia, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language*, *43*(2), 284-309. https://doi.org/10.1017/S0305000915000203

Kokkinaki, T., & Kugiumutzakis, G. (2000). Basic aspects of vocal imitation in infant-parent interaction during the first 6 months. *Journal of Reproductive and Infant Psychology*, *18*(3), 173–187. https://doi.org/10.1080/713683042

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*, *363*(1493), 979-1000. https://doi.org/10.1098/rstb.2007.2154

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Labov, W. (2001). *Principles of Linguistic Change:Social Factors.* Blackwell.

Ladegaard, H. J., & Bleses, D. (2003). Gender differences in young children's speech: The acquisition of sociolinguistic competence. *International Journal of Applied Linguistics*, *13*(2), 222-233. https://doi.org/10.1111/1473-4192.00045

Legerstee, M. (1990). Infants use multimodal information to imitate speech sounds. *Infant Behavior and Development*, *13*(3), 343-354. https://doi.org/10.1016/0163-6383(90)90039-B

Levy, H., and Hanulíková, A. (2019) Variation in children's vowel production: Effects of language exposure and lexical frequency. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1), 9. https://doi.org/10.5334/labphon.131

Lewandowski, N., & Jilka, M. (2019). Phonetic convergence, language talent, personality and attention. *Frontiers in Communication*, *4*, Article 18. https://doi.org/10.3389/fcomm.2019.00018

Lewandowski, E. M., & Nygaard, L. C. (2018). Vocal alignment to native and non-native speakers of English. *The Journal of the Acoustical Society of America*, *144*(2), 620-633. https://doi.org/10.1121/1.5038567

MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., & Munhall, K.G. (2012). Children's development of self-regulation in speech production. *Current Biology, 22*(2), 113-117. doi.org/10.1016/j.cub.2011.11.052

Masur, E. F. (1993). Transitions in representational ability: Infants' verbal, vocal, and action imitation during the second year. *Merrill-Palmer Quarterly*, *39*(4), 437-456. https://www.jstor.org/stable/23087243

Masur, E. F., & Eichorst, D. L. (2002). Infants' spontaneous imitation of novel versus familiar words: Relations to observational and maternal report measures of their lexicons. *Merrill-Palmer Quarterly*, *48*(4), 405–426. https://doi.org/10.1353/mpq.2002.0019

Masur, E. F., & Olson, J. (2008). Mothers' and infants' responses to their partners' spontaneous action and vocal/verbal imitation. *Infant Behavior & Development*, *31*(4), 704-715. https://doi.org/10.1016/j.infbeh.2008.04.005

Masur, E. F., & Rodemaker, J. E. (1999). Mothers' and infants' spontaneous vocal, verbal, and action imitation during the second year. *Merrill-Palmer Quarterly*, *45*(3), 392-412. https://www.jstor.org/stable/23092579

McRoberts, G. W., & Best, C. T. (1997). Accomodation in mean f_0 during mother-infant and father-infant vocal interaction: A longitudinal case study. *Journal of Child Language*, *24*(3), 719-736. https://doi.org/10.1017/S030500099700322X

Messum, P. (2007). Mirroring, not imitation, for the early learning of L1 pronunciation. *The Journal of the Acoustical Society of America*, *122*(5), 2997. https://doi.org/10.1121/1.2942701

Messum, P., & Howard, I. S. (2015). Creating the cognitive form of phonological units: The speech sound correspondence problem in infancy could be solved by mirrored vocal interactions rather than by imitation. *Journal of Phonetics*, *53*, 125-140. https://doi.org/10.1016/j.wocn.2015.08.005

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2013). Is speech alignment to talkers or tasks? *Attention, Perception, & Psychophysics*, *75*, 1817-1826. https://doi.org/10.3758/s13414-013-0517-y

Milroy, L. (1980). *Language and Social Networks*. Basil Blackwell.

Munro, M. J., Derwing, T. M., & Flege, J. E. (1999). Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics*, *27*(4), 385–403. https://doi.org/10.1006/jpho.1999.0101

Munson, B., Crocker, L., Pierrehumbert, J. B., Owen-Anderson, A., & Zucker, K. J. (2015). Gender typicality in children's speech: A comparison of boys with and without gender identity disorder. *The Journal of the Acoustical Society of America*, *137*(4), 1995-2003. https://doi.org/10.1121/1.4916202

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, *21*(4), 422–432. https://doi.org/10.1177/026192702237958

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

Nardy, A., Chevrot, J.-P., Barbu, S. (2014). Sociolinguistic convergence and social interactions within a group of preschoolers: A longitudinal study. *Language Variation and Change*, *26*(3), 273-301. https://doi.org/10.1017/S0954394514000131

Nielsen, K. (2014). Phonetic imitation by young children and its developmental changes. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2065–2075. https://doi.org/10.1044/2014_JSLHR-S-13-0093

Nycz, J. (2015). Second dialect acquisition: A sociophonetic perspective. *Language and Linguistics Compass*, *9*(11), 469-482. https://doi.org/10.1111/lnc3.12163

Olmstead, A. J., Viswanathan, N., Aivar, M. P., & Manuel, S. (2013). Comparison of native and non-native phone imitation by English and Spanish speakers. *Frontiers in Psychology*, *4*, Article 475. https://doi.org/10.3389/fpsyg.2013.00475

Oviatt, S., Darves, C., & Coulston, R. (2004). Toward adaptive conversation interfaces: Modeling speech convergence with animated personas. *ACM Transactions on computer-human interaction*, *11*(3), 300-328. https://doi.org/10.1145/1017494.1017498

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language*, *9*(6), 137–157. https://doi.org/10.1177/014272378900900603

Paquette-Smith, M., Schertz, J., & Johnson, E. K. (2021). Comparing phonetic convergence in children and adults. *Language and Speech.* https://doi.org/10.1177/00238309211013864

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*(1), 190-197. https://doi.org/10.1016/j.wocn.2011.10.001

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, *72*(8), 2254-2264. https://doi.org/10.3758/bf03196699

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language*, *69*(3), 183–195. https://doi.org/10.1016/j.jml.2013.06.002

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception & Psychophysics*, *79*, 637–659. https://doi.org/10.3758/s13414-016-1226-0

Pawlby, S. J. (1977). *A study of imitative interaction between mothers and their infants* [Doctoral dissertation, University of Nottingham]. Proquest Dissertations & Theses Global.

Payne, A. C. (1976). The Acquisition of the Phonological System of a Second Dialect [Doctoral dissertation, University of Pennsylvania]. Proquest Dissertations & Theses Global.

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

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-347. https://doi.org/10.1017/S0140525X12001495

Powell, L. J., & Spelke, E. S. (2018). Human infants' understanding of social imitation: Inferences of affiliation from third party observations. *Cognition*, *170*, 31-48. https://doi.org/10.1016/j.cognition.2017.09.007

Rasilo, H., & Räsänen, O. (2017). An online model for vowel imitation learning. *Speech Communication*, *86*, 1-23. https://doi.org/10.1016/j.specom.2016.10.010

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421-436. https://doi.org/10.1006/jpho.1997.0051

Sankoff, G. (2006). Age: Apparent time and real time. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (2nd ed., pp. 110-116). Elsevier. https://doi.org/10.1016/B0-08-044854-2/01479-6

Sankoff, G., & Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language, 83*(3), 560-588. https://doi.org/10.1353/lan.2007.0106

Schertz, J., Paquette-Smith, M., & Johnson, E. K. (2019). The relationship between perceptual similarity judgments and VOT convergence in a shadowing task. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia* (pp. 3711-3715). Australasian Speech Science and Technology Association Inc.

Seidl, A., Cristia, A., Soderstrom, M., Ko, E.-S., Abel, E. A., Kellerman, A., & Schwichtenberg, A. J. (2018). Infant-mother acoustic-prosodic alignment and developmental risk. *Journal of Speech, Language, and Hearing Research*, *61*(6), 1369-1380. https://doi.org/10.1044/2018_JSLHR-S-17-0287

Shimpi, P. M., Gámez, P. B., Huttenlocher, J., & Vasilyeva, M. (2007). Syntactic priming in 3- and 4-year-old children: Evidence for abstract representations of transitive and dative forms. *Developmental Psychology*, *43*(6), 1334-1346. https://doi.org/10.1037/0012-1649.43.6.1334

Shockey, L. (1984). All in a flap: Long-term accommodation in phonology. *International Journal of the Sociology of Language*, *46*, 87–95. https://doi.org/10.1515/ijsl.1984.46.87

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*, 422–429. https://doi.org/10.3758/BF03194890

Smith, J., & Durham, M. (2019). *Sociolinguistic Variation in Children's Language*: *Acquiring Community Norms*. Cambridge University Press. https://doi.org/10.1017/9781316779248

Siegel, G. M., Cooper, M., Morgan, J. L., & Brenneise-Sarshad, R. (1990). Imitation of intonation by infants. *Journal of Speech, Language, and Hearing Research*, *33*(1), 9–15. https://doi.org/10.1044/jshr.3301.09

Sonderegger, M., Bane, M., & Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language*, *93*(3), 598-640. https://doi.org/10.1353/lan.2017.0038

Stoehr, A., Benders, T., van Hell, J. G., & Fikkert, P. (2019). Bilingual preschoolers' speech is associated with non-native maternal language input. *Language Learning and Development*, *15*(1), 75-100. https://doi.org/10.1080/15475441.2018.1533473

Story, B. H., & Bunton, K. (2016). 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

Street, R. L., Jr. (1983). Noncontent speech convergence and divergence in adult-child interactions. *Annals of the International Communication Association*, *7*(1), 369-395. https://doi.org/10.1080/23808985.1983.11678543

Sundqvist, A., Nordqvist, E., Koch, F.-S., Heimann, M. (2016). Early declarative memory predicts productive language: A longitudinal study of deferred imitation and communication at 9 and 16 months. *Journal of Experimental Child Psychology*, *151*, 109-119. https://doi.org/10.1016/j.jecp.2016.01.015

Tagliamonte, S. A., & Molfenter, S. (2007). How'd you get that accent?: Acquiring a second dialect of the same language. *Language in Society*, *36*(5), 649–675. https://doi.org/10.1017/S0047404507070911

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology*, *6*, Article 546. https://doi.org/10.3389/fpsyg.2015.00546

Welkowitz, J., Cariffe, G., & Feldstein, S. (1976). Conversational congruence as a criterion of socialization in children. *Child Development*, *47*(1), 269-272. https://doi.org/10.2307/1128311

Wynn, C. J., Borrie, S. A., & Sellers, T. (2018). Speech rate entrainment in children and adults with and without autism spectrum disorder. *American Journal of Speech-Language Pathology*, *27*(3), 965-974. https://doi.org/10.1044/2018_AJSLP-17-0134

Wynn, C. J., Borrie, S. A., & Pope, K. A. (2019). Going with the flow: An examination of entrainment in typically developing children. *Journal of Speech, Language, and Hearing*, *62*(10), 3706-3713. https://doi.org/10.1044/2019_JSLHR-S-19-0116

Appendix: Recorded words

Word
baby
ball
bear
bike
bird
boat
bunny
butterfly
cow
dog
duck
elephant
finger
fish
frog
horse
house
monkey
orange
phone
plane
shoe
spoon
squirrel
strawberry
stroller
swing
toothbrush
train
tree
truck
turtle

Figure Captions

Figure 1. Relationship between children's and mother's mean f0 values (left panel) and f0 Range values (right panel). Error ribbons denote +/- 1 standard error.

Figure 2. DID f0 (top left panel), DID Duration (top right panel) and DID f0 Range (bottom left panel) against proportion of perceived similarity between mothers and children in AXB similarity task. Error ribbons denote +/- 1 standard error.

Figure 3. Correlations between children's and mother's mean f0 values (first row) and f0 Range values (second row). Values in the Mother 1^{st} condition are presented in the left column, and values in the Child 1^{st} condition in the right. Error ribbons denote +/- 1 standard error.

Figure 4. Mean proportion similarity between children and their mothers in Experiment 4 by child. Each pair of initials represents the mean of a particular child-mother dyad.